Refining and improving a next-generation sequencing gene panel for clinical diagnosis of hereditary hematologic diseases and development of bioinformatic tools for biomedical analysis

Beatriz Cadenas Sevilla
2019

UVIC
UNIVERSITAT DE VIC
UNIVERSITAT CENTRAL
DE CATALUNYA

TESI DOCTORAL

Refining and improving a next-generation sequencing gene panel for clinical diagnosis of hereditary hematologic diseases and development of bioinformatic tools for biomedical analysis

Beatriz Cadenas Sevilla

Director de tesis: Mayka Sánchez Fernández
Director de tesis: M. Luz Calle Rosingana
Director de tesis: Cristian Tornador Antolín
Tutor: M. Luz Calle Rosingana

Programa de doctorat: Experimental Science and Technology

2019

UNIVERSITAT DE VIC
UNIVERSITAT CENTRAL DE CATALUNYA
Escola de Doctorat

“Some people want it to happen,
some wish it would happen,
others make it happen.”
Michael Jordan

“It is necessary to make a dream of life,
and to make of a dream a reality.”
Marie Curie

Summary

Research & Development and Innovation (R&D&I) are fundamental aspects for the growth and evolution of a company. Innovation is fast becoming a critical factor in the long-term growth development and survival of companies. The industrial doctorate manifested here, serves to bridge the gap between industry and academia, helps to industrialize the valuable knowledge that the academic research has generated, and puts the investigation into a practice in an industrial environment.

In this industrial doctorate, the aim was to fulfill the Whole Genix innovative objectives, including: master Next Generation Sequencing (NGS) protocols; cover a bigger spectrum of diseases; develop new bioinformatic and telemedicine tools for analysis of different omics data; and consequently, progress towards personalized medicine. Four independent projects have been carried out to cover these objectives.

1. Refining and improving an NGS gene panel for clinical diagnosis of hereditary hematologic diseases and focus on diseases with ferritin alterations.

The goal of this project is to improve an already established NGS-gene panel for rare inherited blood diseases (commercialized by BLOODGENTICS SL, a company associated with Whole Genix, SL) in order to accelerate diagnostics. Four design, each more updated and with more genes, have been studied for the diagnosis of twelve groups of hereditary hematological diseases. The implementation of the NGS methodology in clinical practice for the diagnosis of hereditary hematological diseases allows the rapid and effective diagnosis of these cases and to improve the clinical classification of these diseases.

Moreover, additional research in the *FTL* gene has been carried out due to its correlation with five hematological disorders. Two novel mutation in the *FTL* gene have been identified to be associated to two FTL-disorders thanks to the HHD gene panel: hyperferritinemia with cataracts syndrome (Esplugues mutation, c.-164_158del7) and L-ferritin deficiency with autosomal dominant inheritance (c.375+2T>A). Also, the second case of the mutation p.Met1Val in *FTL* gene associate with L-ferritin deficiency autosomal dominant is described.

2. CoDysAn: a web tool for the diagnosis of Congenital Dyserythropoietic Anemia

Congenital Dyserythropoietic Anemias (CDAs) are inherited hematological disorders characterized by chronic hyporegenerative anemia, morphological abnormalities of erythroblasts, and hemolytic component. Six subtypes of CDA has been defined genetically that present different morphological alterations of erythroblasts in the bone marrow. A proficient classification of CDA cases is crucial to provide efficient patient care and treatment.

CoDysAn aims to provide a responsive webpage that brings awareness of congenital dyserythropoietic anemia (CDA) to doctors, patients and general public, and provides a step-by-step diagnostic tool. The diagnostic algorithm is based in hematological parameters of the patients, including: hemoglobin level, Mean Corpuscular Volume of erythrocytes, and reticulocytes and platelets count. The normal ranges are dependent of the gender and age. The diagnostic tool achieved a satisfactory specificity of 89.5% and sensitivity of 78.5%. CoDysAn algorithm is connected to the NCBI Genetic Testing Registry (GTR) in a way to inform medical doctors about the existence of accredited diagnostic centers to perform a complete genetic test, if required.

3. BEA (BioMark Expression Analysis): a web tool to perform relative expression analysis from Fluidigm – BioMark raw data.

The high-throughput technologies for transcriptome analysis are ubiquitously used in modern biology and medical research. Nowadays researchers are employing these methods to analyze multiple genes from multiple samples in parallel, significatively reducing the handling time and the cost. BioMark HD system is a nanofluidic high-throughput qPCR capable to run more than nine thousand reactions at the same time (96 samples x 96 assays). In these high-throughput settings, the relative quantification through the ΔΔCT method and the statistical analysis could be not affordable for non-bioinformatic users.

BEA is a Shiny-based app to help wet-bench biologist with limited knowledge of programming and statistics to perform gene expression analysis from qPCR (quantitative Polymerase Chain Reaction) BioMark-Fluidigm data. BEA was designed to provide an easy step-by-step analysis workflow, with an appropriated statistical pipeline in a user-friendly interface. This enables users with low levels of programming and statistical knowledge to perform differential gene expression analysis for two or three groups of samples. The analysis includes relative quantification (ΔΔCT method) and appropriated statistical testing to identify differentially expressed genes between tested groups. The BEA app is available from http://bloodgenetics.com/otros/ web page.

The bioinformatic tool has been tested using BioMark data from Myelodysplastic syndrome project, whose goal was to identify genes associated with mitochondrial iron overload (ring sideroblasts) in sideroblastic anemia. Twenty-three genes identified to be differentially expressed between patients with myelodysplastic syndrome with ring sideroblasts (MDS-RS), patients with myelodysplastic syndrome without ring sideroblasts (MDS-nonRS) and healthy controls.

4. Pipeline for construction of gene expression signatures for prediction and prognosis using data from patients with muscle-invasive bladder cancer.

The advance and expansion of high-throughput technology have brought the opportunity to discover specific genetic signatures that can help to diagnose and classify a pathogenic medical condition, or predict the prognosis or the pharmacological responses to a specified treatment. Nevertheless, most phenotypes are complex traits resulting from the influence of both genomic factor and external conditions (non-genomic factors). To understand these complex traits and be able to define a gene expression signature of these phenotypes it is essential to study the genomic information and the non-genomic information jointly. However, the integration of these two types of data requires elaborated modeling and engenders some challenges that must be considered.

In this project we have elaborated a statistical pipeline based on an integrative conditional modeling approach in which non-genomic variables are selected first and used subsequently to select and adjust the genetic model. The non-genomic variables are selected by Stepwise regression, and the genomic variables by LASSO penalized regression in a cross-validation process. The statistical modeling has been tested using data from a muscle-invasive bladder cancer project with the objective to identify a gene expression signature that can predict the prognosis and efficacy of platinum-based neoadjuvant chemotherapies. The integrative conditional modeling found three non-genomic variables (tumor morphology type, the presence of hydronephrosis and the taxonomic classification) and nine genes (*RAD51, CXCL9, PARP, 53BP1, HERC2, ERCC2, CHEK1, Ku80* and *RNF168*) with the highest predictive ability (AUC=0.663). However, optimization analysis suggested that the non-genomic variables and the first three or four genes selected by LASSO (*RAD51, CXCL9, PARP* and *53BP1*) may have enough predictive ability to predict response to the treatment (AUC =0.671).

Resumen

La investigación y el desarrollo y la innovación (I+D+i) son unos de los aspectos más indispensables para el crecimiento y la evolución de una empresa. La innovación se está convirtiendo rápidamente en un factor crítico en el desarrollo y crecimiento a largo plazo y la supervivencia de las empresas. El doctorado industrial manifestado aquí sirve para abrir un puente entre la industria y la academia y ayuda a industrializar el valioso conocimiento que ha generado la investigación académica.

En este doctorado industrial el objetivo era cumplir los propósitos innovadores de Whole Genix, que incluyen: dominar los protocolos de secuenciación masiva de próxima generación (en inglés next-generation sequencing, NGS), cubrir un espectro más amplio de enfermedades, desarrollar nuevas herramientas bioinformáticas y de telemedicina para el análisis de diferentes datos ómicos; y consecuentemente, progresar hacia la medicina personalizada. Se han llevado a cabo cuatro proyectos independientes para cubrir estos objetivos.

1. Refinar y mejorar un panel de genes NGS para el diagnóstico clínico de enfermedades hematológicas hereditarias e investigación sobre enfermedades con alteraciones de la ferritina.

El objetivo de este proyecto es mejorar un panel de genes NGS para enfermedades sanguíneas hereditarias raras (comercializado por la empresa BLOODGENETICS SL asociada a la empresa Whole Genix SL) con el fin de acelerar el diagnóstico. Se han estudiado cuatro diseños, cada uno más actualizado y con más genes, para el diagnóstico de doce grupos de enfermedades hematológicas hereditarias. La implementación de la metodología de la secuenciación masiva en la práctica clínica para el diagnóstico de enfermedades hematológicas hereditarias permite el diagnóstico rápido y efectivo de estos casos y mejorar la clasificación clínica de estas enfermedades.

Además, se han realizado investigaciones adicionales en el gen *FTL* debido a su correlación con cinco enfermedades hematológicas. Se han identificado dos nuevas mutaciones en el gen *FTL* asociadas a dos enfermedades distintas, gracias al panel de genes HHD: hiperferritinemia con síndrome de cataratas (mutación de Esplugues, c.-164_158del7) y deficiencia de L-ferritina con herencia autosómica dominante (c.375 + 2T> A). Además, se describe el segundo caso con mutación p. Met1Val en FTL asociado con deficiencia de L-ferritina autosómica dominante.

2. CoDysAn: una herramienta web para el diagnóstico de anemia diseritropoyética congénita

Las anemias diseritropoyéticas congénitas (CDA) son trastornos hematológicos hereditarios caracterizados por anemia hiporgenerativa crónica, anomalías morfológicas de los eritroblastos y componente hemolítico. Se han definido genéticamente seis subtipos de CDA que presentan diferentes alteraciones morfológicas de los eritroblastos en la médula ósea. Una clasificación competente de los casos de CDA es crucial para proporcionar atención y tratamiento eficientes al paciente.

CoDysAn tiene como objetivo proporcionar una página web receptiva que concientice a médicos, pacientes y al público en general sobre la anemia diseritropoyética congénita (CDA) y proporcione una herramienta de diagnóstico de fácil uso. El algoritmo de diagnóstico está basado en parámetros hematológicos de los pacientes que incluye: nivel de hemoglobina, volumen corpuscular medio de eritrocitos y recuento de reticulocitos y plaquetas. Los rangos normales dependen del género y la edad. La herramienta de diagnóstico logró una satisfactoria especificidad del 89,5% y una sensibilidad del 78,5%. El algoritmo CoDysAn está conectado al Registro de Pruebas Genéticas (GTR) de NCBI para

informar a los médicos sobre la existencia de centros de diagnóstico acreditados para realizar una prueba genética completa, si es necesario.

3. BEA (BioMark Expression Analysis): una herramienta web para realizar análisis de expresión relativa de datos provenientes del sistema BioMark-Fluidigm.

Las tecnologías de alto rendimiento para el análisis de transcriptomas se utilizan de manera ubicua en la biología moderna y la investigación médica. Hoy en día, los investigadores están empleando estos métodos para analizar múltiples genes de múltiples muestras en paralelo, reduciendo significativamente el tiempo de manipulación y el coste. El sistema BioMark HD es un qPCR nanofluídico de alto rendimiento capaz de ejecutar más de nueve mil reacciones al mismo tiempo (96 muestras x 96 ensayos). La cuantificación relativa a través del método ΔΔCT y el análisis estadístico no es fácilmente abordable para usuarios con bajos conocimientos de bioinformática.

BEA es una aplicación basada en Shiny para ayudar al biólogo con conocimiento limitado de programación y estadística a realizar análisis de expresión génica a partir de datos de qPCR (reacción cuantitativa en cadena de la polimerasa) del sistema BioMark-Fluidigm. BEA fue diseñado para proporcionar un proceso de análisis de fácil uso estructurado en cómodos e intuitivos pasos, en una interfaz fácil de usar. Esto permite a los usuarios con bajos niveles de conocimientos de programación y estadística realizar análisis de expresión génica diferencial para dos o tres grupos de muestras. El análisis incluye cuantificación relativa (método ΔΔCT) y pruebas estadísticas apropiadas para identificar genes expresados diferencialmente entre los grupos probados. La aplicación BEA está disponible a través de la página web de BloodGenetics (http://bloodgenetics.com/otros/).

La herramienta bioinformática se ha probado utilizando datos de BioMark del proyecto de síndrome mielodisplásico, cuyo objetivo era identificar genes asociados con la sobrecarga de hierro mitocondrial (sideroblastos en anillo) en la anemia sideroblástica. Veintitrés genes han sido identificados que se expresan diferencialmente entre pacientes con síndrome mielodisplásico con sideroblastos en anillo (MDS-RS), pacientes con síndrome mielodisplásico sin sideroblastos en anillo (MDS-nonRS) y controles sanos.

4. *Pipeline* para la construcción de firmas de expresión génica para predicción y pronóstico utilizando datos de pacientes con cáncer de vejiga músculo-invasivo.

El avance y la expansión de la tecnología de alto rendimiento ha brindado la oportunidad de descubrir firmas genéticas específicas, que pueden ayudar a diagnosticar y clasificar una condición médica patogénica, o predecir el pronóstico o las respuestas farmacológicas a un tratamiento específico. Sin embargo, la mayoría de los fenotipos son complejos, resultantes de la influencia tanto del factor genómico como de las condiciones externas (factores no genómicos). Para comprender estos rasgos complejos y poder definir una firma de expresión génica de estos fenotipos es esencial estudiar conjuntamente la información genómica y la información no genómica. Sin embargo, la integración de estos dos tipos de datos requiere un modelado elaborado y genera algunos desafíos que se deben tener en cuenta.

En este proyecto hemos elaborado un *pipeline* estadístico basado en el modelado condicional integrativo, en el que las variables no genómicas se seleccionan primero y posteriormente se usan para seleccionar y ajustar el modelo genético. Las variables no genómicas se seleccionan por regresión *Stepwise* y las variables genómicas por regresión penalizada LASSO en un proceso de validación cruzada. El modelo estadístico se ha probado utilizando datos del proyecto de cáncer de vejiga músculo-invasivo, con el objetivo de identificar una firma de expresión génica que pueda predecir el pronóstico y la eficacia de las quimioterapias neoadyuvantes basadas en platino. El modelo

condicional integrativo formuló que tres variables no genómicas (tipo de morfología tumoral, la presencia de hidronefrosis y la clasificación taxonómica) y nueve genes (*RAD51, CXCL9, PARP, 53BP1, HERC2, ERCC2, CHEK1, Ku80* y *RNF168*) tenían la mayor capacidad predictiva (AUC = 0.663). Sin embargo, el análisis de optimización sugiere que las variables no genómicas y los primeros tres o cuatro genes seleccionados por LASSO (*RAD51, CXCL9, PARP* y *53BP1*) pueden tener suficiente capacidad predictiva para predecir la respuesta al tratamiento (AUC = 0.671).

Content

i. List of figures XVII

ii. List of tables XIX

iii. List of genes XXI

iv. List of Abbreviations XXV

GENERAL INTRODUCTION 1

INTRODUCTION: 3

Industrial doctorate 3

Whole Genix 3

Which organizations are involved in this industrial Ph.D? 4

Objectives of this industrial doctorate 4

Objectives 4

Projects 5

Refining and improving an NGS gene panel for clinical diagnosis of hereditary hematologic diseases and focus on diseases with ferritin alterations. 5

CoDysAn: a web tool for the diagnosis of Congenital Dyserythropoietic Anemia 5

BEA (BioMark Expression Analysis): a web tool to perform relative expression analysis from Fluidigm – BioMark raw data. 5

Pipeline for the construction of gene expression signatures for prediction and prognosis using data from patients with muscle-invasive bladder cancer. 5

Heterogenicity of the projects 5

CHAPTER I: Hereditary hematologic diseases Next Generation Sequencing Panel 7

1.1. INTRODUCTION 9

1.1.1. Hereditary Hematologic Diseases NGS Panel 9

1.1.1.1. Hereditary hematological diseases 9

1.1.1.2. Hereditary Hemochromatosis 9

1.1.1.3. Congenital sideroblastic Anemia and acquired sideroblastic Anemia 11

1.1.1.4. Iron and Copper Related Anemias 16

1.1.1.5. Congenital dyserythropoietic anemia (CDA) 18

1.1.1.6. Congenital Erythrocytosis / Familial Polycythemia 19

1.1.1.7. Hereditary hemolytic anemia including membranopathies and enzymopathies. 20

1.1.1.8. Congenital Erythropoietic Protoporphyria and Congenital Erythropoietic Porphyria 22

1.1.1.9. Bone marrow failure 23

1.1.1.10. Neurodegeneration with Brain Iron Accumulation Disorders 26

1.1.1.11. Hyper and hypoferritinemia 27

1.1.2. NGS and genetic diagnosis 28
1.1.3. HHD Panels 28
1.2. MATERIAL AND METHODS 31
1.2.1. Design of the new panel (version 15) 31
1.2.2. Validation of the NGS panel and diagnosis of new cases 31
1.2.2.1. Validation of the NGS panel 31
1.2.2.2. Diagnosis of patients with HHD using panel v15 31
1.2.2.3. DNA extraction 33
1.2.2.4. Library preparation and Sequencing 33
1.2.2.5. Results, Interpretation, and Reporting 33
1.2.3. External validation (EMQN-EQA) 34
1.2.4. RNA Fold Predictions for 5'UTR FTL variant 34
1.3. RESULTS 35
1.3.1. The final design (genes and diseases) 35
1.3.2. Patients analyzed for validation (Internal validation) 38
1.3.3. External Validation (EQMN – EQA) 42
1.3.4. Diagnosis 42
1.3.5. L-Ferritin: One Gene, Five Diseases. 45
1.4. DISCUSSION 53
1.4.1. Discussion regarding HHD panel 53
1.4.2. Discussion regarding *FTL* gene and diseases 56
CHAPTER II: CoDysAn: A telemedicine tool to improve awareness and diagnosis for patients with Congenital Dyserythropoietic Anemia 59
2.1. INTRODUCTION 61
2.1.1. Congenital Dyserythropoietic Anemias: subtypes and associated genes 61
2.1.2. Aim of CoDysAn 63
2.2. MATERIAL AN METHODS 67
2.2.1. Patients and validation 67
2.2.2. Design of the web server 67
2.2.3. Diagnostic algorithm for CDA 68
2.3. RESULTS 69
2.3.1. CoDysAn scope 69
2.3.2. Webpage structure and design 70
2.3.3. Implementation of the web tool 71
2.3.4. Validation 73

2.4. DISCUSSION 75

CHAPTER III: BEA (BioMark Expression Analysis): 77

3.1. INTRODUCTION 79

3.1.1. Background – transcriptomics methods 79

3.1.2. BioMark HD system (Fluidigm Technology) 80

3.1.3. BEA: BioMark Expression Analysis web tool 82

3.2. MATERIAL AND METHODS 83

3.2.1. Testing data (Myelodysplastic syndromes project) 83

3.2.1.1. MDS patients 83

3.2.1.2. Experiment design 83

3.2.1.3. Real-time qPCR using the BioMark HD System (Fluidigm) 83

3.2.1.4. Data acquisition and pre-processing 83

3.2.2. Implementation 83

3.2.2.1. Programming languages 83

3.2.2.2. Availability of data and materials 84

3.2.2.3. Normalization and statistical workflow 84

3.3. RESULTS 87

3.3.1. BEA scope 87

3.3.2. Implementation 87

3.3.2.1. Raw data preprocessing and acquisition 87

3.3.2.2. Input files 87

3.3.2.3. Tabs 88

3.3.3. Results from the testing data (Myelodysplastic syndromes project) 91

3.3.4. Error correction and running validation of BEA with external datasets 97

3.4. DISCUSSION 99

CHAPTER IV: Pipeline for construction of gene expression signatures for prediction and prognosis. 103

4.1. INTRODUCTION 105

4.1.1. Gene expression signatures in biomedicine 105

4.1.2. Genetic signatures in complex traits and the challenge of integration of genomic and non-genomic data. 105

4.1.3. Statistical strategies for variable selection 107

4.1.4. Bladder cancer study 108

4.1.5. MIBC Project 109

4.2. MATERIAL AND METHODS 111

4.2.1. Data from the MIBC Project 111
4.2.2. Statistical strategy 111
4.3. RESULTS 115
4.3.1. Pipeline to construct genetic signatures integrating genomic and non-genomic data 115
4.3.2. Validation of the pipeline with bladder cancer data 117
4.3.2.1. Filtering of samples and genes. 117
4.3.2.2. Descriptive analysis of the clinical and pathological variables of samples. 117
4.3.2.3. Clinical, pathological and taxonomical modeling and construction of the genetic signature. 118
4.3.2.4. The increasing predictive ability 120
4.3.2.5. Survival analysis 121
4.4. DISCUSSION 127
4.4.1. Pipeline and genetic signatures 127
4.4.2. MIBC project 128
4.4.3. Conditional integrative strategy 128
FINAL DISCUSSION 131
General discussion of all projects 133
CONCLUSIONS 139
BIBLIOGRAPHY 145
APPENDIX 169
1. Supplementary Material 169
2. National and International congress 187
2.1. Congress presentations 187
2.2. Attendance to Courses and congress 188
3. Publications 191
Cadenas et al. 2019 Pharmaceuticals 193
Tornador et al. 2019 Frontiers in Physiology 209
4. Other documents 215
Certificate of participation in 2018 EMQN external quality assessment for NGS and Sanger sequencing. 216
Example of Clinical report 217
Instruction of use of BEA program 221
Example of report performed by BEA program 237
ACKNOLEDGEMENTS 251

i. List of figures

GENERAL INTRODUCTION

Figure 0. 1. Institutions. 4

CHAPTER I:

Figure 1.1 Proteins involved in the regulation of hepcidin and in the development of Hereditary Hemochromatosis (HH).. 11
Figure 1.2. Scheme of the pathways and genes involved in CSA in erythroid cells.. 15
Figure 1. 3 Scheme of the genes associated to the pathogenesis of NBIA. 26
Figure 1.4. Example of two genes (ALAS2 and ALDOA) completely covered by Haloplex probes....36
Figure 1.5. Example of genes (HBA2 and HBB) not covered by Haloplex probes. 37
Figure 1.6. Region of VHL gene (exon 2) not covered by the HHD panel.....41
Figure 1.7. Proportion of patients analyzed in the HHD Panel (version 15) by group of disease (subpanel). ..42
Figure 1.8. Genetic results of the patients analyzed by the HHD panel (v15).. 43
Figure 1.9. Pedigree of the three analyzed families affected from dominant L-ferritin deficiecy and HHCS. Squares indicate males and circles females.. 45
Figure 1.10. Schematic localization of literature reported and new FTL mutations.. 48
Figure 1.11. WT and mutated FTL-IRE fold prediction.. 49
Figure 1.12. Algorithm for diagnosis of diseases caused by defects in FTL gene. 58

CHAPTER II:

Figure 2. 1 Pathogenic mechanism of CDA. CDA is caused by ineffective erythropoiesis during differentiation and proliferation stages, and consequently a faulty production of red blood cells.. 64
Figure 2. 2 Algorithm for differential diagnosis of Congenital Dyserythropoietic Anemia (CDAs). The algorithm is structured in five steps.. 69
Figure 2. 3 Home page from CoDysAn website. 70
Figure 2. 4 Step 1 of CoDysAn diagnostic algorithm: evaluation of anemia. 71
Figure 2. 5 Step 2 of CoDysAn diagnostic algorithm: evaluation of other hematological parameters....72
Figure 2. 6 Step 3 of CoDysAn diagnostic algorithm: exclusion of other diseases with similar manifestations. 72
Figure 2. 7 Step 4 of CoDysAn diagnostic algorithm: evaluation of other clinical features. 73
Figure 2. 8 Results window of CoDySan.. 73

CHAPTER III:

Figure 3. 1 Fluidigm BioMark system workflow. 81
Figure 3. 2 Statistical workflow of BEA program 85
Figure 3. 3 Inputs files required to start the gene expression analysis with BEA. 88
Figure 3. 4 Conceptual architecture of BEA software structured in eight tabs. 88
Figure 3. 5 First tab from BEA program (File Manager tab). 88
Figure 3. 6 Heatmap of gene expression (Ct) of samples with RS-MDS, nonRS-MDS and healthy controls for 96 genes.. 92
Figure 3. 7 Barplot of number of failed reactions per gene. 93
Figure 3. 8 Barplot of number of failed reactions per sample. 93
Figure 3. 9 Boxplot of gene expression of significative genes in the three tested groups.. 96
Figure 3. 10 Heatmap of the 23 significative genes. 97

CHAPTER IV:

Figure 4. 1 Regression techniques according the dependent variable 107
Figure 4. 2. Design of the models to evaluate the clinical and genomic signature from patients with MIBC. 112
Figure 4. 3 Pipeline for construction of genetic signatures by integrating genetic and non-genetic data. 116
Figure 4. 4 Boxplot of the logarithm of the gene expression of the genes selected by Lasso. 120
Figure 4. 5 Increase of the predictive ability (AUC) of pathological response by adding variables to the model. 121
Figure 4. 6 Kaplan-Meier curves for Progression Free survival (PFS) comparison of patients with favorable (in yellow) and unfavorable (in blue) genetic signature for the models. 123
Figure 4. 7 Kaplan-Meier curves for Overall survival (OS) comparison of patients with favorable (in yellow) and unfavorable (in blue) genetic signature for the models. 124
Figure 4. 8 Kaplan-Meier curves for Disease Specific survival (DSS) comparison of patients with favorable (in yellow) and unfavorable (in blue) genetic signature for the models. 125
Figure 4. 9 Kaplan-Meier curves for Event Free survival (EFS) comparison of patients with favorable (in yellow) and unfavorable (in blue) genetic signature for the models. 126

SUPPLEMENTARY MATERIALS:

Supplementary Material Figure S1. Algorithm for differential diagnosis of Congenital Dyserythropoietic Anemia (CDA). (Chapter II) 184

ii. List of tables

CHAPTER I:

Table 1. 1 Hereditary hematological disorders....9
Table 1. 2 Hereditary hemochromatosis sub-types....10
Table 1. 3 Genetic classification of Sideroblastic anemias....16
Table 1. 4 Genetic classification of Iron and Copper deficiency Anemia....17
Table 1. 5 Classification of CDAs....19
Table 1. 6 Familial Erythrocytosis Subtypes....20
Table 1. 7 Membranopathies subtypes....21
Table 1. 8 Enzymopathy subtypes....22
Table 1. 9 Congenital Erythropoietic Porphyria subtypes....23
Table 1. 10 Genetic classification of Fanconi Anemia....24
Table 1. 11 Genetic classification of DBA....25
Table 1. 12 Genetic classification of Dyskeratosis congenita....25
Table 1. 13 Genetic classification of NBIA....27
Table 1. 14 Comparison of the four different versions of the HHD panel designed. 30
Table 1. 15 Genes included in the version 15 of the NGS target panel for the diagnosis of HHD. 32
Table 1. 16 Coverage of the HHD panel 15 respect the total of known causative genes for the time being....38
Table 1. 17 Samples used for validation through NGS panel version 15 including additional findings....39
Table 1. 18 Validation results....41
Table 1. 19 Summary of the genetic diagnosis by HHD Panel (version 15) 44
Table 1. 20 Genetic and clinical features of the probands. 46
Table 1. 21 Summary of Diseases caused by mutations in FTL gene 51

CHAPTER II:

Table 2. 1 Classification of CDAs 65
Table 2. 2 Samples used to test and validate the CoDysAn diagnostic algorithm. 67
Table 2. 3 Normal ranges for hematological values used by the diagnostic CoDysAn algorithm. 68

CHAPTER III:

Table 3. 1 Dynamic Array IFC subtypes 81
Table 3. 2 Packages installed in R for the development of BEA 84
Table 3. 3 Dataset used to validate the program 85
Table 3. 4 Experimental information…. ….....91
Table 3. 5 Reference genes …….. 91
Table 3. 6 Groups of analysis 91
Table 3. 7 Selection of groups for analysis 94
Table 3. 8 Description of reference genes 94
Table 3. 9 Differentially expressed genes between tested groups....95

CHAPTER IV:

Table 4. 1 Increasing gene models for survival analysis....113
Table 4. 2 Genes discarded for the analysis 117
Table 4. 3 Description of the clinical, pathological and taxonomic variables....117
Table 4. 4 Stepwise regression coefficients of Model 1 (clinical-pathological model) 118
Table 4. 5 Stepwise regression coefficients of Model 2 (clinical-pathological and taxonomic model) 118
Table 4. 6 Predictive ability (AUC) of the four models....119

Table 4. 7 Variable selection for Lasso in Model 4.....119
Table 4. 8 Models studied in the survival analysis.....121

SUPPLEMENTARY MATERIALS:

Supplementary Material Table S1. Summary of hereditary hematological disorders covered by HHD panel version 15. (Chapter I) 170
Supplementary Material Table S2. Summary of FTL identified mutations. (Chapter I) 177
Supplementary Material Table S3. DNA sequence for FTL RNA fold predictions. (Chapter I).....181
Supplementary Material Table S4. Biomarkers candidates to predict response to cisplatin treatment (Chapter IV). 182
Supplementary Material Table S5. Univariant analysis (simple logistic regression) of all genes (Chapter IV) 183

iii. List of genes

Gene name	Gene description
ABCB6	ATP binding cassette subfamily B member 6 (Langereis blood group)
ABCB7	ATP binding cassette subfamily B member 7
ABCG5	ATP binding cassette subfamily G member 5
ABCG8	ATP binding cassette subfamily G member 8
ACD	ACD
ACO1	aconitase 1
ACVR1	activin A receptor type 1
ADA	adenosine deaminase
AK1	adenylate kinase 1
ALAS2	5'-aminolevulinate synthase 2
ALDOA	aldolase
ANK1	ankyrin 1
ATP13A2	ATPase cation transporting 13A2
ATP4A	ATPase H+/K+ transporting subunit alpha
ATP7B	ATPase copper transporting beta
ATPIF1	ATP synthase inhibitory factor subunit 1
BMP6	bone morphogenetic protein 6
BPGM	bisphosphoglycerate mutase
BRCA1	BRCA1
BRCA2	BRCA2
BRIP1	BRCA1 interacting protein C-terminal helicase 1
C15orf41	chromosome 15 open reading frame 41
C19orf12	chromosome 19 open reading frame 12
CDAN1	codanin 1
CLPX	caseinolytic mitochondrial matrix peptidase chaperone subunit
COASY	Coenzyme A synthase
CP	ceruloplasmin
CTC1	CST telomere replication complex component 1
CYB5R3	cytochrome b5 reductase 3
DCAF17	DDB1 and CUL4 associated factor 17
DKC1	dyskerin pseudouridine synthase 1
EGLN1	egl-9 family hypoxia inducible factor 1
EPAS1	endothelial PAS domain protein 1
EPB41	erythrocyte membrane protein band 4.1
EPB42	erythrocyte membrane protein band 4.2
EPOR	erythropoietin receptor
ERCC4	ERCC excision repair 4
EXOC6	exocyst complex component 6
FA2H	fatty acid 2-hydroxylase
FANCA	FA complementation group A

FANCB	FA complementation group B
FANCC	FA complementation group C
FANCD2	FA complementation group D2
FANCE	FA complementation group E
FANCF	FA complementation group F
FANCG	FA complementation group G
FANCI	FA complementation group I
FANCL	FA complementation group L
FANCM	FA complementation group M
FECH	ferrochelatase
FTH1	ferritin heavy chain 1
FTL	ferritin light chain
G6PD	glucose-6-phosphate dehydrogenase
GATA1	GATA binding protein 1
GCLC	glutamate-cysteine ligase catalytic subunit
GLRX5	glutaredoxin 5
GNPAT	glyceronephosphate O-acyltransferase
GPI	glucose-6-phosphate isomerase
GPX1	glutathione peroxidase 1
GSR	glutathione-disulfide reductase
GSS	glutathione synthetase
GYPC	glycophorin C (Gerbich blood group)
HAMP	hepcidin antimicrobial peptide
HEPH	hephaestin
HFE	homeostatic iron regulator
HFE2	also known as HJV (nomenclature is going to be officially change soon)
HK1	hexokinase 1
HSPA9	heat shock protein family A (Hsp70) member 9
IREB2	iron responsive element binding protein 2
JAK2	Janus kinase 2
KCNN4	potassium calcium-activated channel subfamily N member 4
KIF23	kinesin family member 23
KLF1	Kruppel like factor 1
LARS2	leucyl-tRNA synthetase 2
MAD2L2	mitotic arrest deficient 2 like 2
mtATP6	mitochondrially encoded ATP synthase 6
NCOA4	nuclear receptor coactivator 4
NDUFB11	NADH:ubiquinone oxidoreductase subunit B11
NHP2	NHP2 ribonucleoprotein
NOP10	NOP10 ribonucleoprotein
NT5C3A	5'-nucleotidase
PALB2	partner and localizer of BRCA2
PANK2	pantothenate kinase 2
PARN	poly(A)-specific ribonuclease
PCBP1	poly(rC) binding protein 1

PFKM	phosphofructokinase
PGK1	phosphoglycerate kinase 1
PIEZO1	piezo type mechanosensitive ion channel component 1
PIM1	Pim-1 proto-oncogene
PKLR	pyruvate kinase L/R
PLA2G6	phospholipase A2 group VI
PUS1	pseudouridine synthase 1
RAD50	RAD50 double strand break repair protein
RAD51C	RAD51 paralog C
RFWD3	ring finger and WD repeat domain 3
RHAG	Rh associated glycoprotein
RPL11	ribosomal protein L11
RPL15	ribosomal protein L15
RPL19	ribosomal protein L19
RPL26	ribosomal protein L26
RPL27	ribosomal protein L27
RPL31	ribosomal protein L31
RPL35A	ribosomal protein L35a
RPL36	ribosomal protein L36
RPL5	ribosomal protein L5
RPL9	ribosomal protein L9
RPS10	ribosomal protein S10
RPS15	ribosomal protein S15
RPS17	ribosomal protein S17
RPS19	ribosomal protein S19
RPS24	ribosomal protein S24
RPS26	ribosomal protein S26
RPS27	ribosomal protein S27
RPS27A	ribosomal protein S27a
RPS28	ribosomal protein S28
RPS29	ribosomal protein S29
RPS7	ribosomal protein S7
RTEL1	regulator of telomere elongation helicase 1
SEC23B	Sec23 homolog B
SF3B1	splicing factor 3b subunit 1
SH2B3	SH2B adaptor protein 3
SLC11A2	solute carrier family 11 member 2
SLC19A2	solute carrier family 19 member 2
SLC25A37	solute carrier family 25 member 37
SLC25A38	solute carrier family 25 member 38
SLC25A39	solute carrier family 25 member 39
SLC2A1	solute carrier family 2 member 1
SLC40A1	solute carrier family 40 member 1
SLC4A1	solute carrier family 4 member 1 (Diego blood group)
SLX4	SLX4 structure-specific endonuclease subunit

SPTA1	spectrin alpha
SPTB	spectrin beta
STEAP3	STEAP3 metalloreductase
TERC	telomerase RNA component
TERT	telomerase reverse transcriptase
TF	transferrin
TFR2	transferrin receptor 2
TFRC	transferrin receptor
TINF2	TERF1 interacting nuclear factor 2
TMPRSS6	transmembrane serine protease 6
TPI1	triosephosphate isomerase 1
TRNT1	tRNA nucleotidyl transferase 1
TSR2	TSR2
UBE2T	ubiquitin conjugating enzyme E2 T
UGT1A1	UDP glucuronosyltransferase family 1 member A1
UROS	uroporphyrinogen III synthase
USB1	U6 snRNA biogenesis phosphodiesterase 1
VHL	von Hippel-Lindau tumor suppressor
WDR45	WD repeat domain 45
WRAP53	WD repeat containing antisense to TP53
XK	X-linked Kx blood group
XRCC2	X-ray repair cross complementing 2
YARS2	tyrosyl-tRNA synthetase 2

iv. List of Abbreviations

Abbreviation	Meaning
ΔΔCt	Delta delta Ct method
ACMG	American College of Medical Genetics
AD	Autosomal dominant inheritance
AIC	Akaike information criteria
AML	Acute myelogenous leukemia
AR	Autosomal recessive inheritance
AUC	Area Under the Curve
BEA	BioMark Expression Analysis
BM	Bone marrow
bp	Base pair
CDA	Congenital dyserythropoietic anemia
cDNA	Complementary DNA
CE	Congenital Erythrocytosis
CEP	Congenital Erythropoietic Porphyria
CMV	Cisplatin, methotrexate vinblastine chemotherapy
CNSHAs	Congenital nonspherocytic hemolytic anemia
CNVs	Copy number variants
CoDysAn	Congenital Dyseroproyetic Anemia project
CSA	Congenital sideroblastic Anemia
csv	Comma separated value (file format)
Ct	Cycle threshold
DBA	Diamond-Blackfan anemia
DC	Dyskeratosis congenital
DHS1	Dehydrated hereditary stomatocytosis type 1
DHS2	Dehydrated hereditary stomatocytosis type 2
DNA	Deoxyribonucleic acid
ds-cDNA	Double-stranded cDNA
DSS	Disease-specific survival
EFS	Event-free survival
EMA	European Medicines Agency
ENET	Elastic Net
EPP	Erythropoietic Protoporphyria
ExAC	Exome Aggregation Consortium
FA	Fanconi anemia
FDA	Food and Drug Administration
fdr	False discovery rate
FFPE	Formalin-Fixed Paraffin-Embedded
FTM	Fundación Teresa Moretó (Teresa Moretó Fundation)
GTR	NCBI's Genetic Testing Registry
HA	Hemolytic Anemia

HB	Hemoglobin
HGSV	The Human Genome Structural Variation Consortium
HGTiP	Hospital Germans Trias i Pujol
HH	Hereditary Hemochromatosis
HHCS	Hereditary Hyperferritinemia Cataract Syndrome
HHD	Hereditary Hematological Diseases
IFC	Integrated fluidic circuits
IRE	Iron-responsive element
IRIDA	Iron-refractory Iron-Deficiency Anemia
IRP	Iron regulatory protein
LASSO	Least Absolute Shrinkage and Selection Operator
MAF	Minor allele frequency
MCV	Mean corpuscular volume
MDS	Myelodysplastic syndrome
MDS-nonRS	Myelodysplastic syndrome without ring sideroblasts development
MDS-RS	Myelodysplastic syndrome with ring sideroblasts development
MDS-RS-MLD	MDS with multilineage dysplasia and ring sideroblasts
MDS-RS-SLD	MDS with single linage dysplasia and ring sideroblasts
MIBC	Muscle invasive bladder cancer
MLASA	Mitochondrial myopathy with lactic acidosis and sideroblastic anemia
MLPA	Multiplex ligation-dependent probe amplification
MR	Magnetic resonance
mtDNA	Mitochondrial DNA
Mut	Mutation
NA	Not available
NAC	No Amplification Control
NBIA	Neurodegeneration with brain Iron Accumulation
NGS	Next Generation Sequencing
NMD	Nonsense mediated decay
nonRS	Myelodysplastic syndrome without ring sideroblasts development
NTC	No Template Control
OMIM	Online Mendelian Inheritance in Man
OS	Overall survival
PCR	Polymerase Chain Reaction
PFS	Progression-free survival
PMPS	Pearson marrow-pancreas syndrome
qPCR	Quantitative Polymerase Chain Reaction
R&D&I	Research and Development and Innovation
REC	Research Ethics Committees
RIN	RNA Integrity Number
RLS	Restless leg syndrome
RNA	Ribonucleic acid
ROC	Ceiver Operating Characteristics)
RS	Myelodysplastic syndrome with ring sideroblasts development
RV	Regular Values

SIF	Sample information file
SIFD	Sideroblastic anemia, B cell immunodeficiency, fevers and developmental delay
SMS	Single molecule sequencing
SNVs	Single nucleotide variants
TGS	Third Generation sequencing
TNM	Malignant tumors classification
TRMA	Thiamine-responsive megaloblastic anemia
tRNA	Transfer RNA
UTR	Untranslated region
VUS	Variants of unknown clinical significance
WES	Whole Exome Sequencing
WG	Whole Genix
WT	Wild-type
XL	X-linked inheritance
XLR	X-link recessive
XLSA	X-linked sideroblastic anemia
XLSA/A	Sideroblastic microcyst anemia with X-linked ataxia

GENERAL INTRODUCTION

INTRODUCTION:

Industrial doctorate

Recent social, economic and cultural changes have brought the increase of the "knowledge-based economy." Research & Development and Innovation (R&D&I) are fundamental aspects for the growth and evolution of a company. Innovation is fast becoming a critical factor in the long-term growth development and survival of companies. Therefore, in order to help solve the development and innovation challenges that the current environment demands, companies require employees with excellent training that respond to business needs. This situation remarks the need for close cooperation between education institutions and companies and the introduction of the industrial doctorate concept.

An industrial doctorate is a process where a doctoral student develops a research project carried out at a company in collaboration with a university or research center. The research project is based on an in-depth investigation into a particular and relevant problem for the company, which is supervised by experts from universities or research institutions. Therefore, the doctoral thesis serves to bridge the gap between industry and academia.

The benefits of the "hybrid" industrial doctoral degree respect traditional PhDs are:

- The doctoral research is conducted in the workplace, so the investigation is put into practice in the industrial environment.
- The company will be able to gain valuable academic knowledge that could be industrialized.

Whole Genix

This industrial Ph.D. program has been carried out in Whole Genix company. Whole Genix (WG) is a company dedicated to providing genetic analyses and to help hospitals and clinics in the diagnosis of different diseases. WG encourages to the improvement of people's health and quality of life utilizing the latest biomedical techniques.

WG offers specialized counseling in clinical genetics to doctors and hospitals regarding new forms of prevention, early diagnosis and the effective treatment of diseases, especially in the field of neoplastic processes. The company is a pioneer in translational medicine and precision medicine on neoplastic diseases. One of the revolutionary services is the analysis of genetic markers, resulting from the study of complete exomes (Whole Exome Sequencing, WES) from clinic Formalin-Fixed Paraffin-Embedded samples (FFPE), selecting those genetic targets that provide the oncologist with the necessary information to apply the most effective personalized medicine to a given patient.

WG offers services not only related to neoplastic diseases but also in other disorders. These services include:

- Development of custom next-generation sequencing panels: WG offers specialized counseling services to doctors and hospitals for the design, development, and production of specified genetic panels for sequencing.
- Specialized analysis for the diagnosis or prevention of the disease: In order to set a specific diagnosis of the disease, WG performs an in-depth analysis of the data obtained from high-throughput technology employing bioinformatics tools and protocols.

- Interpretation of genomic sequencing data: We accomplish a selection and interpretation of clinically significant genetic data (genetic and structural variants, expression change), showing the genetic profile of patients in the framework of approved personalized therapies (FDA, EMA), or in the final phases of clinical trials. Finally, a concise clinical reported is delivered, supervised by experts in the field of clinical genomics and oncology.

Which organizations are involved in this industrial Ph.D?

This Industrial Ph.D. is carried out in Whole Genix company, in collaboration with two academic organizations: Josep Carreras Research Institute and the University of Vic - Central University of Catalonia. Part of the Ph.D. project has been performed with the participation of another company, BloodGenetics, which gave scientific guidance; and with the support of the non-profit entity Teresa Moreto Foundation (FTM), which provided laboratory facilities.

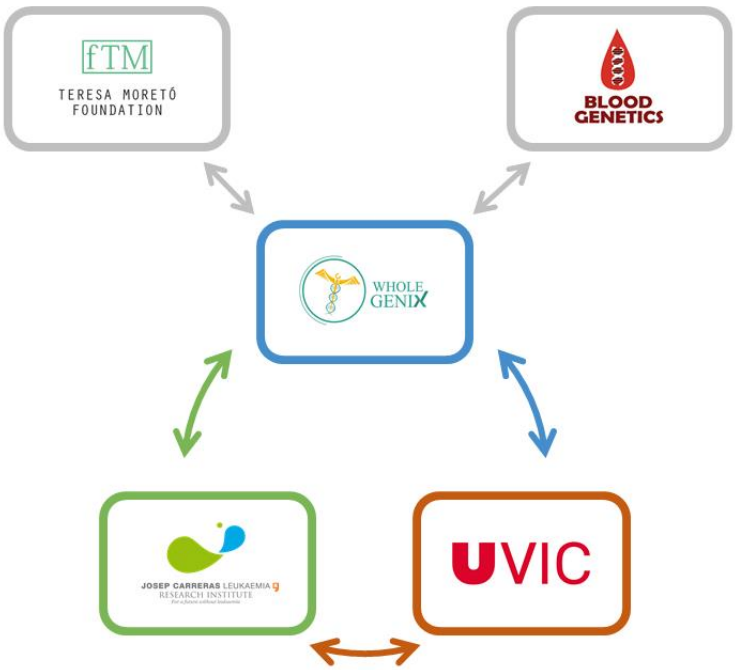


Figure 0. 1. Institutions. Main company: Whole Genix; academic centers: Josep Carreras Research Institute and the University of Vic; other collaborators: BloodGenetics and Teresa Moreto Foundation

Objectives of this industrial doctorate.

Objectives

The industrial research project for Whole Genix is formulated in four ideas:

1. **Acquire better know-how of Next Generation Sequencing (NGS) protocols.** NGS technology is growing up faster, and WG needs to be updated to last technology and protocols.
2. **Cover different types of cancer and expand to other genetic diseases.** WG is specialized in neoplastic diseases.
3. **Develop new bioinformatic and telemedicine tools for analysis of different omics data.** Computational science is the base of this biotech company, and the development of new computational tools and bioinformatics pipelines are crucial for the sustainability of WG.

4. **Progress towards personalized medicine.** WG wants to endorse genetics and genomics in research and integrate this knowledge into clinical medicine to help accomplish the promise of personalized medicine.

Projects

In order to accomplish the WG objectives, four projects have been carried out:

Refining and improving an NGS gene panel for clinical diagnosis of hereditary hematologic diseases and focus on diseases with ferritin alterations.

The goal of this project is to improve an already established NGS-gene panel for rare inherited blood diseases in order to accelerate diagnostics. Moreover, additional research in the *FTL* gene has been carried out due to its correlation with five hematological disorders. (See page 7)

CoDysAn: a web tool for the diagnosis of Congenital Dyserythropoietic Anemia

CoDysAn aims to provide a responsive webpage that provides practical information about congenital dyserythropoietic anemia (CDA) awareness and a step-by-step diagnostic tool. (See page 59)

BEA (BioMark Expression Analysis): a web tool to perform relative expression analysis from Fluidigm – BioMark raw data.

By creating BEA program, the objective is to provide a web tool that helps scientists to analyze qPCR data generated from one high-throughput technology: Fluidigm – BioMark system. The analysis includes relative quantification (ΔΔCT method) and appropriated statistical testing to identify differentially expressed genes between tested groups. (See page 77)

Pipeline for the construction of gene expression signatures for prediction and prognosis using data from patients with muscle-invasive bladder cancer.

In this project, the goal is to build the optimal statistical analysis, to find a clinical-molecular signature in patients with muscle-invasive bladder cancer treated with neoadjuvant chemotherapy, in order to predict the prognosis and efficacy of platinum-based chemotherapies. (See page 103)

Heterogenicity of the projects

The four projects described before cover the business needs of Whole Genix, for what the industrial doctorate was developed. (1) A better knowledge of NGS protocols will be acquired by developing the NGS panel for hereditary hematological diseases. (2) In order to cover other types of cancer and also expand the service to other diseases, we will study samples from bladder cancer, and from different hematological rare diseases. (3) CoDysAn and BEA tools will help to improve new skills in new bioinformatic tool development. (4) Overall, all projects help to acquire more knowledge of different genetic diseases and go forward to a closer development of personalized medicine.

CHAPTER I:

Hereditary Hematologic Diseases Next Generation Sequencing Panel

1.1. INTRODUCTION

1.1.1. Hereditary Hematologic Diseases NGS Panel

In order to achieve the first goal of the thesis, I contributed in the development of NGS targeted gene panel in Hereditary Hematologic Diseases (HDD).

1.1.1.1. Hereditary hematological diseases

One of the largest organs in the human body is blood, a liquid tissue containing several types of specialized cells needed for the correct performance of the human body. A wide variety of blood disorders can appear when one or more of these cell types are damaged. There is a broad spectrum of hematology diseases that are divided into oncological and non-oncological. Hereditary hematological diseases (HHD) are non-oncological disorders that are transmitted from parents to offspring (De Falco et al. 2014; 2013; Caroline Kannengiesser et al. 2011; Sara Luscieti et al. 2013). These diseases are also called benign or non-malignant hematological diseases and cover a broad spectrum of disorders (Table 1.1 and Supplementary Material Table S1).

Table 1. 1 Hereditary hematological disorders

Studied hereditary hematological disorders (HHD) groups		
1	**Hereditary hemochromatosis**	
2	**Congenital sideroblastic anemia**	
3	**Iron Related Anemia and Cooper disease**	Atranferrinemia Aceruloplasminemia Microcytic and hypochromic anemia with iron overload Iron refractory Iron Deficiency Anemia Wilson disease
4	**Congenital dyserythropoietic anemia**	
5	**Congenital Erythrocytosis / Familial Polycythemia**	
6	**Hereditary hemolytic anemia**	Membranopathies Enzymopathies
7	**Congenital Erythropoietic Protoporphyria and Congenital Erythropoietic Porphyria**	
8	**Bone marrow failure**	Fanconi anemia Diamond-Blackfan Anaemia Dyskeratosis congenital
9	**Neurodegeneration with Brain Iron Accumulation Disorders**	
10	**Hyper- and Hypoferritinemia**	

1.1.1.2. Hereditary Hemochromatosis

Hereditary hemochromatosis (HH) is a group of genetic disorder characterized by excessive accumulation of iron in vital organs, due to the disruption of the iron regulation in the body. HH has an estimated prevalence of up to 1 in 100 individuals in northern European populations (Merryweather-Clarke et al. 2000). Iron overload in HH is associated with a variety of genetic conditions. Of these, HFE hemochromatosis (Hemochromatosis Type 1) is by far the most frequent inherited cause (85-90% of affected patients) (European Association for the Study of the Liver 2010)

due to a homozygous missense mutation in *HFE* gene: C282Y (rs1800562). HFE protein is responsible for regulating hepcidin, the primary iron regulatory hormone. When iron is high in the system, hepcidin is secreted by the hepatocytes, decreasing intestinal iron absorption in the enterocytes and decreasing iron release by macrophages. Hepcidin maintains iron levels in a physiologic interval. When the C282Y variant is present, the expression of hepcidin is not stimulated in response to excess iron, resulting in an inappropriate iron loading in the body (Crownover et al. 2013).

Persons with HH usually are clinically asymptomatic, especially in the early stages. Women have less severe disease manifestation due to the menstruation (24-fold less risk of iron-overload respect men) (Allen et al. 2009). When symptoms are presents, the most common are weakness, lethargy, erectile dysfunction and arthralgias (McDonnell et al. 1999).

Beyond the HFE C282Y mutation, other rare forms of genetic iron overload have been identified, presenting pathogenic mutations in other genes. Up to now, five hereditary hemochromatoses have been described discerning on the molecular biology aspects (Table 1.2 and Figure 1.1).

Hereditary hemochromatosis Type 1 (OMIM #235200) is the classical *HFE* gene mutation already described.

Hereditary hemochromatosis Type 2 or juvenile hemochromatosis is caused by homozygous or compound heterozygous mutation on *HFE2* gene (type 2A; HFE2A; OMIM #602390) (Papanikolaou et al. 2004) or homozygous mutation in the *HAMP* gene (type 2B; HFE2B; OMIM #613313)(Roetto et al. 2003). Manifestation appears before 30 years of age.

Hereditary hemochromatosis Type 3 (HFE3; OMIM #604720) is caused by homozygous or compound heterozygous mutation in the transferrin receptor-2 gene (*TFR2*)(Camaschella, Roetto, et al. 2000).

Hereditary hemochromatosis Type 4 (HFE4; OMIM #606069) is an autosomal-dominant condition caused by heterozygous mutation in the *SLC40A1* (ferroportin) gene (Njajou et al. 2001).

Hereditary hemochromatosis Type 5 (HFE5; OMIM #615517) is caused by heterozygous mutation in the *FTH1* gene – Autosomal-dominant inheritance. Only one Japanese family have been reported (J. Kato et al. 2001).

Table 1. 2 Hereditary hemochromatosis sub-types

Type	OMIM	Inheritance	Genes
Hereditary hemochromatosis Type 1	235200	AR	*HFE*
Hereditary hemochromatosis Type 2A	602390	AR	*HFE2*
Hereditary hemochromatosis Type 2B	613313	AR	*HAMP*
Hereditary hemochromatosis Type 3	604250	AR	*TFR2*
Hereditary hemochromatosis Type 4	606069	AD	*SLC40A1*
Hereditary hemochromatosis Type 5	615517	AD	*FTH1*

HFE, Hereditary hemochromatosis; AR, autosomal recessive, AD, autosomal dominant; OMIM, Online Mendelian Inheritance in Man

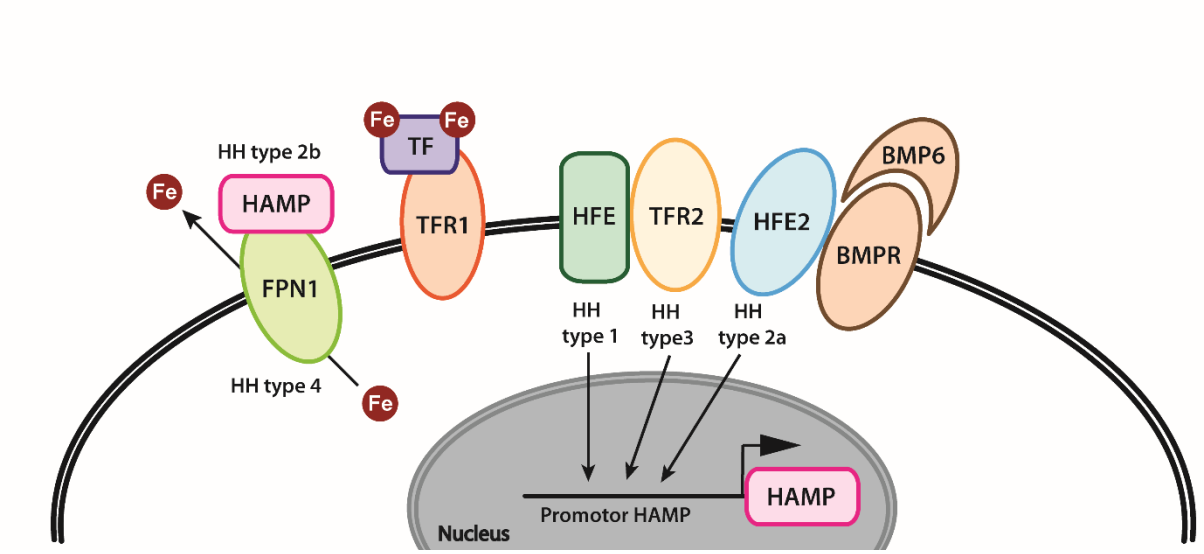


Figure 1.1 Proteins involved in the regulation of hepcidin and the development of Hereditary Hemochromatosis (HH). HFE is decoupled from the transferrin receptor 1 (TFR1) when there are high levels of transferrin saturated with iron (TF-2Fe^{3+}) and interact with the transferrin receptor 2 (TFR2) to activate the transcription of the HAMP gene that encodes to hepcidin. The TFR2-HFE complex also interacts with hemojuvelin (HJV or HFE2), the co-receptor of BMP6 (bone morphogenic protein6), to activate the hepcidin transcript. The hepcidin is produced and secreted by the hepatocytes, and interacts with the iron-regulated transported ferroportin-1 (FPN1) inhibiting its primary function: to prevent iron to be released into the circulation and decrease iron-plasma concentration. In hemochromatosis 1,2 and 3 the signaling pathway of hepcidin production is affected since that its expression is reduced, which leads to an over-absorption of intestinal iron, and an iron overload in the organism. Figure adapted from chapter 30 book *Eritropatología* author Dr. Mayka Sanchez (Sanchez and Altés 2017).

1.1.1.3. Congenital sideroblastic Anemia and acquired sideroblastic Anemia

Sideroblastic anemias are anemic disorders characterized by the presence of ring sideroblasts in the bone marrow aspirate smear. Ring sideroblasts are pathologic iron deposits in erythroblast mitochondria (Patnaik and Tefferi 2017). These disorders are typically divided into two groups, congenital sideroblastic anemia (CSA), and acquired sideroblastic anemia (non-hereditary), which include the well-defined subtype Myelodysplastic syndrome (MDS) with ring sideroblasts: MDS with single lineage dysplasia and ring sideroblasts (MDS-RS-SLD), and MDS with multilineage dysplasia and ring sideroblasts (MDS-RS-MLD). CSA can be classified into four types depending on the affected molecular pathway: heme biosynthesis, iron-sulfur (Fe-S) cluster biosynthesis, mitochondrial protein synthesis, and t-RNA biosynthesis (Table 1.3 and Figure 1.2) (Bottomley and Fleming 2014; Ducamp and Fleming 2019).

CSA caused by defects in heme biosynthesis

Heme is the principal element of hemoglobin, composed by a porphyrin ring with an atom of iron. One heme molecule binds one molecule of oxygen. Several enzymes are involved in the heme biosynthesis. Impairment of any of these enzymes could cause anomaly heme synthesis and affect iron homeostasis, leading to iron overload in mitochondrial erythroblast and the clinical manifestation of sideroblastic anemia (Bottomley 2006; Fontenay et al. 2006). Hitherto, three genes encoding enzymes that participate in the heme biosynthesis have been identified to cause this type of sideroblastic anemia: *ALAS2*, *SLC25A38,* and *SLC19A2*.

X-linked sideroblastic anemia due to ALAS2 mutation

X-linked sideroblastic anemia (XLSA; OMIM #300751) is the most prevalent form of CSA (around 40% of cases (Bergmann et al. 2009), and it is caused by mutations in *ALAS2* gene (Harigae and Furuyama 2010). This gene is located at Xp11.21, so it is developed with X-linked inheritance, and the incidence

of the disease is much higher in men than in women. To date, more than eighty different mutations in *ALAS2* have been reported, including missense, nonsense and frameshift mutation distributed from exon 5 to exon 11 and in regulatory regions (Ducamp and Fleming 2019; Long et al. 2018). Mutation in this gene leads to ALAS2 enzyme deficiency, disturbing the heme synthesis which results in iron overload in mitochondrial erythrocyte precursor and anemia.

SLC25A38 deficiency

The second most prevalent nonsyndromic CSA is caused by a mutation in the *SLC25A38* gene (OMIM #205950), located on chromosome 3p22.1 This gene encodes an erythroid-specific protein of the inner mitochondrial membrane, and is involved in the importation of glycine into mitochondria, which is essential for the ALA synthesis and the generation of heme. Defects of the mitochondrial SLC25A38 transporter causes severe hypochromic microcytic anemia with the accumulation of iron in the mitochondria of erythroblasts, and the formation of ring sideroblasts. This disease is inherited in autosomal recessive manner.

Thiamine-responsive megaloblastic anemia due to SLC19A2 mutation

Thiamine-responsive megaloblastic anemia (TRMA) is a form of syndromic anemia (OMIM #249270), presenting other symptoms including diabetes and sensory deafness (Ricketts et al. 2006; Beshlawi et al. 2014).

More than 30 mutations in *SLC19A2* gene have been identified to cause this condition in autosomal-recessive inheritance. *SLC19A2* gene is located on chromosome 1q23.3 and encodes high-affinity thiamin transporter, which is responsive to the thiamine supplementation. The complete mechanism of pathogenesis of mutations in *SLC19A2* is still unclear, but seems to lead to thiamin deficiency, and affect the generation of succinyl-CoA and consequently, a defective synthesis of heme (Labay et al. 1999; Setoodeh et al. 2013).

CSA caused by defects in iron-sulfur cluster biosynthesis

Iron-sulfur (Fe-S) clusters are inorganic cofactor synthesized in the mitochondria by comprising iron and sulfur. These clusters are transported out of the organelle to regulate ALAS2 transcription and keep the maintenance of iron homeostasis. To date, three genes are involved in the development of sideroblastic anemia due to a defect in iron-sulfur cluster biosynthesis: *ABCB7*, *GLRX5* and *HSPA9*.

X-linked SA with ataxia due to defects in ABCB7

Sideroblastic microcyst anemia with X-linked ataxia (XLSA/A) is a syndromic form XLSA associated with spinocerebellar ataxia. XLSA/A is caused by mutations in *ABCB7* gene (Allikmets et al. 1999) (OMIM #301310), which is located on chromosome Xp13 and encodes a mitochondrial transporter of FE-S cluster. The ABCB7 protein is situated in the inner membrane of mitochondria and is fundamental for hematopoiesis. The deficiency of the Fe-S cluster leads to the activation of the IRP1 protein that inhibits the synthesis of the ALAS2 protein, the first enzyme in the construction of the heme group, and finally leading to anemia. *ABCB7* gene is expressed not only in bone marrow but also in the cerebellum so that iron overload can happen in the mitochondria of the nervous system (Napier 2005).

Glutaredoxin 5 deficiency

GLRX5 gene is localized on chromosome 14q32.13 and encodes glutaredoxin 5, a mitochondrial protein, with an essential role in the formation of iron-sulfur clusters (Fe-S cluster). The deficiency of this enzyme engenders a mild microcytic hypochromic anemia with iron overload in the liver, an extension of spleen and liver and type 2 diabetes (OMIM #616860). Only two mutations have been

reported with recessive inheritance transmission: and homozygous splicing mutation (Camaschella et al. 2007) and two compound heterozygous mutations (G. Liu et al. 2014).

HSPA9 deficiency

Heat shock protein family A member 9 (HSPA9) is a protein constituent of the Fe-S cluster assembly complex, involved in its biogenesis. This protein is encoded by *HSPA9* gene, located on chromosome 5q31.2. Mutation in HSPA9 has been described to generate a severe loss of function allele and resulting in the development of CSA (OMIM #182170) in recessive or pseudo-dominant inheritance manner (Schmitz-Abe et al. 2015).

HSCB deficiency

Recently, one patient has been reported to have two mutations in the *HSBC* gene, a frameshift variant and a nucleotide substitution in the promoter area of the gene, that is predicted to alter a conserved ETS transcription factor binding site (Crispin et al. 2017). HSCB cooperate with HSPA9 to stimulate ATPase activity and promote the transfer of Fe-S cluster to target proteins

CSA caused by mitochondrial protein synthesis defects and mitochondrial respiratory protein mutations

The mammalian mitochondrial respiratory chain compromises of several subunits encoded by mitochondrial DNA (mtDNA) and *NDUFB11*. Dysfunction in the respiratory chain results in a deficiency of cytochrome c, iron reduction inability and impaired heme biosynthesis. As a result, iron is accumulated in the mitochondria and sideroblastic anemia appears. Several genes associated with mitochondrial respiratory chain pathway have been found to be causative of this condition. We have separated these genes in: CSA caused by mitochondrial protein synthesis defects and CSA caused by mitochondrial protein synthesis defects mitochondrial respiratory protein mutations

CSA caused by mitochondrial protein synthesis defects

Pearson marrow-pancreas syndrome (PMPS)

PMPS is a rare syndrome presented in early childhood associated with lactic acidosis, ataxia, hepatic failure and endocrine pancreas dysfunction (OMIM #557000). Nearly half of the patients reported have been detected to carry a 4977-bp deletion in the mitochondrial DNA (mtDNA), affecting essential mitochondrial tRNAs for the translation of proteins (Pearson et al. 1979; Rotig et al. 1989).

Mitochondrial myopathy with lactic acidosis and sideroblastic anemia (MLASA)

Mitochondrial myopathy with lactic acidosis and sideroblastic anemia (MLASA) belongs to the heterogeneous family of metabolic myopathies. MLASA is an autosomal recessive disorder caused by defects in *PUS1* and *YARS2* genes, denominated MLASA1 and MLASA2 respectively.

The MLASA1 subtype is due to mutations in the gene *PUS1*, which is located at 12q24.33 and encodes for nuclear pseudo-uridine synthase 1 (OMIM #600462). MLASA2 form is due to mutations in the *YARS2* gene (Riley et al. 2013) that codes for the mitochondrial enzyme tyrosyl-tRNA synthetase (OMIM #613561). Mutation in PUS1 and YARS2 compromise the translation of respiratory complexes, particularly complex I and IV, leading to mitochondrial respiratory chain dysfunction.

Sideroblastic anemia, B cell immunodeficiency, fevers and developmental delay (SIFD)

SIFD is a severe form of congenital sideroblastic anemia characterized by severe microcytic anemia, B-cell lymphopenia, periodic fevers, and variable neurodegeneration (OMIM #616084). The disease is

caused by mutations in the *TRNT1* gene, which encodes the CCA addition enzyme, an essential enzyme that catalyzes the inclusion of CCAs at the 3 'end of the tRNA precursors (Wiseman et al. 2013; Chakraborty et al. 2014). This reaction is crucial for the tRNAs maturation, so a mutation in *TRNT1* leads to decreased mitochondrial protein synthesis, resulting in mitochondrial respiratory chain dysfunction in erythroid cells. The disease occurs in childhood with recurrent fever episodes, gastrointestinal disorders, developmental delay, convulsions, ataxia and sensorineural deafness.

LARS2 deficiency

LARS2 encodes mitochondrial leucyl-tRNA synthase, which can attach leucine to its cognate tRNA. To date, two mutations have been reported in homozygous and in heterozygous compound form (Riley et al. 2015). Patients exhibit syndromic sideroblastic anemia, including hydrops, lactic acidosis, and "Perrault syndrome" (premature ovarian failure and hearing loss).

CSA caused by mitochondrial respiratory protein mutations

MT-ATP6 sideroblastic anemia (MT-ATP6-SA)

A single nucleotide variation in mtDNA-encoded *ATP6* gene (*MT-ATP6*) has been reported to cause MLASA-like phenotype (OMIM #500011). This mutation causes respiratory chain complex V defect and is inherited in a maternal or sporadic pattern (Burrage et al. 2014; Berhe et al. 2018).

CSA due to defects in NDUFB11

One deletion, no other than c.276_278del, in *NDUFB11* gene has been identified to cause sideroblastic anemia and variable symptoms, including lactic acidosis (Lichtenstein et al. 2016). *NDUFB11* is an X-linked gene, which encodes noncatalytic subunit mitochondrial respiratory complex I protein B11. The underlying molecular pathogenesis of mutation in *NDUFB11* remains unclear.

Other cases:

CSA due to mutation in *STEAP3*

A new form of hypochromic and transfusion-dependent anemia associated with a nonsense mutation of the *STEAP3*/*TSAP6* gene (AHMIO2 OMIM #609671) has been described (Grandchamp et al. 2011)). The clinical presentation of the patients was in some respects similar to that of non-syndromic sideroblastic anemia (CSA) with the presence of sideroblasts and iron overload. However, in this family, protoporphyrin levels are increased, while they are normal or even low in cases of non-syndromic CSA. This gene codes for prostate transmembrane epithelial antigen (steap3), a ferrireductase involved in the reduction of iron in the endosomes of erythroblasts. The heme or the Fe-S biosynthesis may be compromised due to defects in *STEAP3*, although the molecular mechanism of pathogenesis is still not elucidated.

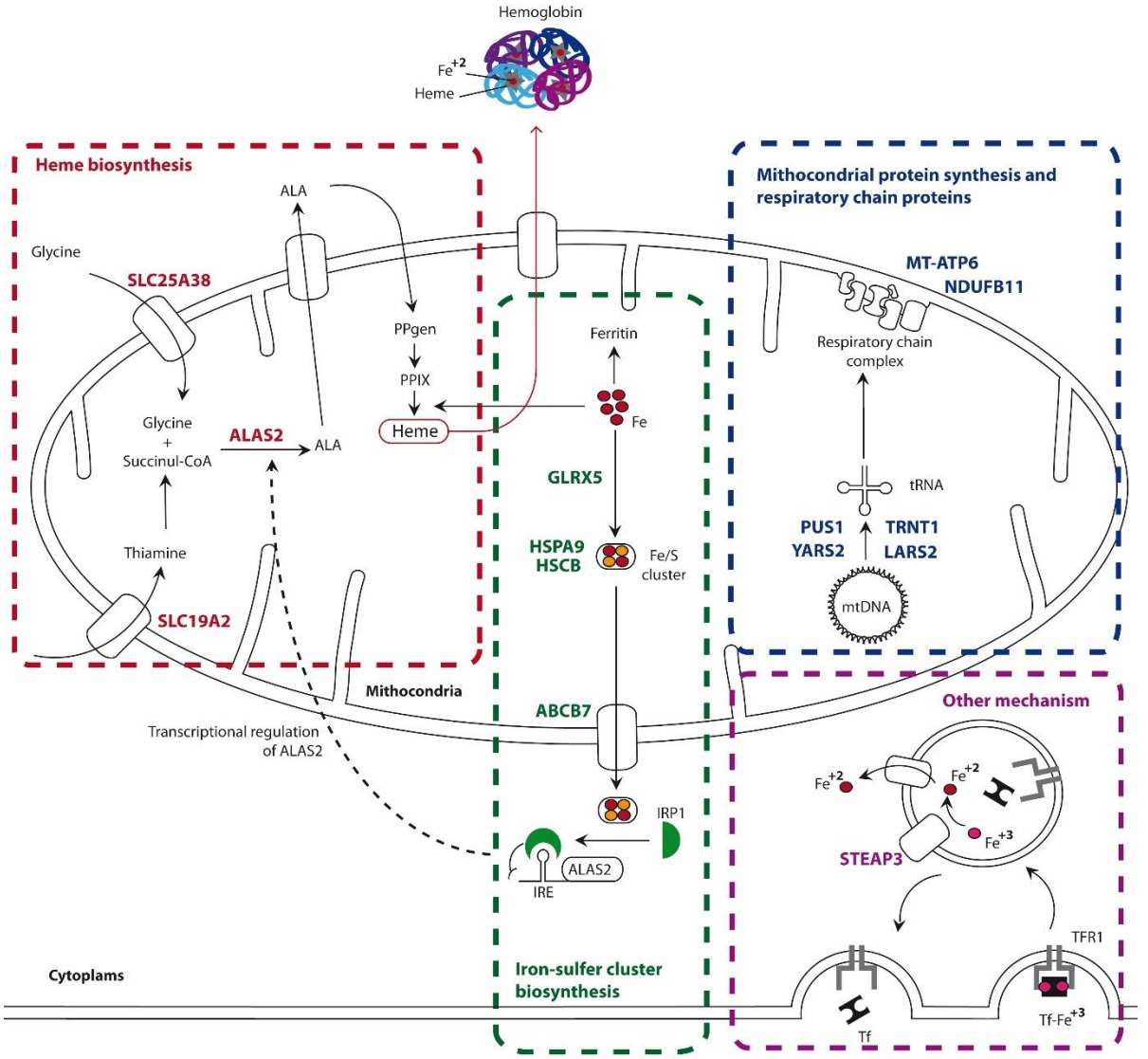


Figure 1.2. Scheme of the pathways and genes involved in CSA in erythroid cells. The pathways implicated in the CSA pathogenesis are diagrammed: Heme biosynthesis pathway genes are in red (ALAS2, SLC25A38, SLC19A2); Iron-sulfur cluster biosynthesis are in green (GLRX5, HSPA9, HSCB, ABCB7); Genes implicated in the synthesis of mitochondrial proteins (PUS1, YARS2, TRNT1, LARS2) and respiratory chain proteins (MT-ATP6, NDUFB11) are in blue. The gene STEAP3 (in purple) is classified as other mechanism since its function occurs outside the mitochondria. ALA, delta-aminolevulinate synthase; CoA, succinyl-coenzyme A; PPgen, protoporphyringen; PPIX, protoporphyrin IX; IRP1, Iron regulatory protein 1; IRE, Iron regulatory protein; Fe/S, iron-sulfur; Tf, transferrin; TFR1, transferrin receptor 1. This figure is original and was designed by Beatriz Cadenas and adapted from different literature sources (Ducamp and Fleming 2019; Long et al. 2018).

Acquired sideroblastic anemia, myelodysplastic syndrome with ring sideroblasts

Myelodysplastic syndromes (MDS) are a heterogeneous group of clonal disorders of hematopoietic stem cells characterized by peripheral blood cytopenias and the increased risk to progress to acute myelogenous leukemia (AML). MDS with single lineage dysplasia and ring sideroblasts (MDS-RS-SLD) and MDS with multilineage dysplasia and ring sideroblasts (MDS-RS-MLD) care two phenotypically well-defined subtypes of MDS that are characterized by ring sideroblasts (Arber et al. 2016).

Iron overload is common in patients with MDS and has a negative influence on survival in these patients (worse overall survival and worse leukemia-free survival). Recently, somatic mutations in the *SF3B1* gene, which is involved in RNA splicing machinery, have been reported to be particularly common (60-80%) in patients with MDS-RS-SLD or MDS-RS-MLD (Papaemmanuil et al. 2011).

Table 1. 3 Genetic classifications of Sideroblastic anemias

Pathway	Disorder		Gene	Inheritance	CSA class
Heme biosynthesis	X-linked sideroblastic anemia	XLSA	*ALAS2*	X-linked	Non-syndromic
	SLC25A38 deficiency	SA	*SLC25A38*	AR	Non-syndromic
	Thiamine-responsive megaloblastic anemia	TRMA	*SLC19A2*	AR	Syndromic
Iron-sulfur cluster biosynthesis	X-linked SA with ataxia	XLSA/A	*ABCB7*	X-linked	Syndromic
	Glutaredoxin 5 deficiency	SA	*GLRX5*	AR	Non-syndromic
	HSPA9 deficiency	SA	*HSPA9*	AR/D?	Non-syndromic
	HSCB deficiency	SA	*HSCB*	AR	Non-syndromic
Mitochondrial protein synthesis defects	Pearson marrow-pancreas syndrome	PMPS	*mtDNA*	SP/M	Syndromic
	MLASA1	MLASA	*PUS1*	AR	Syndromic
	MLASA2	MLASA	*YARS2*	AR	Syndromic
	Sideroblastic anemia, B cell immunodeficiency, fevers and developmental	SIFD	*TRNT1*	AR	Syndromic
	LARS2 deficiency	SA	*LARS2*	AR	Syndromic
Mitochondrial respiratory protein mutations	MLASA-like	MLASA	*MT-ATP6*	SP/M	Syndromic
	CSA due to defects in NDUFB11	XLSA	*NDUFB11*	X-linked	Syndromic
Other cause	CSA due to defects in *STEAP3*	SA	*STEAP3*	AD	Syndromic
Acquired SA	MDS-RS-SLD		*SF3B1*	SOM	N/A
	MDS-RS-MLD		*+ JAK2, MLP, CARL*	SOM	N/A

AD, autosomal dominant; AR, autosomal recessive; SP, sporadic; M, maternal; SOM, somatic; CSA, congenital sideroblastic anemia; SA, sideroblastic anemia, MLSASA, mitochondrial myopathy with lactic acidosis and sideroblastic anemia; N/A, Not applicable. MDS-RS-SLD, MDS with single lineage dysplasia and ring sideroblasts; MDS-RS-MLD, MDS with multilineage dysplasia and ring sideroblasts; MDS, Myelodysplastic syndrome

1.1.1.4. Iron and Copper Related Anemias

Defects in genes involved in the metabolism of iron (globin genes, heme synthesis, maintenance of iron, or the acquisition of iron by erythroid precursors) genes can lead to hypochromic and microcytic anemias (Table 1.4).

Atransferrinemia

Atransferrinemia (OMIM #209300) is a rare autosomal recessive disorder, characterized by microcytic anemia and by iron loading, caused by mutations in the gene that codes for Transferrin (*TF*), a plasma protein that transports iron in the blood. Values of TF are usually half the standard values (204-360 mg/dl) in carriers and extremely low in affected patients (lack of TF is incompatible with life). The decrease in TF levels leads to a reduction in the transfer of iron to the bone marrow, resulting in reduced development of erythroid precursors, and the accumulation of iron in peripheral tissues. Affected patients show severe microcytic hypochromic anemia in childhood, grow retardation and frequent infections. Mainly, iron overload occurs in the liver, joints, heart, pancreas, thyroid, kidney and bone, leading to symptoms such as liver problems, heart malfunction, arthropathy and hypothyroidism (Athiyarath et al. 2013).

Aceruloplasminemia

Aceruloplasminemia (OMIM #604290) is an autosomal recessive disorder due to mutations in the gene encoding ceruloplasmin (*CP*), which is involved in the release of iron from macrophages and other cells to the blood. Clinical manifestation of the disease includes reduced levels or absence of serum ceruloplasmin, low levels of serum iron, high serum ferritin, mild to moderate microcytic anemia, iron overload in the tissues, diabetes mellitus, and late-onset neurological symptoms (Kono 2013).

Microcytic and hypochromic anemia with iron overload

Mutations in the *SLC11A2* gene generate a deficiency of the divalent metal transporter 1 (DMT1), a major transporter involved in the absorption of duodenal iron and for the transfer of iron from the endosomes to the cytosol. Abnormal DMT1 may disturb the development of erythroid cells and leads into microcytic and hypochromic anemia with progressive iron overload (AHMIO1; OMIM # 206100). This disease is inherited in an autosomal recessive manner (Iolascon et al. 2006).

Grandchamp et al. identified a nonsense mutation of the *STEAP3/TSAP6* gene in three siblings with a new form of hypochromic anemia and iron overload (AHMIO2 OMIM #609671) (Grandchamp et al. 2011). The clinical presentation of the patients was chronic hypochromic anemia with the presence of sideroblasts and iron overload, similar to that of non-syndromic sideroblastic anemia (CSA), but with increased level of protoporphyrin. STEAP3 gene codes for prostate transmembrane epithelial antigen (steap3) that is involved in regulation of cell cycle, apoptosis, and in the secretion of non-classical proteins, including exosomes. In iron metabolism, the *STEAP3/TSAP6* gene codes for a ferrireductase involved in the acquisition of iron by red blood cells.

Table 1. 4 Genetic classification of Iron and Copper deficiency Anemia

Disease	OMIM	Gene/s	Inheritance
Atransferrinemia	209300	*TF*	AR
Aceruloplasminemia	604290	*CP*	AR
AHMIO1	206100	*SLC11A2*	AR
AHMIO2	609671	*STEAP3*	AR
Iron-refractory Iron-Deficiency Anemia (IRIDA)	206200	*TMPRSS6*	AR
Wilson disease	277900	*ATP7B*	AR
IRIDA-like (new form)		*TMPRSS6 + ACVR1*	Digenic

AHMIO1, microcytic and hypochromic anemia with iron overload type 1; AHMIO2, microcytic and hypochromic anemia with iron overload type 2; AR, autosomal recessive; AD, autosomal dominant

Iron-refractory Iron-Deficiency Anemia (IRIDA)

Protein matriptase-2, is a serine protease with an important function in the absorption of iron which encoded by *TMPRSS6* gene. Mutations in this gene produce a reduction in the activity of matriptase-2 in hepatocytes, increase the hepcidin activity inhibiting iron absorption and leads to moderate hypochromic microcytic anemia from birth. This disorder is named Iron-refractory Iron- Deficiency Anemia (IRIDA; #206200) and is inherited in autosomal recessive pattern (De Falco et al. 2013; 2014). Pagani et al, reported in 2017 a new form of IRIDA due to combined heterozygous mutations in *TMPRSS6* and *ACVR1* (Pagani et al. 2017).

Wilson disease

Wilson disease (OMIM #277900) is an autosomal recessive disorder caused by mutations in the *ATP7B* gene and is characterized by severe accumulation of intracellular hepatic copper with subsequent hepatic and neurologic abnormalities (Thomas et al. 1995).

1.1.1.5. Congenital dyserythropoietic anemia (CDA)

Congenital dyserythropoietic anemia (CDA) is a heterogeneous group of inherited conditions characterized by abnormal production of red blood cells, due to erythropoietic maturation arrest phenomenon, that could lead to anemia of variable severity. Five different forms of CDA have been defined: type I, II, III, IV and XLTDA (Table 1.5). The shared symptoms include variable severity anemia, jaundice, hepatomegaly, splenomegaly, and progressive iron overload. Bone marrow examination reveals polychromatic erythroblasts with abnormal shape and size (Iolascon, Esposito, and Russo 2012; Iolascon et al. 2013; Russo et al. 2014).

Type I CDA may be caused by mutations in the *CDAN1* gene (Dgany O et al., Am J Hum Genet, 2002 Dec; 71 (6): 1467-74) (OMIM #224120), which codes for a protein involved in maintaining the integrity of the nuclear envelope. Type I CDA is characterized by mild to severe anemia generally diagnosed in childhood or adolescence. Iron overload can lead to arrhythmia, congestive heart failure, diabetes, and chronic liver disease (cirrhosis). Variants in gene *C15ORF41* has also been associated with being causative of CDA I (Babbs et al. 2013).

Type II CDA is due to mutations in the *SEC23B* gene (Schwarz et al. 2009) (OMIM #224100), and is characterized by the presence of binucleated erythroblasts (diplo erythroblasts) with a double membrane derived from endoplasmic reticulum residues that are visualized by electron microscopy.

CDA Type III (OMIM #105600) is a rare form of dyserythropoietic anemia characterized by moderate to mild and non-progressive hemolytic iron deficiency, dyserythropoiesis, large multinucleated erythroblasts in bone marrow, and macrocytosis in the peripheral blood. The clinical presentation is variable. Recently, the *KIF23* gene has been described as the gene that causes the autosomal dominant form of CDA III. This gene codes for the mitotic kinesin-like protein 1 (MKLP1), which is a crucial protein for cytokinesis (Liljeholm et al. 2013).

CDA IV (ORPHA293825, OMIM #613673) is caused by mutations in the *KLF1* gene that encodes an erythroid transcription factor that plays a crucial role in the development of the erythroid lineage, including the expression of globin and other erythropoiesis genes. CDA IV is characterized by inefficient erythropoiesis and hemolysis that leads to severe anemia at birth with the requirement of multiple repeated transfusions. Patients showed increased levels of fetal hemoglobin, and vast multi-nucleated erythroblasts with morphologic abnormalities.

XLTDA, X-linked thrombocytopenia and dyserythropoietic anemia, (OMIM #300367) is a rare hematological disorder characterized by moderate to severe thrombocytopenia and abnormal platelet morphology and activity due to damage in the maturation of platelets. XLTDA is caused by mutations in the erythroid transcription factor *GATA1*, which play an essential role in the development of megakaryocytes and erythroid precursors. The disease affects mainly males as females are usually asymptomatic or have only mild symptoms (X-linked inheritance pattern). It is presented in infancy as a bleeding disorder with further manifestations of thrombocytopenia including epistaxis, petechiae, ecchymoses, or splenomegaly.

Other rare syndromes have been found to be associated with mutations in GATA1: X-linked thrombocytopenia (XLT), X-linked thrombocytopenia with beta-thalassemia (XLTT; OMIM #314050) and X-linked anemia with or without neutropenia and/or platelet abnormalities (XLANP, OMIM #300835) (Nichols et al. 2000; Hollanda et al. 2006).

Table 1. 5 Classification of CDAs

Subtype	OMIM	Gene	Inheritance	Cellular or hematological abnormalities
CDA I	224120	*CDAN1, C15ORF41*	AR	Abnormal chromatin structure, chromatin bridge
CDA II	224100	*SEC23B*	AR	Bi- or multi-nuclearity of mature erythroblasts
CDA III	105600	*KIF23*	AD	Giant multi-nucleated erythroblasts
CDA IV	613673	*KLF1*	AD	Multi-nucleated erythroblasts
XLDA	300367	*GATA1*	X-linked	CDA I or CDA II-like with thrombocytopenia

AD, autosomal dominant; AR, autosomal recessive; CDA, Congenital dyserythropoietic anemia

1.1.1.6. Congenital Erythrocytosis / Familial Polycythemia

Congenital erythrocytosis or familial polycythemia is a hereditary hematological disorder, characterized by a high absolute mass of red blood cells caused by uncontrolled production of red blood cells. Clinical symptoms, if developed, include headache, dizziness, epistaxis, and exertional dyspnea. Although the hematologic disorder is present from birth, clinical symptoms can be detected at any time during childhood or adulthood. Hematologic manifestations include the presence of erythrocytosis without platelets or leukocytes incrementation and progression to leukemia. Low levels of serum EPO is found in type 1 familial erythrocytosis, while type 2 familial erythrocytosis shows normal or high EPO. Familial polycythemia is triggered by mutations in different genes (Table 1.6).

Type 1 familial erythrocytosis or primary familial polycythemia (OMIM #133100) is due to mutations in the EPO receptor gene (*EPOR*), located on chromosome 19p13.3-p13.2 (Juvonen et al. 1991; Chapelle et al. 1993). EPOR mutations lead to hypersensitivity to EPO and cause the receptor to be permanently activated to stimulate red blood cells production. Few cases have been described following an autosomal dominant inheritance (Kralovics et al. 1997). The diagnosis is based on the presence of isolated erythrocytosis without splenomegaly, low levels of serum EPO, the normal affinity of hemoglobin for oxygen, and erythroid progenitors in the bone marrow exhibiting hypersensitivity to EPO.

Polycythemia Vera is the most common form of primary polycythemia and is due to a somatic mutation in *JACK2* and *SH2B3* genes (Prchal 2005; Tefferi and Barbui 2017; Bento et al. 2014) [refs]. These genes need to be evaluated to exclude polycythemia Vera in the differential diagnosis of congenital erythrocytosis.

Type 2 familial erythrocytosis (also known as Von Hippel-Lindau syndrome or Chuvash polycythemia; OMIM #263400) is an autosomal recessive disorder due to mutations in the *VHL* gene (3p25), characterized by an increased volume of erythrocytes, high serum EPO levels, and normal oxygen affinity (Ang et al. 2002). This type of familial erythrocytosis has aspects of both primary and secondary erythrocytosis: increased circulating levels of EPO, consistent with a secondary; and erythroid precursors are also hypersensitive to EPO, consistent with a primary.

Type 3 familial erythrocytosis (OMIM #609820) is a disorder caused by heterozygous mutations in the *EGLN1* gene (an autosomal dominant inheritance), located on chromosome 1q42. Phenotypic manifestations are similar than ECYT1 but showing average levels of EPO in blood (Ladroue et al. 2008).

Type 4 familial erythrocytosis (OMIM #611783) is an autosomal dominant disorder caused by mutations in the *EPAS1* gene, that encodes the protein HIF2A (Percy et al. 2008). It is characterized by increased serum red blood cell mass and elevated serum hemoglobin and erythropoietin (EPO).

Type 5 familial erythrocytosis (OMIM #617907) has been reported to be caused by heterozygous mutation in *EPO* gene. Deletion of single nucleotide in exon 2 of *EPO* gene produces a frameshift mutation that disturbs the translation of the EPO mRNA transcript (Zmajkovic et al. 2018).

Type 6 familial erythrocytosis (OMIM #617980) and **type 7 familial erythrocytosis** (OMIM #617981) are autosomal dominant forms of erythrocytosis due to mutations in beta (*HBB*) and alpha globin genes (*HBA1* or *HBA2*) (Charache, Weatherall, and Clegg 1966; González Fernández et al. 2009).

Type 8 familial erythrocytosis is (OMIM #222800) caused by compound heterozygous mutation in the *BPGM* gene (Lemarchandel et al. 1992).

Table 1. 6 Familial Erythrocytosis Subtypes

Familial Erythrocytosis Type	OMIM	Gene/s	Inheritance
ECYT1	133100	*EPOR*	AD
		JAK2; SH2B3	SOM
ECYT2	263400	*VHL*	AR
ECYT3	609820	*EGLN1*	AD
ECYT4	611783	*EPAS1*	AD
ECYT5	617907	*EPO*	AD
ECYT6	617980	*HBB*	AD
ECYT7	617981	*HBA1; HBA2*	AD
ECYT8	222800	*BPGM*	AR

AD, autosomal dominant; AR, autosomal recessive

1.1.1.7. Hereditary hemolytic anemia including membranopathies and enzymopathies

Hemolytic anemia (HA) is a heterogeneous disorder represented by premature destruction of red blood cells, and is manifested in hereditary and acquired forms. Among hereditary hemolytic anemia, the most common causes are due to defects in the membrane of erythrocytes (membranopathies) and in the glycolytic pathways enzymes (enzymopathies). HA is commonly presented with severe expression in infancy, although milder forms are also shown in adulthood. Clinical manifestations include anemia, pallor, jaundice and in some cases splenomegaly. Other laboratory presentations are anemia of variable degree, mild-to-moderate chronic hemolysis and increased reticulocytes (Haley 2017; Gallagher 2015).

Membranopathies are a subgroup of hereditary hemolytic anemias, characterized by defects in red blood cells membrane due to abnormal cytoskeletal proteins. Peripheral blood smear analysis shows atypical erythrocyte shape. Mutations in genes encoding proteins that are implicated in the red cell membrane skeleton generate this type of hemolytic anemias (Table 1.7). The two most common types of membranopathy are hereditary spherocytosis (spherocytes) and hereditary elliptocytosis (elliptical-shaped erythrocyte) (Bolton-Maggs et al. 2012).

Table 1. 7 Membranopathies subtypes

	OMIM	GENE	Inheritance
Hereditary spherocytosis			
Type 1	182900	*ANK1*	AD
Type 2	616649	*SPTB*	AD
Type 3	270970	*SPAT1*	AR
Type 4	612653	*SLC4A1*	AD
Type 5	612690	*EPB42*	AR
Hereditary elliptocytosis			
Type 1	611804	*EPB41*	AR, AD
Type 2	130600	*SPAT1*	AD
Type 3	617948	*SPTB*	AD
Type 4	166900	*SLC4A1*	AD

AD, autosomal dominant; AR, autosomal recessive

The primary metabolic functions of erythrocytes include sustaining membrane proteins, conserving iron hemoglobin in the redox state and modulating the oxygen affinity of hemoglobin. These capacities are dependent upon the production of ATP, NADH, NADPH, and 2,3 diphosphoglycerate (Koralkova, van Solinge, and van Wijk 2014). **Enzymopathies** are the result of deficiencies in the glycolytic pathway enzymes, and they are also denominated as congenital nonspherocytic hemolytic anemia (CNSHAs), as these conditions do not present deficiencies in morphology in RBC. The two more common enzymatic defects are glucose 6-phosphate dehydrogenase (G6PD) deficiency and pyruvate kinase (PK) deficiency. Less frequent defects in other enzymes have been reported (Table 1.8).

Glucose 6-phosphate dehydrogenase (G6PD) deficiency is a common disorder (more than 400 million people affected worldwide) caused by mutations in *G6DP* gene. This gene is located on the X chromosome (X-linked disorder), so males are more likely to be affected. Defects in this enzyme, reduce the production of NADPH, leading to the oxidation of hemoglobin sulfhydryl groups, hemoglobin precipitation and damages of the RBC membrane, and finally, to hemolysis (Beutler 2008).

Pyruvate Kinase deficiency is the second most popular enzymopathy. Defects of PK results in the depletion of ATP and 2,3-DPG, but the molecular mechanism of hemolysis is not entirely understood (Zanella et al. 2005; Prchal and Gregg 2005).

Table 1. 8 Enzymopathy subtypes

Enzymopathies	OMIM	Gene/s	Inheritance
Pyruvate kinase deficiency	266200	*PKLR*	AR
Hemolytic anemia, G6PD deficient	300908	*G6PD*	XLD
Hemolytic anemia due to hexokinase deficiency	235700	*HK1*	AR
Hemolytic anemia, nonspherocytic, due to glucose phosphate isomerase deficiency	613470	*GPI*	AR
Diphosphoglycerate mutase deficiency of erythrocyte, DPGM Deficiency (Erythrocytosis, familial, 8)	222800	*BPGM*	AR
Hemolytic anemia due to glutathione reductase deficiency		*GSR*	AR
Hemolytic anemia due to glutathione synthetase deficiency	231900	*GSS*	AR
Hemolytic anemia due to glutathione peroxidase deficiency	614164	*GPX1*	AR
Hemolytic anemia due to gamma-glutamylcysteine synthetase deficiency	230450	*GCLC*	AR
Anemia, hemolytic, due to UMPH1 deficiency	266120	*NT5C3A*	AR
Hemolytic anemia due to adenylate kinase deficiency	612631	*AK1*	AR
Phosphoglycerate kinase 1 deficiency	300653	*PGK1*	XLR
Hemolytic anemia due to triosephosphate isomerase deficiency	615512	*TPI1*	AR
Glycogen storage disease VII	232800	*PFKM*	AR
Glycogen storage disease XII	611881	*ALDOA*	AR
Hemolytic anemia due to elevated adenosine deaminase	102700	*ADA*	AD
Methemoglobinemia, type I and type II	613213	*CYB5R3*	AR

AD, autosomal dominant; AR, autosomal recessive; XLR, X-link recessive

1.1.1.8. Congenital Erythropoietic Protoporphyria and Congenital Erythropoietic Porphyria

Erythropoietic porphyrias are metabolic disorders caused by defects in heme synthesis occurring in red blood cells, resulting in the aggregation of porphyrins and porphyrin precursors in erythrocytes. Beside increased levels of porphyrins in the urine, feces and blood, clinical manifestations include photosensitivity to visible light.

Erythropoietic Protoporphyria (EPP) is an autosomal-recessive trait caused by mutation in *FECH* gene, which encodes the enzyme ferrochelatase (Kieke et al. 2019). EPP is clinically distinguished by photosensitivity to visible light initiated in early childhood, and biochemically by high levels of protoporphyrin in red blood cells. Two more genes have been identified to cause EPP: *ALAS2* gene (X-link inheritance)(Brancaleoni et al. 2016; Whatley et al. 2008) and *CLPX* (autosomal dominant) (Yien et al. 2017) (Table 1.9).

Congenital Erythropoietic Porphyria (CEP) or Gunther disease is caused by a deficiency of the enzyme in the heme biosynthesis pathway (URO-S protein) that leads to a massive accumulation of isomeric porphyrins I (uro and coproporphyrins) in the bone marrow. The enzymatic deficiency appears due to mutations of the *UROS* gene, which encodes URO-S, in an autosomal recessive inheritance (Xu, Astrin, and Desnick 1996).

Table 1. 9 Congenital Erythropoietic Porphyria subtypes

Type	Acronym	OMIM	Gene	Inheritance
EPP	EPP1	177000	*FECH*	AR
	XLEPP	300752	*ALAS2*	X-Linked
	EEP2	618015	*CLPX*	AD
CEP	CEO	263700	*UROS*	AR

AD, autosomal dominant; AR, autosomal recessive; XLR, X-link recessive

1.1.1.9. Bone marrow failure

The inherited bone marrow failure (BMF) syndromes are a heterogeneous group of disorders characterized by bone marrow failure usually in association with one or more somatic abnormalities. These diseases are often presented in infancy but may not do so until adulthood in some cases, and can be caused by defects in DNA repair-FA/BRCA pathway; telomere maintenance- dyskeratosis congenita-related genes; and ribosome biogenesis-Diamond-Blackfan anemia (DBA) genes.

Fanconi anemia (FA) is a rare congenital heterogeneous disorder characterized by congenital progressive bone marrow failure, malformations, hematological problems and predisposition to malignancies in children. To date, 22 genes involved with the cross-linked DNA repairment, have been reported to be associated with FA (Table 1.10). Among these, *FANCA* is the most frequent (60-65%) (Solomon et al. 2015) The disorder is autosomal-recessive inherited in most of de cases, except for those rare cases with a mutation in *FANCB* (X-linked) and *FANCR* (autosomal dominant) (Knies et al. 2017; Bagby 2018).

Diamond-Blackfan anemia (DBA), or congenital hypoplastic anemia, is a rare congenital red blood cell aplasia that belongs to bone marrow failure disorders (Ulirsch et al. 2018; Da Costa et al. 2018). DBA is usually manifested in the first year of life with normochromic macrocytic anemia, reticulocytopenia, and a scarcity of erythroid precursors in the bone marrow. Patients show growth restriction, and about 30-50% have physical anomalies (craniofacial, upper limb, and heart malformations). DBA is considered a ribosomopathy, as this disorder is almost exclusively driven by haploinsufficient mutations in a ribosomal protein (RP) genes (Table 1.11).

Dyskeratosis congenita (DC) is a rare multisystem disease characterized by defective telomere maintenance. Clinical aspects are heterogeneous and include BMF, cancer predisposition, and pulmonary and hepatic fibrosis (Kelmenson and Hanley 2017; Dokal 2011; Fernández García and Teruya-Feldstein 2014). In addition, patients can present abnormal skin pigmentation, leukoplakia, and nail dystrophy, although it is not always observed.

DC is a genetically heterogeneous disorder, presented in three modes of inheritance: autosomal recessive, autosomal dominant and X-linked, and is caused by mutations in genes involved in the telomere maintenance. *DKC1* was the first gene identified to be associated with DC, and it is the most frequent presentation of the (30%) (Heiss et al. 1998; Kelmenson and Hanley 2017). *DKC1* encodes for the nuclear dyskerin protein and is X-linked inherited. In total, 12 genes have been reported to cause DC (Table 1.12).

Table 1. 10 Genetic classification of Fanconi Anemia

FA type	*Gene*	OMIM	Inheritance
FANCA	*FANCA*	227650	AR
FANCB	*FANCB*	300514	XL
FANCC	*FANCC*	227645	AR
FANCD1	*BRCA2*	605724	AR
FANCD2	*FANCD2*	227646	AR
FANCE	*FANCE*	600901	AR
FANCF	*FANCF*	603467	AR
FANCG	*XRCC9*	614082	AR
FANCI	*FANCI*	609053	AR
FANCJ	*BRIP1*	609054	AR
FANCL	*FANCL*	608111	AR
FANCM	*FANCM*	NA	AR
FANCN	*PALB2*	610832	AR
FANCO	*RAD51C*	613390	AR
FANCP	*SLX4*	613951	AR
FANCQ	*ERCC4*	615272	AR
FANCR	*RAD51*	617244	AD
FANCS	*BRCA1*	617883	AR
FANCT	*UBE2T*	616435	AR
FANCU	*XRCC2*	617247	AR
FANCV	*MAD2L2*	617243	AR
FANCW	*RFWD3*	617784	AR

AD, autosomal dominant; AR, autosomal recessive; XL, X-linked

Table 1. 11 Genetic classification of DBA

Subtype	Gene	OMIM	Inheritance
DBA1	*RPS19*	105650	AD
DBA2	*8p23.3-p22*	606129	
DBA3	*RPS24*	610629	AD
DBA4	*RPS17*	612527	AD
DBA5	*RPL35A*	612528	AD
DBA6	*RPL5*	612561	AD
DBA7	*RPL11*	612562	AD
DBA8	*RPS7*	612563	AD
DBA9	*RPS10*	613308	AD
DBA10	*RPS26*	613309	AD
DBA11	*RPL26*	614900	AD
DBA12	*RPL15*	615550	AD
DBA13	*RPS29*	615909	AD
DBA14	*TSR2*	300946	XLR
DBA15	*RPS28*	606164	AD
DBA16	*RPL27*	617408	AD
DBA17	*RPS27*	617409	AD
DBA18	*RPL18*	618310	AD
DBA19	*RPL35*	618312	AD
DBA20	*RPS15A*	618313	AD

AD, autosomal dominant; AR, autosomal recessive; XLR, X-link recessive

Table 1. 12 Genetic classification of Dyskeratosis congenita

Subtype	OMIM	Gene	Inheritance
Dyskeratosis congenita, X-linked	305000	*DKC1*	XLR
Dyskeratosis congenita, autosomal dominant 1	127550	*TERC*	AD
Dyskeratosis congenita, autosomal dominant 2	613989	*TERT*	AD, AR
Dyskeratosis congenita, autosomal recessive 4	613989	*TERT*	AD, AR
Dyskeratosis congenita, autosomal dominant 4	615190	*PARN*	AD, AR
Dyskeratosis congenita, autosomal recessive 5	615190	*PARN*	AD, AR
Dyskeratosis congenita, autosomal recessive 6	616353	*PARN*	AR
Dyskeratosis congenita, autosomal recessive 2	613987	*NHP2*	AR
Dyskeratosis congenita, autosomal dominant 3	613990	*TINF2*	AD
Dyskeratosis congenita, autosomal dominant 6	616553	*ACD*	AD, AR
Dyskeratosis congenita, autosomal recessive 7	616553	*ACD*	AD, AR
Dyskeratosis congenita, autosomal recessive 1	224230	*NOP10*	AR
Dyskeratosis congenita, autosomal recessive 3	613988	*WRAP53*	AR
Dyskeratosis congenita, autosomal dominant 4	615190	*RTEL1*	AD, AR
Dyskeratosis congenita, autosomal recessive 5	615190	*RTEL1*	AD, AR

AD, autosomal dominant; AR, autosomal recessive; XLR, X-link recessive

1.1.1.10. Neurodegeneration with Brain Iron Accumulation Disorders

Neurodegeneration with brain Iron Accumulation (NBIA) is a clinically and genetically heterogeneous group of neurodegenerative disorders. This disease is characterized by the accumulation of iron in the basal glia regions (*globus pallidus* and *subsantia nigra*), although other regions like the cortex and the cerebellum can be affected. The hallmarked clinical manifestations of this disorder are extrapyramidal symptoms, including progressive dystonia, spasticity, parkinsonism, and neuropsychiatric anomalies. To date, 15 genes have been reported as causative for this neurodegenerative disorder, of which only two are associated with iron metabolism: *CP* and *FTL* (Figure 1.3*)*. This rare monogenetic disease is typically inherited in AR trait, although it has also been presented in AD and X-linked pattern (Table 1.13) (Levi and Tiranti 2019).

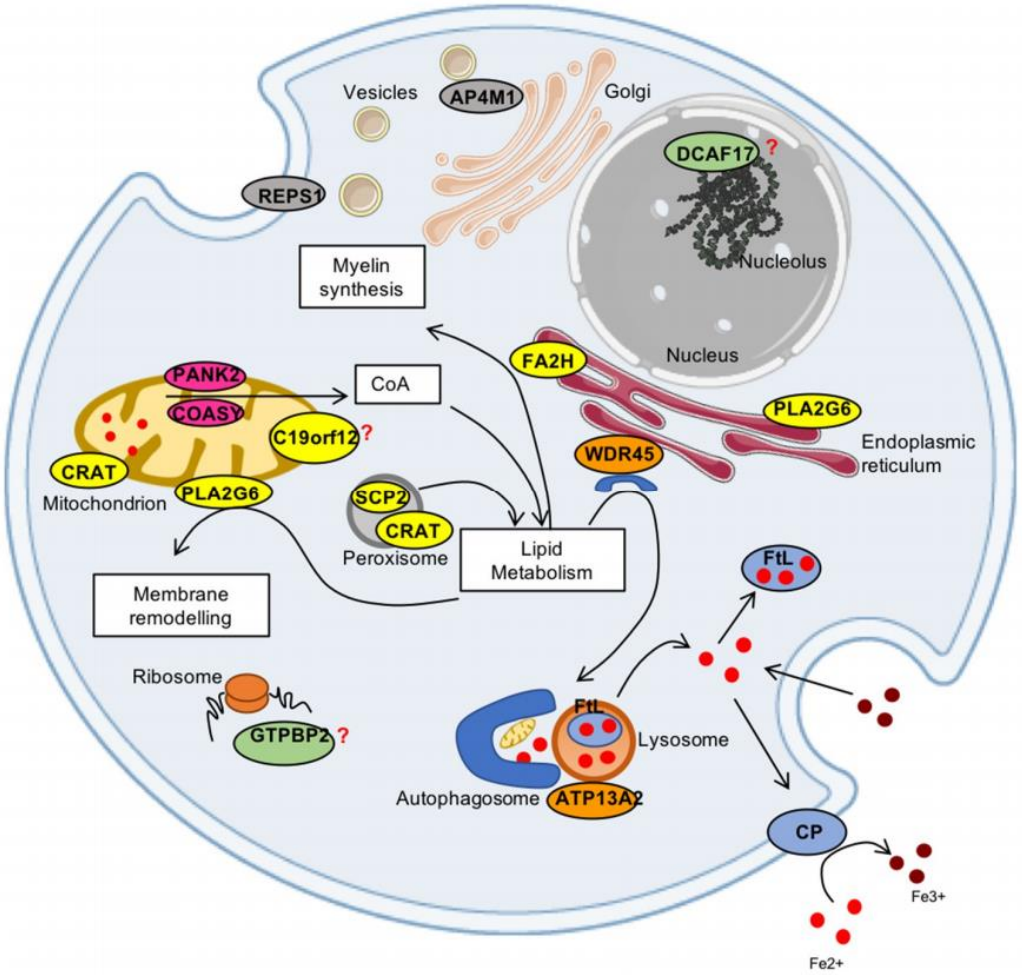


Figure 1. 3 Scheme of the genes associated to the pathogenesis of NBIA. The iron-proteins (CP and FTL) are represented in light blue; the proteins (PANK2 and COASY) involved in CoA synthesis are in pink; in yellow are the proteins linked to lipid metabolism (PLA2G6, FA2H,SCP2, CRAT, C19orf12); in orange are the proteins (WDR45, ATP13A2) implicated in autophagy; in grey are the proteins (RESP1 and AP4M1) related to vesicle trafficking; and the proteins in green (DCAF17 and GTPBP2) in green still have unknown functions. Picture reproduced with permission from Levi et al. 2019.

Table 1. 13 Genetic classification of NBIA

Gene	Disease	Inheritance	OMIM	NBIA Type
PANK2	Pantothenate Kinase-associated neurodegeneration (PKAN)	AR	234200	NBIA1
PLA2G6	PLA2G6-associated neurodegeneration (PLAN)	AR	256600	NBIA2
FTL	Neuroferritinopathy (NF)	AD	606159	NBIA3
C19orf12	Mitochondrial membrane protein-associated neurodegeneration (MPAN)	AR	614298	NBIA4
WDR45	B-propeller-associated neurodegeneration (BPAN)	XL	300894	NBIA5
COASY	COASY protein-associated neurodegeneration (CoPAN)	AR	615643	NBIA6
REPS1	Neurodegeneration with brain iron accumulation 7	AR	617916	NBIA7
CRAT	Neurodegeneration with brain iron accumulation 8	AR	617917	NBIA8
CP	Aceruloplasminemia	AR	604290	
FA2H	Fatty acid hydroxylase-associated neurodegeneration	AR	612319	
SCP2	Leukocenphalopathy with dystonia and motor neuropathy	AR	613724	
ATP13A2	Kufor-Rakeb disease (KRS)	AR	606693	
AP4M1	AP4 deficiency	AR		
DCAF17	Woodhouse-Sakati syndrome (WSS)	AR	241080	
GTPBP2	Jaberi-Elahi syndrome	AR	617988	

AD, autosomal dominant; AR, autosomal recessive; XL, X-linked

1.1.1.11. Hyper and hypoferritinemia

Five diseases have been identified as directly resulting from mutations in the *FTL* gene: hereditary hyperferritinemia with cataract syndrome (HHCS), neuroferritinopathy (NBIA3), benign hyperferritinemia (or hyperferritinemia without iron overload), autosomal dominant L-Ferritin deficiency and autosomal recessive L-ferritin deficiency (Cadenas et al. 2019).

Hereditary Hyperferritinemia Cataract Syndrome (HHCS) (OMIM #600886) is associated with excess production of ferritin and its accumulation in the lens of the eye, resulting in the development of cataracts (Mumford et al. 2000). More than 40 mutations in the iron-responsive element (IRE) at the 5' untranslated region (UTR) of the *FTL* gene have been reported as causative of this disorder (Sara Luscieti et al. 2013; Cadenas et al. 2019). Defects in the FTL-IRE results in the disability to linkage with iron regulatory proteins (IRPs) and the disturbance of the IRP-IRE post-transcriptional regulatory system (Beaumont et al. 1995; Girelli et al. 1995; Muckenthaler et al. 2017).

Benign hyperferritinemia, also known as genetic hyperferritinemia without iron overload (OMIM #600886), is another disorder associated with mutations in the *FTL* gene. To date, only three mutations have been reported (Caroline Kannengiesser et al. 2009). This disorder is characterized by high (greater than 90%) glycosylated serum ferritin levels (Ravasi et al. 2017), but not other harmful clinical manifestations.

L-ferritin deficiency or hypoferritinemia (OMIM#615604) is a rare disorder that presents different manifestation whether the inheritance is autosomal dominant or recessive. Autosomal recessive inheritance of this disorder is associated with more severe symptoms, including undetectable levels of serum ferritin, seizures and restless leg syndrome (Cozzi et al. 2013). However, patients with autosomal dominant pattern of L-ferritin deficiency, the sole sign is low level of ferritin in blood, but no other manifestations (Cremonesi et al. 2004). The *ATP4A* gene have been also associated with low levels of ferritin in serum in pediatric patients with type 1 diabetes (Chobot et al. 2018).

1.1.2. NGS and genetic diagnosis

Inherited blood diseases affect millions of people, and cause substantial morbidity and mortality, in addition to an immense burden on those affected (Engert et al. 2016). In recent decades, the diagnosis and treatment of congenital blood disorders have received enormous progress. In the past, physical examination, the patient´s medical history, and laboratory testing from blood and bone marrow have been used by clinicians to diagnose hematologic diseases. However, the diagnosis of rare inherited hematological diseases using these methods is complicated in most of the patients and may involve a significant delay in the diagnosis and high healthcare costs. An accurate diagnosis is crucial for optimal clinical management.

Since the identification of rare congenital anemia syndromes is laborious, genetic diagnosis is achieving an essential role in patient care. Early on, Sanger sequencing was the chosen clinical testing method for disorders for which a single causative gene was wholly responsible (when there is a clear phenotypic indication of a classic Mendelian disorder) (Williams et al. 2015). However, since that most of congenital anemias can have several genes implicated, and may present similar clinical and laboratory presentation, Sanger sequencing is getting unproductive and outdated. The ability to sequence multiple genes and multiplexing several patients allows NGS to be ideally suited for addressing the limitations of traditional Sanger sequencing.

The introduction of next-generation sequencing (NGS) in genomic laboratories was established about 10 years ago. This revolutionary high-throughput technology provides massive parallel sequencing data, and has the ability to sequence multiple genes and multiplexing several patients. After improvements in management, robustness and accuracy, this revolutionary technology is becoming widely used for genetic diagnosis as an alternative to the single causative gene-Sanger sequencing approach (Di Resta et al. 2018; Vrijenhoek et al. 2015; Rehm 2013).

Since there is a good knowledge of the genes involved in hereditary hematological diseases, the development of a targeted NGS panel is considered an excellent tool for the genetic diagnosis of these diseases. The aim of this project is to refine and improve an already established NGS targeted gene panel for rare inherited blood diseases (HHD panel) in order to accelerate diagnosis.

1.1.3. HHD Panels

During the progress of this project and in the time covered by this doctorate, four versions of the HHD panel have been developed: v13, v14, v15 and v16 (Table 1.14). The version v13 of the panel comprised 58 genes for 8 groups of hematological diseases. The target region size was 382,058 bases with a total coverage of 98.99 % (15,729 amplicons). The read length was predefined to 250 bases. This panel v13 was validated with 26 patients suffering from different congenital hematological

disease. Subsequently, this panel was updated increasing the number of genes to 74 for the same hematological diseases (version 14, v14). The target region size was 374,978 bases with a total coverage of 99.67 % (18,454 amplicons). The version v15 has been designed covering 153 genes for 12 groups of hematological diseases, with 31,193 amplicons covering 99,63 % of the total target region size (618,664 bases). Finally, the version v16 has been designed covering 203 genes for 12 groups of hematological diseases, with 44,656 amplicons covering 99.16 % of the total target region size (898,561 bases).

The following project will be focused in the design and implementation of version 15 of the panel, where I was more involved. However, the other versions (v13, v14 and v16) will be also mentioned in some substantial points, such as the validation of the HHD panel. Finally, a comparison of the four designs used will be made to verify the improvement in the diagnosis of HHD with the refining and improvements of the new panels.

Table 1. 14 Comparison of the four different versions of the HHD panel designed.

			VERSION 13 (2015)	VERSION 14 (2016)	VERSION 15 (2017)	VERSION 16 (2019)
Panel design	**Total number of genes**		58	74	153	203
	Design		Haloplex	Haloplex	Haloplex HS	Haloplex HS
	Read Length		250	150	150	150
	Target Region Size		382,058	374,978	618,664	918,841
	Total Amplicon		15,729	18,454	31,193	44,656
	Total Coverage (%)		98.99	99.67	99.50	99.16
	Tier		1	1	2	2
		DISEASES	**N genes**	**N genes**	**N genes**	**N genes**
Group of diseases covered	**1**	**Hereditary hemochromatosis and Hyperferritinemia**	9	9	10	11
	2	**Congenital sideroblastic anemias and acquired sideroblastic anemia**	14	14	14	16
	3	**Iron related Anemias (Aceruloplasminemia, Atransferrinemia, IRIDA, DMT1- deficiency anemia, Hypochromic microcytic Anemia with Iron overload 2)**	5	5	8	9
	4	**Congenital dyserythropoietic anemia**	6	6	6	9
	5	**Congenital Erythrocytosis/Familiar Polycythemia**	5	5	7	11
	6	**Hereditary hemolytic anemias due to RBC enzymopathies or glycogen storage disease**	17	17	17	17
	7	**Hereditary hemolytic anemias Membranopathies and Gilbert syndrome**	5	13	16	16
	8	**Erythropoietic Protoporphyria and Congenital Erythropoietic Porphyria**	2	3	4	6
	9	**Fanconi Anemia**	0	0	21	22
	10	**Diamond- Blackfan-Anemia, DBA1**	0	0	23	30
	11	**Dyskeratosis congenita**	0	0	12	15
	12	**NBIA, Neurodegeneration with brain iron accumulation**	0	0	10	32
	-	**Candidates genes**	-	9	12	14

1.2. MATERIAL AND METHODS

1.2.1. Design of the new panel (version 15)

An extensive search in several databases, including USCS Genome Browser, 1000 Genomes, dbSNP, HGMD, OMIM, and ORPHANET, has been done to obtain the most updated list of genes that are known to be mutated in anemia syndromes till the date of the design of the gene panel. Our custom-made targeted NGS panel (version 15) covers 153 genes known to be mutated in anemia syndromes, as detailed in Table 1.15. The panel includes 12 candidate genes, that are known to be related with HHD, although they have not been reported as a disease causative gene yet. Genes associated with hemoglobinopathies (HBB, HBA1, HBA2) were not included in the panel, due to the high homology and the consequent difficulty to design efficient probes that discriminate the sequencing between these genes. The target regions for sequencing include the exons, intron-exon boundaries (50pb), and 5'UTRs. The 3'UTR was entirely covered for 44 genes, while for the rest of the genes (n = 109) only 75pb was sequenced.

The panel was designed using SureDesign software (Agilent Technologies, Santa Clara, CA, USA) and optimized for HaloPlex for high-sensitivity sequencing technology (HaloPlex^{HS}), Illumina Platform and a read length of 150bp. The custom design was constructed under the hg19 human reference genome (Homo sapiens, UCSC hg19, GRCh37, February 2009). The custom panel includes a total of 31,193 amplicons with a target coverage of 99.63% of the total target region size (618.664 kbp). Some regions of interest with difficulties to design regular probes, were designed as Formalin-fixed paraffin-embedded (FFPE) to enrich the target sequencing and increase the target length coverage.

1.2.2. Validation of the NGS panel and diagnosis of new cases

All subjects gave their informed diagnostic and research consent for inclusion before they participated in the genetic analysis. The research was conducted in conformity with the Declaration of Helsinki, and the Ethics Committee approved the protocol on 10th July 2015.

1.2.2.1. Validation of the NGS panel

Twenty-six patients with inherited hematological diseases were used to carry out internal validation of the HHD panel version 13. All patients were previously genetically diagnosed by Sanger methodology. Samples were distributed as following: six for hereditary hemochromatosis (10010); four patients with CSA (10020); five samples from Iron and Cooper related diseases (10030); four from CDA (10040); 1 from CE (10050) and 6 from HA (10070).

1.2.2.2. Diagnosis of patients with HHD using panel v15

Moreover, the study included 42 patients, referred to the molecular laboratory of BloodGenetics (Esplugues de Llobregat, Spain) for genetic diagnosis due to the manifestation of any of the congenital hematologic diseases described before. The clinical data provided by the referring medical doctors included: the presumed diagnosis, current age, age at presentation, ethnicity, description of the peripheral blood and bone marrow morphology, other symptoms and the administered treatment (including blood transfusion, medical treatment and splenectomy).

Table 1. 15 Genes included in version 15 of the NGS target panel for the diagnosis of HHD.

Panel ID	Group	DISEASES	N genes					
10010	1	Hereditary hemochromatosis and Hypo/hyperferritinemia	10	*HFE*	*HFE2*	*HAMP*	*TFR2*	*ATP4A*
				SLC40A1	*BMP6*	*FTL*	*FTH1*	*GNPAT*
10020	2	Congenital sideroblastic anemias and acquired sideroblastic anemia	14	*ALAS2*	*SLC25A38*	*GLRX5*	*ABCB7*	*STEAP3*
				TRNT1	*PUS1*	*YARS2*	*SLC19A2*	*SF3B1*
				HSPA9	*LARS2*	*NDUFB11*	*MT-ATP6*	
10030	3	Iron and Copper related Anemias	8	*TF*	*CP*	*TMPRSS6*	*SLC11A2*	*STEAP3*
				ATP4A	*ACVR1*	*ATP7B*		
10040	4	Congenital dyserythropoietic anemia	6	*CDAN1*	*C15ORF41*	*SEC23B*	*KIF23*	*KLF1*
				GATA1				
10050	5	Congenital Erythrocytosis	7	*EPOR*	*VHL*	*EGLN1*	*EPAS1*	*BPGM*
				SH2B3	*JAK2*			
10061	6.1	Hereditary hemolytic anemias due to RBC enzymopathies or glycogen storage disease	17	*PKLR*	*G6PD*	*HK1*	*GPI*	*BPGM*
				GSR	*GSS*	*GPX1*	*GCLC*	*NT5C3A*
				AK1	*PGK1*	*TPI1*	*PFKM*	*ALDOA*
				ADA	*CYB5R3*			
10062	6.2	Hereditary hemolytic anemias Membranopathies	16	*ANK1*	*SPTB*	*SPTA1*	*SLC4A1*	*EPB42*
				UGT1A1	*EPB41*	*GYPC*	*PIEZO1*	*KCNN4*
				SLC2A1	*ABCB6*	*RHAG*		
				ABCG5	*ABCG8*	*XK*		
10070	6.1+6.2	Hereditary hemolytic anemias including membranopathies and enzymopathies	33	*PKLR*	*G6PD*	*HK1*	*GPI*	*BPGM*
				GSR	*GSS*	*GPX1*	*GCLC*	*NT5C3A*
				AK1	*PGK1*	*TPI1*	*PFKM*	*ALDOA*
				ADA	*CYB5R3*	*ANK1*	*SPTB*	*SPTA1*
				SLC4A1	*EPB42*	*UGT1A1*	*EPB41*	*GYPC*
				PIEZO1	*KCNN4*	*SLC2A1*	*ABCB6*	*RHAG*
				ABCG5	*ABCG8*	*XK*		
10080	8	Erythropoietic Protoporphyria and Congenital Erythropoietic Porphyria	4	*FECH*	*ALAS2*	*UROS*	*CLPX*	
10090	9	Fanconi Anemia	21	*FANCA*	*XRCC2*	*FANCI*	*RAD51C*	*FANCM*
				FANCB	*FANCE*	*BRIP1*	*SLX4*	*MAD2L2*
				FANCC	*FANCF*	*FANCL*	*ERCC4*	*FANCD2*
				BRCA2	*FANCG*	*PALB2*	*UBE2T*	*RFWD3*
				BRCA1				
10100	10	Diamond- Blackfan-Anemia, DBA1	23	*RPS19*	*RPL11*	*RPL15*	*RPS28*	*RPS27A*
				RPS24	*RPS7*	*RPS29*	*RPL27*	*RPL36*
				RPS17	*RPS10*	*TSR2*	*RPL31*	*RPL9*
				RPL35A	*RPS26*	*GATA1*	*RPS27*	*RPS15*
				RPL5	*RPL26*	*RPL19*		
10110	11	Dyskeratosis congenita	12	*TERC*	*NHP2*	*NOP10*	*RTEL1*	*USB1*
				TERT	*TINF2*	*WRAP53*	*DKC1*	*CTC1*
				PARN	*ACD*			
10120	12	NBIA, Neurodegeneration with brain iron accumulation	10	*PANK2*	*FTL*	*WDR45*	*FA2H*	*CP*
				PLA2G6	*C19orf12*	*COASY*	*ATP13A2*	*DCAF17*
		Candidate genes	12	*ACO1*	*ATPIF1*	*EXOC6*	*HEPH*	*IREB2*
				Pim1	*SLC25A37*	*SLC25A39*	*TFRC*	*NCOA4*
				PCBP1	*BMP2*			
		TOTAL		**153 unique genes* (141 without candidate genes)**				

*Note: some genes are implicated in different subpanels. The total number of unique genes is displayed at the end of the table.

1.2.2.3. DNA extraction

Genomic DNA was extracted from peripheral blood using the Maxwell® Blood DNA Purification (Promega), FlexiGene DNA kit (Qiagen, Hilden, Germany) or QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

DNA samples were quantified by using the QuantiFluor® dsDNA System (Promega), and fluorescence was detected using Quantus™ Fluorometer (Promega). Nanodrop concentration was also calculated and DNA quality was accepted if the A260/280 ratio and A260/230 ratio were between 1.7 and 1.9

Integrity evaluation of DNA extracts was performed by electrophoresis in 0.8 % agarose gels. 50 ng of the DNA sample was loaded on the gels individually. Electrophoresis was run for 60 min at 80 V in Tris-Borate-EDTA (TBE) buffer. Gels were stained with 2.5 µl MIDORI^{GREEN} Advance (NIPPON Genetics Europe, Dueren, Germany) and captured under UV light (NuGenius, Sygene, Cambridge, United Kingdom).

1.2.2.4. Library preparation and Sequencing

The capture of genomic regions was conducted starting from 225 ng of gDNA using a custom design HaloPlex HS Target Enrichment kit (Agilent Technologies, Santa Clara, CA) Tier 2 (501-2599 bp; up to 200k probes) according to the manufacturer's instructions. The quality of the library was determined using the Agilent 2100 Bioanalyzer and high sensitivity DNA chips (Agilent Technologies Inc., Palo Alto, CA, USA). Libraries were sequenced using MiniSeq Mid Output Kit (300 cycles) (Illumina, San Diego, CA) on an Illumina MiniSeq sequencer (Illumina, San Diego, CA), generating paired-end reads of 150bp length. Samples were aligned using the GRCh37/hg19 reference human genome and data analysis was performed employing our algorithms. Conventional Sanger sequencing was performed to confirm the mutations detected by our NGS panel.

1.2.2.5. Results, Interpretation, and Reporting

Sequencing reads were aligned against the reference genome GRCh37/hg19 and variants were called and annotated using the SureCall software (v.3.5.1.46; Agilent Technologies). Variant filtering was established using our own variant database. Filtering cut-off values were as follows: variant call quality threshold >100; minor allele frequency (MAF) < 3%; deep read > 50; variant allele frequency of > 25%; primary effect: missense, stop gained, stop lost, initiator codon, frameshift, in frame-insertion, in-frame deletion and splicing. Filtered variants were visualized in genomic context IGV Viewer tool integrated in SurCall software.

BloodGenetics Laboratory classifies sequencing variants following the guidelines of the American College of Medical Genetics (ACMG) Laboratory Practice Committee Working Group (Richards et al. 2015; Ogino et al. 2007). The clinical significance of each variant was estimated by the presence of the variant in known databases (ClinVar, HGMD, dbSNP), but also in our own database based on deep search on publications about HHD. The degree of evolutionary conservation of the encoded amino acid, and the MAF was also considered to determine pathogenicity. *In silico* analysis was also done to predict the pathogenicity of a variant according to its potential to modify the protein function. These bioinformatics tools include: SIFT (https://sift.bii.a-star.edu.sg/), PROVEAN (http://provean.jcvi.org/index.php), and Polyphen2 (http://genetics.bwh.harvard.edu/pph2/). Variant of Unknown significance (VUS) were reported if found in genes relevant to the primary indication for testing (Richards et al. 2015; Ogino et al. 2007). Progeny Laboratory Information Management System (LIMS, Progeny, USA) was used to annotate and organize patients' samples and to include all the obtained results.

1.2.2.6. Variant validation with Sanger

Variants identified by the HHD panel reported as pathogenic or likely pathogenic were validated by Sanger. The region of DNA containing the variant (~300pb) was amplified using 50-100 ng of genomic DNA. The resulting amplification products were verified on a 2% agarose gel, dyed with MIDORIGREEN. PCR products were purified by mixing 5ul of PCR product with 1 µl of FastAP Thermosensitive Alkaline Phosphatase and 0.5 µl Exonuclease I (Thermo Fisher) and incubating the mixture 15 minutes at 37ºC and 15 minutes at 95 ºC. The purified PCR products were sequenced using a Sanger sequencing service provided by GATC Biotech (Konstanz). Sequencing results were analyzed using Mutation Surveyor software (SoftGenetics LLC, PA, USA).

All primers were designed using Primer 3 software (http://bioinfo.ut.ee/primer3-0.4.0/), acquired in 25nmole, desalted and dry from Invitrogen (Thermo Fisher Scientific) and diluted in MilliQ water to a 100µM working solution.

1.2.3. External validation (EMQN-EQA)

HHD NGS Panel was externally validated by participating in the external quality assessment (EQA) scheme for germline NGS analysis provided by the European Molecular Genetics Quality Network (EMQN). The “DNA Sequening – NGS (vGermline)” scheme is designed specifically for labs doing NGS based germline testing. One germline DNA sample was provided by the organization to test the NGS strategy.

1.2.4. RNA Fold Predictions for 5’UTR FTL variant

The RNA folds were analyzed with the Sfold web server (http://sfold.wadsworth.org/) (Ding, Chan, and Lawrence 2004) to predict IRE structure of wild-type (WT) as well as mutated FTL-IRE. Supplementary Material Table S2 shows the reference DNA sequences use to make the predictions. Ion-responsive elements were also predicted with SIREs web server tool (Campillos et al. 2010).

1.3. RESULTS

1.3.1. The final design (genes and diseases)

The HHD Panel compromise a total of 153 genes, including exons, exon/intron boundaries, 5'UTR regions and 3'UTR regions (except some genes that have been shorted). The total target region size is 618.664 kbp, being covered the 99.63% for 31193 amplicons. The design covers the exons of all isoforms of the genes. As an illustrative example, Figure 1.4 shows the covered regions and the amplicons that cover all the exon from three isoforms of *ALAS2* gene and the five isoforms of *ALDOA* gene.

Not all genes described in the introduction were included in the panel v15. Genes involved in the development of Familial Erythrocytosis due to defects in globin proteins, were finally not included in the panel. Because of the high homology of the genes, that alpha-globin and beta-globin proteins (*HBB*, *HBA1* and *HBA2* genes), the SureCall software is not able to design probes for these genes. The Figure 1.5 presents in red the missing regions for sequencing of HBB and HBA2, which exon sequences are almost completely missed.

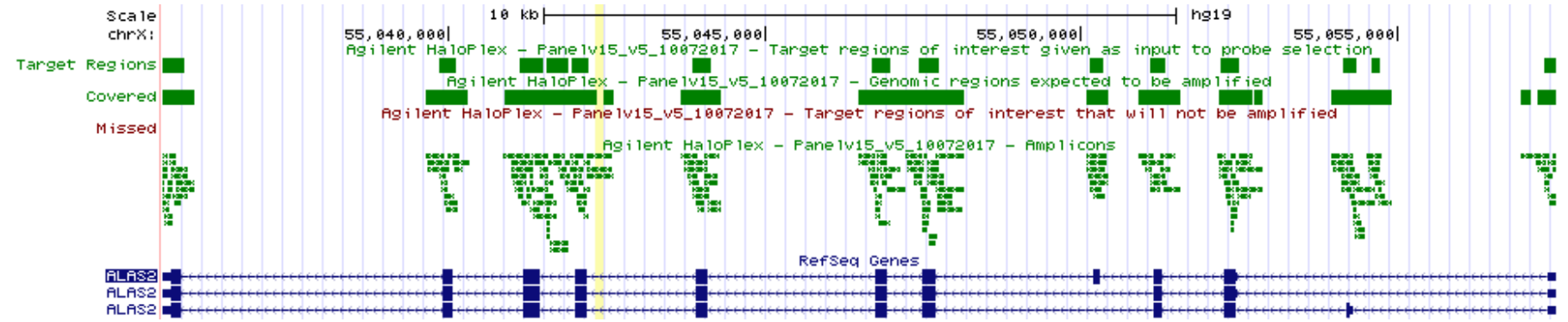
Some genes known to be causative genes of HHD were not included in our NGS panel since the publication of these genes was posterior to the date of design of the panel. Those genes are included in the version 16 of the panel. Other genes have been identified to be associated with HHD, but they remain confidential until they are published in a scientific paper.

Our version 15 of the NGS panel for HHD cover 75.36% (153 out of 203 genes) of the genes that are known to be associated with HDD to present day and are part of the NGS panel version 16 (Table 1.16). Apart from the genes described in the introduction, new genes have been identified to be causative of HHD, but their confidentiality must be maintained until publication.

The less covered subpanel of the panel 15 compared with the version 16 is NBIA with 31.25 % (10 out of 32 genes) of coverage (Table 1.16). DBA and FA are the two-following diseases with less gene coverage (88.46% and 95.45% respectively). Four subpanels were uncovered with less than 80% compared with version 16: the congenital erythrocytosis with a 63.63% of coverage; CDAs with 66.67%; erythropoietic protoporphyria with 63.63% and DBA with 76.67%. The rest of disease were highly covered (more than 80%), and only the hemolytic anemias due to enzymopathies or membranopathies were equally covered in the versions 15 and 16 (100%).

Example of genes completely covered by Haloplex probes

A – *ALAS2* gene



B- *ALDOA* gene

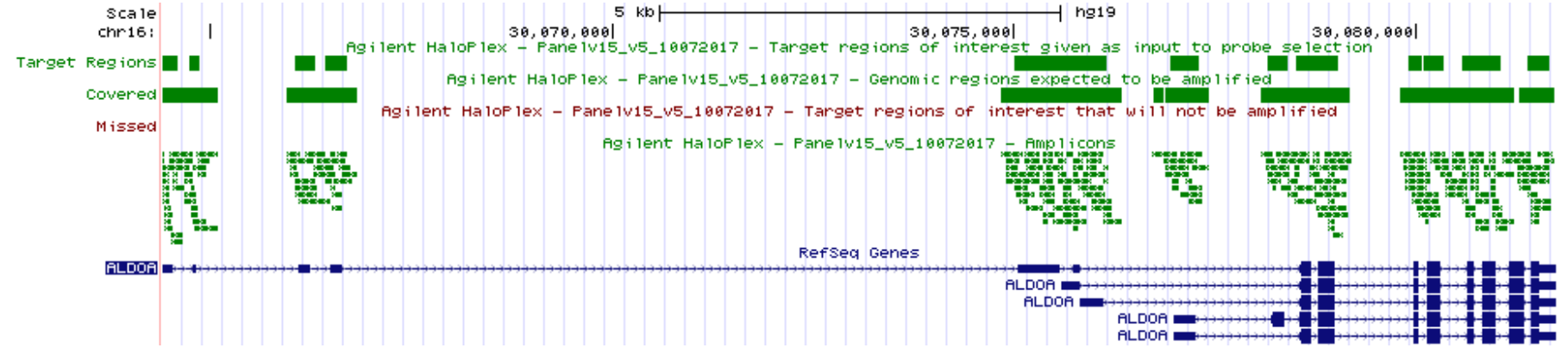
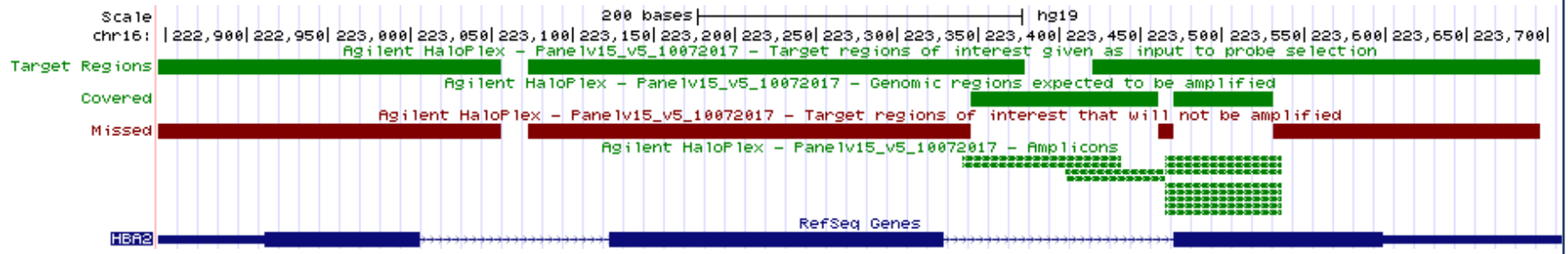


Figure 1.4. Example of two genes (*ALAS2* and *ALDOA*) entirely covered by Haloplex probes. In green are the "Target Regions", regions of interest to be sequenced; the "Covered" regions, the region covered by the panel; and the "Agilent Haloplex – Amplicons". In these two genes, the Haloplex probes cover all the exons and intron boundaries of the genes.

Example of genes not covered by Haloplex probes

A – *HBA2* gene



B – *HBB* gene

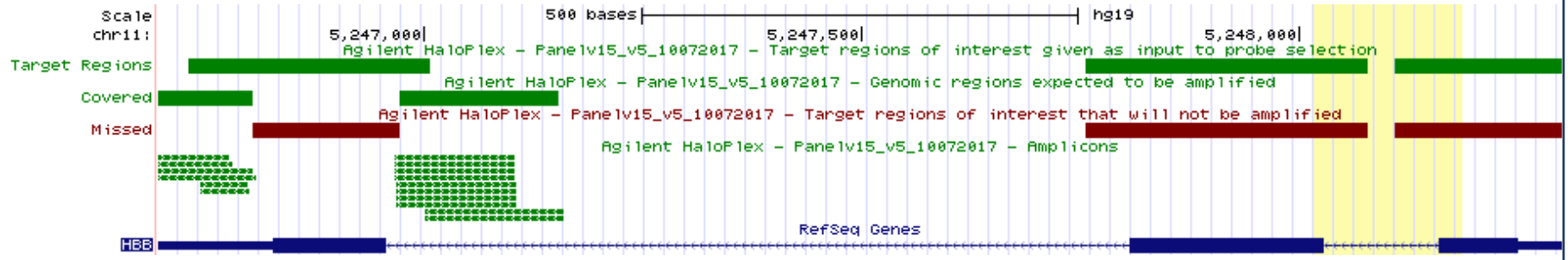


Figure 1.5. Example of genes (HBA2 and HBB) not covered by Haloplex probes. In green are the "Target Regions", regions of interest to be sequenced; the "Covered" regions, the region covered by the panel; and the "Agilent Haloplex – Amplicons". In red are the "Missed" regions, the target regions of interest that will not be covered by the panel. In these two genes, a large part of the region of interest is not covered by the amplicons designed by Agilent Haloplex.

Table 1. 16 Coverage of the HHD panel 15 respect the total of known causative genes for the time being.

Panel	Group	DISEASES	Nº genes in HHD panel (v15)	Nº genes in updated HHD panel (v16)	% coverage
10010	**1**	**Hereditary hemochromatosis and Hyperferritinemia / hypoferritinemia**	10	11	90,90%
10020	**2**	**Congenital sideroblastic anemias and acquired sideroblastic anemia**	14	16	87,50%
10030	**3**	**Iron and Copper related Anemias**	8	9	88,89%
10040	**4**	**Congenital dyserythropoietic anemia**	6	9	66,67%
10050	**5**	**Congenital Erythrocytosis/Familiar Polycythemia**	7	11	63,63%
10061	**6.1**	**Hereditary hemolytic anemias due to RBC enzymopathies or glycogen storage disease**	17	17	100,00%
10062	**6.2**	**Hereditary hemolytic anemias membranopathies and Gilbert syndrome**	16	16	100,00%
10070	**6.1+6.2**	**Hereditary hemolytic anemias including membranopathies and enzymopathies**	33	33	100,00%
10080	**8**	**Erythropoietic Protoporphyria and Congenital Erythropoietic Porphyria**	4	6	66,69%
10090	**9**	**Fanconi Anemia**	21	22	95,45%
10100	**10**	**Diamond- Blackfan-Anemia, DBA1**	23	30	76,67%
10110	**11**	**Dyskeratosis congenita**	12	15	80,00%
10120	**12**	**Neurodegeneration with brain iron accumulation (NBIA)**	10	32	31,25%
		Candidates genes	12	14	85,71%
		Total number of unique genes (no candidate genes included)	**153**	**203**	**75,36%**

1.3.2. Patients analyzed for validation (Internal validation)

The NGS panel version 13 was validated by including 26 cases with known mutation by Sanger methodology (Table 1.17). In 25 out of 26 cases analyzed (96.15%), the previously described variant was detected (Table 1.18). In one case, the mutation detected by sanger (NM_000551.3 c.[376G>A]:[376G>A]; NP:000542.1 p.(Asp126Asn);(Asp126Asn)) was not validated by the panel, since the variant was located in a GC-rich region (~60%), and is not covered by the Haloplex panel (see Figure 1.6). This data was reported as a selected poster at the EHA congress 2018 Stockholm (see Appendix-2.1. Congress presentations).

Table 1. 17 Samples used for validation through NGS panel version 15 including additional findings

PANEL	SAMPLE	DISEASE	GENE	Ref GENE	Ref PROTEIN	SANGER RESULT (HGVS nomenclature)	VALIDATED BY THE PANEL	ADDITIONAL RELEVANT FINDINGS USING THE PANEL
10010	P-00003	Hereditary Hemochromatosis	TFR2	NM_003227.3	NP_003218.2	Mut 1: c.[916C>T];[=]; p.(Gln306Ter);(=)	Yes	
			TFR2	NM_003227.3	NP_003218.2	Mut 2: c.[2014C>T];[=]; p.(Gln672Ter);(=)		
10010	P-00017	Hereditary Hemochromatosis	HFE	NM_000410.3	NP_000401.1	Mut 1: c.[845G>A];[845G>A]; p.(Cys282Tyr);(Cys282Tyr)	Yes	
10010	P-00021	Hereditary Hemochromatosis	HFE	NM_000410.3	NP_000401.1	Mut 1: c.[845G>A];[=]; p.[(Cys282Tyr)];[(=)]	Yes	
10010	P-00020	Hereditary Hemochromatosis	SLC40A1	NM_014585.5	NP_055400.1	Mut 1: c.[*1114T>A];[=]; p.(?) 3'UTR variant	Yes	
10010	P-00032	Hereditary Hemochromatosis	HFE2	NM_213653.3	NP_998818.1	Mut 1: c.[445delG];[=]; p.(Asp149Thrfs*97);(=)	Yes	
10010	P-00018	Hyperferritinemia with cataracts	FTL	NM_000146.3	NP_000137.2	Mut 1: c.[36_42delCAACAGT];[=]; p.(?) IRE mutation	Yes	
10020	P-00005	Congenital sideroblastic anemia	SLC25A38	NM_017875.2	NP_060345.2	Mut 1:c.[1552T>C];[=]; p.(Leu 51Pro);(=)	Yes	
			SLC25A38	NM_017875.2	NP_060345.2	Mut 2: c.[559C>T];[=]; p.(Arg187Ter);(=)	Yes	
10020	P-00016	Congenital sideroblastic anemia	ALAS2	NM_000032.4	NP_000023.2	Mut 1: c.[497A>T];[0]; p.(Lys166Met);(0)	Yes	
10020	P-00023	Acquired sideroblastic anemia	SF3B1	NM_012433.1	NP_036565.2	Mut1c.[2098A>G];[=]; p.(Lys700Glu);(=)	Yes	
10020	P-00008	Acquired sideroblastic anemia	SF3B1	NM_012433.1	NP_036565.2	NO MUTACION (negative control)	Yes	
10030	P-00004	Aceruloplasminemia	CP	NM_000096.3	NP_000087.1	Mut 1: c.[1783_1787delGATAA];[=]; p.(Asp595Tyrfs*2);(=)	Yes	
			CP	NM_000096.3	NP_000087.1	Mut 2: c.[2520-2523_delAACA];[=]; p.(T841Rfs*52);(=)	Yes	
10030	P-00014	Atransferrinemia	TF	NM_001063.3	NP_001054.1	Mut 1: c.[665_668delCCTT];[=] ; p.(Ala222Valfs*32);(=)	Yes	
10030	P-00015	Atransferrinemia	TF	NM_001063.3	NP_001054.1	Mut 1: c.[665_668delCCTT];[=] ; p.(Ala222Valfs*32);(=)	Yes	
			TF	NM_001063.3	NP_001054.1	Nothing else found by Sanger		Var 1: c.[*53A>G];[=] ; p.(?);(=). In 3'UTR
10030	P-00010	Anemia with iron-overload	SLC11A2	NM_000617.2	NP_000608.1:	Mut 1: c.635G>T; [=]; p.Gly212Val; [=]	Yes	Mut 3: c.[1336C>T]; [=]; p.(Gln446Ter);(=)
			SLC11A2	NM_001174125.1	NP_001167596.1	Mut 2: c.49+3A>T Splicing defect	Yes	
10030	P-00002	IRIDA	TMPRSS6	NM_153609.2	NP_705837.1	Mut 1: c.[76_80delGGTGA];[=]; p.(Gly26Trpfs*14);(=)	Yes	
			TMPRSS6	NM_153609.2	NP_705837.1	Mut 2: c.[1817T>G];[=]; p.(Leu606Arg);(=)	Yes	

(continuation Table 1.17)

PANEL	SAMPLE	DISEASE	GENE	Ref GENE	Ref PROTEIN	SANGER RESULT (HGVS nomenclature)	VALIDATED BY THE PANEL	ADDITIONAL RELEVANT FINDINGS USING THE PANEL
10040	P-00001	Congenital dyserythropoietic anemia	CDAN1	NM_138477.2	NP_612486.2	Variant1c.[2407+21G>A];[=]; p.(?)	Yes	
10040	P-00035	Congenital dyserythropoietic anemia	C15ORF41	NM_032499.5	NP_115888.1	Variant1c.[423-56T>C]Yes	Yes	
10040	P-00031	Congenital dyserythropoietic anemia	SEC23B	NM_006363.4	NP_006354.2	Mut 1: c.[1254T>G];[=]; p.(Ile418Met);(=)	Yes	
			SEC23B	NM_006363.4	NP_006354.2	Mut 2: c.[1512-2A>G];[=]; p.(?) Splicing defect	Yes	
10040	P-00034	Congenital dyserythropoietic anemia	SEC23B	NM_006363.4	NP_006354.2	Mut 1: c.[2073dupGATC];[=]; p.(Asp692Glyfs*2);(=)	Yes	
10050	P-00019	Familiar erythrocytosis/Polycythemia	VHL	NM_000551.3	NP_000542.1	Mut 1: c.[376G> A];[376G>A]; p.(Asp126Asn);(Asp126Asn)	No	Region not covered by the HHD panel
10070	P-00011	Hemolytic anemia due to GPI deficiency	GPI	NM_000175.3	NP_000166.2	Mut 1: c.[341A>T]; [=]; p.(Asp114Val);(=)	Yes	
			GPI	NM_000175.3	NP_000166.2	Mut 2: c.663T>G; p.(Asn221Lys);(=)	Yes	
10070	P-00024	Hemolytic anemia due to GPI deficiency	GPI	NM_000175.3	NP_000166.2	Mut 1: c.[145G>C];[=]; p.(Gly49Arg);(=)	Yes	
			GPI	NM_000175.3	NP_000166.2	Mut 2: c.[921C>A];[=]; NP_000166.2: p.(Phe307Leu);(=)	Yes	
10070	P-00025	Hemolytic anemia	BPGM	NM_001724.4	NP_001715.1	Mut 1: c.[269G>A];[=]; p.(Arg90His);(=)Yes	Yes	
10070	P-00026	Hemolytic anemia	NT5C3A	NM_016489.12	NP_001002009.1	Mut 1: c.[576delG];[576delG]; p.(Asp193Ilefs*15);(Asp193Ilefs*15)	Yes	
10070	P-00080	Hereditary hemolytic anemia =-NADH diaphorase deficiency	CYB5R3	NM_000398.6	NP_000389.1	Mut 1: c.[637G> A];[637G>A]; p.(Glu213Lys);(Glu213Lys)	Yes	
10070	P-00081	Hereditary eliptocytosis with Hereditary spherocitosys	SPTA1	NM_003126.2	NP_003117.2	Mut 1: c.[460_462dupTTG];[=]; p.(Leu155dup);(=)	Yes	
			SLC4A1	NM_000342.3	NP_000333.1	Nothing else found by Sanger		Var 1: c.[1530C>G];[=] ; p.(Ser510Arg);(=)
			EPB42	NM_000119.2	NP_000110.2	Nothing else found by Sanger		Var 2: c.[1654C>A];[=]; p.(leu552Ile);(=)

Table 1. 18 Validation results

Clinical diagnosis	Subpanel	N	Patients with genetic diagnosis	Variants		Identified gene mutations
				Sanger	Panel	
HH + HHCS	10010	6	6 (100%)	7	7	*FTL; HFE; TFR2; HFE2; SLC40A1*
CSA	10020	4	4 (100%)	4	4	*ALAS2; SLC25A38; SF3B1*
Fe/Cu metabolism	10030	5	5 (100%)	8	10	*CP; TF; SLC11A2; TMPRSS6*
CDA	10040	4	4 (100%)	5	5	*CDAN1; C15ORF41; SEC23B*
Erythrocytosis	10050	1	0 (0%)	1	0	*VHL*
HA	10070	6	6 (100%)	8	10	*GPI; BPGM; NT5C3A; CYB5R3; SPTA1; SLC4A1; EPB42*
Total cohort		26	25 (96.15%)	32	37 (115%)	

F, Female; M, Male; VUS, Variants of Unknown significance

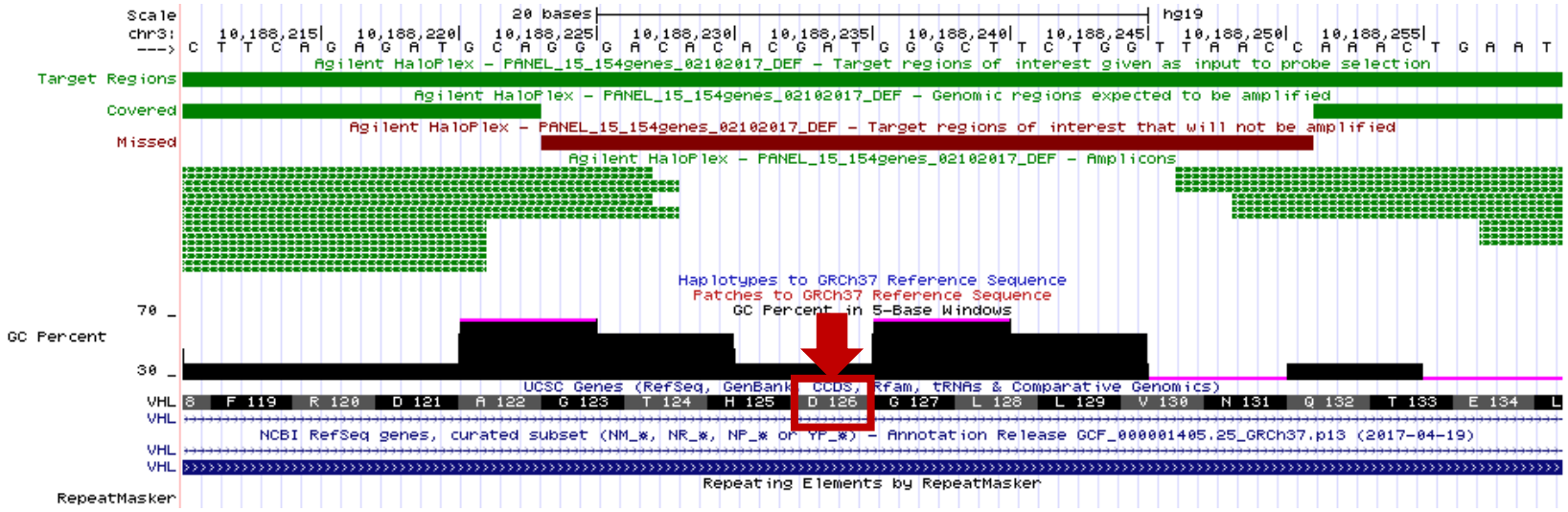


Figure 1.6. Region of *VHL* gene (exon 2) not covered by the HHD panel. The mutation NM_000551.3: c.[376G> A]; p.(Asp126Asn) is situated in a missed region of the panel v15, and this region is not amplified and sequenced probably due to the high percentage of GC (~60%).

1.3.3. External Validation (EQMN – EQA)

The HHD NGS panel version 15 was satisfactorily validated by participating in the 2018 European Molecular Genetic Quality Network (EMQN) external quality scheme of DNA sequencing – NGS (vGermline). The certificate is attached in Appendix – Other documents page 216 of this dissertation.

1.3.4. Diagnosis

A total of 42 patients were analyzed using the HHD Panel, including 13 females and 29 males. Patients were referred from various medical centers mainly from Spain but also from other countries. The mean age was 46.97 years with a range of 1-82 years.

The initial diagnosis before genetic testing was hereditary hemochromatosis in 18 patients, hyperferritinemia with cataract syndrome in five patients (therefore, 23 patients from subpanel 10010), congenital sideroblastic anemia in five patients (10020), iron-related anemias in five patients (10030), congenital erythrocytosis in two patients (10050), one patient for hemolytic anemia (10062), one case for Hemolytic anemia (10070), one case for Fanconi Anemia (10090) and two cases with NBIA (10120). More than 50% of the cases were patients with hereditary hemochromatosis or hyperferritinemia (subpanel 10010) (see Figure 1.7).

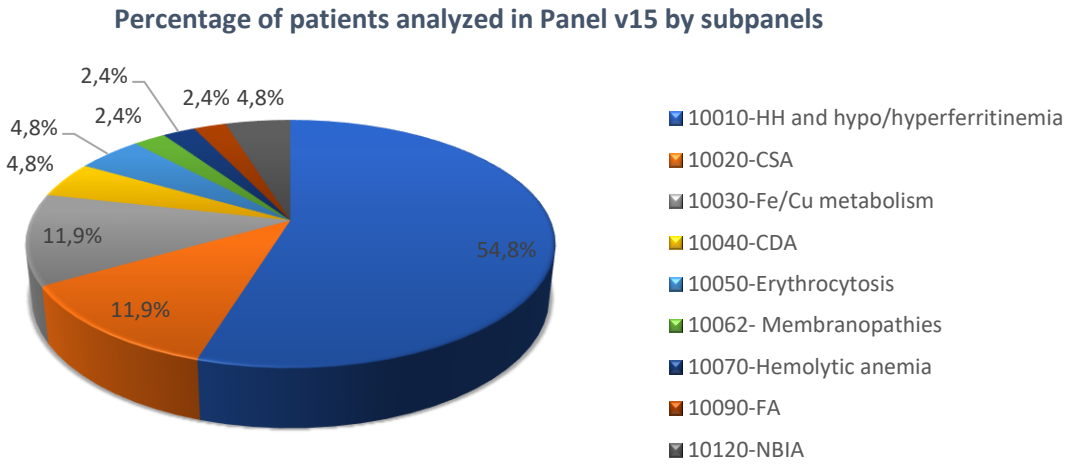


Figure 1.7. Proportion of patients analyzed in the HHD Panel (version 15) by group of disease (subpanel). The percentage of patients analyzed for each subpanel respect the total number of patients analyzed in the HHD panel version 15 is represented in the pie chart. The different subpanels are represented in different colors. The largest group belongs to subpanel 10010 (in blue) with 54.8% of the patients analyzed for HH or hypo/hyperferritinemia. CSA (subpanel 10020 in orange) and anemias due to Fe/Cu metabolism (subpanel 10030 in grey) are the following most popular groups of disease (11.9%).

Overall, 27 out of the 42 patients (64.3%) who underwent HHD panel sequencing, were found a significant variant, but only 18 (42.9%) cases were solved by obtaining a pathogenic or likely pathogenic variant agree with the mode of inheritance of the disease that can explain the cause of the disorder (category A, see Figure 1.8). In 3 patients (7.1%), only one pathogenic variant was found in heterozygous state, but the disease is autosomal recessive, so the genetic diagnosis is not resolved (category B). In 14.3% of the cases (6 patients), we found variants of unknown significance (category C). VUS variants are rare variants (present in less than 3% of the population), but the evidence that we have is insufficient to classify them as pathogenic or likely pathogenic variants. However, they

could be implicated in the disease as mutations in those genes cause the disease that is present in the patient (there is a correlation with the clinical presentation of the disease and a genetic defect in the gene where we detect the VUS). These variants need to be further investigated with functional analysis. The 35.7% remaining were negative results: no variant was found in the analyzed genes, or benign or likely benign variant was detected.

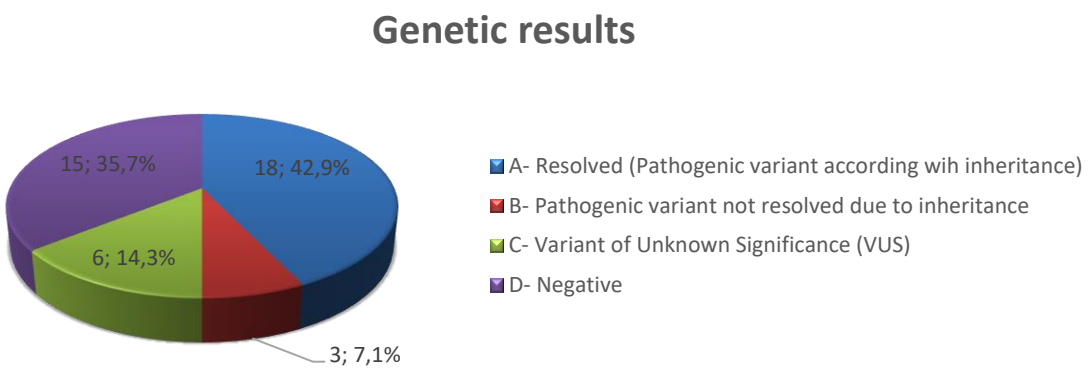


Figure 1.8. Genetic results of the patients analyzed by the HHD panel (v15). The results were categorized in four groups: A, patients with pathogenic or likely pathogenic variants that fit with the mode of inheritance of the disease; B, patients with pathogenic variants that do not agree with the mode of inheritance; C, Variants of Unknown significance (VUS); and D, Negative results (no variants found, or benign or likely benign variant detected).

Six runs were performed to evaluate the 42 diagnostic samples and seven samples were analyzed per run on average (min=5, max=11), using the Mid Output Kit and MiniSeq Instrument both from Illumina. The average output size is 4.29 Gb, and the mean Qscore (>=30) was 89. The HHD panel v15 acquires high read depth of the regions of interest, with an average depth coverage of 388x, and a mean read length of 115 pb. On average, the total number of reads detected per run is 20,498,214 (2,729,438 per sample) with a cluster density of 368 K/mm^2 (the first run was overclustered with 1037 K/mm^2, and the rest of runs ranged from 147-358). The total number of variants found per sample was 562 on average, including all the genes from all the subpanels. Between 3 and 6% of the total variants had low depth coverage (<50).

The subpanel 10010 (HH and hypo/hyperferritinemia) was the group with more genes identified (see Table 1.19), probably because it was the most demanded with the highest number of samples analyzed. Thirteen out of the 23 patients from this subpanel (56.52%) had a pathogenic variant, likely pathogenic variant or a VUS. The rest of subpanels were more dispersed, with mutation detection ranging from 0 to 100% due to the low number of samples. The total variant detection is 57.14%, highly influenced by the subpanel 10010.

42.9% of the patients obtained a genetic result with a pathogenic or likely pathogenic variant that can explain the disease phenotype. Some patients, mostly from Hereditary hemochromatosis, have no mutations or only a VUS was identified. However, negative results are also very relevant. First, negative results are used to discard the presence of a genetic disease in an individual (disease could be acquired and without a genetic component). Second, in negative result cases other genes of the panel, including candidate genes, can be tested to try to identify new genes implicated in human diseases. If no variants are found with the panel, whole exome sequencing is performed for some cases (research studies), including the mother and the father if it is possible (trio analysis), to try to find the causative gene of the disease in the proband.

Table 1. 19 Summary of the genetic diagnosis by HHD Panel (version 15)

Clinical diagnosis	Patients	Gender		Range age	Average age	Categories				Patients with genetic result (A + B + C)	Identified gene mutations
		F	M			A	B	C	D		
10010-HH and hypo/hyperferritinemia	23	7	16	10-82	56.27	9	1	3	10	13 (56.52%)	FTL; HFE; TFR2; HAMP; BMP6
10020-CSA	5	1	4	25-38	31.5	3	0	1	1	4 (80%)	ALAS2; SF3B1
10030-Fe/Cu metabolism	5	2	3	1-78	32.4	2	0	1	2	3 (60%)	SLC11A2; TMPRSS6
10040-CDA	2	1	1	31-38	34.5	0	2	0	0	2 (100%)	SEC23B
10050-Erythrocytosis	2	0	2	1-30	15.5	1	0	0	1	1 (50%)	EPAS1
10062- Membranopathies	1	0	1	47	47	1	0	0	0	1 (100%)	SPTA1
10070-Hemolytic anemia	1	1	0	72	72	0	0	0	1	0 (0%)	
10090-FA	1	0	1	37	37	1	0	0	0	1 (100%)	FANCA
10120-NBIA	2	1	1	7-59	33	1	0	1	0	2 (100%)	CP; DCAF17
Total cohort	**42**	**13**	**29**	**1-82**	**46.97**	**18**	**3**	**6**	**15**	**27 (64,28%)**	

F, Female; M, Male; A, Patients with genetic diagnosis resolved (pathogenic or likely pathogenic variant according with inheritance); B, Pathogenic variant not resoled due to inheritance; C, Variant of Unknown Significance (VUS); D, Negative result (no variant found or benign or likely benign variant detected

1.3.5. L-Ferritin: One Gene, Five Diseases.

We have focused part of the research on studying patients with variants in *FTL* gene, in order to have an overview and update of the different diseases that mutations in this gene cause. The results here exposed were published in the journal *Pharmaceuticals* (see Appendix-3. Publications, page 193 for a copy of the manuscript).

Pharmaceuticals (Basel). 2019 Jan 23;12(1). pii: E17. doi: 10.3390/ph12010017.
"L-Ferritin: One Gene, Five Diseases; from Hereditary Hyperferritinemia to Hypoferritinemia-Report of New Cases."
***Cadenas B**, Fita-Torró J, Bermúdez-Cortés M, Hernandez-Rodriguez I, Fuster JL, Llinares ME, Galera AM, Romero JL, Pérez-Montero S, Tornador C, Sanchez M.*

Three probands were studied due to the suspicion of mutation in *FTL* gene (hyperferritinemia or hypoferritineamia). We completely sequenced the entire coding region, intron-exon boundaries and 5' and 3' regulatory regions for the *FTL* gene either by Sanger sequencing or by next generation sequencing (NGS).

Family 1 - a case with autosomal dominant L-ferritin deficiency

Proband II.1 from family 1 shown in Figure 1.9 A is a Spanish female aged four who was referred to the Pediatric Oncohematology Unit of the Virgen de la Arrixaca University Hospital due to refractory hypoferritinemia, with serum ferritin levels of 4-9 ng/mL (Table 1.20). The proband was not responsive to the oral administration of iron capsules but exhibited no further symptoms. A physical examination determined her weight and size were normal for her age. When she was six years old, the proposita began to experience severe migraines. CT and MRI scans of the brain uncovered a small subcortical area of gliosis located in the right frontal lobe; everything else was normal. This discovery was incidental since pediatric neurologists had initially diagnosed the proband with a primary headache and suggested flunarizine as treatment.

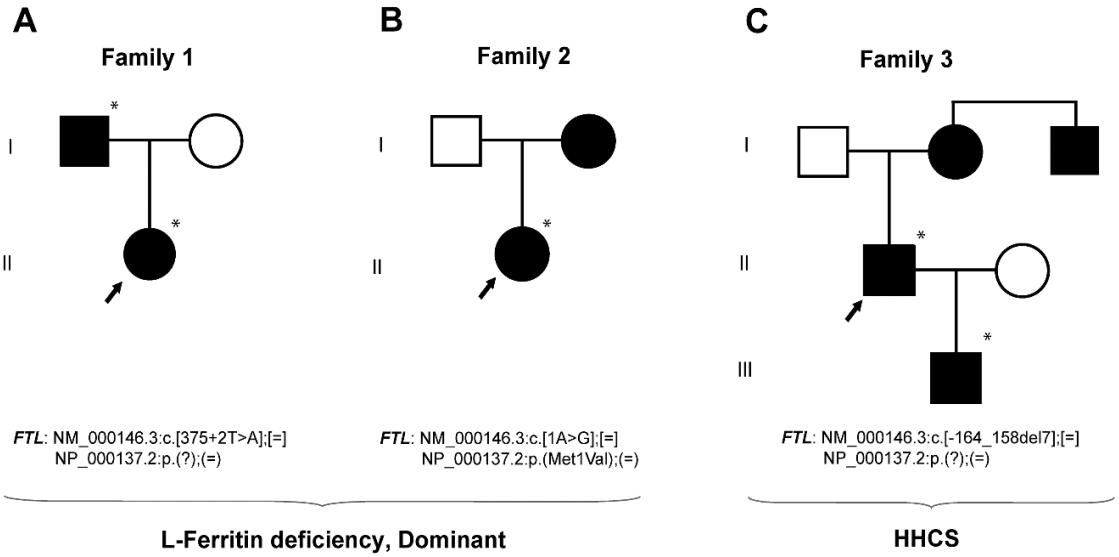


Figure 1.9. Pedigree of the three analyzed families affected from dominant L-ferritin deficiency and HHCS. Squares indicate males and circles females. Filled symbols indicate affected members and asterisks indicate subjects with genetic studies done at BloodGenetics SL. Probands are pointed with an arrow. Mutations are named according the HGVS nomenclature.

Mutation testing unveiled an A>T substitution in intron 3 of the *FTL* gene, specifically at position 375 + 2 (NM_000146.3:c.[375 + 2T > A][=], HGSV nomenclature). Proband I.1, the father of proposita II.1, also exhibited this mutation in a heterozygous state. As predicted by Human Splicing Finder, the variant disrupts mRNA splicing by modifying the wild type splicing donor site. This mutation has not been mentioned in literature or in public databases (e.g. NCBI, ENSEMBL, HGMD, and 1000Genomes). Nevertheless, the SNP database has identified another variant at the same position (c.375 + 2T > C) known as rs1371561306. According to the TOPMed database, the allele frequency of this other mutation is very low (MAF = 0.000008), but its precise clinical implications remain a mystery.

Table 1. 20 Genetic and clinical features of the probands.

Case	Family 1	Family 2	Family 3	Reference Values
Patient	II.1	II.1	II.1	-
Gender	F	F	M	-
Age at diagnosis (years)	4	2	67	-
Hb (g/dL)	12.2-13.3	13.1-13.7	14.0	13.5-17.5 (M); 12.1-15.1 (F)
MCV (fL)	78-84	80	90.2	80-95
Ferritin (ng/mL)	4-9	2-7	3037	12-300 (M); 12-200 (F)
Transferrin sat (%)	12.9	17.2-26.2	22.0-41.0	25-50
Iron (µL/dL)	n/a	61.95	46	49-226
Mutation	c.375+2T>A	p.Met1Val	c.-164_158del7	-
	Novel	Previously reported (Cremonesi et al. 2004)	Novel	-
Disease	L-ferritin deficiency	L-ferritin deficiency	HHCS	-
Inheritance	AD	AD	AD	-

The following abbreviations were used: HB, hemoglobin; MCV, mean corpuscular volume; TF sat, transferrin saturation; F, female; M, male; n/a, not available. The mutation nomenclature used follows the HGVS guidelines.

Family 2 - a case with autosomal dominant L-ferritin deficiency

Proband II.1 from family 2 (Figure 1.9 B and Table 1.20) was referred to the Pediatric Oncohematology Unit of the Virgen de la Arrixaca University Hospital for an in-depth analysis of her asymptomatic mild neutropenia and eosinophilia. A small ventricular septal defect had been previously detected in the patient's heart. A subsequent medical checkup failed to find hematological anomalies, but rather marked hypoferritinemia without anemia. The proband's condition did not improve with oral iron supplementation. She experienced mild asthenia and sporadic mild headaches, which a pediatric neurologist attributed to tension headaches after examination. The serum ferritin levels and transferrin saturation percentages of the proposita's mother were low (<6 ng/mL and 9.6% respectively), but her transferrin levels were normal.

Sequencing analysis of this proband (Figure 1.9 B, II.1) showed an A>G substitution at position 1 in the heterozygous state. This caused methionine, the start codon, to be replaced with valine. The SNP database identified this variant—rs139732572—in 2004, with a very low allele frequency (MAF = 0.000008 according to the ExAC database). The ClinVar database has classified this mutation (Variation ID 96689) as pathogenic because it is linked to L-ferritin deficiency in dominant inheritance mode. Here, we report the second case of a patient with hypoferritinemia and this same mutation in the *FTL* gene (NM_000146.3:c[1A>G];[=], NP_000137.2:p(Met1Val);(=), HGSV nomenclature).

Family 3 - a case with HHCS

The proband I.1 in family 3 (Figure 1.9 C and Table 1.20) is a 65-year-old man with a history of alcoholism and dyslipidemia, showing high levels of serum ferritin (>3000 ng/ml) motive for what he was referred to the Hematology Service at the University Hospital Germans Trias i Pujol (HGTiP). He previously had cataract surgery, when aged 45, and had three therapeutic phlebotomies which had to be stopped due to anemia advancement. Magnetic resonance (MR) imaging showed normal deposits of iron in the liver, with his levels being at 30µmol/g. There is reasonable evidence in the family's history to support the presence of HHCS. The proband's son and uncle had hyperferritinemia and his mother had cataracts. The proband's son (II.1) is a 39-year-old male with history of stage 1 orchiectomized and disease-free seminoma, but no surgically associated cataracts. He was contacted by the same Hematology Service (HGTiP) under suspicion of HHCS because of hyperferritinemia (>2000 ng/ml) and cataracts. The hematological evaluation was normal, with normal amounts of hepatic iron (20 µmol/g), but higher ferritin levels.

The genetic studies performed on family 3 showed the presence a heterozygous deletion c.[-164_-158del7] located in the 5' FTL-IRE in the proband (I.1) and his son (II.1), both affected with hereditary hyperferritinemia with cataracts syndrome (Figure 1.9). Genetic analyses were not available for the mother and the uncle of the proband. This variant consists of a deletion of seven nucleotides (CAACAGT), excising part of the hexanucleotide loop and upper stem of the FTL-IRE (Figure 1.10). Following the traditional nomenclature for FTL-IRE mutations, we refer to these mutations as Esplugues +36_42del7 mutation (HGVS nomenclature as NM_000146.3:c.[-164_158del7];[=]). This deletion is predicted to impair the IRE structure. RNA secondary structure modelling of WT and mutated FTL 5' IRE sequences was performed using the Sfold web server (Ding, Chan, and Lawrence 2004), which predicted that -164_-158del7 mutation, located at the hexanucleotide loop, is likely to disturb the WT IRE conformation (Figure 1.11). In addition, the SIREs web server prediction (Campillos et al. 2010) indicate loss of the IRE structure, as the mutated query returned no results. This mutation has not been previously described in the literature, but other similar IRE deletions have been previously demonstrated to be pathogenic for HHCS (Cadenas et al. 2019; Sara Luscieti et al. 2013). The location and severity of this mutation, together with the clinical manifestations of HHCS present in the affected individuals of this family, indicates that this variant is most probably the genetic cause of disease in this family.

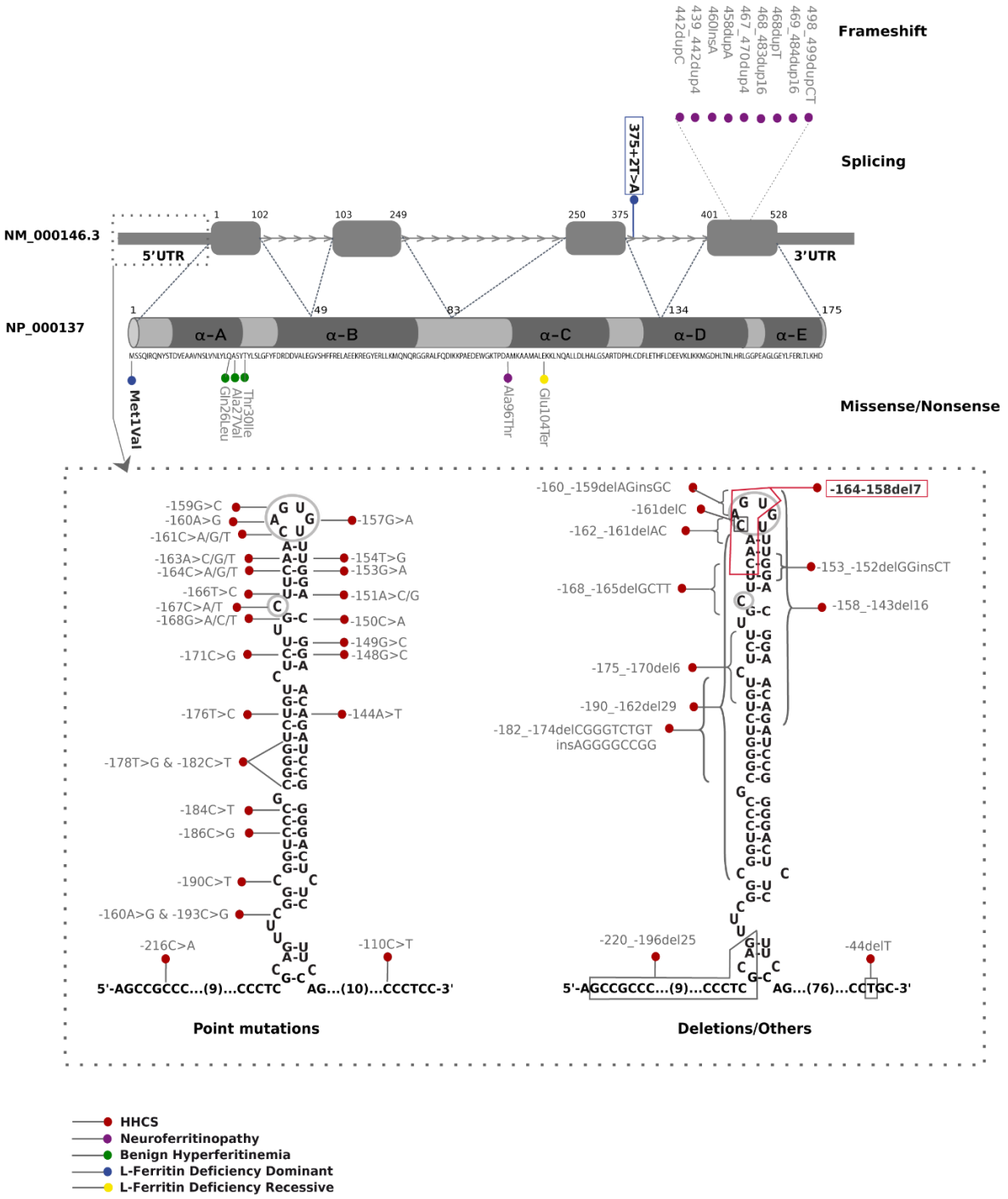


Figure 1.10. Schematic localization of literature reported and new *FTL* mutations. Mutations described in this work are in bold and new mutations are boxed. The domains of the five alpha helices (A to E) are represented in the protein (NP_000137.2). Mutations are classified as nonsense, frameshift, missense or splicing. We report here FTL protein changes using the three-letter amino acid code.

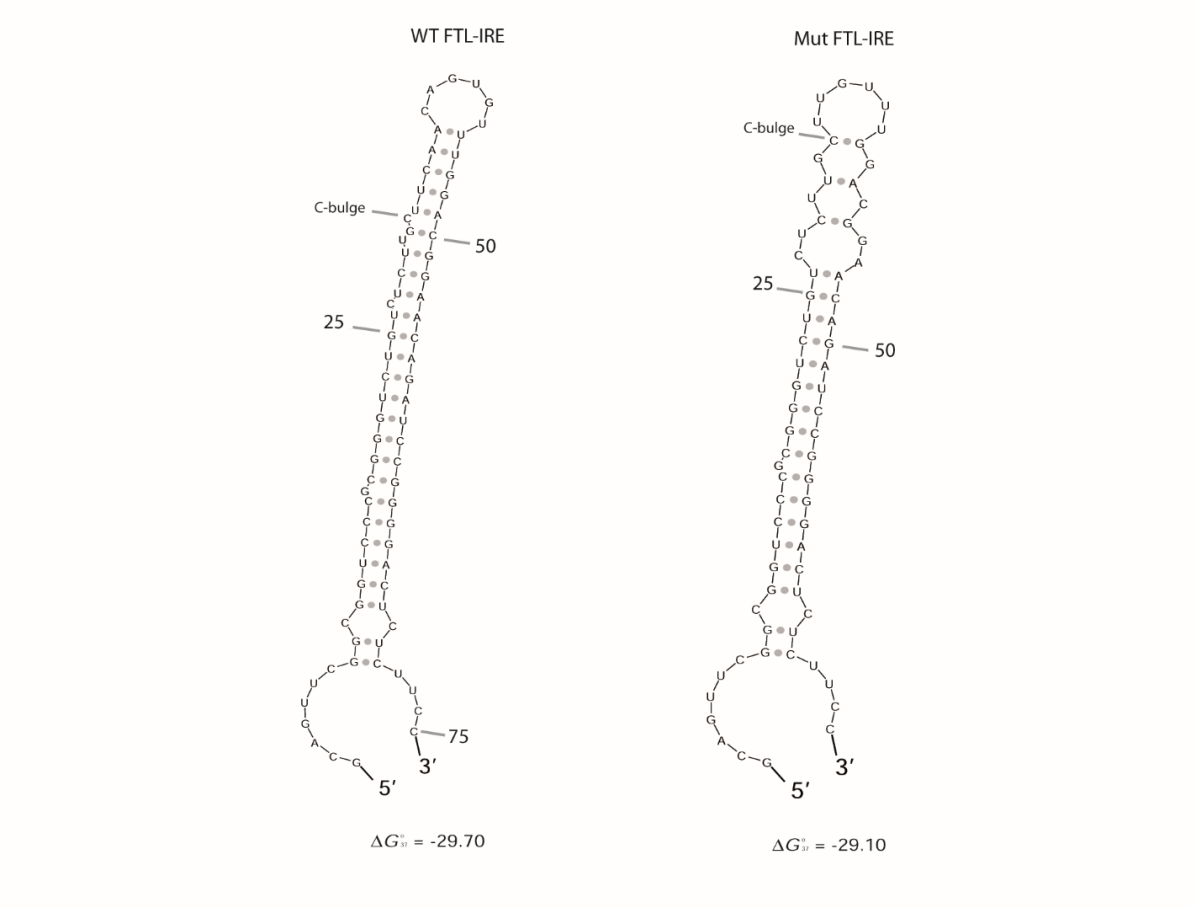


Figure 1.11. WT and mutated FTL-IRE fold prediction. Predicted secondary structure of WT and mutated FTL-IRE using Sfold web server (Ding, Chan, and Lawrence 2004). Deletion in hexanucleotide loop (c.-164_158del7) is expected to disturb completely the IRE structure. Nucleotides are numbered from the transcription starting site. Free energy (ΔG) is detailed.

Update on L-ferritin mutations and diseases

Table 1.21 summarizes the five phenotypic diseases that are triggered by L-ferritin gene mutations. A thorough review of literature yielded 63 different known variants for these five diseases, which includes the two novel mutations identified the present work (Figure 1.10 and Supplementary Material: Table S2).

Most of *FTL* gene mutations described above are linked to hereditary hyperferritinemia with cataracts syndrome (HHCS), including 36 point mutations, nine deletions, and two insertion-deletions. All HHCS-associated mutations occur at the 5'UTR location of the *FTL* gene (chr19:49468566-49468764), which disrupts the primary sequence and structure of FTL-IRE (Figure 1.10). HHCS is classified as an autosomal dominant disorder, and all reported variants are in heterozygous state except for three cases (Sara Luscieti et al. 2013; Alvarez-Coca-Gonzalez J et al. 2010; Van de Sompele et al. 2017) where homozygous mutations have been described. HHCS patients exhibit congenial bilateral nuclear cataracts and high serum ferritin levels but manifest no additional signs or symptoms.

Neuroferritinopathy is the only NBIA disorder with an autosomal dominant inheritance. It is caused by mutation in the FTL gene. Up to now, nine insertions from one to 16 nucleotides located in the exon 4 of FTL gene have been reported to cause this neurodegenerative disorder (Figure 1.10). These insertions alter the reading frame of the C-terminal region, generating a longer protein with additional aberrant amino acids. In addition, a missense Ala96Thr mutation (rs104894685) has been also

described to cause neuroferritinopathy; amino acid 96 is predicted to be situated at the same tertiary structure region as the pathogenic insertions (Maciel et al. 2005). Mutations in the C-terminal region of FTL disrupt α-helix D or E, which seems to be essential for the stability of the peptide (Kubota et al. 2009; Baraibar et al. 2010). Clinical and biochemical manifestations of this disease include low serum ferritin levels, iron accumulation in the basal ganglia and progressive and severe neurological dysfunctions with subtle cognitive deficits in some cases.

Three heterozygous mutations in FTL exon 1 have been associated with hyperferritinemia without iron overload where cataracts were absent; this condition has also been named benign hyperferritinemia. The Thr30Ile mutation (rs397514540, ExAC MAF=0.000008) has been identified in French and British families (Caroline Kannengiesser et al. 2009; Bhuva et al. 2018). Two further pathogenic mutations - p.(Ala27- Val) and p.(Gln26Ile) - have been reported in two additional patients (Thurlow et al. 2012). Mutations in exon 1 of *FTL* are associated with higher than normal serum ferritin glycosylation. These three variants alter the A α-helix near the N terminus of L-ferritin; it has been hypothesized that the aberrant peptide extends the length of the hydrophobic cluster of amino acids at the N terminus, increasing the secretion of L-ferritin (Caroline Kannengiesser et al. 2009). However, the reason for the development of hyperferritinemia and hyperglycosylation associated with these mutant ferritin forms is still not fully elucidated (Bhuva et al. 2018).

We have described here a novel intronic splicing mutation in the *FTL* gene (NM_000146.3:c.[375+2T>A];[=]) associated also with autosomal dominant L-ferritin deficiency. Including this novel splicing mutation, two heterozygous mutations have been described to cause L-ferritin deficiency with autosomal dominant transmission. Cremonesi and collaborators identified in 2004 a heterozygous A>G substitution in the first nucleotide of FTL, which change the ATG start codon (methionine) into a valine (Cremonesi et al. 2004). This mutation is predicted to encode a non-functional and unstable protein. Despite the low serum levels of L-ferritin, the proband presenting this mutation does not show either serious neurological problems (other than headaches) indicating that the molecular mechanism of L-ferritin deficiency with autosomal dominant inheritance is haploinsufficiency.

Finally, only one Italian case has been reported to cause L-ferritin deficiency with autosomal recessive inheritance; a homozygous substitution at nucleotide 310 G to T that produces a premature stop codon (E104X) (Cozzi et al. 2013). This amino acid change is predicted to be located in the middle of the α-helix C domain, a critical region for the stability of the protein. *In silico* analysis predicted that a stop codon at this position produces a truncated protein unable to fold into a functional peptide and, therefore, leads to the generation an L-ferritin subunit with a complete loss of function. This homozygous mutation is associated with a more aggressive phenotype, which is characterized by undetectable ferritin levels, idiopathic generalized seizures and atypical restless leg syndrome.

Table 1. 21 Summary of Diseases caused by mutations in FTL gene

DISEASE	HEREDITARY HYPERFERRITINAEMIA CATARACT SYNDROME	BENIGN HYPERFERRITINAEMIA	NEURODEGENERATION WITH BRAIN IRON ACCUMULATION 3	L-FERRITIN DEFICIENCY, DOMINANT	L-FERRITIN DEFICIENCY, RECESSIVE
FIRST PUBLICATION	(Beaumont et al. 1995),(Girelli et al. 1995)	(Caroline Kannengiesser et al. 2009)	(Curtis et al. 2001)	(Cremonesi et al. 2004)	(Cozzi et al. 2013)
INHERITANCE	**AUTOSOMAL DOMINANT**	**AUTOSOMAL DOMINANT**	**AUTOSOMAL DOMINANT**	**AUTOSOMAL DOMINANT**	**AUTOSOMAL RECESSIVE**
MECHANISM	LOST of IRP REGULATION	(Do not proceed)	DOMINANT NEGATIVE EFFECT	HAPLOINSUFICIENCY	TOTAL LOSS OF FTL
MUTATION/S	Many in the 5' IRE	Missense in exon 1	Frameshift in exon 4	p.(M1V; =)	p.(E104X; E104X)
TYPE	5' UTR	Affects the A α-helix near the N-terminus	Predicted to cause loss of the C-terminal secondary structure	Start loss	Nonsense
HAEMATOLOGICAL FEATURES	**High serum ferritin** Normal serum iron Normal transferrin saturation Normal red cell counts Normal hematologic parameters	**High serum ferritin** Serum ferritin hyperglycosylation Normal hematologic parameters	**Low serum ferritin**	**Low serum ferritin** Low transferrin saturation (17%) Normal serum iron Normal hematologic parameters	**Undetectable serum ferritin** Normal Transferrin saturation Normal hematologic parameters
OTHER FEATURES	Congenital bilateral nuclear cataract		Severe neurological dysfunction Gastrointestinal dysphagia		Idiopathic generalized seizures Atypical RLS Progressive hair loss

The following abbreviations were used: IRP, Iron Regulatory Protein; IRE, Iron Responsive Element; UTR, Untranslated region; RLS, Restless leg syndrome; A α-helix, first domain alpha helix of FTL protein.

1.4. DISCUSSION

1.4.1. Discussion regarding HHD panel

The recent advances in genetic have enabled to simultaneous sequence many genes for many samples at the same time. Here, we developed a 154-gene targeted NGS panel in order to evaluate the genetic basis of hereditary hematological diseases in 42 patients. If we only consider genetically pathological or likely pathological mutations, we found that 42.9% of the patients were genetically diagnosed, finding a genetic cause that can explain the clinical features that they present (category A). If we include Variants of unknown significance (VUS) the rate of diagnosis is 57.2%. These results are in agreement with others where they report a genetic diagnostic rate of 44-60% (Shefer Averbuch et al. 2018; Muramatsu et al. 2017).

1.4.1.1. Limitations

Although targeted NGS is a widely used method to detect clinically significant genetic changes, this technology still has some limitations that can hinder the analysis. These limitations include the interpretation of VUS and negative results, the areas of the DNA that are uncovered by NGS technology such as deep intronic region or promoter regions, the limitation in databases and in medical knowledge to classify variants, and the limitation in the detection of structural variants and copy number variants.

Interpretation of VUSs

Six out of the 42 patients (14.3%) were variants of unknown clinical significance (VUS). These variants represent a challenge to interpret, and the American College of Medical Genetics (ACMG) developed a guideline to help to pathogenically classify them (Green et al. 2013).

In order to categorize the pathogenicity of this type of variants, it is important to consider the allele frequency in the population (from 1000 Genomes, ExAc or other databases), the amino acid conservation, and the effect on the protein (Richards et al. 2015). In order to better understand a VUS, a functional analysis help to observe the effect of the mutation in the protein; however, functional studies are very tedious and are not available in many VUSs. Also, a segregation analysis of trio families, or large families if it is possible, can help to determine if the mutation is present in other member of the family and confirm the mode of inheritance, or if the variant is a *de novo* variant. When *in vitro* analysis is not possible, a variety of *in silico* tools (SIFT, PolyPhen-2, PROVEAN, Human Splicing Finder…) are widely used to interpret missense, indels and splicing variations, and predict the effect of the variant in the protein. However, these tools have some intrinsic limitation and caveats that can lead to potential false-positive and false-negative interpretations (Di Resta et al. 2018).

In the HHD panel v15 the exons of each transcript, 50bp of the intron boundaries and the 3 'and 5' regions of specific genes are sequenced. The clinical report does not contain information on uncovered areas of genes such as deep intronic regions or any currently uncharacterized alternative exon, so we may be missing variants in these regions. Nevertheless, the fact that deep intronic regions are not analyzed, decrease the number of variants that are difficult to interpret, as bioinformatics tools for prediction of the effect of variants situated in regulatory regions are very limited.

Negative results

Fifteen patients (38.5%) were classified in the D category (negative result), since no mutation was found in subpanel genes, or benign or likely benign variant was detected. However, negative results are also meaningful. Some negative cases belong to patients that are sent to our laboratory to discard a particular congenital disease, and this is a valuable information for the medical doctor. In other cases, patients with negative results present a clear family history of a particular disease or a clear phenotype of a particular genetic disease; however, no significant variant was found; therefore in those cases, an exhaustive investigation is performed. First, candidate genes included in the HHD panel are screened to check the existence of mutations in new genes. These candidates' genes have been selected from the literature (i.e. they might be involved in mouse in iron-related anemias but have never been involved in human diseases) or from pathway analysis, that indicate that they might be associated with the diseases. If still noting relevant is found, WES analysis is performed for selected samples, including the parents and other members of family if it possible (trio approach). The WES application may allow the identification of new genes responsible of HHD. However, one limitation of WES is that 10% of the sequenced bases do not have the recommended depth coverage for clinical diagnosis (>50x), and other method may be necessary to confirm the presence of a potential causative variant (Di Resta et al. 2018).

Database and Knowledge

The interpretation of the results is limited by the information currently available in databases, medical literature, and clinical experience. Every year, new genes are identified to be associated with HHD. The difference between the versions of the HHD panel that have been designed in the study evidence this point: as more recent is the design of the panel, more genes are incorporated. Databases are not always updated with the information supplied by novel publications, and this may result in two issues: variants described in publications but not found in databases; and variants with conflict of interpretation between databases and publications (e.g. variant reported as likely benign in databases and a novel publication that demonstrates the pathogenicity). Furthermore, the knowledge about digenic or multigenic influence in HHD is limited. A better interpretation will be possible in the future as data and knowledge about human genetics and specific diseases increase.

Difficult-to-sequence areas

Also, there are areas of some genes that are not covered by the panel such as homologous regions (Figure 1.5) and GC-rich regions (Figure 1.6). Homologous regions are areas of the genome with high sequence similarity that may differ between them by only few base pairs. The alpha and beta globin sequences have a high gene sequence similarity (HBA2 and HBB have almost 97% of similarity) (Moradkhani et al. 2009; Fan et al. 2019), so design probes to sequence these genes efficiently is a challenge for the NGS technology. The Agilent SureDesign was not able to design reliable amplicons for these genes, a difficulty that may be solved by the use of longer reads. Moreover, since the Haloplex library is dependent on restriction enzyme, several Single Nucleotide Variants (SNVs) may be missed generating potential false negatives, possibly because their location in the proximity of restriction-enzyme digestion sites (Samorodnitsky, Datta, et al. 2015; Samorodnitsky, Jewell, et al. 2015).

Besides, amplification technologies such as PCR, are sensitive to high content of G and C nucleotides, so GC-rich regions are typically uncovered by NGS technology based in amplification (such as Illumina). In figure 1.6 we can see an area of the VHL gene (~20 nucleotide) that is not covered by the Haloplex design probably due to high GC-content (around 60%). Unfortunately, one patient with mutation in

this region was not able to be genetically diagnosed using NGS targeted panel. It is essential to know the conflicting regions in the NGS panel in order to cover them with alternative methods (i.e. Sanger sequencing).

Moreover, few regions (ideally <10%) may be poorly depth covered with less than 50 reads, the minimum depth coverage for genetic diagnosis. To get this target area covered, missing targets could be re-sequenced through Sanger sequencing. These limitations must be considered, and the use of other techniques such as Sanger sequencing can be applied to complete the genetic diagnosis of the uncovered regions, although this could be labor-intensive and time-consuming.

Even though NGS has some limitations, advantages are also present (sequencing of multiple genes at once and multiplexing patients) what makes of this technology a superior one compare with the still gold standard sequencing method (Sanger Sequencing).

Structural Variant and Copy Number Variant

NGS platforms based in short length reads, such as Illumina, are reasonable efficient to detect SNVs and small insertions or deletions (indels). However, structural variants (large deletions, insertions, duplication, translocations and inversions) and copy number variants (CNVs) are not detected with ease, especially with NGS enrichment systems, and require a particular bioinformatics analysis. Since HHD are primarily caused by SNVs and indels, structural variants and CNVs have been not examined in our patients. Therfore, we cannot discard that some cases present a CNV or large reorganizations as causative of their disease. Alternative techniques such as MLPA should be used to discard this possibility.

Some of the technical limitations described here, including the detection of CNV and structural variants, and the coverage of areas “difficult to sequence”, may be solved with the development of Third Generation sequencing (TGS) instruments, based in long sequencing reads or single-molecule sequencing (SMS). The TGS instruments that are available until now are the PacBio SMRT (single molecule real time) and Oxford Nanopore (Oxford, United Kingdom). Illumina has developed the True-seq Synthetic Long-Read system, but some authors do not accept its inclusion in the TGS methodology since the length of the reads (~ 10 kb) are no as longer as the PacBio and Oxford Nanopore (>80 Kb) [Van Dij,2018]. The use of TGS instruments, that are exempt of amplification steps, may allow overcoming the limitations of the SNV and structural variants detection, and to cover areas that are difficult to sequence by amplification methods (GC-rich areas, homologous genes and repetitive areas) (Di Resta et al. 2018; Yohe and Thyagarajan 2017; van Dijk et al. 2018). Also, TGS help to estimate the haplotype pashing, and reconstruction of the parental homologs (van Dijk et al. 2018). Said so, nowadays these new technologies are still in development and have not been validated and standardized for clinical diagnosis.

Clinical reports and patient care

Finally, the genetic results found in the patients are communicated to the clinician in an understandable clinical report to facilitate the medical care of the patients, including the genetic testing results and also clinical recommendations written by hematologic medicine experts for management and treatment of the patient (see Appendix- 4.Other documents- page 217 for an example of genetic report generated by BloodGenetics).

1.4.2. Discussion regarding *FTL* gene and diseases

L-ferritin disease is a clear case of Mendelian disease-related genes that are associated with multiple diseases, including Hereditary Hyperferritinemia Cataract Syndrome (HHCS), isolated hyperferritinemia (benign hyperferritinemia), hypoferritinemia associated with other symptoms (neurological symptoms as in neuroferritinopathy, or muscular symptoms in autosomal recessive L-ferritin deficiency) or isolated hypoferritinemia (Autosomal dominant L-ferritin deficiency). For improving the diagnosis of these different diseases, we have created a clinical diagnostic algorithm of diseases caused by mutations in *FTL* gene (Figure 1.12).

The underlying mechanisms for these disease patterns are not fully elucidated for all these diseases. In HHCS, it is known that the pathological molecular mechanism is linked to the disruption of the IRP-IRE post-transcriptional regulatory system, with the de-repression of L-ferritin mRNA and the subsequent overproduction of L-ferritin protein that precipitates and deposits in the lens of the eyes, producing cataracts. In HHCS, authors have argued that an association exists between the clinical severity of the disease and the location of the mutation in FTL-IRE (Sara Luscieti et al. 2013; Faniello et al. 2009; Allerson, Cazzola, and Rouault 1999). Higher serum ferritin levels are mostly associated with mutations in the apical loop or the C-bulge area, compared with mutations in the upper or lower stems. These findings are consistent with our cases in family 3, i.e. a father and son with a seven-nucleotide deletion in the IRE of the *FTL* gene (5'-CAACAGU3'). The deletion partially affects the hexanucleotide loop of the IRE, and patients show considerably high serum ferritin levels (2071-3037 ng/ml) and develop early bilateral cataracts (Table 1.20).

Mutations in exon 4 of *FTL* gene that alter the C-terminal sequence and the length of the protein have been reported to cause neuroferritinopathy. These mutations affect the D- or E α-helix (see Figure 1.10), leading to significant disruption of the tertiary and quaternary structure of the FTL protein and producing unstable and leaky ferritin. Studies have shown that mutant ferritin maintains the normal spherical shell structure and size (Baraibar et al. 2008). However, mutant FTL C-termini, rich in amino acids with iron-binding properties, may be extended and unravel out of the shell. These mutant structures could initiate aggregation, forming ferritin inclusion bodies/precipitates (Baraibar et al. 2008; R. Vidal et al. 2004). Previous *in vitro* functional analysis and *in vivo* mouse models have revealed that these molecular-level defects have two main consequences: (1) the loss of normal protein function, reducing iron incorporation; (2) and the acquisition of toxic function through radical production, ferritin aggregation and reactive oxygen species (ROS) generation (Muhoberac and Vidal 2013; Garringer et al. 2016; Ruben Vidal et al. 2008; Muhoberac, Baraibar, and Vidal 2011). It has been demonstrated that mutations in the C terminus have a dominant-negative effect, which explains the dominant transmission of the disorder (S. Luscieti et al. 2010).

In benign hyperferritinemia, missense variations in exon 1 of FTL increase the hydrophobicity of the ferritin N-terminal site. The A-α-helix of the protein is altered and hyperglycosylated, although the precise molecular mechanism is not clear yet.

L-ferritin deficiency is characterized by haploinsufficiency of the FTL protein in autosomal dominant L-ferritin deficiency, or complete loss of function in autosomal recessive L-ferritin deficiency. Mutations associated with this condition are predicted to affect protein expression. The inactivation of one allele of the *FTL* gene, which occurs in the two described heterozygous cases of autosomal dominant L-ferritin deficiency, produces a significant reduction of L-ferritin protein in serum. The phenotype of this dominant condition is normal apart from low levels of serum ferritin and low transferrin saturation. However, the homozygous mutation Glu104Ter causes a total loss of function of L-ferritin and leads to ferritin missing the FTL chain. Previous mutational studies have suggested that the

presence of H homopolymer ferritin in fibroblasts is associated with reduced cellular iron availability and increased ROS production, which triggers cellular damage. These findings were also found in neurons derived from patient fibroblasts and correlate with the neurological phenotype of this more severe condition (Cozzi et al. 2013).

In this study, we have identified two new cases of autosomal dominant L-ferritin deficiency, one due to a new mutation in intron 3 of the FTL gene that most probably affects splicing. The T>G substitution in nucleotide 375+2 (genomic coordinates g.49469665 in Assembly GRCh37.p13) modifies a highly conserved dinucleotide GT donor site, a key sequence recognized by the spliceosome during splicing. Mutations in splice site junctions are likely to lead to exon skipping or total or partial intron retention in the mRNA transcript in most cases (Ruben Vidal et al. 2008). Our *in silico* analysis using Human Splicing Finder suggests that our variant would activate a new alternative donor site onwards (375+5G and +6T), leading to the insertion of four additional nucleotides in the mRNA sequence between exon 3 and 4 and altering the reading frame of the transcript (Muhoberac, Baraibar, and Vidal 2011). Therefore, this change would lead to a premature stop codon formation and the putative generation of a truncated FTL protein of 128 amino acids, if the aberrant mRNA is not detected and degraded by the nonsense mediated decay (NMD) system. According to our bioinformatics prediction, the c.375+2T>A mutation will not generate a dominant-negative version of FTL protein, but a truncated FTL protein completely missing the E- α-helix domain and partially lacking the D- α-helix domain. Supported by the clinical manifestations found in this family (low serum ferritin levels, low transferrin saturation and lack of serious neurological or movement abnormalities), the molecular mechanism, in this case, is most probably due to the loss of function of FTL and it will be not expected that the disease derives in neuroferritinopathy. However, we cannot totally and completely exclude the later development in life of brain iron overload and neuroferritinopathy in these patients by a yet unknown and novel mechanism.

Patients with hereditary hyperferritinemia could be misdiagnosed as patients suffering from hereditary haemochromatosis, liver dysfunction or inflammation. Some patients have received unnecessary invasive diagnostic techniques, such as liver biopsy, and are inappropriately treated with venesections and phlebotomies that can cause severe iron-deficiency anemia. On the other hand, clinical manifestations of neuroferritinopathy including tremor, parkinsonism, psychiatric problems and abnormal involuntary movements, the presence of ferritin-iron precipitation in glia cells and neurons (R. Vidal et al. 2004), are often misdiagnosed and treated as Huntington's or Parkinson's disease. Potential treatment targets for neuroferritinopathy may include an optimized iron chelator to induce the re-solubilization of iron aggregations, in combination with radical scavengers to prevent oxidative ferritin damage (Baraibar et al. 2012). Patients with benign hyperferritinemia could be misdiagnosed as patients with HHCS due to the presence of high ferritin levels and the possible late-onset appearance of cataracts. Apart from the surgical removal of cataracts in HHCS, HHCS and L-ferritin deficiency have no specific therapy.

These facts emphasize the importance of a correct and early genetic diagnosis for the subsequent implementation of proper treatment, avoiding detrimental or inappropriate treatments. Following clinical algorithms, such as the one included in this publication (Figure 1.12) and also HIGHFERRITIN Web Server (http://highferritin.imppc.org/) (Altes, Perez-Lucena, and Bruguera 2014) will surely help in this goal.

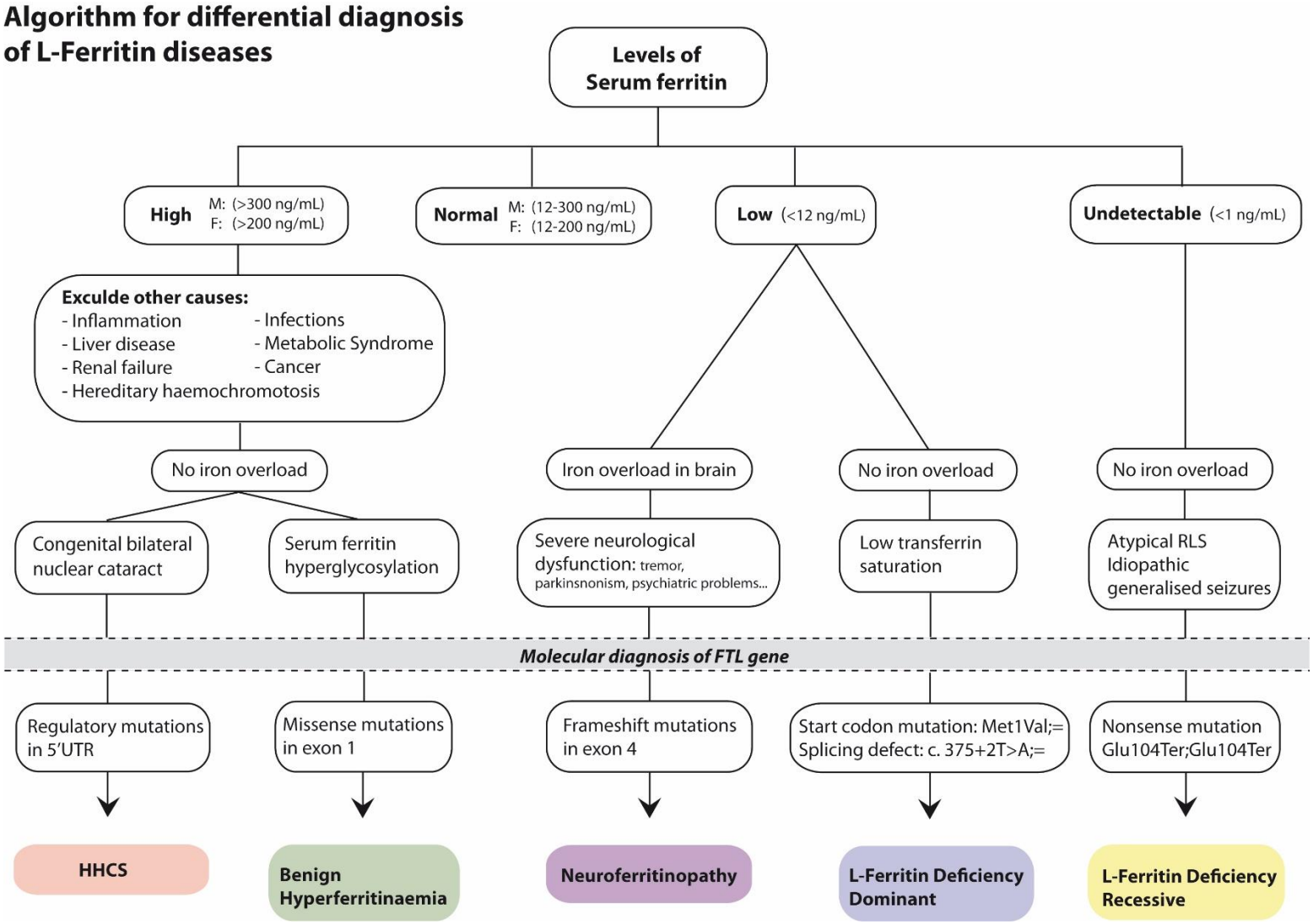


Figure 1.12. Algorithm for diagnosis of diseases caused by defects in FTL gene.
The following abbreviations were used: F, female; M, male. The mutation nomenclature used follows the HGVS guidelines.

CHAPTER II:

CoDysAn: A telemedicine tool to improve awareness and diagnosis for patients with Congenital Dyserythropoietic Anemia

2.1. INTRODUCTION

The aim of CoDysAn web page is to provide a freely accessible web site where the general public, patients and medical doctors can better understand and learn more about Congenital Dyserythropoietic Anemia. The results of this work have been recently published in Frontiers Physiology (Tornador et al. 2019). See Appendix-3.Publication, page 209 for a copy of the manuscript.

Frontiers in Physiology - Red Blood Cell Physiology section; doi: 10.3389/fphys.2019.01063
CoDysAn: A telemedicine tool to improve awareness and diagnosis for patients with Congenital Dyserythropoietic Anemia
C. Tornador, E. Sánchez-Prados, **B. Cadenas**, R. Russo, V. Venturi, I Andolfo, I. Hernández-Rodriguez, A Iolascon, M. Sánchez*

2.1.1. Congenital Dyserythropoietic Anemias: subtypes and associated genes

Congenital Dyserythropoietic Anemias (CDAs) are inherited hematological disorders characterized by chronic hyporegenerative anemia, morphological abnormalities of erythroblasts, and hemolytic component (Iolascon et al. 2012). CDA is a subtype of inherited bone marrow failure syndromes (IBMFS), hallmarked by ineffective erythropoiesis during differentiation and proliferation stages, and consequently a faulty production of red blood cells. Patients with CDA present congenital and chronic anemia of variable degree with relative reticulocytosis, jaundice, and frequently splenomegaly and/or hepatomegaly (Iolascon, Esposito, and Russo 2012; Iolascon et al. 2013).

Five classical types of CDAs (I-II-III-IV and XLTDA) have been defined based on the morphological alteration of erythroblasts in the bone marrow. Despite the fact that the morphological classification is still widely used, a most recent genetic reclassification has been introduced in clinical practice. So far, six subtypes of CDA exist depending on the causative gene (Table 2.1 and Figure 2.1).

CDA I

Patients with CDA type 1 present macrocytic anemia with mean cell volume of variable range (100-120 fL) and variable values of Hb, although the disorder can be normocytic in infancy. Inadequately decreased reticulocytes count is typically observed in these patients. All the cases develop splenomegaly, although 4-20% patients can manifest other phenotypic anomalies: pigeon chest, short stature, neurologic deficiency or malformation in hands or feets like syndactyly, polydactyly, or absence of nails.

Bone marrow (BM) exploration in optical microscopy shows hypercellularity and hyperplasia of erythroid precursors. The ratio of erythropoietic to granullo-poietic cells (E:G ratio) is four to eight times more elevated (3-8/1) than the normal range (0.2-1/1) (Heimpel, Matuschek, et al. 2010). A typical feature of CDA type I is the presence of internuclear chromatin bridges in erythroblasts. Electron microscopy shows spongy heterochromatin, also called “swiss cheese” appearance, due to the widening of nuclear pores and presence of invaginations of the nuclear membrane (Heimpel, Matuschek, et al. 2010; Heimpel et al. 2006).

The inheritance of CDA type I is autosomal recessive and is generated due to mutations in *CDAN1* (CDA type Ia; OMIN #224120) (Dgany et al. 2002) and *C15ORF41* (CDA type Ib; OMIM #615631) (Babbs et al. 2013). More than 40 pathogenic variants have been identified in *CDAN1* (chr15q15.2) so far. This gene encodes for codain-1 protein, a nuclear ubiquitously expressed protein which levels are

increased during S phase of the cell cycle, due to transcriptomic regulation by E2F1 (Figure 2.1). The function of codain-1 is not totally elucidated, although different studies suggest that this protein is implicated in the organization of the heterochromatin during DNA replication and the nucleosome assembly and disassembly (Noy-Lotan et al. 2009; Ask et al. 2012).

C15ORF41 (chr15q14) gene was the second gene identified as causative of type I CDA (CDA Ib;). Two missense pathogenic variants were identified by Babbs et al. on three unrelated families: L178Q and Y94C (Babbs et al. 2013). This gene encodes metal dependent restriction endonuclease, and although the function is still unknown, it has been hypothesized in different studies to have an essential role in the chromosome segregation between metaphase and anaphase directly involved with the Holliday junction solvates (Figure 2.1) (Gambale et al. 2016), and in the DNA replication and chromatin assembly due to the interaction with Asf1b protein (Babbs et al. 2013).

CDA II

CDA type II (OMIM #224100) is the most common form of CDA, presented with moderate normocytic anemia (HB: 8.5-9.5 g/dL) with inadequate low reticulocyte count (ineffective erythropoiesis), accompanied with jaundice, splenomegaly, and cholelithiasis due to hemolytic component (Iolascon, Esposito, and Russo 2012; Iolascon et al. 2013). The mean age of diagnosis is around 22 years old, as a consequence of the moderate symptoms or the misdiagnosis with other diseases as hereditary spherocytosis.

BM biopsy shows hypercellularity and erythroid hyperplasia, resulting in raised E:G ratio. More than 10% of erythroblasts are binucleated, and rarely multinucleated. The nucleus of erythroblasts is typical of equal size with condensed chromatin (pyknosis). Electronic microscopy examination exhibits the appearance of double plasma membrane in mature erythroblasts, caused by the excess of endoplasmic reticulum cisternae (Heimpel, Kellermann, et al. 2010; Alloisio et al. 1996).

CDA type II is inherited in an autosomal-recessive manner and is caused by mutations in *SEC23B* gene (chr 20p11.23). More than 80 different mutations have been described so far (Russo et al. 2014). This gene encodes for coat complex II component (COP II), which participate in protein vesicle trafficking through endoplasmic reticulum (Schwarz et al. 2009). The mechanism of pathogenesis is still not totally understood, but *SEC23B* could play an important role in the assembly and deconstruction of the midbody during cytokinesis (Figure 2.1), and defects in *SEC23B* could lead impairment of cell division (Skop et al. 2004).

CDA III

Type III (OMIM #105600) CDA is a rare dyserythropoietic anemia characterized by giant erythrocytes or giant multinucleated (up to ten nuclei) erythroblasts. Patients show vague to moderate anemia, and the number of reticulocytes is normal or low. Hemoglobinuria with hemosiderinuria is presented due to the presence of hemolytic component. Peripheral blood smear exhibit macrocytes, some gigantocytes and poikilocytes. At optical and electron microscopy, BM shows hyperplasia of erythroid precursors and giant multinucleated erythroblasts (Sandström and Wahlin 2000; Björkstén et al. 1978).

Familiar CDA III is an autosomal dominant condition caused by a mutation in *KIF23* gene (chr 15q21) which encodes a mitotic kinesin protein, MKLP1, involved in the formation of the central spindle and the midbody during telophase, that is essential for the cell division (Figure 2.1) (Liljeholm et al. 2013).

CDA IV

Only four cases of CDA type IV (OMIM #613673) have been described to date. Patients with CDA IV suffer from severe normocytic anemia (HB range between 5 and 9.5 g/dL), with normal or slightly increased reticulocyte count and splenomegaly. Other physical manifestations include short stature, hydrops fetalis, thalassemic face and female genitalia in an XY Karyotype patient (Jaffray et al. 2013; Ravindranath et al. 2011)

The morphological evaluation of peripheral blood smears of patients with CDA IV reveals the presence of anisopoikilocytosis and polychromasia in erythroblasts and erythrocytes. The examination of bone marrow aspirate at optical microscopy shows tri- and multinucleated erythroblasts, erythroid hyperplasia, and karyorrhexis (Jaffray et al. 2013). At electron microscopy, we can find invagination of the nuclear membrane and intranuclear precipitates.

The four described patients with CDA IV evidence the same heterozygous missense mutation (E325K) in *KLF1* gene (chr 19p13.2), which encodes the Erythroid-Krüppel-like factor (EKLF), essential for the activation of erythroid-specific genes (Figure 2.1). Erythrocytes of these patients have reduced expression of CD44 and AQP1. The disorder is inherited in an autosomal dominant manner (Wickramasinghe, Illum, and Wimberley 1991; Arnaud et al. 2010; Jaffray et al. 2013; Ravindranath et al. 2011).

XTLDA

X-linked thrombocytopenia with or without dyserythropoietic anemia (OMIM #300367) is a variant of CDA, marked by anemia with a paucity of platelets (thrombocytopenia) and erythrocytes with abnormal size and shape. Bone marrow aspiration exhibits dyserythropoiesis and dysplastic megakaryocytes. Electron microscopy of patient's bone marrow showed abnormal megakaryocytes with clusters of smoothy endoplasmic reticulum and paucity of alpha-granules (Nichols et al. 2000; Mehaffey et al. 2001).

This variant of CDA is originated by a mutation in *GATA1* (chr Xp11.23), which encodes for the zinc finger DNA binding protein GATA1. This protein is involved in the regulation of hematopoiesis and the development and maintenance of megakaryocytes and erythrocytes cells (Figure 2.1). This disorder has x-linked inheritance (K. Freson et al. 2001; Kathleen Freson et al. 2002).

Different treatments have been established depending on the CDA subtype. Allogenic bone marrow transplantation has been successfully employed in a few severe cases of CDA I and CDA II. Patients with CDA III may require a blood transfusion only during times of extreme anemia such as pregnancy or surgery. Treatment focuses on hemoglobin normalization with the administration of interferon (IFN) alpha is used with success in CDA I patients with CDAN1 mutations; however, patients bearing a mutation in a different gene i.e. *C15ORF41* were unresponsive to this same treatment. Severe cases of fetal anemia associated with CDAI, CDAII and XLTDA may require intrauterine transfusions. Blood iron levels should be closely monitored in CDA I, CDA II and other CDA patients undergoing regular transfusions. In these cases, morbidity may be severe due to iron overload complications that can be fatal if left untreated (Gambale et al. 2016; Palmer et al. 2018). Therefore, it is imperative to monitor iron overload and induce iron depletion by iron chelation when needed.

2.1.2. Aim of CoDysAn

Although the morphological diagnosis is still in use in clinical practice, the identification of the mutated genes involved in the majority of CDA subgroups will improve the diagnostic possibilities and allow a better classification of CDA patients. In many cases of CDA, a final diagnosis is not achieved due to unawareness of the disease or unfamiliarity with advanced genetic diagnostic procedures. In addition,

there are cases that fulfill the general definition of CDA, but do not conform to any of the classical CDA classification. Therefore, it is very plausible that new forms of CDA exist, which may be possible to identify if a proper genetic diagnosis is performed in each case with suspected CDA.

We have developed a new telemedicine tool named as CoDysAn (Congenital Dyserythropoietic Anemia) for improving the management and the diagnosis of patients with this condition. The aim of CoDysAn web page is to provide a freely accessible web site where general public, patients and medical doctors can better understand and learn more about this disease. Moreover, CoDysAn web page includes a diagnosis algorithm tool, to ease the classification and diagnosis of CDA types.

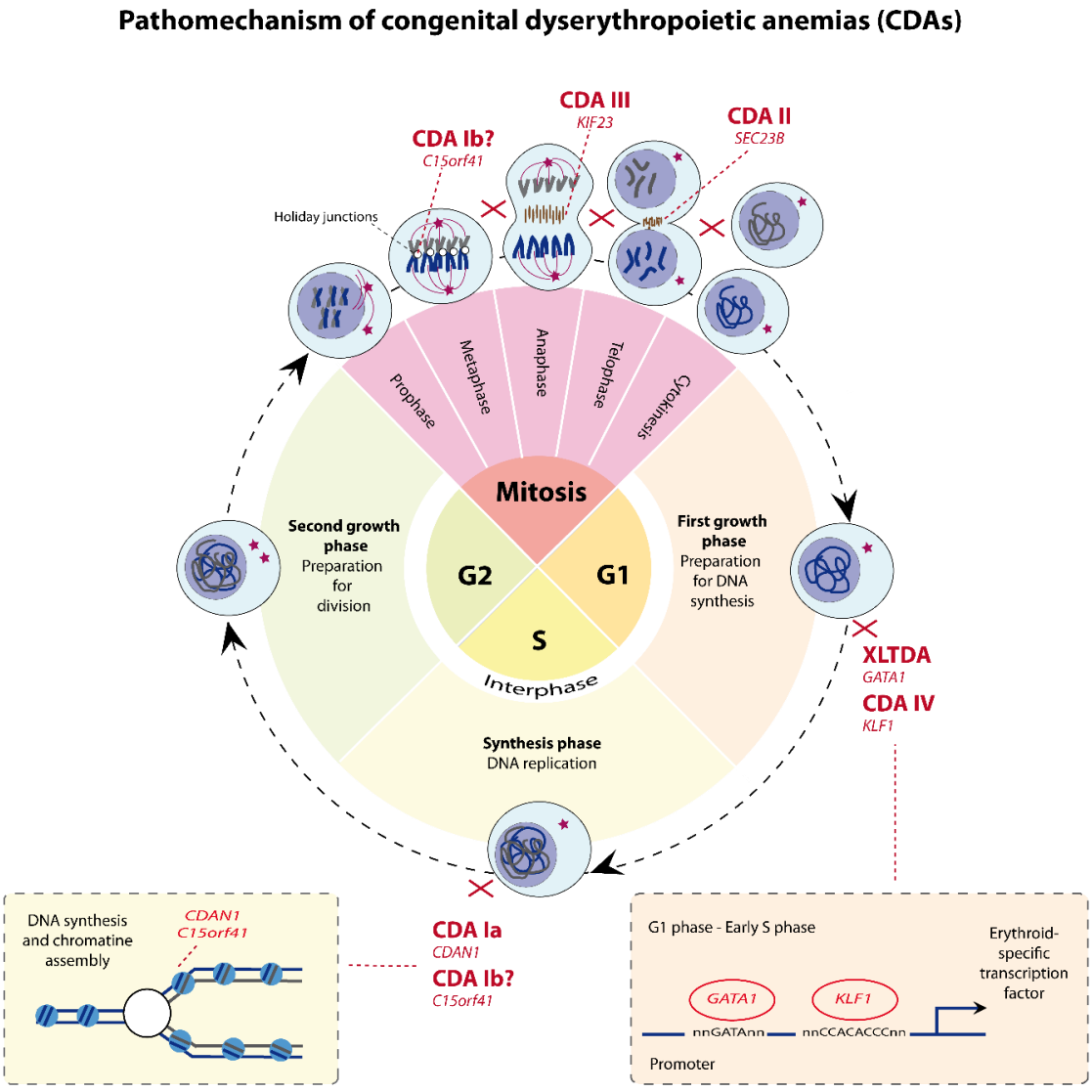


Figure 2. 1 Pathogenic mechanism of CDA. CDA is caused by ineffective erythropoiesis during differentiation and proliferation stages, and consequently a faulty production of red blood cells. Mutation in *CDAN1* cause CDA Ia, and may affect the organization of heterochromatin during DNA replication and the nucleosome assembly. CDA Ib is caused by mutation in *C15ORF41*, and seems to influence the chromosome segregation between metaphase and anaphase due to its homology to the Holliday junction resolvases. The protein encoded by *SEC23B* is a component of the midbody, and a defect in this protein impairs the cell division during cytokinesis. (CDAII). The cell division is also disable due to mutation in *KIF23*, that is also involved in the formation of the midbody and the central spindle during telophase (CDAIII). The gene *KLF1* is essential for the activation of erythroid-specific genes (CDA IV). Finally, XTLDA is produced by a mutation in *GATA1*, that is involved in the regulation of hematopoiesis and the development of erythrocytes cells. This figure is original and was designed by Beatriz Cadenas and adapted from different literature sources (Gambale et al. 2016).

Table 2. 1 Classification of CDAs

CDA subtype	Phenotype	OMIM	Gene	*Locus*	Inheritance	Phenotype	Bone Marrow Examination	
							OM	EM
CDA Ia	Congenital dyserythropoietic anemia type 1a	224120	*CDAN1*	15q15.2	AR	Macrocytic anemia; Malformations in fingers and toes.	Incompletely divided cells; binucleated erythroblasts (3-7%); thin internuclear bridges between erythroblasts	Erythroblasts with "Swiss cheese appearance" heterochromatin
CDA Ib	Congenital dyserythropoietic anemia type 1b	615631	*C15orf41*	15q14	AR			
CDA II	Congenital dyserythropoietic anemia type 2	224100	*SEC23B*	20p11.23	AR	Normocytic anemia; jaundice, splenomegaly, and cholelithiasis	>10% of erythroblasts are binucleated, and rarely multinucleated.	Double plasma membrane appearance in erythroblasts
CDA III	Congenital dyserythropoietic anemia type 3	105600	*KIF23*	15q21	AD	Normocytic anemia; reticulocytes count is normal or low	Giant multinucleated (up to 12 nuclei) erythroblasts;	
CDA IV	Congenital dyserythropoietic anemia type 4	613673	*KLF1*	19p13.2	AD	Severe normocytic anemia; normal or slightly increased reticulocyte count; splenomegaly; short stature, hydrops fetalis, thalassemic face and female genitalia in an XY Karyotype	Tri- and multinucleated erythroblasts	Invagination of the nuclear membrane and intranuclear precipitates.
XLTDA	Thrombocytopenia X-linked with or without dyserythropoietic anemia	300367	*GATA1*	Xp11.23	XLR	The paucity of platelets; Abnormal erythrocytes (poikilocytosis and anisocytosis)	Dysplastic megakaryocytes; thrombocytopenia	Abnormal megakaryocytes with smooth endoplasmic reticulum; platelets with a paucity of alpha granules

NOTE: OM, optical microscopy; EM, electron microscopy; AR, autosomal recessive; AD, autosomal dominant; XLR, X-link recessive.

2.2. MATERIAL AN METHODS

2.2.1. Patients and validation

CoDysAn web algorithm has been tested with a set of 48 patients, 24 of which were genetically diagnosed as CDA, counting: 18 CDA type II, 1 CDA type Ib, 4 CDA type Ia and 1 XLTDA. The 19 remaining patients were genetically diagnosed of non-CDA congenital anemias (negative controls), including: 8 hereditary spherocytosis, 4 patients with pyruvate kinase defects, 1 patient with pyruvate kinase defect and a beta thalassemia trait, 1 patient with defects in Hemolytic anemia due to adenylate kinase deficiency (*AK1* gene), 1 patient with X-linked sideroblastic anemia, 8 patients with dehydrated hereditary stomatocytosis type 1 (DHS1) and 1 patient with dehydrated hereditary stomatocytosis type 2 (DHS2). A second set of 23 CDA II patients was used to validate the sensitivity of the algorithm. Patients where previously reported (Schwarz et al. 2009; Iolascon et al. 2010; Russo et al. 2011; 2013; 2014; Unal et al. 2014; Pierro et al. 2015; Andolfo et al. 2015; Russo et al. 2016; 2018; Andolfo et al. 2018) and diagnosed at Medical Genetics Unit of A.O.U. Federico, CEINGE – Biotecnologie Avanzate (Napoli). See Table 2.2.

Table 2. 2 Samples used to test and validate the CoDysAn diagnostic algorithm.

Disease		N Test set	N Validating set
CDA	Ia	4	
	Ib	1	
	II	18	23
	XLTDA	1	
	Total CDA	**24**	**23**
Non-CDA	Hereditary spherocytosis	8	
	Pyruvate kinase defects	4	
	Pyruvate kinase defect and a beta thalassemia trait	1	
	Hemolytic anemia due to adenylate kinase deficiency	1	
	X-linked sideroblastic anemia	1	
	Dehydrated hereditary stomatocytosis type 1 (DHS1)	3	
	Dehydrated hereditary stomatocytosis type 2 (DHS2)	1	
	Total Non-CDA	**19**	
	TOTAL OF SAMPLES	**43**	**23**

CDA, Congenital Dyserythropoietic Anemia, N, number of samples

2.2.2. Design of the web server

CoDysAn is implemented in PHP, HTML5, CSS and Javascript. Coding was performed by bioinformaticians from Mayka Sanchez research group (Iron Metabolism group, UIC), and Whole Genix company. The web server is executed in a XAMPP. Network visualization and interactive exploration modules are based on several open-source projects: Bootstrap, jQuery and Filezilla. The source code

of the diagnostic tool algorithm is implemented in php at http://www.codysan.eu/diagnostics-tool.html. It is integrated within this web page between lines 661 and 1204 in four steps corresponding to the four steps of the form. The code can be checked by typing in a browser: "view-source: http: //www.codysan.eu/diagnostics-tool.html".

2.2.3. Diagnostic algorithm for CDA

CoDysAn algorithm has been created based on different publications and databases (OMIM, ORPHANET, GARD) and it has been approved and corroborated by experts in hereditary hematological diseases and CDA. This algorithm is based on hematological parameters depending on age and gender (Table 2.3). Three different ranges of age have been considered: infancy (from 0 months to 6 months), childhood (from 6 months to 12 years), and adulthood (older than 12 years old). The differential diagnosis of CDA is translated in the CoDysAn algorithm in 4 steps (Figure 2.2). (1) The first required criteria for CDA is the evidence of anemia. CDA patients, independently of the subtype, present hemoglobin levels (Hb) lower than the reference values. (2) Abnormal levels of other hematological parameters, including mean corpuscular volume (MCV), reticulocytes count and platelets count, are presented in CDA subtypes differently. The reticulocyte count is normal or decreased in all CDA subtypes, MCV is increased in CDA type I, and a shortage of platelets count is presented in XLTDA. (3) Exclusion of other congenital anemias with similar hematological parameters must be considered. (4) And finally, evidence of other typical manifestations of the CDA subtype is shown. References values for hematological data are obtained from general hematological reference books (Hoffman et al. 2018; Wakeman et al. 2007; Rabinovitch 1990).

Table 2. 3 Normal ranges for hematological values used by the diagnostic CoDysAn algorithm.

Parameter		0 to 6 months	6 months to 12 years	>12 years	Units
Hemoglobin	**M**	9.5-18	11-15.5	13-17.5	g/dL
	F	9.5-18	11-15.5	12-16	
MCV*	**M**	77.5-111.5	74-89.5	80-100	fL
	F	77.5-111.5	74-89.5	80-100	
Reticulocytes	**M**	61-134	24-114	29-95	$X10^9/L$
	F	67-142	40-162	27-91	
Platelets	**M**	145-450	145-450	145-450	$X10^9/L$
	F	145-450	145-450	145-450	

MCV: Mean Corpuscular Volume; M: male; and F: female

2.3. RESULTS

2.3.1. CoDysAn scope

We have developed CoDysAn web tool, a user-friendly webpage for better awareness of congenital dyserythropoietic anemia (CDA), including a step-by-step diagnostic algorithm based on hematological parameters (Figure 2.2). The website is freely available at URL http://www.codysan.eu in four languages (Catalan, Spanish, Italian, and English).

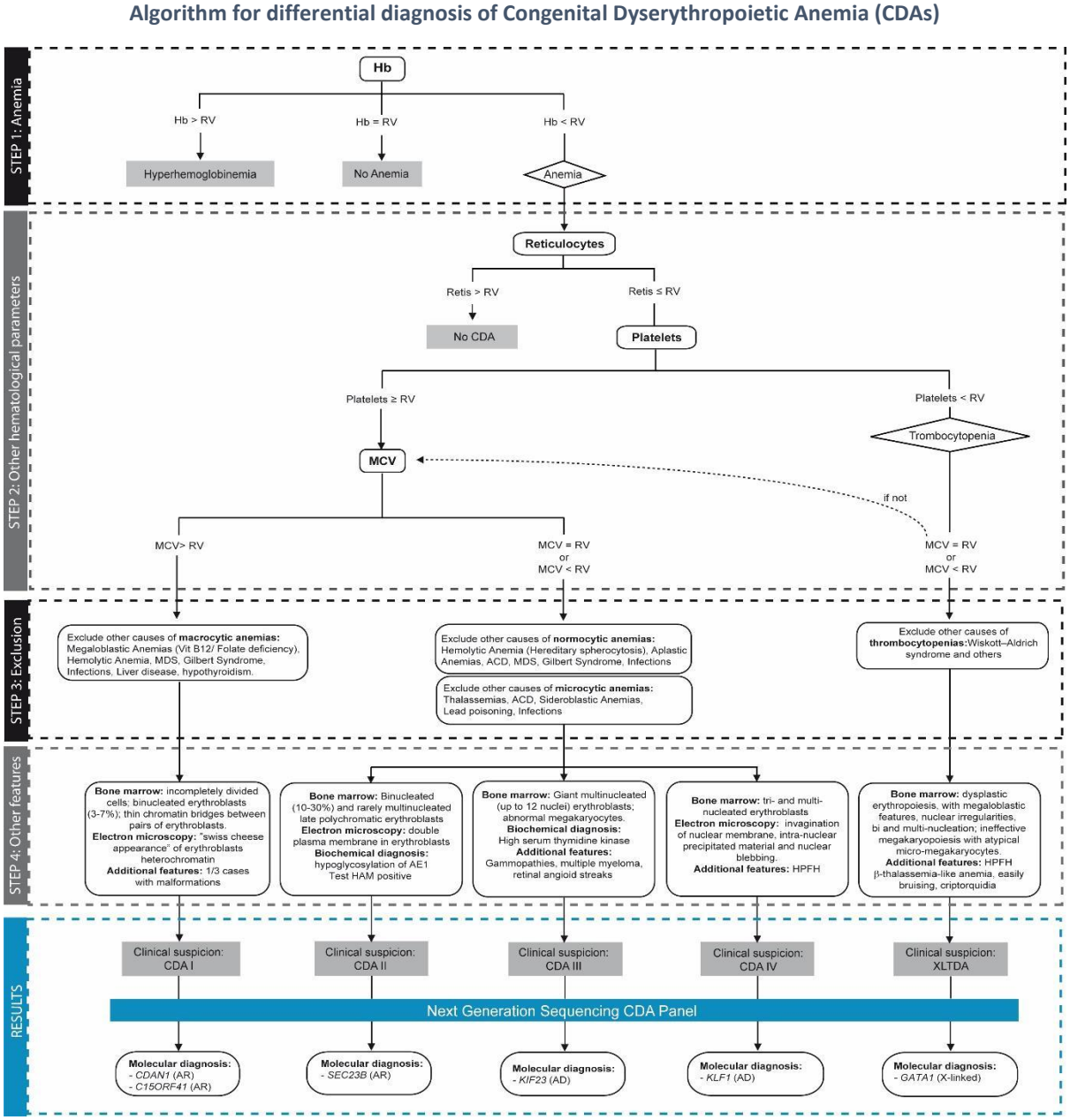


Figure 2. 2 Algorithm for differential diagnosis of Congenital Dyserythropoietic Anemia (CDAs). The algorithm is structured in five steps. (1) Anemia, the hemoglobin levels are considered for the presence of anemia. (2) Other hematological parameters are evaluated: reticulocytes, platelets, and Mean Corpuscular Volume (MCV). (3) Exclusion of other diseases with similar characteristics is performed. (4) Other clinical features are examined. (5) The clinical suspicious of CDA is shown, including the gene that is involved in the CDA subtype. This figure is also shown in supplementary material Figure S1. in two pages and in bigger dimension for better visualization. RV, regular values.

2.3.2. Webpage structure and design

The CoDysAn website is configurated in 8 sections.

1- The **home section** is the welcome page that describes the goal of the website, and contains links to the most important sections, including; description of CDA, an interactive algorithm of diagnosis, information about CoDysAn project, and guidance to collaborators of the project.

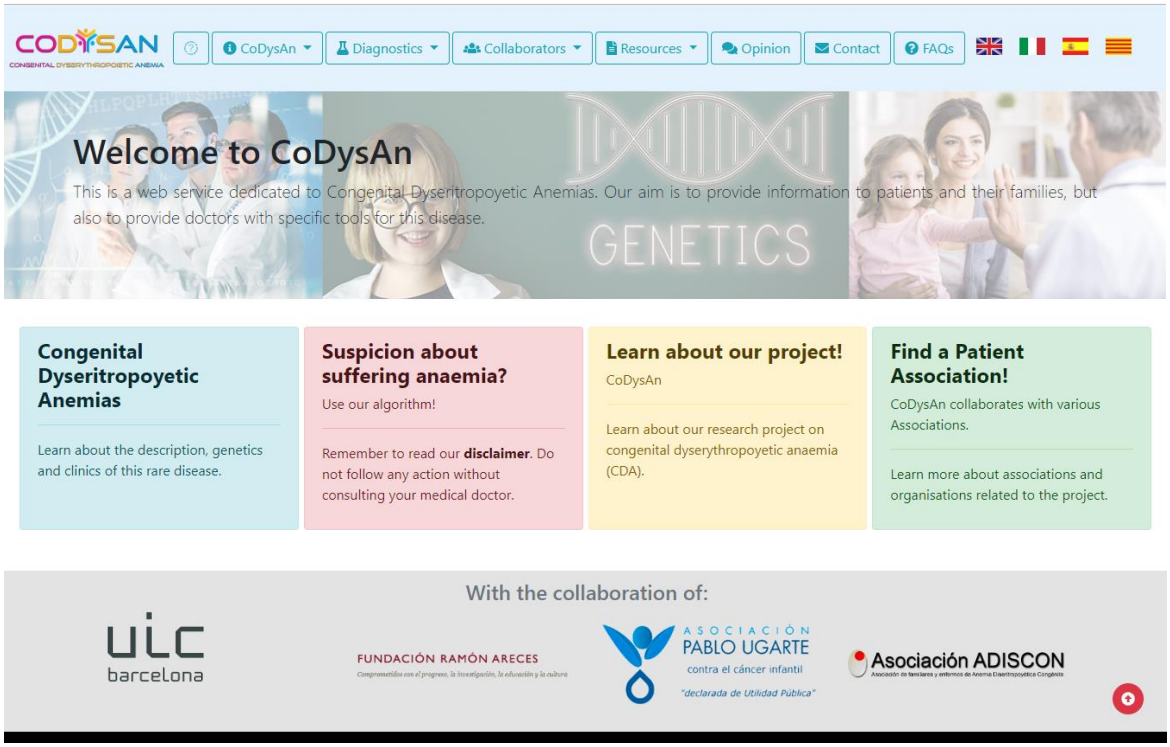


Figure 2. 3 Home page from CoDysAn website.

2- The **CoDysAn section** encloses five subdivisions.

a. *"About the disease"* subsection gives some information about CDA (description of the disease, subtypes, and genetic and clinical information

b. *"What is CoDysAn?"* subsection describes the research project that is behind the CoDysAn acronym, also titled as "Towards the improvement of diagnosis and treatment in Congenital Dyserythropoietic Anemias". The page informs about the intention and the goal of the research project and gives some details of the foundation research grant.

c. *"Privacy policy"* subsection informs about the storage, processing and manage the user´s data.

d. *"Cookies"* subsection describes the used cookies in CoDysAn webpage.

e. *Disclaimer* subsection specifies the warnings and expectations to users in order to fulfill the duty of care. CoDysAn is a telemedicine diagnostic tool intended to be used as a preliminary diagnostic test. A specialist medical doctor should verify a conclusive diagnosis.

3- The **diagnostic section** contains a step-by-step CDA diagnostic tool, the algorithm flowchart for the diagnosis of CDA (Figure 2.2), and specific instructions on how to use the telemedicine tool.

4- The **collaborator section** includes links to the contributors for the CoDysAn project, patient associations (ADISCON, http://www.adiscon.es/; and APU, http://www.asociacionpablougarte.es/index.htm) and a link to a similar telemedicine tool, the HIGHFERRITIN web server (http://highferritin.imppc.org/) (Altes, Perez-Lucena, and Bruguera 2014).

5- In the **resource section**, the most relevant bibliographical references related to CDA can be found, but also news on the CoDysAn project and reference values used for the diagnostic algorithm (Table 2.3).

6- The **opinion section** sends the user to a questionnaire to express its opinion and degree of satisfaction with the website. This information will lead us to future improvements and updates for the website.

7- In the **contact section**, users can directly contact with CoDysAn developers through a web form, to address any doubt, suggestion or complain regarding the webpage.

8- The **FAQs sections** contain the frequently asked questions.

2.3.3. Implementation of the web tool

The algorithm used for setting up the CoDysAn diagnostic tool is depicted in Figure 2.2. A user-friendly and four-steps form will progressively ask to the user for relevant information from the patient. The first step evaluates the level of hemoglobin according to age and gender (Figure 2.4). Three different ranges of age have been considered: infancy (from 0 months to 6 months), childhood (from 6 months to 12 years), and adulthood (older than 12 years old). Patients with hemoglobin levels higher than normal range (for the indicated gender and age, table 2.3), will be diagnosed as hyperhemoglobinemia. If normal levels of hemoglobin are provided, the web tool will indicate that the patient is not suffering from anemia.

Figure 2. 4 Step 1 of CoDysAn diagnostic algorithm: evaluation of anemia.

If anemia is detected, the telemedicine tool drives the user to the second step, where three additional hematological parameters must be provided: the mean corpuscular volume (MCV), the absolute

reticulocytes count and the platelets count. Varied units for the hematological parameter are facilitated (Figure 2.5).

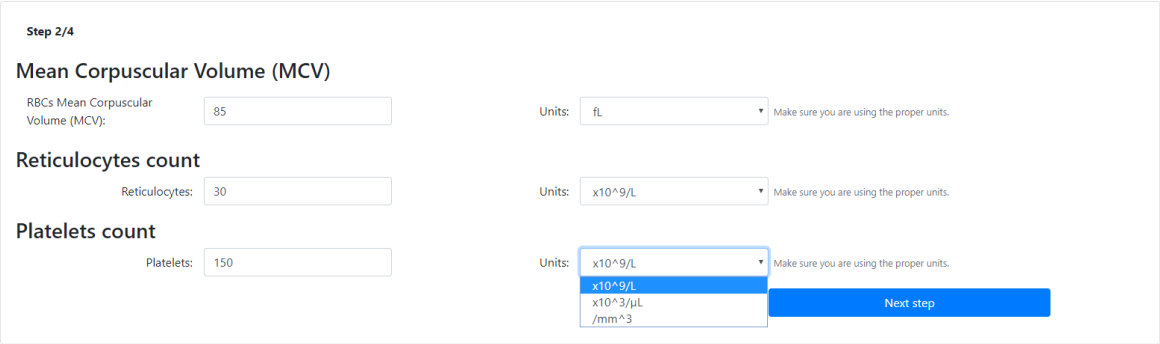


Figure 2. 5 Step 2 of CoDysAn diagnostic algorithm: evaluation of other hematological parameters

The algorithm will evaluate the hematological parameters, which will be compared with the normal ranges (table 2.3) and make a proper subclassification of the CDA. Depending on the data provided, a new form will appear asking to exclude for specific possible causes of macrocytic, normocytic or microcytic anemia (figure 2.6). At least one alternative cause of anemia should be excluded to continuous to next step of the diagnostic tool. In the fourth and final step, the user is asked to select other clinical manifestations or biochemical factors that are present in the patient, such as physical malformations, peripheral blood smear features or bone marrow evaluation by optical or electron microscopy (figure 2.7).

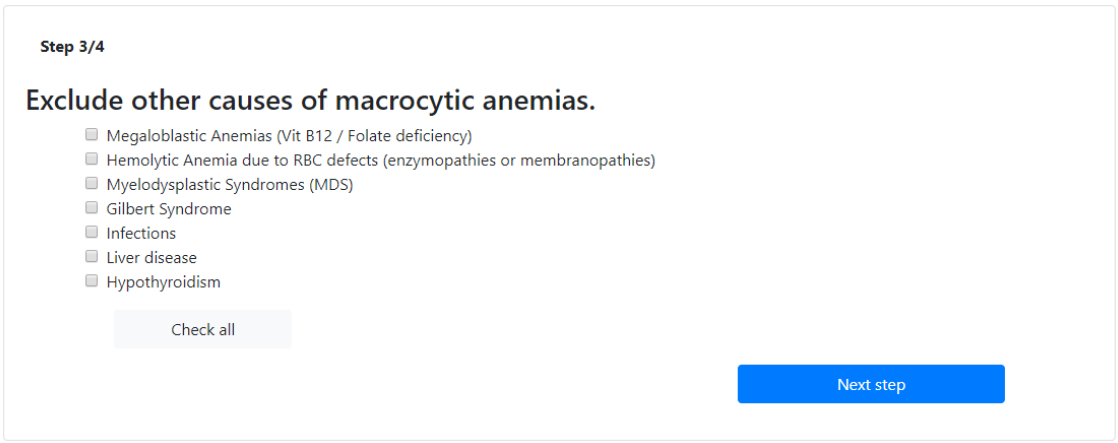


Figure 2. 6 Step 3 of CoDysAn diagnostic algorithm: exclusion of other diseases with similar manifestations.

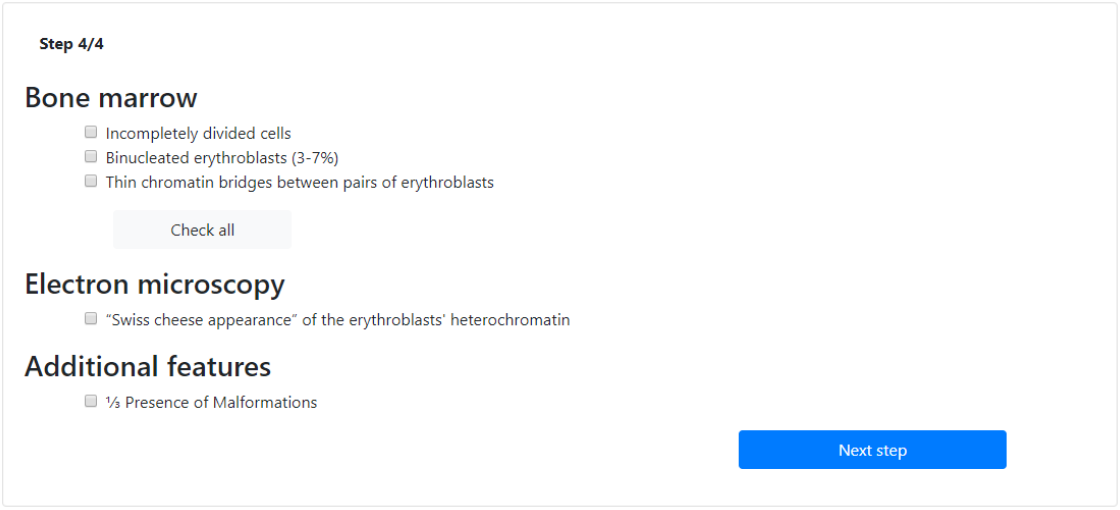


Figure 2. 7 Step 4 of CoDysAn diagnostic algorithm: evaluation of other clinical features.

Finally, depending on the information provided, the user will obtain the clinical suspicion: the CDA subtype if it applies, or "not suffering CDA" outcome if there is no clinical suspicion of CDA, including a brief explanation (figure 2.8).

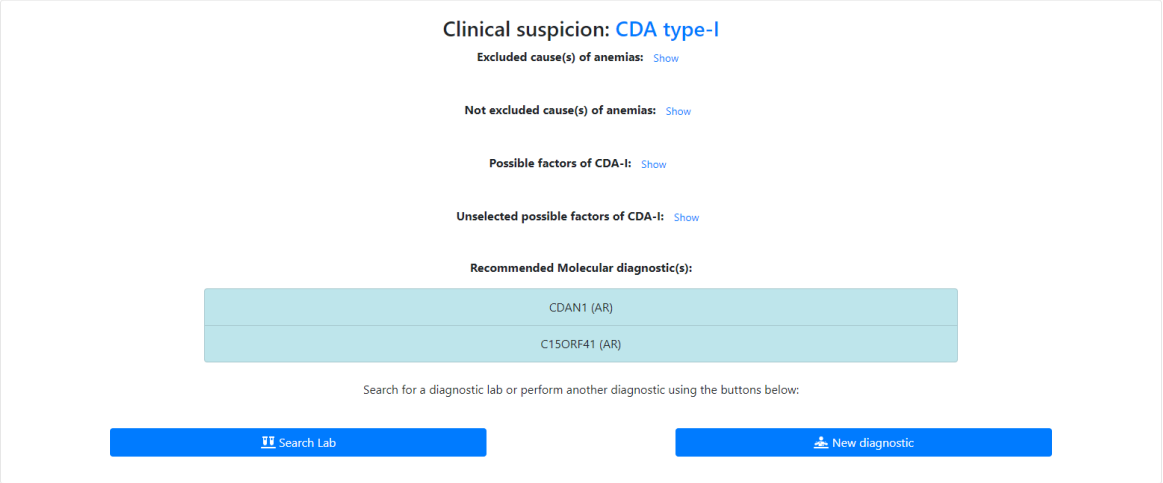


Figure 2. 8 Results window of CoDySan. The results include the clinical suspicious, the recommended molecular diagnosis and two links: one for search a laboratory which offers the genetic test, and a second button to start a new diagnosis.

If the clinical suspicion indicates one of the CDA subtypes, the associated genes for the particular CDA subtype are shown, and a genetic diagnosis test is recommended in order to verify the clinical suspicion. The user has the option to automatically search for a molecular laboratory that provides genetic tests for CDA ("Search Lab" button). The list of genetic laboratories is obtained from the NCBI's Genetic Testing Registry (GTR) webpage (Rubinstein et al. 2013). A new diagnostic test is possible to perform by pressing the "New diagnostic" button.

2.3.4. Validation

Forty-three patients with congenital anemia have been used to test the CodysAn algorithm, including 24 patients genetically diagnosed of different types of CDA (18 CDA type II, 1 CDA type Ib, 4 CDA type IIa and 1 XLTDA) and 19 additional patients genetically diagnosed of non-CDA congenital anemias. The algorithm reported a specificity of 89.5% (17 out of the 19 nonCDA patients were correctly diagnosed)

and a sensitivity of 87.5% (21 out of 24 true positives). An additional set of 23 patients (all CDA II) was used to validate the algorithm, which returned a sensitivity of 87% (20 out of 23).

2.4. DISCUSSION

Telemedicine webpages and tools are significantly changing the way medical doctors and patients approach health care and diagnosis (Dinesen et al. 2016). Here, we have developed CoDysAn, a telemedicine tool intended to increase awareness about the rare disease CDA as, currently, patients suffering from this disease are under-diagnosed (Russo et al. 2014). The content of the webpage serves as an informative and training resource for the general public, patients and medical doctors. In addition to the brief explanation of the disease and subtypes, this web page includes a user-friendly step-by-step diagnostic algorithm based on hematological parameters and other clinical features. The website is freely available at URL http://www.codysan.eu in four languages (Catalan, Spanish, Italian, and English).

The use of this tool presents limits. Patients should be considered as a whole entity and multiple biochemical determinations are needed due to time fluctuation of biochemical parameters in the same subject. Although hematological reference ranges are useful in results interpretation and in clinical decision-making, it should be borne in mind that variations within the population may affect some outcomes.

Other specific limitation of the web tool is that the CoDysAn algorithm is not fully able to distinguish between CDA type II, III and IV if test results on specific techniques or technologies (i.e. electron microscopy), that are not always available in routine hematological services, are not provided. This CDA subtypes present overlapping hematological parameters and cytologic features. The three subtypes hold high or normal platelets level, low or normal MCV, and bi- or multinucleated erythroblasts. The program recommends to genetically test *SEC23B*, *KIF23*, *KLF1* genes to confirm the subtype, and perform an extensive medical examination.

In the addition to previous limitations, the program is not able to classify the CDA subtype when no other disease is discarded in the third step ("Exclusion of other causes of anemia"), and the patient is categorized as not suffering from CDA. At least one disease must be excluded for the program to work properly, and this issue implies that patients should be examined with caution.

Despite the limitations previously described, CoDysAn is the first web tool dedicated to the diagnosis of CDA, with a high specificity of and sensitivity (89.5% and 87-87.5% respectively). CoDysAn incorporates a diagnostic algorithm that proved to be useful for a preliminary diagnostic. It will help medical doctors to know which molecular diagnosis they should request, reducing time and effort necessary for the diagnosis of CDA and allowing a direct implementation of a proper treatment once reached a definitive molecular diagnosis. Few reference centers are now offering genetic diagnostic panels screening the six known genes causing CDA. CoDysAn algorithm is connected to the NCBI Genetic Testing Registry (GTR) in a way to inform medical doctors about the existence of these accredited diagnostic centers to perform a complete genetic test, if required. This telemedicine tool aims to inform the general public and aid in the diagnosis of CDA. It is not intended as an attempt to practice medicine or provide specific medical advice and it should not be used to replace or overrule a qualified health care provider's judgment. Users should not rely upon this website for self-medication. We believe that CoDysAn webpage will positively contribute to improve medical and scientific communication on the anemia field.

CHAPTER III:

BEA (BioMark Expression Analysis): a web-based tool for gene expression Analysis from BioMark-Fluidigm Dynamic Arrays data.

3.1. INTRODUCTION

BEA (BioMark Expression Analysis software) is a Shiny-based app to help wet-bench biologist with poor knowledge of programming and statistics, to perform gene expression analysis from qPCR (quantitative Polymerase Chain Reaction, or real-time PCR) BioMark-Fluidigm data. BEA was designed to provide an easy step-by-step analysis workflow, with an appropriated statistical pipeline in a user-friendly interface. This enables users with low levels of programming and statistics knowledge to perform differential gene expression analysis for two or three groups of samples. The results exposed here are in preparation for a future publication.

3.1.1. Background – transcriptomics methods

The high-throughput technologies for transcriptome analysis is ubiquitously used in modern biology and medical research. Nowadays researchers are employing these methods to analyze multiple genes from multiple samples in parallel, significatively reducing the handling time and the cost. There are different platforms available in the market based in different technologies, that share the same objective: quantify the expression of genes in tissues. Hybridization (microarrays), sequencing (RNA-seq) and high-throughput qPCRs (such as BioMark) technologies are the most commonly used (Fassbinder-Orth 2014; Olwagen, Adrian, and Madhi 2019; Rao et al. 2019).

Microarray / hybridization principle

Microarrays consist of a large number of oligonucleotide sequences, known as "probes", which are immobilized on a solid surface (e.g. glass slides or nylon membranes) (Bumgarner 2013; Wang, Gerstein, and Snyder 2009). One microarray can have between 10.000-50.000 probes, with several probe sets per gene. The principle of microarrays technologies consists in the hybridization of extracted cDNA, previously fluorescently labeled, to these probes. The expression of the transcript is determined by the fluorescence intensity at each probe spotted on the microarray. There are two types of microarrays: single- or dual- channel design. Single-channel microarrays detect only one mRNA sample, detecting absolute gene expression levels. Affymetrix GeneChip array (Santa Clara, CA) is the most popular single-channel microarrays. Double-channel microarray design uses different fluorophores for test and control samples, and the relative expression is calculated by measuring the ratio of fluorescence (Bumgarner 2013).

Microarrays measure the expression of a large number of genes in parallel, and hence they are appropriate for a screening study. Nevertheless, this methodology entails several limitations (Eszlinger et al. 2007). Firstly, lowly/highly expressed genes may not be detected properly due to the insufficient sensitivity of the analysis. Second, cross-hybridization may result in non-specific transcript-probe binding, increasing the rate of false-positive results. And third, the detecting dynamic range for expression analysis is limited, and therefore, microarray results require validation using more sensitive and quantitative assays, like qRT-PCR.

Nanostring Technologies (nCounter platform) is another high-throughput technology based in hybridization for the analysis of gene expression (Margaret H. Veldman-Jones et al. 2015; M. H. Veldman-Jones et al. 2015). Nanostring presets excellent robustness and reproducibility, with high sensitivity even with very low input mRNA abundance (Margaret H. Veldman-Jones et al. 2015).

RNA-Seq

RNA-Seq is an approach that uses next-generation sequencing methodology in combination with computational methods to perform transcriptomic analysis. Mature RNA is fragmented and copied

into stable double-stranded cDNA (ds-cDNA) through reverse-transcription. cDNA is then sequenced using a high-throughput method, originating short sequencing reads that can range from 30 bp to over 10,000 bp, depending on the sequencing technology used. The transcriptome is computationally re-built by either de novo or reference-based assembly. The level of expression for each gene is quantified by the abundance of reads (sequencing deep coverage).

RNA-Seq can be used to profile the whole transcriptome or to analyze only target genes. There are several high-throughput sequencing platforms commonly used for RNA-Seq: Illumina (Illumina, San Diego, CA, USA), SOLiD (Thermo Fisher Scientific, Waltham, MA, USA), Ion Torrent (Thermo Fisher Scientific, Waltham, MA, USA), PacBio (Pacbio, Menlo Park, CA, USA) (Chu and Corey 2012).

RNA-Seq has some significant advantages to respect microarrays. First, the technique is more precise and is not subjected to cross-hybridization, reducing the false positive rate. Second, RNA-Seq has higher accuracy and wider dynamic range, and therefore, offers a more accurate capture level. Third, previous knowledge of the genes is not required, when analyzing the whole transcriptome, and this technology allows the possibility to identify novel genes. On the other hand, the labor intensity is higher during sample preparation and when analyzing the huge dataset generated. The data analysis commonly requires a high-performance computer system and bioinformatics specialists. The cost per sample is also higher than other transcriptomic methods, although it is continually decreasing.

RT-qPCR

RT-qPCR has been considered as the gold standard for transcription analysis by reason of high sensitivity, reproducibility and large dynamic range. The traditional format of RT-qPCR (96 and 384 microwell plate) is low-throughput, significatively increasing time and cost when analyzing several samples and genes. However, “novel” technologies have developed RT-qPCR arrays that conserve the benefits of the conventional RT-qPCR, such as high sensitivity and a large dynamic range, while allowing gene expression analysis in a high throughput format (Korenková et al. 2015). High-throughput RT-qPCR can analyze a larger number of reactions per run, making the technique more cost-effective and less time-consuming. This method is not suitable to perform general expression profiling, but it is ideal to measure genes that are involved in a specific pathway of interest.

Up to date, three high-throughput RT-qPCR platforms are available in the market: Dynamic Array BioMark HD system (Fluidigm, South San Francisco, CA, USA) (Spurgeon, Jones, and Ramakrishnan 2008), TaqMan OpenArray Real-Time PCR Assay (Applied Biosystems, Carlsbad, CA, USA), and SmartChip (Wafergen Bio-systems Inc., Fremont, CA, USA). These three qPCR platforms use nanofluidic scale technology, in which thousands of reactions are performed in a single run, reducing significatively the consumption of reagents and the hands-on time. BioMark system was the first high-throughput qPCR instrument available in the market, and is at present, the qPCR platform that higher number of reactions can be processed in a single run (9216) with the lowest volume per reaction (7-9 nl). OpenArray is able to process 3072 reactions, each one with a volume of 33 nl, and SmartChip 5,184 reactions, with a reaction volume of 100 nl (Korenková et al. 2015).

3.1.2. BioMark HD system (Fluidigm Technology)

BioMark HD system is a nanofluidic automated RT-qPCR based on microfluidic technology from Fluidigm. This system uses Dynamic Arrays, a microfluidic device with integrated fluidic circuits (IFCs) that contain thousands of interconnected channels controlled by valves (Olwagen, Adrian, and Madhi 2019). This system allows the automatic combination of biological samples and reagents in a central matrix with individual reaction chambers. There are four different Dynamic Array formats that differ in the number of samples and assays that the chip can analyze (Table 3.1). The most popular Dynamic

Arrays are the 48.48 Dynamic Array IFC (48 samples x 48 assays) and the 96.96 Dynamic Array IFC (96 samples and 96 assays) that can run 2304 and 9616 reactions respectively in a single run.

Table 3. 1 Dynamic Array IFC subtypes

IFC	48.48 Dynamic Array IFC	96.96 Dynamic Array IFC	192.24 Dynamic Array IFC	FLEXsix IFC
Assays	48	96	24	6×12
Samples	48	96	192	6×12
Reactions	2304	9616	4608	864
Pipetting steps	96	192	216	Variable
Reaction volume (nl)	9	7	8	9

IFC, integrated fluidic circuits

The workflow protocol is defined by the following four simple steps (Figure 3.1). First, samples and assays are manually pipetted in the inlets of the Dynamic Array chip. This is the only human-manipulated step. Second, the array is loaded in the IFC Controller, where samples and reagents are automatically loaded in the chambers of the IFC chip. Third, the Dynamic array is loaded in the BioMark HD System (qPCR) where real-time PCR amplification and signal detection are performed. At the end of each PCR cycle, the chip is photographed to detect the emitted signal that is proportional to the abundance of DNA amplified. Finally, the data is extracted from the BioMark HD system and it is preprocessed using the Real-Time PCR software (Fluidigm), where PCR cycles are generated for each reaction, and the cycle threshold (Ct) values are calculated. The cycle threshold (Ct) value of a reaction is defined as the cycle number when the fluorescence of a PCR product can be detected above the background signal.

Figure 3. 1 Fluidigm BioMark system workflow. Figure adapted from *www.Fluidigm.com.*

Users can obtain the results in few hours after completion of the described simple assay workflow. The big challenge is to analyze the high-throughput data. Up to date, several tools are available to analyze qPCR data, but not so many are focused on high-throughput datasets and BioMark-Fluidigm output. The most commonly used is the software developed by Fluidigm, the "Fluidigm Real-Time PCR Analysis software". This software is used to extract the Ct values (after determining fluorescence threshold, quality threshold and the accepted quantification cycle), define reference genes and reference samples, and finally calculate the Fold Changes using the -2 ΔΔCt method described by Livak

and Schmittgen (Livak and Schmittgen 2001). However, this program have some limitations. First, when several housekeeping genes have been included in the experiment, the software cannot provide additional information to select the most efficient reference genes to perform an appropriate ΔΔCt normalization. And second, the proper statistical analysis when 3 groups of samples need to be tested gets much more complicated by using this software. For this reason, most of the users utilize this software to preprocess the results, obtain the Ct values of the experiment, and perform custom normalization and statistical analysis with other programs. Some researches use some statistical programs that do not require programming knowledge, to perform the statistical analysis (e.g. Excel, GraphPad, SPSS, STATA) (Olwagen, Adrian, and Madhi 2019; Jang et al. 2011). However, these methods are not the most appropriate for high-throughput outputs, leading to a waste of time for the research. Custom analysis often requires and expert in bioinformatics that can handle that amount of data by using programming languages such as perl, pyhton or R. For this reason, there is a need to develop a user-friendly tool for performing normalization and statistical analysis of BioMark output data, without programming and advance statistics knowledge requirement.

3.1.3. BEA: BioMark Expression Analysis web tool

To meet this need, we are developing BEA, BioMark Expression Analysis, an R/Shine-based desktop application designed to analyze Real-Time qPCR data from BioMark HD System across multiple conditions and replicates. The analysis includes relative quantification using normalization against a reference gene (ΔΔCT method), and appropriate statistical testing to identify differential expressed genes between tested groups, taking into account test of normality and number of groups. Also, BEA program will help the user to select suitable reference gene for reliable qPCR data analysis, an essential step that has a significant impact on the final outcome (Vandesompele et al. 2002; Song et al. 2012).

BEA will help scientists perform gene expression analyses by interactively visualizing each step with customizable options and intuitive graphical user interface. The program provides comprehensive results visualization, downloadable plots and a summarizing report, and it does not require programming knowledge. BEA will be an open source software freely available online at http://bloodgenetics.com/otros/ web page.

3.2. MATERIAL AND METHODS

3.2.1. Testing data (Myelodysplastic syndromes project)

BEA has been tested using BioMark 96.96 Dynamic Array dataset acquired from a secondary research project: DJCLS R 14/04 Myelodysplastic syndromes, cause of mitochondrial iron overload in sideroblastic anemia (Principal Investigators: Dr. Mayka Sanchez (Spain) and Dr. Nobert Gattermann (Germany)) and financed by the Deutsche Josep Carreras Leukämie-Stiftung). The aim of this project was to identify genes associated with iron regulatory proteins (IRPs), that are differentially expressed between patients with myelodysplastic syndrome with ring sideroblasts development (MDS-RS), patients with myelodysplastic syndrome without ring sideroblasts development (MDS-nonRS) and healthy controls.

3.2.1.1. MDS patients

RNA samples from bone marrow were provided by Dr. Nobert Gattermann inside the research project DJCLS R 14/04 Myelodysplastic syndromes. Twenty-eight samples where from the MDS Registry in Dusseldorf (Germany) (14 were from MDS-RS patients, 14 from MDS-nonRS patients) and 4 from healthy control people. RNA integrity and concentration were evaluated using 2200 TapeStation system. The 32 samples obtained an RNA Integrity Number (RIN) higher than 7.2 and a nanodrop concentration of 5ng/µl with adequate 260/280 ratios ranged from 1.25 to 2.14.

3.2.1.2. Experiment design

The 96.96 Dynamic Array IFC was used to analyze 96 genes, of which four were housekeeping genes previously identified as an efficient normalization genes (*ACTB, B2M, HPRT1* and *TBP*)(Dheda et al. 2004; Lazarini et al. 2013; Dolatshad et al. 2015), three positive control genes (*ALAS2, GDF15* and *ARG2*)(del Rey et al. 2015; Pellagatti et al. 2006), and 89 genes identified to be associated with IRP-targets (Sanchez et al. 2011). No negative control genes were added.

3.2.1.3. Real-time qPCR using the BioMark HD System (Fluidigm).

RNA samples were reverse transcribed into cDNA and analyzed in the qPCR Fluidigm-BioMark instruments in the Fluidigm Platform from Josep Carreras Research Institute according to manufacturer instructions.

3.2.1.4. Data acquisition and pre-processing

Data was pre-processed by automatic threshold for each assay using BioMark Real-Time PCR Analysis Software 4.1.3 version (Fluidigm). The quality threshold was set at the default setting of 0.65, the baseline correction method was linear derivative and the Ct threshold method was automatically detected. The Ct values for each reaction were extracted in a csv (Export Type: Table Results).

3.2.2. Implementation

3.2.2.1. Programming languages

BEA has been developed through the Shiny software-development platform, using R and Java Script languages. The R packages used to develop the program are described in Table 3.2.

Table 3. 2 Packages installed in R for the development of BEA

Libraries	Version	Repository
data.table	1.12.0	CRAN
dplyr	0.7.8	CRAN
DT	0.5	CRAN
ggplot2	3.1.0	CRAN
gplots	3.0.1	CRAN
grid	3.5.2	CRAN
gridBase	0.4-7	CRAN
gridExtra	2.3	CRAN
kableExtra	0.9.0	CRAN
knitr	1.23	CRAN
NormqPCR	1.28.0	Bioconductor
RColorBrewer	1.1-2	CRAN
reshape	0.8.8	CRAN
rgl	0.99.16	CRAN
shiny	1.2.0	CRAN
shinyBS	0.61	CRAN
shinyjs	1.0	CRAN
rsconnect	0.8.13	CRAN

3.2.2.2. Availability of data and materials

The BEA program including the required secondary package and programs, the manual, and the example datasets are together compiled in zip folder and available to download in the project webpage.

- **Project name:** BEA
- **Project web page:** http://bloodgenetics.com/otros/
- **Operating systems**: Windows
- **Programming language:** R, JavaScript and css.
- **License:** GPLv3
- **Intellectual property (IP):** The program has been registered in the intellectual property of Catalunya (B-2222-19).
- **Any restriction to use by non-academics:** None

3.2.2.3. Normalization and statistical workflow

The normalization and the statistical workflow are done with BEA program. Although the BEA statistical analysis may be accommodated to user preference, the by default workflow is described in Figure 3.2. Ct values from the Fluidigm BioMark run are imported into BEA program. The gene expression data is first refined removing those genes and samples with low performance. FC values are calculated through the $-2^{\Delta\Delta Ct}$ method described by Livak and Schmittgen (Livak and Schmittgen 2001) by using the selected normalization or housekeeping gens. Delta Ct data, the data calibrated by housekeeping genes, is used for the statistical analysis. The statistical analysis is determined by the number of groups and the normality distribution of our data. Shapiro-Wilk test is utilized to test for normality the expression of each gene and perform adequate test: ANOVA or Kruscal Wallis for three

groups; T Student test for normal distribution, or Wilcoxon test for non-normal distribution data for two groups.

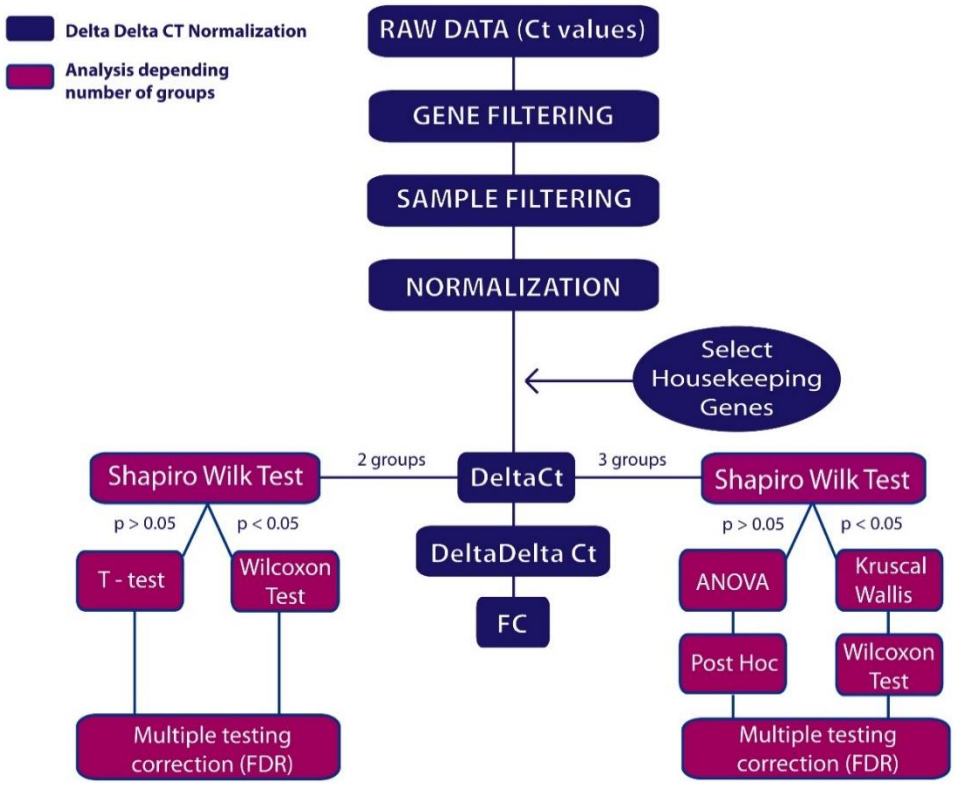


Figure 3. 2 Statistical workflow of BEA program

3.2.3. Error and running validation of BEA with external datasets

The performance of the BEA program has been validated using four datasets: one 96.96 dataset (**dataset 1**) from public sources (https://github.com/jpouch/qPCR-BioMark) and 3 datasets from three different Dynamic Array formats (**dataset 2**, 12.12; **dataset 3**, 48.48, **and dataset 4**, 96.96) coming from the collaboration with the Fluidigm Platform from Josep Carreras Leukemia Research Institute (Table 3.3). The four validating datasets were previously preprocessed as detailed in the BEA manual before being analyzed. A sample information file was created for each dataset, obtaining sample names from the results file, and organizing the samples in two or three groups. The four datasets contained negative control genes which were not used in the analysis. Dataset 2 and 3 contains some blank genes, and only 60 out of 96 and 13 out of 48 genes were included in the analysis respectively. In dataset 4, four samples in triplicate were analyzed.

Table 3. 3 Dataset used to validate the program

	Source	IFC format	Samples (n)	Genes (n)	Observations	N groups analyzed
Dataset 1	Public	96.96	96	96	Negative control genes (NTC)	2 & 3
Dataset 2	IJC	96.96	96	60	Negative control genes (NTC), Contains blank genes	2 & 3
Dataset 3	IJC	49.49	48	13	Negative control genes (NTC) Contains blank genes	2 & 3
Dataset 4	IJC	12.12	4	12	Negative control genes (NTC)	2

3.3. RESULTS

3.3.1. BEA scope

BEA is a ready-to-use desktop program based in R/Shiny language. The program is hassle-free in installation and can be downloadable from http://bloodgenetics.com/otros/, along with user´s manual, and sample datasets. The program is implemented in an interactive desktop application (R Shiny) using a portable R and a portable Chrome browser, and all R packages and secondary programs (pandoc) required are already included. Therefore, users do not require to install R, integrated development environments (RStudio), R packages nor other web browser. Users can operate the program immediately after download and decompress the BEA zip. The program is sustained by 64-bit Windows. BEA is an open-source program available under the GNU General Public License v3.0 (GPLv3, https://www.gnu.org/licenses/gpl-3.0.html), and registered in the Spanish intellectual property (2019). Video tutorial is available at https://youtu.be/GL2a1BLOnM8.

3.3.2. Implementation

3.3.2.1. Raw data preprocessing and acquisition

Before starting the BEA program for the gene expression analysis, raw data (Fluidigm ChipRun.bml file) must be preprocessed using the Fluidigm Real-Time PCR Analysis Software. The steps required for a correct preprocessing of the raw data are described in the user`s manual (Supplementary material 4. Other documents page 221 - *Instruction of use of BEA program*). User must specify the reference genes that have been introduced in the experiment. The procedure for the selection of the reference genes are also explained in the manual. The final preprocessed data must be extracted from the Real-Time PCR Analysis software by selectin the “Table results (*.csv)” from the “Save as” type menu. The exported ChipRun_TableResults.csv file can then be used for analysis with the BEA program.

3.3.2.2. Input files

Two csv input files are required before running Bea (Figure 3.3): the BioMark raw data file and the sample information file.

A “BioMark raw data file” (Figure 3.3 a) is a coma separated value file containing the results (Ct values) from BioMark Fluidigm platform after preprocessing the raw data using the Real-Time PCR Analysis software (“Table results” file, see Fig. 3.3). Sample names must have no spaces, start with a letter, and can contain letter, numbers and certain special characters (dot or underscore). Replicates or triplicates must have exactly the same sample name. The data must be separated by semicolon (;), and number are quoted.

A “Sample information file” (or SIF) (Figure 3.3 b) is a comma separated value file containing three columns:

- SampleName: Sample name used in the experiment. Remember that if you include replicates, those samples must be named identically. The program is letter case sensitive.
- SampleType: Sample Type as described in the BioMark raw data file.
- GROUP: Group of analysis (i.e. “Case” vs. “Control”; “Healthy” vs.” Treated” vs.” No-treated”). Minimum two groups and maximum three groups can be assigned. Group names only can contain letters, numbers and some special characters (dot or underscore). First character must be a letter, and no spaces are permitted.

a

	A	B	C	D	E	F	G	H	I	J
1	Chip Run Info	\\filer1.imppc.loc	1,362E+09	96.96 (136x)	GE 96x96 Fa	ROX	EvaGreen	########	0:31:49	BIOMARKHD3
2	Application Version	4.1.3								
3	Application Build	###############								
4	Export Type	Table Results								
5	Quality Threshold	0,65								
6	Baseline Correction Method	Linear (Derivative)								
7	Ct Threshold Method	Auto (Detectors)								
8										
9										
10	Experiment Information	Experiment Information	Experiment Information	Experiment Information	Experiment Information	Experiment Information	EvaGreen	EvaGreen	EvaGreen	EvaGreen
11	Chamber	Sample	Sample	Sample	EvaGreen	EvaGreen	Ct	Ct	Ct	Ct
12	ID	Name	Type	rConc	Name	Type	Value	Quality	Call	Threshold
13	S96-A01	C4 - H6 - CD34+	Reference	1	AAGAB	Test	12,32	0,997	Pass	0,02
14	S96-A02	C4 - H6 - CD34+	Reference	1	ARG2	Test	16,50	1	Pass	0,02
15	S96-A03	C4 - H6 - CD34+	Reference	1	SIMC1	Test	12,56	1	Pass	0,01
16	S96-A04	C4 - H6 - CD34+	Reference	1	CXCL12	Test	12,62	1	Pass	0,02
17	S96-A05	C4 - H6 - CD34+	Reference	1	ELF2	Test	10,88	1	Pass	0,01
18	S96-A06	C4 - H6 - CD34+	Reference	1	GDF15	Test	19,68	1	Pass	0,01

b

	A	B	C
1	SampleName	SampleType	GROUP
2	C4 - H6 - CD34+	Reference	control
3	R9 - ID11 - RCMD	Unknown	nonRS
4	R24 - ID31 - RCMD	Unknown	nonRS
5	R1 - ID1 - RCMD	Unknown	nonRS
6	R16 - ID19 - RAEB	Unknown	nonRS
7	R25 - ID32 - RAEB	Unknown	nonRS
8	R8 - ID9 - RAEB	Unknown	nonRS
9	R17 - ID20 - RCMD	Unknown	nonRS
10	C3 - H5 - CD34+	Reference	control
11	R10 - ID13 - >15%RS	Unknown	RS
12	R23 - ID30 - >15%RS	Unknown	RS
13	R2 - ID2 - >15%RS	Unknown	RS
14	R15 - ID18 - >15%RS	Unknown	RS
15	R26 - ID33 - >15%RS	Unknown	RS
16	R7 - ID8 - RCMD	Unknown	nonRS
17	R18 - ID22 - >15%RS	Unknown	RS
18	C2 - H3 - MNC	Reference	control

Figure 3. 3 Inputs files required to start the gene expression analysis with BEA. Two CSV format files must be prepared before starting the analysis: (a) the BioMark raw data file, with the preprocessed data of the BioMark gene expression experiment; and (b) the Sample Information File (SIF) with the names and group of analysis of all samples tested.

3.3.2.3. Tabs

The program is structured in eight steps that are incorporated in eight tabs respectively (Figure 3.4): file manager, heatmap, gene filtering, sample filtering, normalization, results, boxplots, and final report. The tabs are activated when the previously step is completed.

Figure 3. 4 Conceptual architecture of BEA software structured in eight tabs.

"File Manager" tab: The two required input files are uploaded to BEA using this interface. An example of the sample information file can be downloaded through a link that redirects the user to another webpage. If the input files are uploaded properly, some tables appear automatically: head of the raw data input file from BioMark-Fluidigm preprocessed output; a table with information about the experiment and the chip; a table with the reference genes included; a table showing the sample information file uploaded; a table with groups names and number of sample per group.

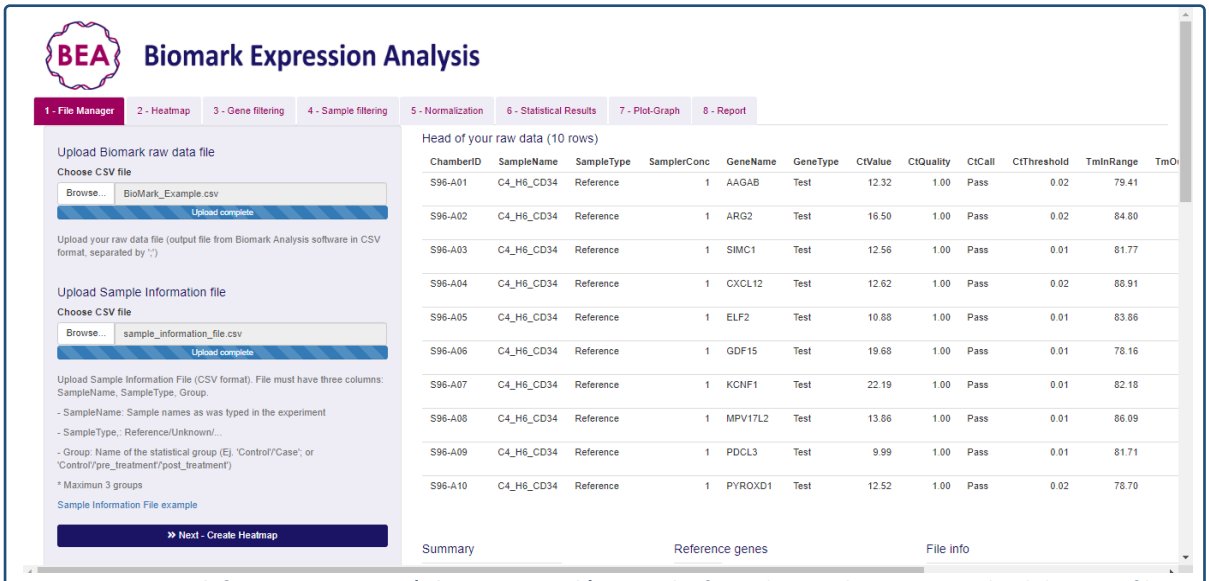


Figure 3. 5 First tab from BEA program (File Manager tab). From the first tab page the user can upload the input files, check out the data of these files, and redirect to two repositories: one to download a SIF example and, and other to download toy example data to try the program.

From this page, users can also obtain a link to download a toy example to train themselves and understand the analysis. The toy example is a zip folder containing the two required input files (table results data, the sample information file), a standard protocol, a final report ,a video tutorial and a README file with all information about the repository content. Although these documents are also attached in the downloadable zip folder, this links were done in anticipation of having the program available in a web-server side, which is the purpose for the following version of the program.

"Heatmap" tab: BEA provides in the second step a general visualization of the experiment with a gene expression heatmap. The Ct value for all genes and samples (individual replicates) are plotted ranged from high expression or low Ct value (red) to low expression or high Ct value (bluish). A black square indicates no Ct value or a value outside of the spectrum range. The heatmap is reconfigurable and axes can be sorted by user preferences. Sample order can be sorted by alphabetic sample name or clustered by the groups itemized in the sample information file. Gene axis can be sorted by alphabetic gene name or by mean gene expression. Heatmap plot can be downloaded in a pdf format by pressing in the "Download" button.

"Gene filtering" tab: Some genes with low efficiency can be filtered out of the analysis to reduce noise in the analysis. The number of fail assays per gene is exposed in a barplot. The fail flag is assigned when the amplification curve for a specific assay does not pass the quality threshold (due to low amplification, out of range Ct value). Genes with high abundance of fails should be removed to reduce noise that degrades the quality of signals and data. The filtering threshold can be selected by the user, by default is 75%. User can discard other genes if desire by clicking over the plot, or by accessing to the "List of genes" and unselect its box.

The barplot with the number of fails per gen can be rearranged by sorting gene axis alphabetically (A to Z) or numerically (low to highest abundance of fails). Discarded genes are colored in red on the barplot and typed under the plot.

"Sample filtering" tab: The filtration of samples with high count of fails can be done in this tab. The filtering process is performed in the same way that in "Gene filtering" tab. A barplot is displayed, and a list of samples can be checked.

"Normalization" tab: In this tab user select all the parameters required to perform delta delta Ct normalization (Livak). First, the groups of samples defined in the sample information file are shown. User will have to select which groups wants to include in the analysis, and assign the appropriate flag: "Reference" for control group; "Target_1" for one group to test; "Target_2" for second group to test when three groups have been added; "Dont_study" if user doesn´t want to include that group in the analysis. Only one group can be assigned as a "Reference", and at least one group has to be assigned as a "Target_1". The aggregation of two different groups in one is possible for tested groups (two groups as a "Target_1).

Secondly, the reference (or housekeeping) genes must be selected. The interface provides a list of the normalization genes that have been annotated in the pre-processing step with the Real Time PCR software, or have been selected by the user in the first step. The choice of the reference gene has a significant impact in the qPCR analysis and gene expression outcome. Hence, the selection of appropriate reference genes is critical for a reliable result [Song J,2012]. In the "Normalization" tab, a table that includes the mean, the standard deviation (SD) and the Gene Stability Measure (M) (Vandesompele et al. 2002) of used references genes is shown, to help the user to select adequate genes for normalization (see an example on Table 3.8). The Gene Stability Measure determine the expression stability of control genes on the basis of non-normalized expression levels and is calculated

by the average pairwise variation for the control gene with all other tested reference genes (Vandesompele et al. 2002). Genes whose mean Ct differs more than 5-fold to all the genes analyzed are recommended to be avoided for normalization (Ma et al. 2016; Kozera and Rapacz 2013). Also, genes with lower SD and M-value are considered more stable reference genes (Dai et al. 2017; Lee et al. 2010). More than one reference gene can be selected for normalization.

The user has the possibility to see the gene expression matrix with all the results. The Ct values for all samples (in columns) and genes (in rows) are shown in this matrix. When the experiment contains replicates, the gene expression matrix displays the median Ct value.

"Statistical Results" tab: Genes are analyzed following the statistical pipeline described in Figure 3.2, and results are presented in a table (Table 3.9). The results table contain the statistical results of the comparison between tested groups for each gene, including Shapiro-Wilk test, statistical test used according to normality, Fold Change, and adjusted p.value among others. Significant p.values are printed in bold. Genes that are differentially expressed in at least one group are printed in pink and marked with stars (*** p<0.005; **p<0.01; *p<0.05 or p< p-threshold). Genes are considered differentially expressed based on fold change and p.value. The minimum fold and the p.value thresholds can be chosen by the user (by default are 2 and 0.05, respectively).

If users prefer to perform a parametric or a non-parametric test for all genes, they can easily change the statistical pipeline by selecting another method of analysis. If user selects "Parametric test", analysis of variance (ANOVA) and Post Hoc Tukey´s tests are performed when analyzing 3 groups, and T-test is performed when analyzing 2 groups. Kruskal-Wallis and Wilcoxon test are carried out if user select "Non-parametric test". Users are recommended to choose the non-parametric test if the gene expression data do not follow a normal distribution (Shapiro-Wilk > 0.05), and or/ the total sample size is small. P.values are corrected by false discovery rate (fdr) by default, but users can select Bonferroni's multiple comparison method, or avoid multiple testing correction ("none").

The parameters used in this tab are shown in the results table. Final table can be download in csv format, containing the data displayed in the "Statistical Results" tab, plus other information of interest as the mean ΔCt and the ΔΔCt per group.

"Boxplots" tab: Samples selected in the results table from previous tab are represented here by boxplots. One plot per gene is displayed, showing normalized fold change expression of the gene by tested groups. Boxplots can be downloaded in pdf format file (Figure 3.9).

"Final report" tab:

In this tab users have the option of downloading a final report in a pdf format. The report resumes the whole analysis process performed with BEA from all tabs, but the heatmap. Due to dimensionality problems, the heatmap representing the expression of all samples and all genes from step 2 of BEA workflow, is not added in the final report. Instead, a heatmap of the significantly expressed genes is included. A download bar gives to the user information about how long the process will take to generate the report. See Supplementary Materials. See Appendix – 4. Other documents (page 237) for an example of a report executed with the BEA program and the testing data.

3.3.3. Results from the testing data (Myelodysplastic syndromes project)

Biomark HD Fluidigm platform was used to analyze the expression profile of 96 genes (including 4 calibration genes (Table 3.5), 3 positive control genes, and 89 IRP-target genes) in 32 bone marrow RNA samples in triplicate. Samples were from MDS-RS patients ("RS"; n=14), MDS-noRS patients ("nonRS"; n=14) or control healthy persons ("controls"; n=4). The relative gene expression using the ΔΔCt normalization method and the statistical analysis to identify the differential expressed genes between the three groups were performed using the BEA program.

The program provides three tables with information about the data uploaded (Table 3.4, 3.5 and 3.6). In the gene expression analysis, 96 total genes have been tested, for 32 samples in triplicates (96 samples in all). A total of 9216 reactions have been done, of which 835 were failed and 845 reactions had low quality. No blanks or negative samples were included. From the 96 samples studied, 12 were selected as a "reference" (4 unique samples in triplicates) and the rest were "unknown".

Tables from the final report

Table 3. 4 Experimental information

	n
Number of Genes	96
Number of samples	32
Replicates	3
Total reactions	9216
Total Fails	835
Low Quality	845
Samples Blank	0
Samples NAC	0
Samples NTC	0
Samples Standard	0
Samples Unknown	84
Samples Reference	12

NAC, No Amplification Control;
NTC, No Template Control;
N, number of assays

Table 3. 5 Reference genes

Ref genes
ACTB
B2M
HPRT1
TBP

Table 3. 6 Groups of analysis

Group	n
control	4
nonRS	14
RS	14

A heatmap involving all the 96 samples and the 96 genes studied in the BioMark experiment was performed in the second step of BEA program (Figure 3.6). The range of gene expression of the assay is represented in different colors from yellow (low Ct value and high expression) to blue (high Ct value and low expression). Black squares indicate failed assays. The plot was sorted by sample groups, and gene-axis were arranged according to the mean expression. We can see that in general the triplicates have similar expression, implying a low dispersion between replicates and good performance of the analysis. The heatmap also suggests that there are some genes with unsuccessful results (black squares) that have low performance. Moreover, there is one sample which three replicates seem to have also failed ("R12_ID15_RCMD").

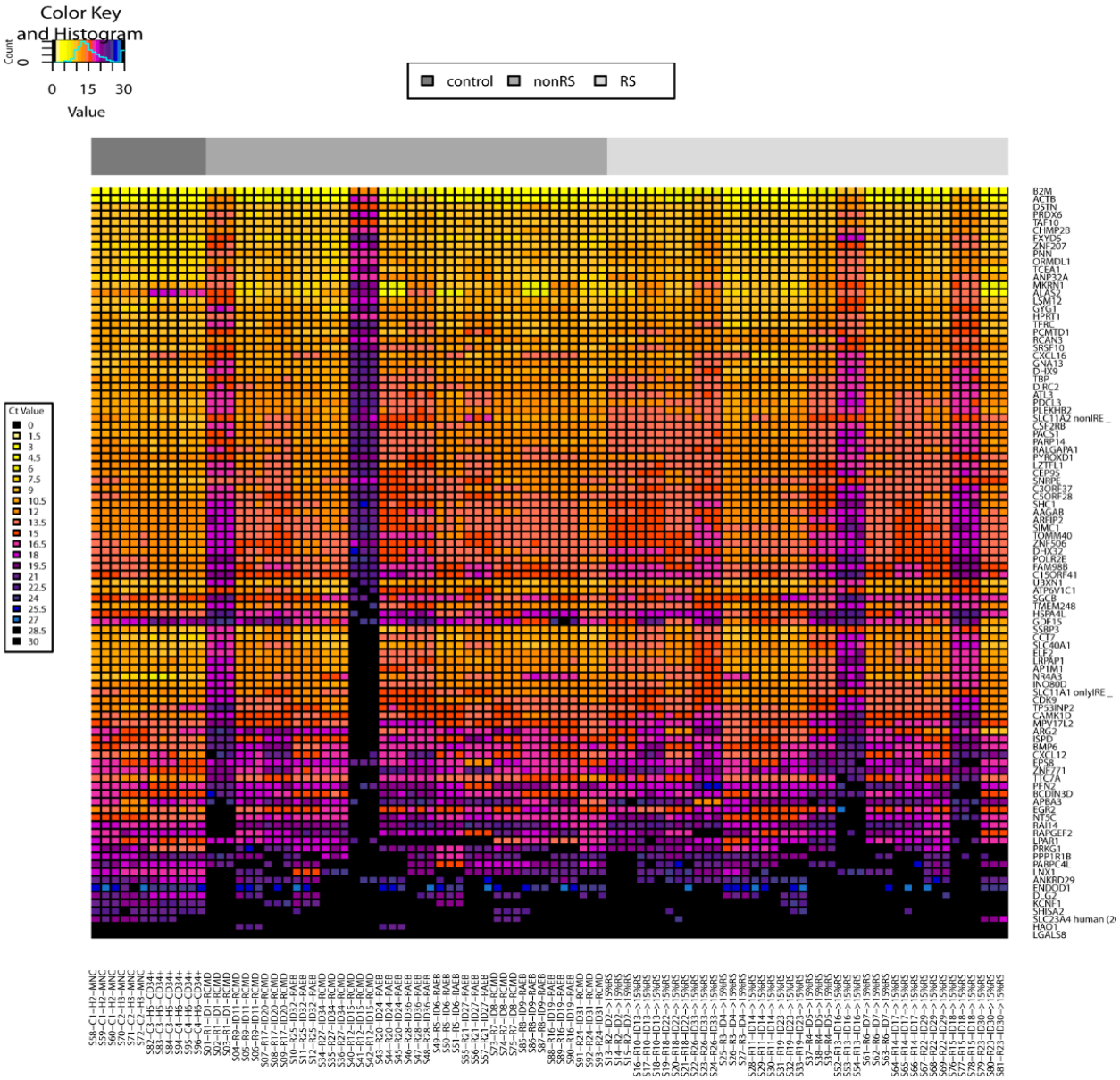


Figure 3. 6 Heatmap of gene expression (Ct) of samples with RS-MDS, nonRS-MDS and healthy controls for 96 genes. The Ct value tested with BioMark Fluidigm system is represented in the heatmap in a color range: yellow for low Ct values (high expression) and blue-purple for high Ct values (low expression). Failed reactions are colored in black. The heatmap is sorted by group of samples: dark grey for healthy controls, medium gray for patients with MDS with ring sideroblasts (RS) and light gray for MDS patients without ring sideroblasts (nonRS).

A barplot with the number of failed assays per gene (Figure 3.8) confirm that 5 genes (*KCNF1*, *SHISA2*, *SLC23A4*, *HAO1* and *LGALS8*) must be filtered out due to low performance since more than 75% of the reactions (72/96 fails as threshold) do not pass the quality control process of Fluidigm system, and are classified as "fail". The same process was performed for the samples, and one sample ("R12_ID15_RCMD" in triplicate) was discarded from the analysis due to slightly elevated number of fails, around 37% (35.5/96) of the reactions failed (Figure 3.9).

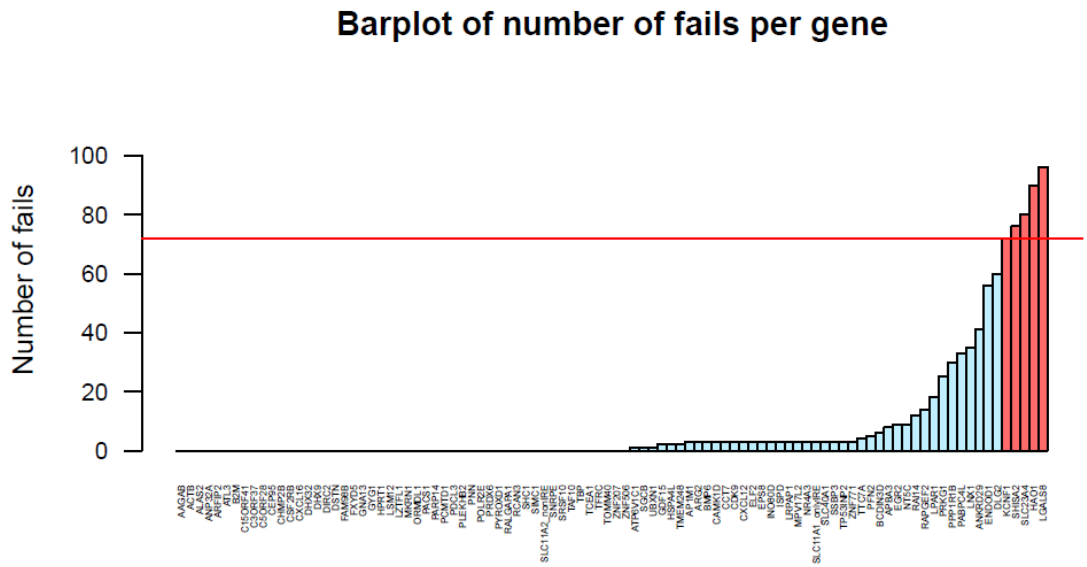


Figure 3. 7 Barplot of the number of failed reactions per gene. The graph shows the number of failed assays (reactions that do not approve the quality control process of the Fluidigm system) per gene. In red are represented those genes with a greater number of fails than the selected threshold (75% or 72 reactions of the total of 96) which are filtered out from the analysis.

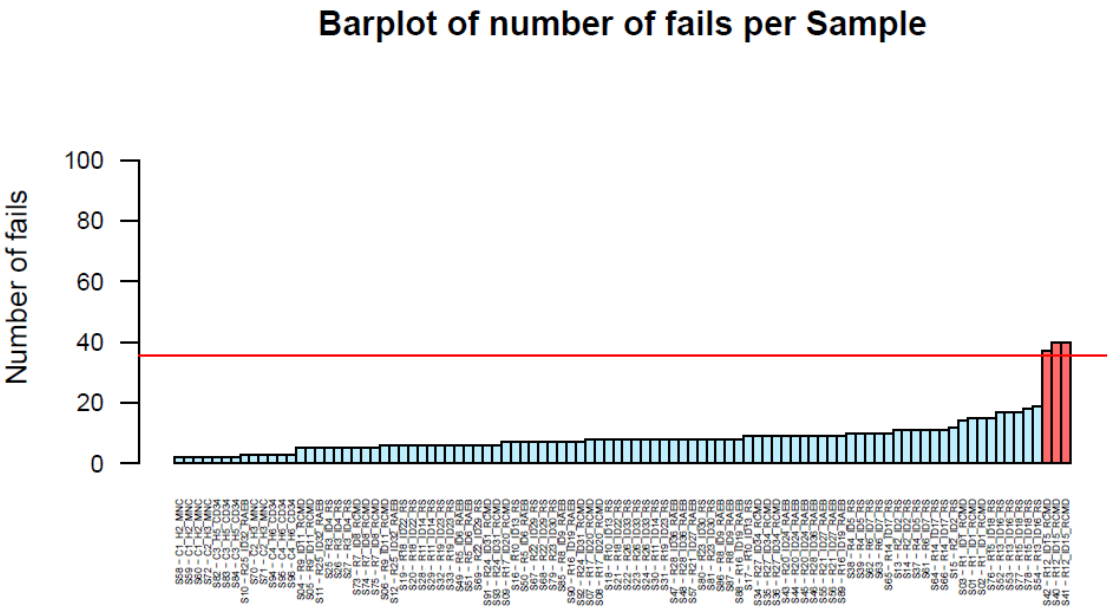


Figure 3. 8 Barplot of the number of failed reactions per sample. The three samples colored in red ("S42 - R12_ID15_RCMD", "S40 - R12_ID15_RCMD "and "S41 - R12_ID15_RCMD") are replicates of the same sample, and are filtered out of the analysis due to the slightly high number of fails compared with the rest of samples.

The ΔΔCt normalization was executed taking the control group (Group 1 of Table 3.7) as a reference, and using HPRT1 and TBP as a normalization genes due to the low standard deviation, low gene stability measure and a similar expression with respect the rest of the tested genes (Table 3.8).

Table 3. 7 Selection of groups for analysis

Groups	Group_Name	Analysis
Group 1	control	Reference
Group 2	nonRS	Target_1
Group 3	RS	Target_2

Table 3. 8 Description of reference genes

	ACTB	B2M	HPRT1	TBP	All.genes
Mean	6.941313	6.1038731	11.4914038	12.893182	14.4009
Standard Deviation	2.062040	1.1623020	1.2473722	1.603699	
Gene-stability measure	1.156331	0.8878971	0.8842936	0.817371	

To evaluate the differential expression between control, RS and nonRS groups (Table 3.7), an appropriate statistical test was used per gene according to normality. Twenty-three genes were differentially expressed between at least one comparison of the three groups of analysis. Significance was established with an absolute fold change larger than 1.5 and a cutoff p value of 0.05. Results were not corrected by multiple comparison testing (Table 3.9 and Figure 3.10). Six of the genes were identified using nonparametric test, meanwhile parametric test was applied for the rest of genes. Five genes were differentially expressed between control group and nonRS-MDS patients, sixteen genes between control groups and RS MDS patients and twelve genes between RS and nonRS-MDS patients.

The three positive control genes including in the experiment (*ALAS2*, *GDF15* and *ARG2*), resulted to be significative in the differential expression analysis. These genes have been previously reported to be more expressed in RS-MDS compared with other cases. In our study, *ALAS2* gene was higher expressed in patients with MDS with or without the development of ring sideroblasts, compared with control group (FC = 90-128 and $p < 0.001$). However, *ARG2* and *GDF15* were highly expressed in RS-MDS compared with nonRS-MDS patients and controls ($FC_{ARG2(RS\ vs\ nonRS)} = 6$, and $p_{ARG2(RS\ vs\ nonRS)} = 0.014$; $FC_{GDF15(RS\ vs\ nonRS)} = 7.5$, and $p_{GDF15\ (RS\ vs\ nonRS)} = 0.012$). These genes have been previously reported to be differentially expressed in ring sideroblasts MDS patients (del Rey et al. 2015; Pellagatti et al. 2006; Malcovati and Cazzola 2016). Differential expression of these three genes can be easily detected in the heatmap of Figure 3.10, along with *NR4A3* and *PABPC4L*, due to the high fold change.

NR4A3 and *PABPC4L* were the genes with higher fold change and significant p.value among the rest of the genes and excluding the positive control genes. In our data, the gene *NR4A3* was lower in both MDS groups (including RS and non-RS) compared with healthy group ($FC_{NR4A3\ (nonRS\ vs\ control)} = -19$, and $p_{NR4A3\ (nonRS\ vs\ control)} = 0.004$; $FC_{NR4A3\ (RS\ vs\ control)} = -10$, and $p_{NR4A3\ (RS\ vs\ control)} = 0.004$). No differences were found between the RS and non-RS MDS patients. This results support the Ramirez-Herrick et al. study, which found the *NR4A3* gene to be down-regulated in hematopoietic stem cells from MDS mice model (Ramirez-Herrick et al. 2011).

PABPC4L genes was found to be downregulated in RS-MDS compared with nonRS-MDS group ($FC_{PABPC4L\ (RS\ vs\ nonRS)} = -6.8$, and $p_{PABPC4L\ (RS\ vs\ nonRS)} = 0.002$). No previous study identified *PABPC4L* as a possible biomarker of RS-MDS.

Table 3. 9 Differentially expressed genes between tested groups

				Ref vs Target 1			Ref vs Target 2			Target 1 vs Target 2			
Gene	Shapiro	Method	P.val	FC	P.val	*	FC	P.val	*	FC	P.val	*	Significative
ALAS2	0.228	One-Way ANOVA	0.000	128.198	0.000	***	90.030	0.000	***	-1.424	0.861		*
ARG2	0.444	One-Way ANOVA	0.012	-1.002	1.000		5.986	0.118		5.997	0.014	*	*
CAMK1D	0.020	Kruskal-Wallis	0.025	-1.030	0.871		-1.979	0.044	*	-1.922	0.044	*	*
CEP95	0.633	One-Way ANOVA	0.004	-1.018	0.995		-1.587	0.060		-1.559	0.006	**	*
CHMP2B	0.786	One-Way ANOVA	0.007	2.179	0.059		3.025	0.005	**	1.388	0.307		*
DHX32	0.959	One-Way ANOVA	0.000	-1.163	0.800		-2.229	0.005	**	-1.917	0.001	***	*
DLG2	0.683	One-Way ANOVA	0.016	-1.075	0.982		-3.700	0.028	*	-3.443	0.024	*	*
DSTN	0.304	One-Way ANOVA	0.010	1.844	0.040	*	2.186	0.007	**	1.186	0.545		*
FAM98B	0.858	One-Way ANOVA	0.014	-1.607	0.096		-1.976	0.011	*	-1.230	0.354		*
FXYD5	0.000	Kruskal-Wallis	0.014	2.524	0.034	*	1.169	0.277		-2.159	0.034		*
GDF15	0.945	One-Way ANOVA	0.003	2.345	0.652		17.468	0.015	*	7.450	0.012	*	*
LPAR1	0.443	One-Way ANOVA	0.029	1.541	0.762		-2.260	0.357		-3.482	0.025	*	*
MPV17L2	0.187	One-Way ANOVA	0.027	-1.062	0.979		-1.838	0.138		-1.730	0.037	*	*
NR4A3	0.010	Kruskal-Wallis	0.006	-19.086	0.004	***	-10.387	0.004	***	1.838	0.239		*
NT5C	0.116	One-Way ANOVA	0.019	-1.788	0.303		-2.988	0.021	*	-1.672	0.152		*
PABPC4L	0.954	One-Way ANOVA	0.003	3.113	0.192		-2.191	0.450		-6.819	0.002	***	*
PARP14	0.007	Kruskal-Wallis	0.001	1.025	0.956		-1.958	0.012	*	-2.006	0.001	***	*
PDCL3	0.008	Kruskal-Wallis	0.002	-1.367	0.102		-2.179	0.004	***	-1.594	0.016	*	*
PYROXD1	0.435	One-Way ANOVA	0.027	1.548	0.130		1.855	0.021	*	1.198	0.446		*
SLC11A2_IRE	0.047	Kruskal-Wallis	0.013	-1.013	0.785		1.786	0.152		1.809	0.012	*	*
SRSF10	0.930	One-Way ANOVA	0.018	-1.327	0.524		-1.973	0.033	*	-1.486	0.075		*
TAF10	0.856	One-Way ANOVA	0.003	2.183	0.005	**	2.333	0.002	***	1.069	0.902		*
ZNF771	0.209	One-Way ANOVA	0.033	-1.845	0.391		-3.253	0.040	*	-1.763	0.179		*

Note:

FC, fold change; Ref, Control group; Target 1, nonRS group; Target 2, RS group (see Table 3.7)

*** p<0.005; **p<0.01; *p<0.05

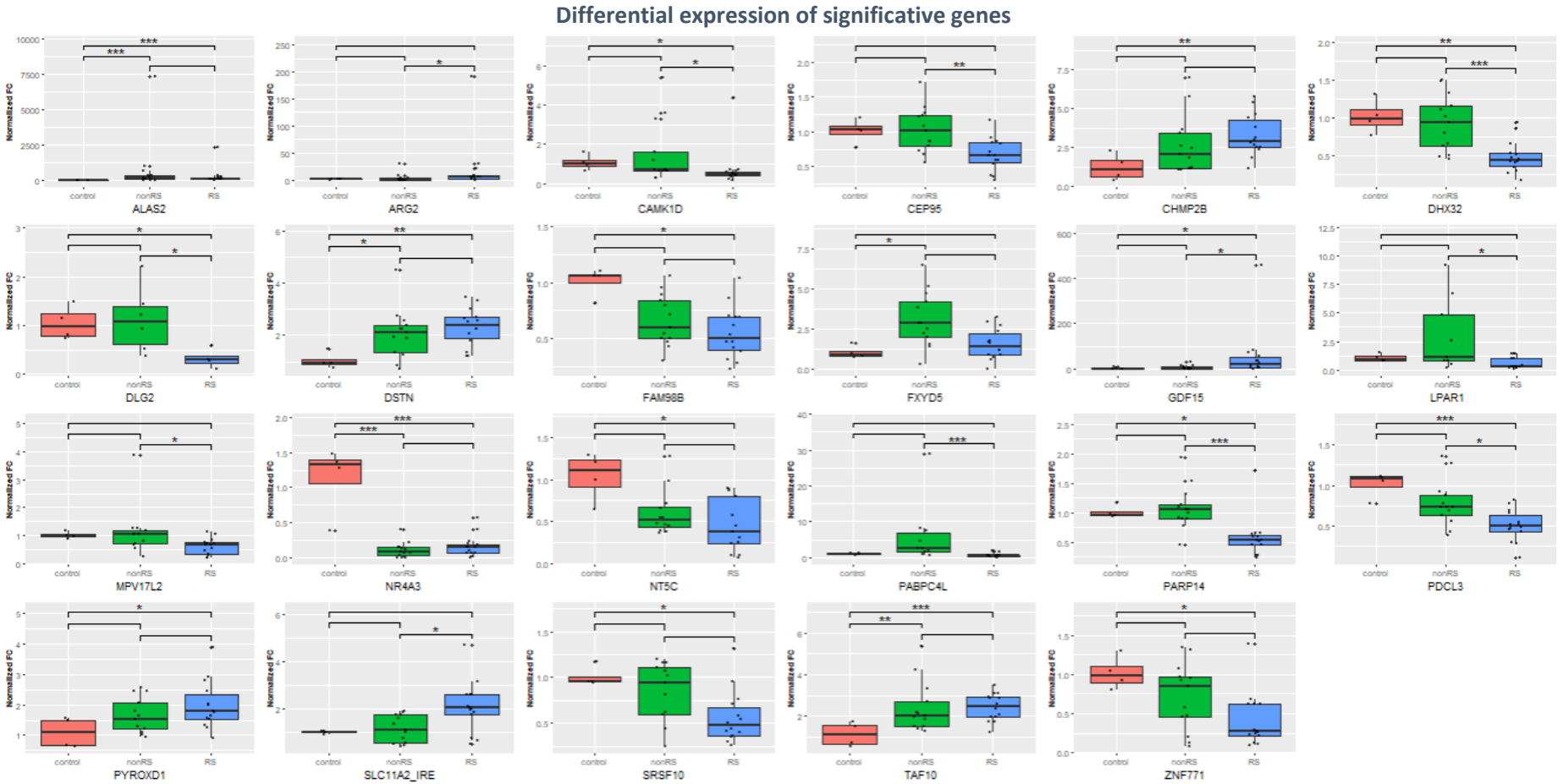


Figure 3. 9 Boxplot of gene expression of significative genes in the three tested groups. The expression (x-axis) is the normalized fold change (FC). The three groups are respresented in three different colors: red for healthy controls, green for patiens of MDS without ring sideroblast (nonRS), and blue for MDS patients with ring sideroblast (RS). Significance is indicated as: **** p<0.005; **p<0.01; *p<0.05.*

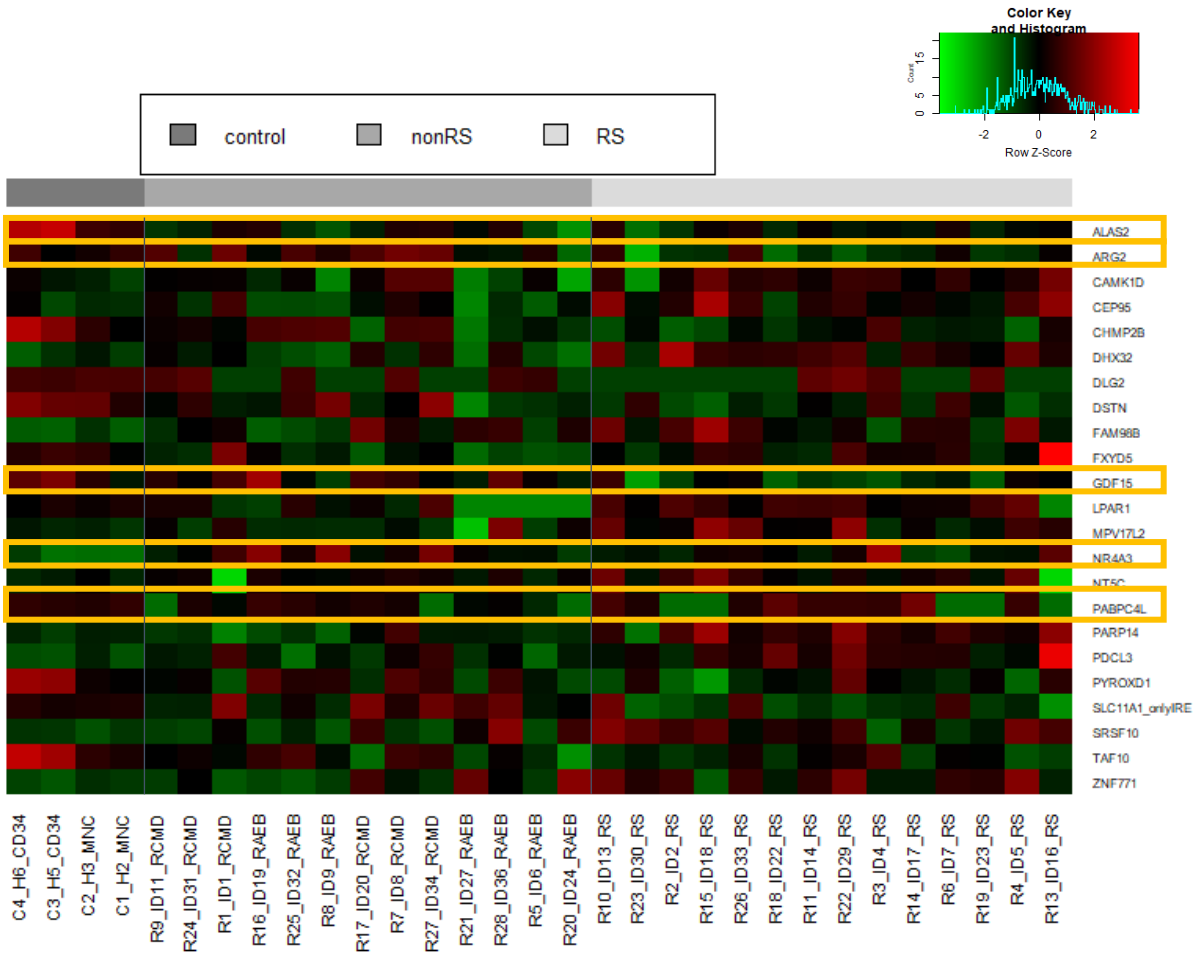


Figure 3. 10 Heatmap of the 23 significative genes. The heatmap is sorted by groups of samples: dark grey for healthy controls, medium gray for patients with MDS with ring sideroblasts (RS) and light gray for MDS patients without ring sideroblasts (nonRS). The significative genes with higher FC are indicated with yellow squares.

3.3.4. Error correction and running validation of BEA with external datasets

Four datasets from BioMark-Fluidigm analysis with three different array formats (96.96, 48.48, and 12.12 Flex formats) have been used to validate the performance of the program, and control or correct the software program errors that were unnoticed with the initial test dataset (MDS). The sample names were edited if they contain spaces or special characters others than dots, underscore or hyphen. The csv results table files were also edited with excel to ensure that the data was separated by a semicolon (;), and the files were saved as csv format. The SIFs were created for dataset 1, 2, 3 and 4 assigned to the samples two or three groups.

The four datasets were correctly analyzed with the BEA program. The different array formats do not influence the performance of the program, although, as it is expected, they have an effect on the statistical results: the smaller the sample size the lower the statistical significance. Datasets with incomplete assays (blank genes) were processed without error. Results from blank genes were removed from the analysis. None of the datasets contain information about the reference genes, so the genes for normalization were selected from the list supplied for the program with all analyzed genes. Widely used standard reference genes were used (e.g. *GAPDH*, *ACTB* or *HPRT1*) (Kozera and Rapacz 2013; Radonić et al. 2004; Dheda et al. 2004; Lee et al. 2010). The program responds

adequately even if only one gene is selected by the user, and when more than one reference genes are selected. However, when only one gene is selected, the gene-stability-measure cannot be calculated, since it needs the comparison with other reference genes, and an “NA” outcome is given.

The program can analyze data from two and three groups only and must be defined in the SIF file. When two groups are investigated, in the normalization tab step the third group is assigned as “NA” and this group is blocked as “Dont_study”. Negative control samples (sample type “NTC” or “NAC”) are described in the SIF as NA in the GROUP column and are excluded for the analysis in the BEA program. Rest of steps and analysis were carried out without issues, and final report with statistical analysis were obtained from all of them.

3.4. DISCUSSION

BEA is a ready-to-use desktop program based in R/Shiny language that is easy to install and deploy, whose purpose is to provide an interactive statistic tools for the analysis of BioMark-Fluidigm data in an intuitive step-by-step process. BEA is ideal for biologists with not much experience in bioinformatics, since it allows the user to benefit from the power of R/Bioconductor packages without having programming knowledge. BEA has been designed to analyze different IFC array formats, multiple conditions and replicates. The analysis includes relative quantification using ΔΔCT normalization and appropriate statistical testing to identify differentially expressed genes between tested groups, considering test of normality and number of groups. Also, BEA program will help the user to select suitable reference gene for reliable normalization and a qPCR data analysis, by providing three essential parameters: mean Ct value, deviation of the data, and gene-stability measure value (Thellin et al. 1999; Vandesompele et al. 2002).

BEA has been tested using data from three groups of individuals: patients of MDS with development of ring sideroblasts (MDS-RS), patients of MDS which did not develop ring sideroblasts, and healthy group control. Using the BEA program, twenty-three genes were identified as differentially expressed between at least one group. The three positive control genes included in the experiment (*ALAS2*, *GDF15* and *ARG2*), were statistically significant in the differential expression analysis, confirming the good performance of the experiment and the analysis pipeline.

Twelve out of the twenty-three identified genes were significative differentially expressed between RS and nonRS-MDS patients. *ARG2*, *GDF15*, and *SLC11A2_IRE* genes were overexpressed in RS MDS, meanwhile the rest of significant genes (*CAMK1D, CEP95, DHX32, DLG2, LPAR1, MPV17L2, PABPC4L, PARC14* and *PDCL3*) were down regulated. Conte et all identified that the *CAMK1D* gene was down-regulated in RS-MDS bone marrow compared with healthy bone marrow samples (Conte et al. 2015) and del Rey et al. found that *SLC11A2* gene is overexpressed in RS patients (del Rey et al. 2015), confirming our findings. We indeed identify that is the IRE isoform of *SLC11A2* the one that is overexpressed and not the *SLC11A2* non-IRE isoform that was also included in the list of 96 genes.

Except for *CAMK1D, SLC11A2* and the positive control genes *ARG2* and *GDF15*, no studies were found to support the differential expression of rest of the twelve genes identified to be differentially expressed in RS-MDS.

Four external datasets were used to validate the correct performance of the program, and correct possible errors generated by the training of the program with the test dataset. The preprocessed data converted in the “Ct-values results table” file can differ from the different experiment due to several factors (i.e. number of assays, type of experiment: EvaGreen or Taqman workflow, description of reference genes…), and these differences may cause problem in the program performance. For this reason, the correct performance of BEA program was validated by analyzing four external datasets with differences in the IFC design, the sample size and number of genes analyzed, and the chemistry workflow (EvaGreen or Taqman).

BEA is the only bioinformatic tool available for the analysis of BioMark-Fluidigm data for non-bioinformatics that can perform relative quantification and statistical analysis for two and three groups in an easy, intuitive and interactive step-by-step process. Other programs for transcriptomic analysis have been developed, but were focused in low-throughput analysis (regular qPCR) or other high-throughput analysis such as RNA-seq or microarray (Mallona et al. 2017; K. Choi and Ratner 2019; Kallio et al. 2011). Other web tools were created for the analysis of other BioMark-Fluidigm data, such as SCExV, that was designed for the analysis of single cell BioMark data (Lang et al. 2015).

Current limitations and future improvements

Although BEA is an useful software, it is in its early stage (alpha version) and we are aware that the program has some limitations that need to be improved in future versions. Firstly, BEA is designed to perform analysis only when two or three groups of samples are present. The program is not able to process data with more than three groups, and although most of the relative quantification studies reported in the past are from two groups (case vs control studies) (Igci et al. 2016; Tabur et al. 2015) or three groups of patients (Yang et al. 2017; Li et al. 2019), the high-throughput technology is bringing more complex analysis including several groups (Kebschull et al. 2017).

Other important limitation is that the program is not able to perform background correction by using negative control samples ("NTC" or "NAC" sample types). The program automatically removes these negative control samples, and performs the relative quantification by using the ΔΔCT normalization method (Livak and Schmittgen 2001). Background calibration could be made using BioMark qPCR software in the preprocessing step, before exporting the Ct-value results table. However, in order to provide all the utilities that the user may require, the background calibration will be implemented in a future version of the program.

BEA can only be executed from Windows operating system, so users with Linux and MacOS computers are not able to use the program. In order to make the program completely available, two possibilities have been considered for future version of the program: create compiled package for Linux and macOS also and recommend to download the required program based in user operating system (K. Choi and Ratner 2019), or create a web page server-side so BEA would be available for any operating system computer with internet connection and web browser (Q. Liu and Gregory 2019). The creation of compiled zip folder with the program (as is currently done in the present version for Windows) for all operating systems is used in other published bioinformatics programs. Choi et al. developed a Shiny based desktop application for the analysis of RNA-seq data. Same as BEA, the program can be downloaded in a compiled folder and executed in a portable R and portable chrome browser. However, iGEAK is available both for windows, Linux and macOS, so three options of download are available from the IGEAK project web page (K. Choi and Ratner 2019). This option makes the programming process simpler than the creation of a web server-side. Some advantages include that user do not require internet connection, or separately web browser, but could be less attractive and accessible than a server-side web service. On the other hand, the creation of a web-server page as other shiny application (Koeppen, Stanton, and Hampton 2017; Ramirez et al. 2019; Mallona et al. 2017), make the program on hand from commodity computer or mobile devices, as far as they have internet connection and updated web browser installed.

BEA is designed to analyze gene expression data from high-throughput BioMark Fluidigm system, and technology is at present widely used to identify molecular biomarkers that help to better classify of different phenotypes. However, results from the analysis of gene expression data alone may be misleading since other possible factor (i.e. environmental, socio-economics, population ancestry…) might be obviated. In order to make the analysis more complete, an integrative analysis that include the gene expression data and other non-genomic data (clinical, pathological, environmental data…) should be performed (López de Maturana et al. 2019). Future improvements of the program include the possibility to include non-genomic data in the analysis, and the implementation of an integrative modelling approach (see chapter 4 for more details about integration of genomics and non-genomics data).

To sum up, BEA is the unique open source program freely available dedicated to help the wet-lab users with limited knowledge in bioinformatics to analyze the high-throughput data originated by BioMark

Fluidigm system in an intuitive and interactive eight-tab process (video tutorial available at https://youtu.be/GL2a1BLOnM8). In this research, we described an alpha version of the program, fully functional and available in http://bloodgenetics.com/otros/. However, further work must be carried out to eliminate the limitations and implement the improvements discussed here.

CHAPTER IV:

Pipeline for construction of gene expression signatures for prediction and prognosis

4.1. INTRODUCTION

4.1.1. Gene expression signatures in biomedicine

The advance and expansion of high-throughput technology has brought the opportunity to increase our knowledge of the molecular biology of different cellular conditions (Macgregor and Squire 2002) and to identify specific genetic signatures (Colombo et al. 2011; Tonella et al. 2017). A gene expression signature is a characteristic patter of gene expression for a group of genes that occurs in a normal or altered biologic process, pathogenic medical condition or pharmacological responses to a specified treatment (Itadani, Mizuarai, and Kotani 2008). These types of biomarkers are mostly studied in oncology to understand the signaling pathways involved in the development of cancer and drug sensitivity (Itadani, Mizuarai, and Kotani 2008; Chibon 2013).

Genetic signatures have three main clinical application: diagnosis, prognosis or prediction of therapeutic response. Diagnostic gene signatures refer to biomarkers that have the potential to better identify and/or classify a disease, and also to understand the biology of the disease. These biomarkers could help practitioners to verify the clinical diagnosis and provide more accurate healthcare and therapeutic options (Nguyen, Welty, and Cooperberg 2015). Prognostic gene signatures are biomarkers for prediction of the likely outcome or development of a disease. Identification of subgroups of tumor with different biology can provide a prognostic biomarker for the disease or the associated phenotype (Tonella et al. 2017). And finally, the predictive gene signature intends to predict the response and the effect to treatment. The objective of this type of genetic signature, that is highly popular in pharmacogenomics, is to identify the group of subjects who respond efficiently to a specific treatment. These signatures are involved in the evolution and growth of the personalized medicine by identifying novel therapeutic targets and the most qualified patients for specific treatments.

4.1.2. Genetic signatures in complex traits and the challenge of integration of genomic and non-genomic data.

Most phenotypes are complex traits, resulting from the influence of internal alterations (genomic factors) and external conditions such as environmental circumstances (non-genomic factors). To understand these complex traits and be able to define a genetic signature of these phenotypes, it is essential to study the genomic information and the non-genomic information jointly. However, the integration of genomic data with clinical/epidemiological data requires elaborated modeling and engenders some challenges that must be considered. One of the challenges in integrating genomic and non-genomic data are associated with the constitution of each type of data. The clinical/epidemiological data are complex to define and subjective. Some clinical variables have not yet a standardized protocol, and are subject to different clinical procedures and the knowledge of practitioners. This subjectivity may produce inconsistent and heterogeneous data that may influence the quality and the grade of comparability of the data, and as a result, the inaccuracy in the outcome prediction. Also, in contrast with genomic data, clinical/epidemiological variables are very heterogeneous, from qualitative (i.e., gender, presence of pathogenic events such as hydronephrosis, lymphovascular invasion) to quantitative variables of different scales (i.e., age, number of nodes, hematological levels). Data require to be transformed and normalized before the integration analysis.

Other challenges of the integration of genomic and non-genomic data are due to the relationship between them: ascertainment bias or the presence of interaction between the genomic data with phenotypic variables.

Three modeling methods have been characterized for integrative modeling: independent, conditional and joint modeling (López de Maturana et al. 2019). In the independent modeling, the genetic and non-genetic models are built independently. The clinical/epidemiological variables are selected separately to fit the clinical model, and the genetic variables are selected in parallel to build the genetic model. The clinical model and the genetic model are combined finally, and the predictive power is estimated. The conditional modeling strategy is based in a two steps process. First, the clinical model is constructed only using the clinical/epidemiological variables. And second, the genomic variables are selected according to a model that adjusts for the clinical variables. As exposed in López de Maturana et.al., "the key point of this conditional modeling approach is to decide which omics variables should be added to the clinical model" (López de Maturana et al. 2019). Finally, the joint modeling approach build the final model by using the genomic and clinical data in conjunction.

In order to build the genetic signature, regression analysis through generalized linear models is the most commonly used statistical modelling technique. Regression analysis is a parametric statistical technique that expresses the relationship between a dependent variable (response) and a set of independent variables (also called explanatory variables or "predictors") through some regression coefficients (Gordis L. 2013). The aim of regression analysis is to build the model that best fits the observed data set according to some loss function (MSE, deviance, etc.). When building genetic signatures, the genetic data and the clinical/epidemiological data are the explanatory variables and their association with the response variable (e.g., survival, response to treatment, tumor growth) is evaluated. Three different generalized linear models may be used depending on the nature of our outcome: linear regression, logistic regression and Cox regression (Figure 4.1). Linear regression is used when the dependent variable is a quantitative continuous outcome with approximately normal distribution. We refer to simple regression when a single independent variable ("the regressor") is analyzed with the dependent variable, and multiple linear regression when more than one explanatory variable is evaluated. Logistic regression provides a setting for modeling dichotomous outcomes based on multiple categorical or continuous predictors (Hosmer DW, Lemeshow S, Sturdivant RX. 2013). Cox regression model or proportional hazards regression is used for investigating the relationship between time-to-event (survival-time) outcomes on one or more predictors. The three regression models have a unified formulation as generalized linear models where a transformation of the mean of the response variable is expressed as a linear function of the explanatory variables:

$$g\big(\mu(Y)\big) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$$

For the linear regression model $g()$ is the identity function, for the logistic regression model $g()$ is the logistic function, that is, $g(x) = \log\left(\frac{x}{1-x}\right)$, and for the Cox model, $g()$ is the hazard ratio with respect to the baseline hazard.

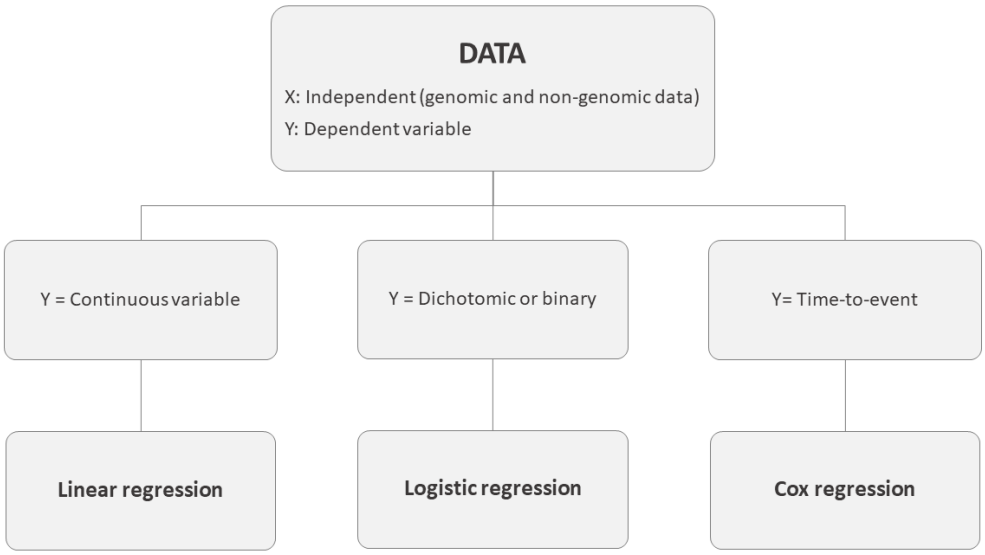


Figure 4. 1 Regression techniques according to the dependent variable. This figure is original and was designed by Beatriz Cadenas

4.1.3. Statistical strategies for variable selection

Different statistical strategies are available for selection of variables. The simplest strategy is the univariate analysis, which evaluates individually the association of each covariate (or gene in the genetic signature) with the response variable. Univariate selection performs poorly because it ignores possible correlations between the predictors. This can generate misleading results when several covariates are tested, and the multivariate analysis is more appropriate. The multivariate strategy is used to analyze the relationship of more than one independent variable with the dependent variable at a time. This technique considers the effect of all covariates on the response variable. Logistic regression is the standard method for modeling binary outcomes since it allows to calculate and easily interpret the odd ratios (Rothman KJ, Greenland S, Lash TL. 2012; Hosmer DW, Lemeshow S, Sturdivant RX. 2013; Gordis L. 2013). However, the classical logistic regression does not perform properly when the number of variables is larger than the number of observations, and problems of multicollinearity and overfitting may arise. This may lead to loss of predictive power of the model and biased risk estimates, or inability of estimation at all (Doerken et al. 2019). An alternative approach to avoid these challenges is to implement penalized regression models such as LASSO, ridge or elastic-net that performs dimensionality reduction within the inference process (Ishizawa et al. 2018). In penalized or regularized regression models, a penalty term is added to the linear regression to reduce the complexity of the model and mitigate multi-collinearity.

Ridge regression adds an L2 penalized least squares criterion to the regular lineal regression (Hoerl and Kennard 1970):

$$+\lambda \sum_{j=0}^{p} \beta_j^2$$

If lambda is equal to zero, the regression has no penalization and it reduces to the standard linear regression model. The larger is lambda, the larger is the shrinkage of the coefficients that trend

towards zero. This method reduces the complexity of the model by decreasing the variance and the error, but it does not reduce the number of variables.

LASSO (Least Absolute Shrinkage and Selection Operator) is a regularization method that uses an L1 penalized criterion (Tibshirani 1996):

$$+\lambda \sum_{j=0}^{p} |\beta_j|$$

Same as in ridge, when lambda is zero, a linear regression is obtained. When increasing lambda, some coefficients are driven to zero. The larger is lambda, more variables are shrunk to zero. With this regularization method, some variables are eliminated, reducing the complexity of the model and the multi-collinearity. Lasso is a popular variable selection method widely used in genomics (Ogutu, Schulz-Streeck, and Piepho 2012).

Elastic Net (ENET) is an extension of LASSO that uses a mixture of the L1 (LASSO) and L2 (Ridge) penalties (Ogutu, Schulz-Streeck, and Piepho 2012).

$$+\lambda \left(\frac{1-\alpha}{2} \sum_{j=0}^{p} \beta_j^0 \; + \; \alpha \sum_{j=0}^{p} |\beta_j| \right)$$

Additionally, to the estimation of lambda, ENET uses also the parameter alpha. The ENET simplifies to simple ridge regression when α =1 and to the lasso when α =0 (Ogutu, Schulz-Streeck, and Piepho 2012). This method is useful when there is correlation between the covariates; ENET tends to eliminate together groups of correlated variables that are not informative (Zou and Hastie 2005).

4.1.4. Bladder cancer study

Bladder cancer is the seventh most frequent tumor worldwide in men and the tenth for both sexes. In 2018, 549,393 new cases of bladder cancer and 176,963 deaths were registered around the world, and more than 75% of the cases and deaths were men. The incidence is 9.6 annual rate per 100,000 persons at risk (ASRs per 100,000 person-years) in men and 2.4 in women, and the mortality rate is 3.2 deaths in men and 0.9 in women (Wong et al. 2018; Bray et al. 2018).

Tobacco smoking is the most important risk factor for bladder cancer. Smokers are at least 3 times as likely to get bladder cancer as non-smokers. More than 50% of bladder cancer cases are associated to smoking in both sexes in United States. Other causes as certain exposures to chemicals during work, or water contaminants as arsenics, or some analgesics, can increase the risk to develop bladder cancer (Saint-Jacques et al. 2014). Moreover, some genetic studies have confirmed that the incidence of bladder cancer can be influenced by genetic susceptibility factor (Murta-Nascimento et al. 2007; Rothman et al. 2010; Kiemeney et al. 2008; Garcia-Closas et al. 2013).

A 25% of all patients with bladder cancer develop cancer in the thick detrusor muscle of the bladder, a condition that is known as muscle-invasive bladder cancer (MIBC). This type of cancer has higher risk to spread and metastasis, increasing the likelihood to death. Between 20-25% of affected MIBC who have received radical cystectomy develop microscopic spread to the lymph nodes. Hence, the treatment targets are the cancer cells in both bladder and lymph nodes (Lerner, Schoenberg, and Sternberg 2015).

If the bladder cancer has invaded the muscle of the bladder wall, then there is a very high risk of death for the patient unless radical treatment is applied. Radical treatment includes either radical cystectomy or radiotherapy, and although both of them offers the best chance of cure, 50% of patients die due to tumor progression despite the treatment received. The probability of relapse can be reduced or delayed by administrating chemotherapy before the radical surgery or radiotherapy (Neoadjuvant chemotherapy). Treatment with neoadjuvant chemotherapy based in cisplatin (NAC) followed by radical cystectomy is the reference standard for MIBC, increasing the survival ratio a 5% to 10% compared to cystectomy alone (Grossman et al. 2003; Griffiths et al. 2011; Hensley et al. 2019).

Several biomarkers involved in different metabolic pathways have been identified to be associated with prediction response of neoadjuvant cisplatin-based chemotherapy. These biomarkers are involved in regulation of apoptosis and cell differentiation, DNA repair pathways, receptor tyrosine kinases, chromatin remodeling targets, taxonomic classification genes and immune response pathways (Supplementary material Table S4) (Brown et al. 2017; Buttigliero et al. 2017; Font et al. 2011; Pitroda et al. 2014; Yoshida et al. 2019). The identification of reliable biomarkers that allow the identification of patients who will really benefit from neoadjuvant chemotherapy is a significant challenge. This identification could lead to precision medicine, setting an individualized therapy in order to optimize the response and thus avoiding the negative impact of neoadjuvant treatment.

4.1.5. MIBC Project

The main objective of this research is to establish a statistical pipeline to build genetic signatures integrating genomic and non-genomic data. This pipeline has been constructed by using genomic and clinical/epidemiological data from MIBC patients treated with NACs followed by cystectomy (part of other project titled “Invasive bladder cancer: toward a precision medicine”). The intention of this study was to establish a clinical-molecular signature to predict the efficacy of the chemotherapeutic treatment. Details of this project are described below.

Name of the project: “Invasive bladder cancer: toward a precision medicine”
Coordinating institute: CENTRO NACIONAL DE INVESTIGACIONES ONCOLÓGICAS (CNIO)
Scientific coordinator: Francisco X. Real Arribas
Participant organisms:
1. SOGUG (GRUPO ESPAÑOL DE ONCOLOGÍA GENITOURINARIA). (Grupo II). Daniel Castellano
2. INSTITUT CATALÀ D´ONCOLOGIA (ICO)-HOSPITAL GERMANS TRIAS I PUJOL-IGTP (Grupo III). Albert Font
3. CENTRO NACIONAL DE INVESTIGACIONES ONCOLOGICAS (Grupo IV). Nuria Malats
Financing: ASOCIACIÓN ESPAÑOLA CONTRA EL CÁNCER (AECC)
Project id: SOG-CVI-2014-05

4.2. MATERIAL AND METHODS

4.2.1. Data from the MIBC Project

Genomic and clinical/molecular data was obtained from 123 MIBC patients treated with neoadjuvant chemotherapy (NAC). Three NAC schemes were used: CMV (cisplatin, methotrexate vinblastine); cisplatin with gemcitabine (CDDP + GMZ); and carboplatin with gemcitabine (CARBO + GMZ). All subjects gave their informed consent for inclusion before they participated in the study. The study was approved by the corresponding Research Ethics Committees (REC).

Clinical and pathologic anatomy history of all patients were registered, including age, sex, presence of, presence of lymphovascular invasion, the morphology of the tumor, TNM stage, presence of hydronephrosis or neoadjuvant chemotherapy used. The taxonomic classification of the tumor was estimated by evaluating the expression of four genes: *FOXA1*, *GATA3*, *KRT5* and *KRT14* (W. Choi et al. 2014).

Gene expression data was extracted from the analysis of expression of 41 genes in the transurethral resection of bladder tumor of these 123 patients. Gene expression analysis was performed by using nCounter® assay (NanoString Technologies, Seattle, WA). The complete list of genes included in the assay can be found in Supplementary material Table S4. Background correction and sample normalization was performed in R using custom pipeline according to manufacturer instructions (Gene Expression Data Analysis Guidelines from Nanostring). The data were logarithmically transformed to ensure normal distribution.

4.2.2. Statistical strategy

1. **Preprocessing**
    a. **Filtering process**

In the genomic dataset, those samples with expression values lower than the negative control in more than 80% of the genes were filtered out. In the same way, those genes with more than 80% of their values lower than the negative control were discarded. Values below negative control indicate low performance or deficiency assays.

Regarding the non-genomic dataset, those samples with incomplete clinical/molecular registry were excluded.

    b. **Descriptive analysis of the clinical/molecular variables**

A descriptive analysis of non-genomic variables was performed for the selected samples to study the distribution of the samples for each covariate.

2. **Construction of genetic signatures integrating the genomic and non-genomic datasets**

The purpose was to build a gene expression signature of the non-response risk to neoadjuvant treatment in patients with MIBC. A conditional modeling approach was followed for construction of the genetic signature model. The clinical model was defined only including the non-genomic variables. Secondly, the genomic covariates were included to the preset clinical model to build the genetic

signature. Two multivariate logistic strategies were considered for the selection of the non-genomic and genomic variables and a cross-validation process to measure the predictive ability of the models.

a. Construction of the clinical model

A multivariate strategy (stepwise regression method) was used to identify the non-genomic variables that are associated to the response variable (cisplatin response). Among the non-genomic data one variable was the molecular classification of the tumor (cluster), which was established based on the genomic information. Since this variable is genomic dependent, we considered to evaluate two different models: a model with clinical and pathological variables but not the taxonomic classification (model 1) and a second model (model 2) including all the variables included in the first model and also the taxonomic classification variable.

b. Construction of the clinical-pathological-genetic model

Multivariate logistic penalized regression (LASSO and Elastic-Net) was applied to identify genetic biomarkers associated with cisplatin response, considering the associated variables studied in model 1 and model 2. The linear predictors from model 1 and model 2 were used as scores representing the associated variables of each model and these score were attached to the gene expression data to build model 3 (genetic variables with score of model 1) and 4 (genetic variables with score of model 2) respectively (Figure 4.2). The penalization of the model (lambda) and the optimal number of variables for the model was calculated with a cross-validated LASSO.

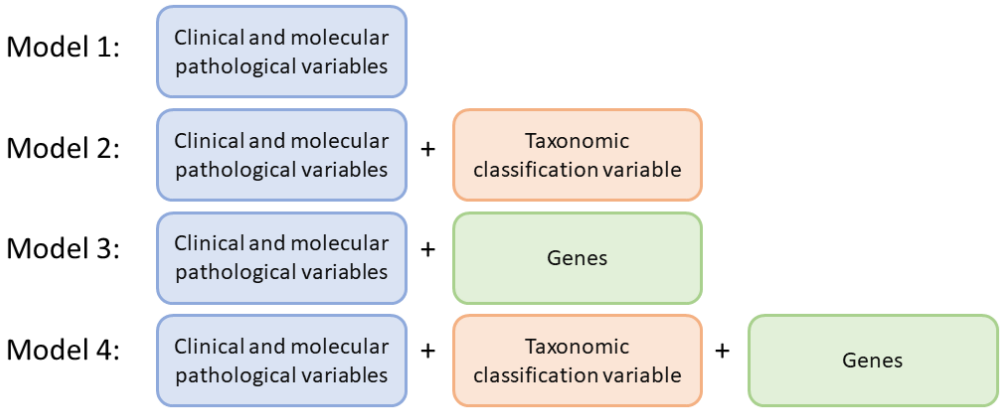


Figure 4. 2. Design of the models to evaluate the clinical and genomic signature from patients with MIBC.

To reduce overfitting, the sample set was split in a training set (70% of samples) and a test set (30% remaining). The genetic signature was modeled with the training set, and the predictive ability was calculated with the test set by estimating the AUC (Area Under the Curve) of the ROC (Receiver Operating Characteristics) curve for the four models previously described. In order to obtain more robust results, the sampling step, the modelling process and the AUC measurement were cross-validated in a hundred cycles (k=100). The mean AUC of the cross-validation was finally determined for the four models.

3. Estimation of the increasing predictive ability with the selected genes

Additionally, the increase of the predictive ability was evaluated with the addition of the genes selected by LASSO one by one. The purpose of this step is to evaluate the discriminatory information given for each gene of the genetic signature.

4. Univariant analysis

A univariant analysis (simple logistic regression) was performed to assess the independent association of each gene with the pathological response.

5. Survival analysis

Survival analysis of the model with better prediction power was performed to evaluate the prognosis of patients with favorable and unfavorable genetic signature. These two groups of patients were formulated by qualifying the linear predictors of the final model. Samples with a linear predictor value higher than the median were assigned as patients with favorable response to the cisplatin treatment ("Firm 1"); and on the contrary, samples with lower linear predictor value than the media were assigned as an unfavorable response ("Firm 2"). Survival curves were studied using the Kaplan-Meier estimator, and the survival discrepancy between the two groups of signatures was calculated using Log-rank test. Four types of survival were studied: progression-free survival (PFS), time during and after the treatment in which the disease being treated does not worsen; overall survival (OS) , time to death from any cause; disease-specific survival (DSS), time to death from the specific disease studied; and event-free survival (EFS), time to death or progression of the disease.

To evaluate how the information of each gene that is part of the genetic signature is reflected in the prognosis and survival of the patients, the survival analysis was performed in different models: from only the clinical model, to the complete clinical-genetic model, adding the genes selected by LASSO one by one.

Table 4. 1 Increasing gene models for survival analysis

	Model	Non-genomic variables	Genonmic variables
1	Clinical model only	Clinical model	NA
2	Clinical + 1 gene	Clinical model	1st gene (selected by LASSO)
3	Clinical + 2 genes	Clinical model	1st gene + 2nd gene
4	Clinical + 3 genes	Clinical model	1st gene + 2nd gene + 3rd gene
N	Clinical + N genes	Clinical model	1st gene + 2nd gene + 3rd gene + … + N gene

4.3. RESULTS

4.3.1. Pipeline to construct genetic signatures integrating genomic and non-genomic data

A pipeline has been developed to construct genetic signatures integrating genomic and non-genomic data following a conditional integrative strategy (Figure 4.3). The pipeline is structured in 5 steps: preprocessing, clinical model construction, genetic signature construction, estimation of the increasing AUC, and survival analysis.

In the **preprocessing** step, samples with incomplete non-genomic variables are discarded. Also, the normalized genomic data set is filtered according to the gene expression values. Samples or genes with lower value than the negative control are excluded.

In the **clinical model construction** step, the multivariate stepwise regression test is used to build the clinical model that provides the best goodness-of-fit. The non-genomic variables with a higher association to the dependent variable are selected to construct the clinical model. The linear predictor (clinical score) of the clinical model are extracted as a single and numerical representation of the clinical model. The clinical score attached to the genomic dataset for the next step of the conditional strategy.

In the **genetic signature construction** step, the genetic variables with a higher association to the dependent variable are selected while adjusted by the clinical score. A cross-validation process is performed to avoid overestimation of the predictive ability of the model. The data is split in training (70%) and test (30%) sets. The clinical-genetic model is built with the training set using LASSO methodology, and the prediction capacity (AUC) of the model is estimated using the test-set. The sampling process and the variable selection is performed 100 times to obtain more robust and reliable results. The genetic signature is finally chosen depending on the selection rate of the genes by LASSO. The mean AUC of all AUC calculated in the loop is obtained. Univariant logistic regression analysis is performed in order to describe the independent association of the gene with the dependent variable.

The following step is the estimation of the **increasing classification (AUC)** as the different selected genes are added to the clinical score. The selected genes are ranked according to their LASSO selection rate. Then, sequential models are considered by adding a new gene to the clinical score: (1) only the clinical variable; (2) the clinical score and the most selected gene by LASSO, (3) the clinical score and the two most selected genes and so on. The AUC is calculated for these models to observe the increase in the predictive power with each variable added.

Finally, **survival analysis** is done for the previously described models (adding the variables one by one), to evaluate the prognosis effect of the genetic signature. The linear predictors are extracted from the models as a representation of the risk of non-response of the samples. Linear predictors with values higher than the median are associated with one specific genetic signature profile (Firm 1), and linear predictors which values are lower than the median are assigned to an opposite genetic signature (Firm 2). The Kaplan Meier curves, and the Log-Rank test are studied for each model to evaluate the survival discrimination of these two groups of genetic signatures.

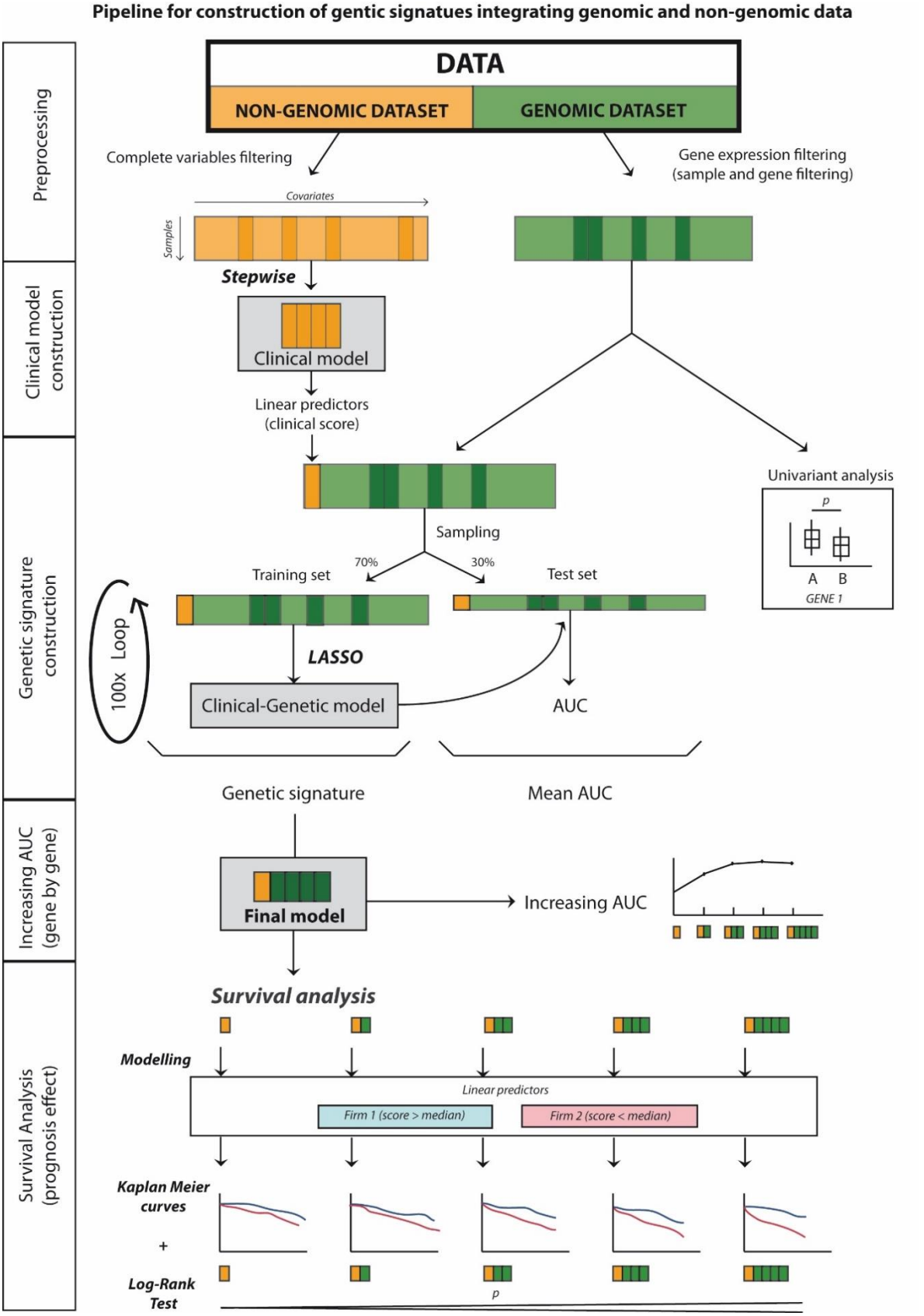


Figure 4. 3 Pipeline for the construction of genetic signatures by integrating genetic and non-genetic data.

4.3.2. Validation of the pipeline with bladder cancer data

Genomic and non-genomic (clinical/pathological/taxonomical) data from patients with MIBC have been used to validate the statistical pipeline for the construction of genetic signatures integrating genomic and non-genomic data. The objective of this study was to establish a clinical-molecular signature to predict the efficacy of the neoadjuvant chemotherapeutic treatment. The dependent variable here is the response to the chemotherapy (complete vs partial/non-response).

4.3.2.1. Filtering of samples and genes.

Only 112 samples out of the 123 samples analyzed in the Nanostring were selected for statistical analysis. Four samples were discarded due to bad Nanostring performance and 7 because of incomplete clinical history. Three genes (*PD1*, *FANCC* and *IFNG*) were filtered out of the analysis since more than 80% of the values were null (Table 4.2). So finally, 112 samples and 39 genes were used for the statistical analysis.

Table 4. 2 Genes discarded for the analysis

	% null values	Number of correct (no null) values	
		N (Complete)	N (partial/No response)
PD1	99.10	1	0
FANCC	83.92	6	12
IFNG	83.03	9	10

4.3.2.2. Descriptive analysis of the clinical and pathological variables of samples.

A descriptive analysis of the clinical, pathological and taxonomic variable was performed for the 112 selected samples (Table 4.3).

Table 4. 3 Description of the clinical, pathological and taxonomic variables

	Variable	**n**	**%**
Gender	Male	104	92,86
	Female	8	7,14
Age	min	41	
	max	80	
	mean	65,53	
Limphovacular invasion (TUR)	No	98	87,50
	Yes	14	12,50
Urothelial morphology	Urothelial	96	85,71
	Squamous	12	10,71
	Other	4	3,57
Clinic TNM	(A) <T2N0M0	10	8,93
	(B) T3a-4aN0M0	83	74,11
	(C) >N1Mx	19	16,96
Hydronephrosis	No	67	59,821
	Yes	45	40,179
CT Scheme	CDDP+GMZ	68	60,714
	CARBO+GMZ	12	10,714

	CMV	28	25
	Others	4	3,5714
Previous superficial bladder tumor	No	93	83,036
	Yes	19	16,964
Taxonomy / cluster	BASQ-like	44	39,286
	Luminal-like	38	33,929
	Mixed	30	26,786
Pathological response	Complete	36	32,143
	Partial or No response	76	67,857

4.3.2.3. Clinical, pathological and taxonomical modeling and construction of the genetic signature.

Stepwise regression analysis was used to analyze which clinical, pathological and taxonomical variables are associated to response to cisplatin treatment. The selected variables were those that minimize the Akaike information criteria (AIC). The results of the multivariate logistic models 1 (all clinical and pathological covariates) and model 2 (covariates from model 1 and the taxonomic classification variable) are shown in Table 4.4 and 4.5 respectively. Limphovascular invasion, urothelial morphology and TNM scheme were selected by stepwise regression of model 1, meanwhile morphology, presence of hydronephrosis, and cluster taxonomy were selected for model 2.

Table 4. 4 Stepwise regression coefficients of Model 1 (clinical-pathological model)

	β coefficient	Std. Error	z value	Pr(>|z|)
(Intercept)	0.626	0.255	2.450	**0.014**
Limphovacular invasion (Yes vs No)	1.250	0.717	1.744	**0.081**
Morphology (Squamous vs Urothelial)	0.979	0.812	1.205	**0.228**
TNM scheme ('>N1Mx' vs 'T3a-4aN0M0')	-0.288	0.575	-0.502	**0.616**
TNM scheme ('>N1Mx' vs '<T2N0M0')	-1.506	0.642	-2.346	**0.019**

Table 4. 5 Stepwise regression coefficients of Model 2 (clinical-pathological and taxonomic model)

	β coefficient	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.559	0.421	-1.329	**0.184**
Morphology (Squamous vs Urothelial)	1.6555	0.874	1.893	**0.058**
Hydronephrosis (Yes vs No)	0.698	0.470	1.485	**0.137**
Cluster (Luminal-like vs BASQ-like)	0.939	0.510	1.842	**0.065**
Cluster (Mixed vs BASQ-like)	2.218	0.665	3.334	**0.0009**

The linear predictors of both stepwise regression models, referred as model 1 score and model 2 score, were attached to the genetic data to construct the model 3 and 4, respectively. The dataset was split in the training set (70% of the data) and test set (30% remaining). LASSO regression modeling (least absolute shrinkage and selection operator) was performed using the training set, to select those genetic variables with higher association to pathological response adjusted by the clinical model. The test set was used to estimate the prediction capacity of the LASSO model. The process was repeated 100 times (5-fold cross validation with 20 iterations), and the genetic signature was built extracting those genes with higher percentage of selection of LASSO. Genes selected for model 3 and 4 are shown in Table 4.6. The model 4 results had higher prediction capacity (AUC = 0.663). The genes selected by

Lasso for model 4 are shown in table 4.7 sorted according to the percentage of selection of each gene in the 100 iterations performed by cross-validation. For genes with a negative coefficient, higher expression is associated with a higher probability of response, while for genes with a positive coefficient the association is in the opposite direction. The variable containing the clinical, pathological and taxonomy information (model 2 score) was selected in 100% of the cases. The differential expression of the selected genes for complete response vs partial or non-response are presented in figure 4.4. Only *CXCL9* was significant for univariant analysis (see Supplementary material Table S5 for the univariant results of all genes).

Table 4. 6 Predictive ability (AUC) of the four models

MODELS		Stepwise variable selection	AUC	Lasso variable selection (genes)
Model 1	Clinical-pathological variables	Limphovacular invasion, Morphology, TNM scheme	0.52	
Model 2	Clinical-pathological and taxonomic variables	Morphology, Hydronephrosis, Cluster	0.58	
Model 3	Clinical-pathological variables (model 1 score) and genes	Limphovacular invasion, Morphology, TNM scheme	0.642	*RAD51, CXCL9, PARP, HERC2, 53BP1, ERCC2, RNF168, Ku80, ATR*
Model 4	Clinical-pathological and taxonomic variables (model 2 score), and genes	Morphology, Hydronephrosis, Cluster	0.663	*RAD51, CXCL9, PARP, 53BP1, HERC2, ERCC2, CHEK1, Ku80, RNF168*

NOTE: AUC = Area Under the Curve.

Table 4. 7 Variable selection for Lasso in Model 4.

		LASSO		Mean	
		Selection (%)	β Coefficient	Complete	Partial/noResp
1	**Clinical-path-taxon variable (model 2 score)**	100	-	-	-
2	**RAD51**	92	-0.257	0.976	0.569
3	**CXCL9**	88	-0.108	4.748	3.606
4	**PARP**	70	-0.198	5.716	5.483
5	**53BP1**	51	0.102	5.206	5.618
6	**HERC2**	51	0.149	5.514	5.816
7	**ERCC2**	44	0.099	2.944	3.375
8	**CHEK1**	37	-0.078	3.724	3.404
9	**Ku80**	37	-0.339	7.379	7.335
10	**RNF168**	33	0.105	3.910	4.255

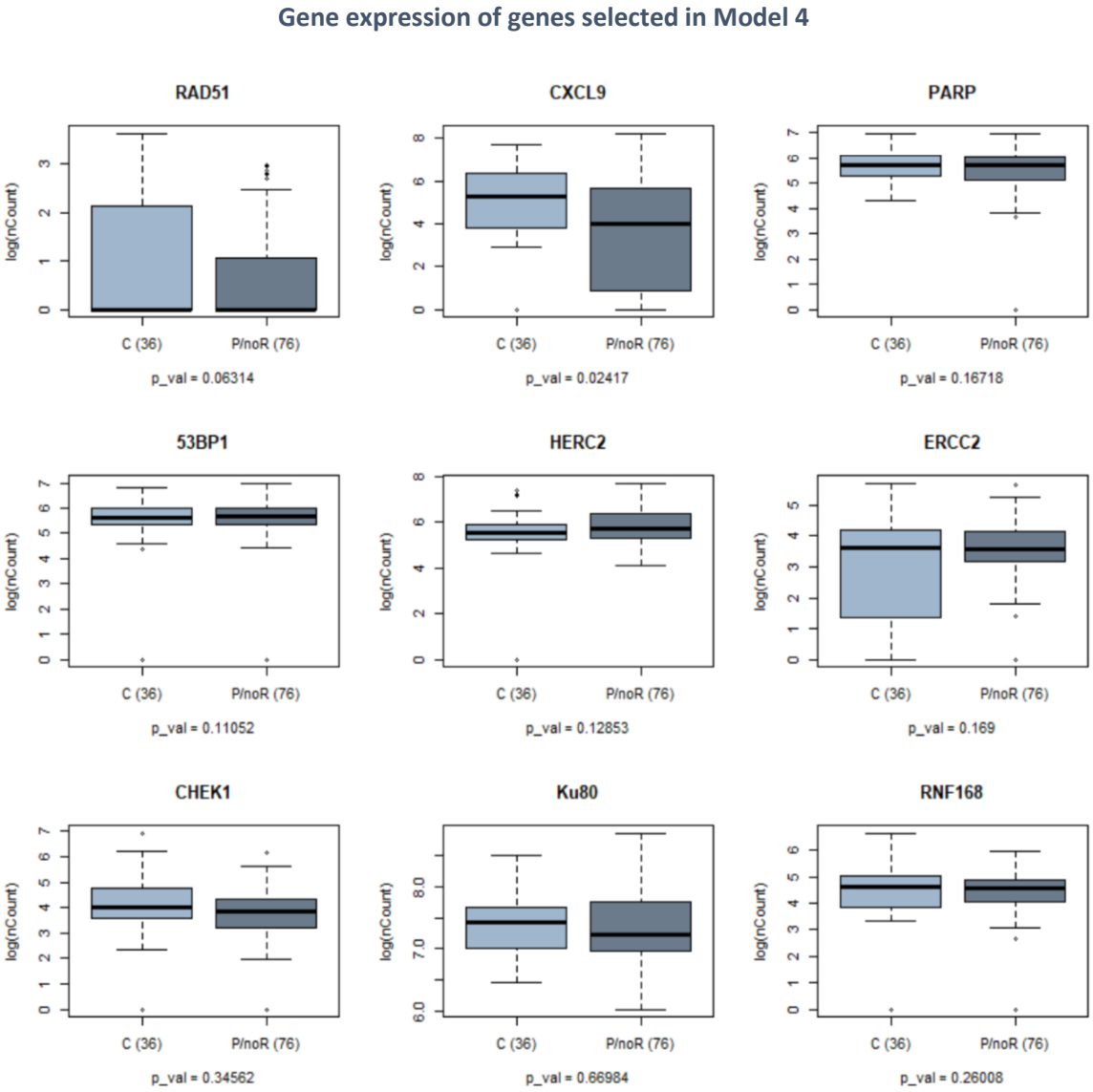


Figure 4. 4 Boxplot of the logarithm of the gene expression of the genes selected by Lasso. Patients with complete response are in light grey and patients with partial response or no response in the dark. In brackets, the number of samples per group for each gene (n) is shown. Below each graph, the p-value obtained in the univariate analysis is specified.

4.3.2.4. The increasing predictive ability

The increasing predictive ability was estimated when adding the taxonomical and genetic variables to the clinical-pathological model (model 1 with AUC of 0.52). First, the taxonomy information was added, and the AUC increased to 0.58. Subsequently, the 9 genes selected by LASSO were added in order of selection, one by one. Thus, the increase in predictive ability was observed with each gene that was added (Figure 4.5).

The predictive ability reaches the maximum value when adding the cluster variable and the first two genes selected by LASSO (*RAD51* and *CXCL9*) to the clinic-pathological model ("+*CXCL9*" in Figure 4.5). The incorporation of the rest of the genes does not increase the predictive ability.

	AUC
Clinical	0.52
+CLUSTER	0.58
+RAD51	0.617
+CXCL9	0.671
+PARP	0.659
+53BP1	0.671
+HERC2	0.664
+ERCC2	0.664
+CHEK1	0.652
+Ku80	0.665
+RNF168	0.664

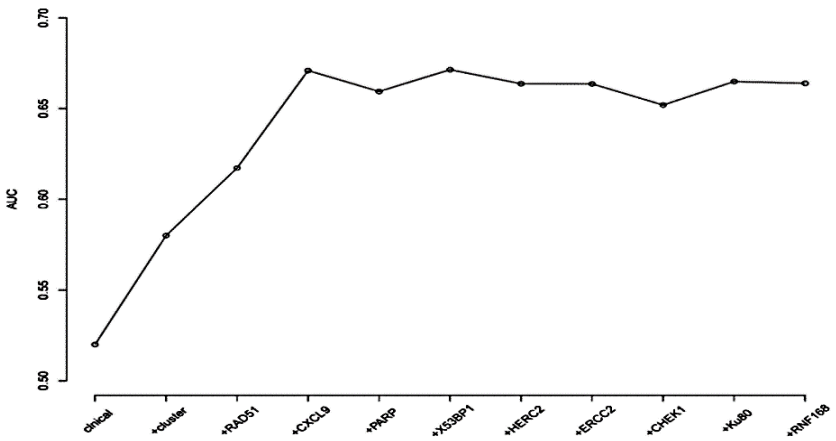


Figure 4. 5 Increase of the predictive ability (AUC) of pathological response by adding variables to the model. The first (clinical) model belongs to the clinical and pathological model. The second model (+ cluster) adds the taxonomy variable. Subsequently, the information of the genes is added one by one in the order of selection of Lasso.

4.3.2.5. Survival analysis

Kaplan-Meier curves were studied based on the variables selected by model 4 to compare the survival risk in patients with favorable and unfavorable genetic signature. In order to evaluate the survival information added for each gene, ten models were studied (described in Table 4.8) for four types of survivals: Progression-free survival (PFS – Figure 4.6), overall survival (OS , Figure 4.7), disease-specific survival (DSS Figure 4.8) and event-free survival (EFS, Figure 4.9). Samples with favorable and unfavorable genetic signature for chemotherapy response were assigned according to the linear predictor of the model: favorable signature for linear predictors higher than the median for ("Firm 1"), and unfavorable signature for linear predictors lower than the median ("Firm 2").

Table 4. 8 Models studied in the survival analysis

Model	Clinical/pathological variables	Genes
Clinical model 2 (no genes)	Morphology, Hydronephrosis and cluster	-
Clinical model 2 + 1 gene	Morphology, Hydronephrosis and cluster	*RAD51*
Clinical model 2 + 2 genes	Morphology, Hydronephrosis and cluster	*RAD51 +CXCL9*
Clinical model 2 + 3 genes	Morphology, Hydronephrosis and cluster	*RAD51 +CXCL9 +PARP*
Clinical model 2 + 4 genes	Morphology, Hydronephrosis and cluster	*RAD51 +CXCL9 +PARP +X53BP1*
Clinical model 2 + 5 genes	Morphology, Hydronephrosis and cluster	*RAD51 +CXCL9 +PARP +X53BP1 +HERC2*
Clinical model 2 + 6 genes	Morphology, Hydronephrosis and cluster	*RAD51 +CXCL9 +PARP +X53BP1 +HERC2 +ERCC2*
Clinical model 2 + 7 genes	Morphology, Hydronephrosis and cluster	*RAD51 +CXCL9 +PARP +X53BP1 +HERC2 +ERCC2 +CHEK1*
Clinical model 2 + 8 genes	Morphology, Hydronephrosis and cluster	*RAD51 +CXCL9 +PARP +X53BP1 +HERC2 +ERCC2 +CHEK1 +Ku80*
Clinical model 2 + 9 genes	Morphology, Hydronephrosis and cluster	*RAD51 +CXCL9 +PARP +X53BP1 +HERC2 +ERCC2 +CHEK1 +Ku80 +RNF168*

The survival analysis showed an increase in discrimination between the two groups of patients as the genes were added, in the four types of survival (PFS, OS, DSS and EFS). The model with the first 3 genes added (*RAD51* + *CXCL9* + *PARP*) obtained the highest discrimination with the least number of genes incorporated. The Log-rank test p.value is significant ($p<0.05$) for OS, DSS and EFS, and it is relatively close to the significance for PFS (p.value = 0.059).

In the analysis of the PFS, none of the models reached the statistical significance. The closest model to significance corresponded to the model with the nine genes ($p = 0.056$). In the OS analysis, the highest statistical significance was obtained with seven genes ($p = 0.0058$), although the model with 3 genes was powerful enough to discriminate the two signatures ($p = 0.01$). For the DSS, the greatest statistical significance was achieved with the 9 genes ($p = 0.026$), but as in OS, the model with 3 genes obtained a significant Log-rank test p.value ($p = 0.033$).

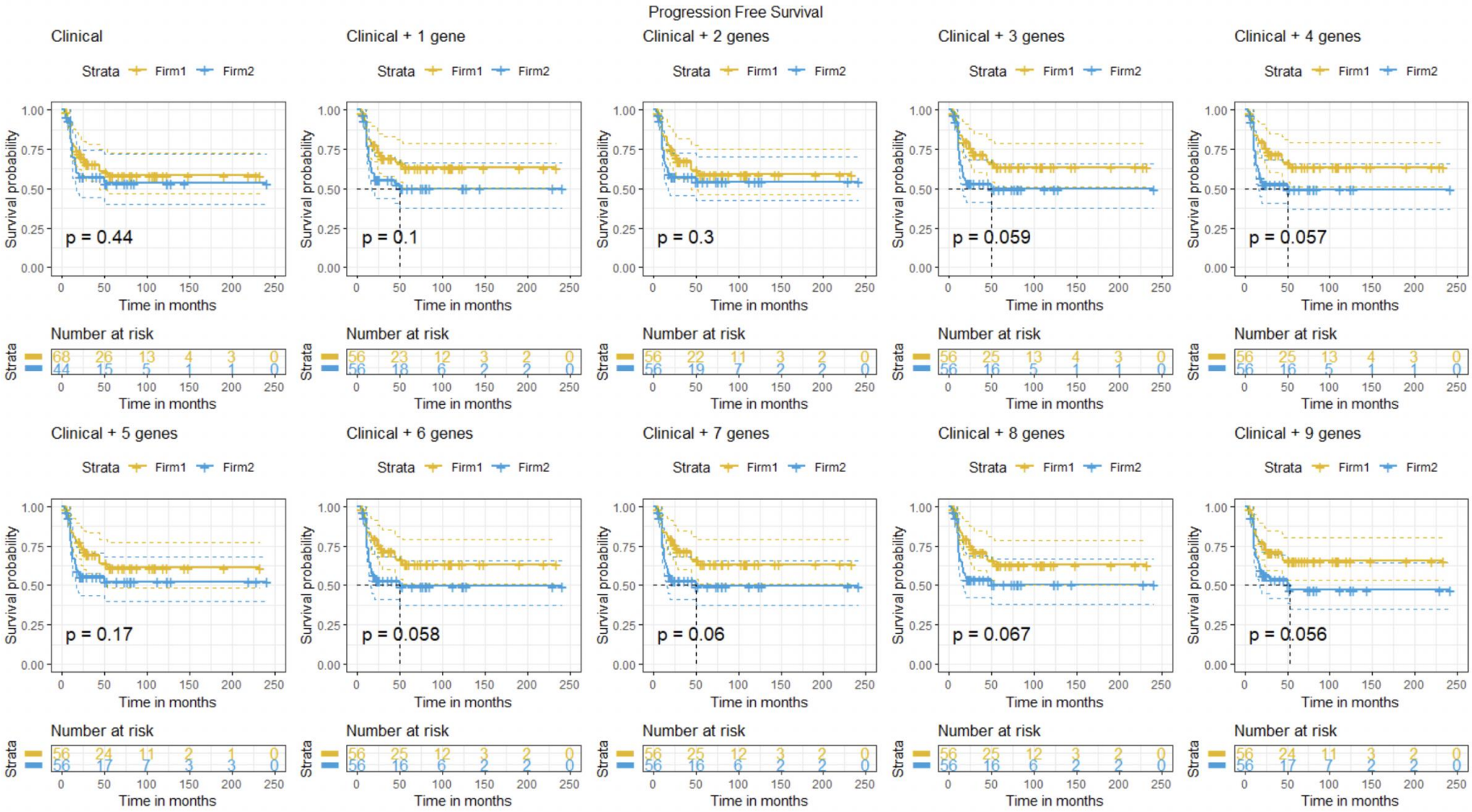


Figure 4. 6 Kaplan-Meier curves for Progression Free survival (PFS) comparison of patients with favorable (in yellow) and unfavorable (in blue) genetic signature for the models.

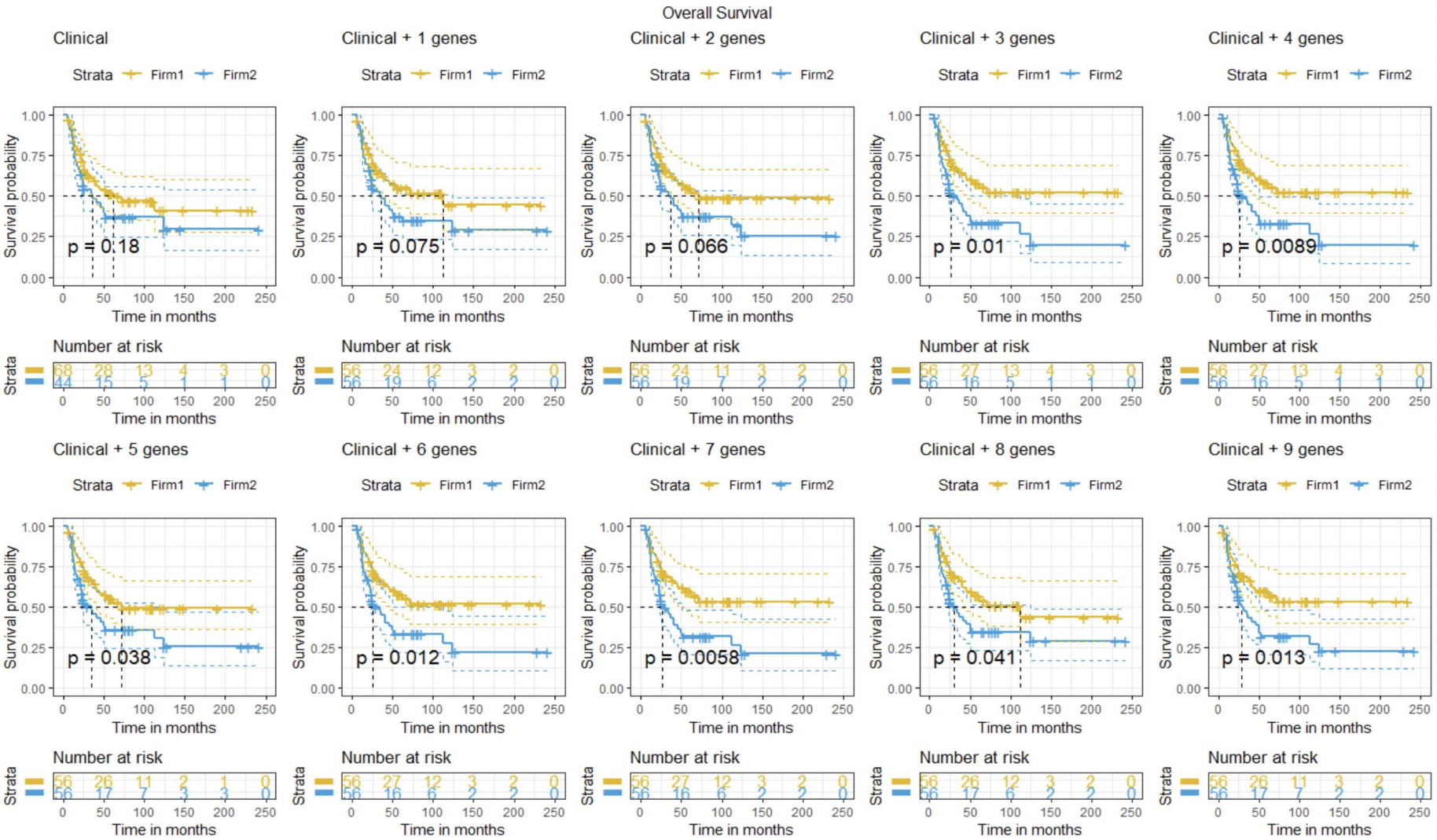


Figure 4. 7 Kaplan-Meier curves for Overall survival (OS) comparison of patients with favorable (in yellow) and unfavorable (in blue) genetic signature for the models.

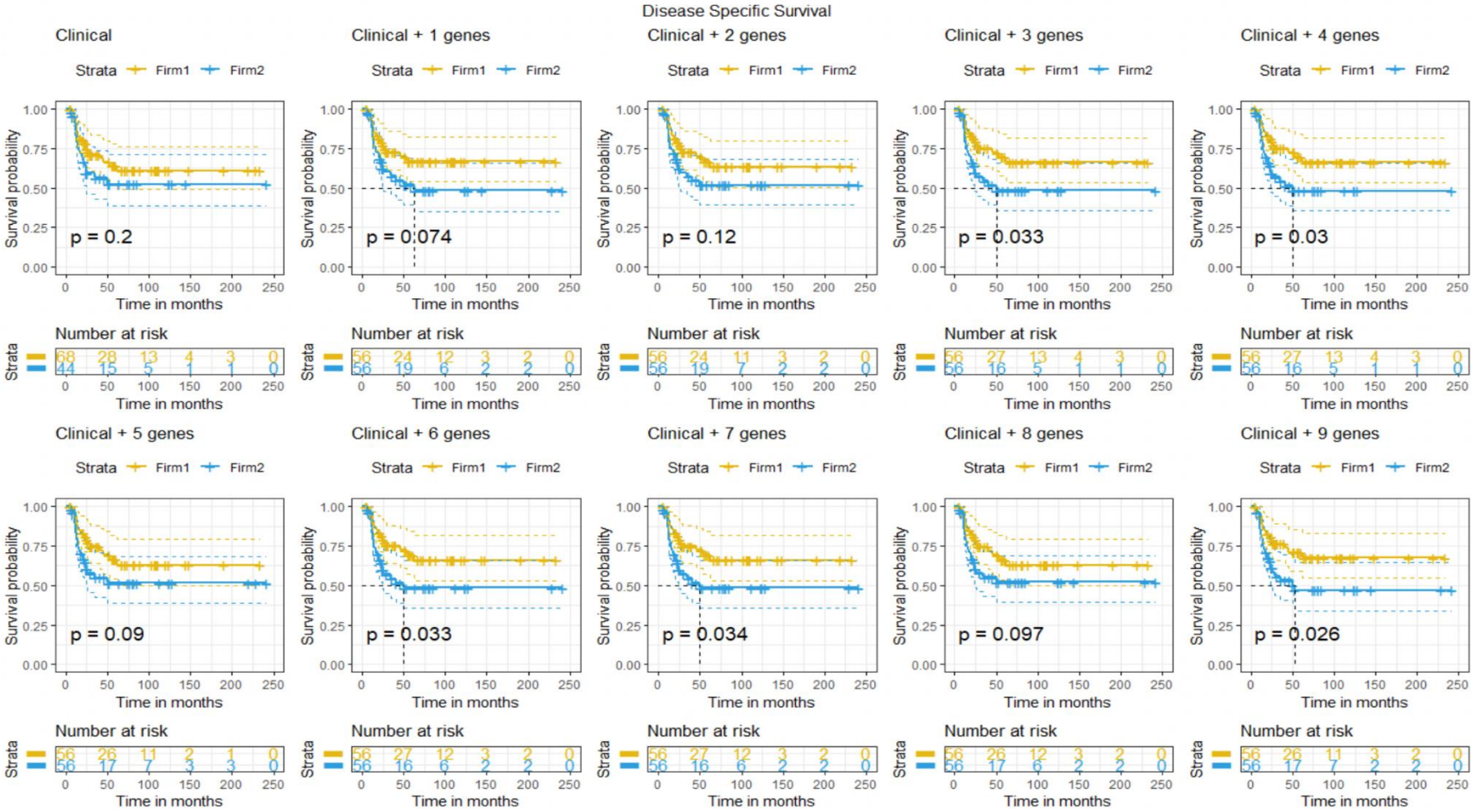


Figure 4. 8 Kaplan-Meier curves for Disease Specific survival (DSS) comparison of patients with favorable (in yellow) and unfavorable (in blue) genetic signature for the models.

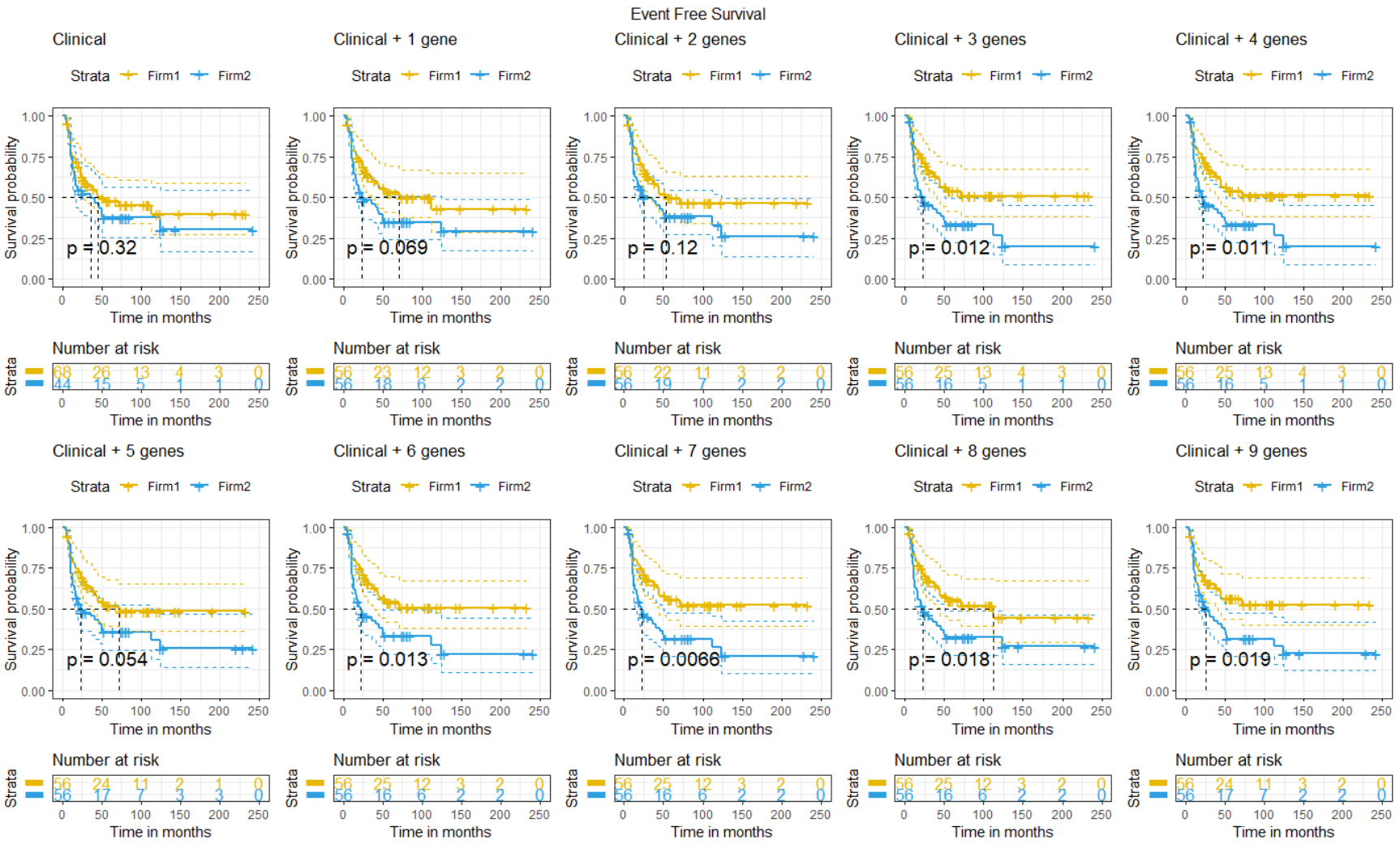


Figure 4. 9 Kaplan-Meier curves for Event Free survival (EFS) comparison of patients with favorable (in yellow) and unfavorable (in blue) genetic signature for the models.

4.4. DISCUSSION

4.4.1. Pipeline and genetic signatures

Here, we have developed a statistical pipeline to construct gene expression signatures integrating genomic and non-genomic data based on a conditional integrative strategy (Figure 4.3). The pipeline is structured in 5 steps: preprocessing, clinical model construction, genetic signature construction, estimation of the increasing AUC, and survival analysis. The preprocessing step helps to improve the quality of the data and to reduce noise before the analysis. Following the conditional modeling strategy, the non-genomic variables are selected from the primary estimated model (clinical model) using stepwise regression. The stepwise regression performs model selection, successively applying regression analysis with the addition and removal of the variables in each step. The set of variables with higher goodness of fit constitute the final non-genomic model.

Once the non-genomic variables are selected, the genomic data is aggregated to construct the final model (the genomic-and-non-genomic model). The genomic variables are selected adjusted by the non-genomic data. To join these two different types of data, the non-genomic variables (mostly qualitative date) are transformed into a numerical score (quantitative) by using the linear predictors of the non-genomic model. This score is attached to the genomic variables as one more variable and represents the set of non-genomic variables. The genetic signature is estimated by using LASSO in a cross-validation process, to avoid overfitting. The penalized regression Elastic-net was also tested, but the model provided lower capacity of prediction than LASSO.

To analyze more in detail the genetic information given by each gene of the signature, an increasing predictive ability analysis was performed. The objective of this step was to build different models increasing the number of genes and estimate the AUC for each model. The clinical or non-genomic model was analyzed. Then, the genes selected by LASSO were incorporated one by one, and the predictive ability was estimated for each model to test the variability of the AUC with each added gene.

Finally, survival analysis was performed to verify if the differential expression in the genetic signature was reflected in the survival. This analysis was also performed for different models including the selected genes one by one. Besides the increasing capacity, we obtained the effect on the survival for each gene. Therefore, the final genetic signature can be optimized according to the results from the increased capacity and the survival analyses.

The main objective in genetic or gene expression signatures for predictive purpose, is to obtain the lowest number of genes with higher predictive ability. The pipeline exposed here, helps shorten the number of genes without jeopardizing the predictive ability. One of the utilities of genetic signatures is the development of molecular tests that can profile the expression of selected genes to perform molecular diagnosis, prediction of the response to a specific treatment, or the prognosis of a determined disease (Chen 2013). The lower the number of the genetic signature, the simpler is the molecular test.

Moreover, gene expression signatures are useful for the identification of biomarkers that guide in pharmacogenomics to drug discovery (Bai et al. 2013). The genetic signature can be converted into possible targets for the development of new treatments. The gene expression signatures may be also useful in later stages of the drug development process, as in the identification of pharmacodynamic markers to evaluate the response to drug exposure (Bai et al. 2013)]. The expression of these

biomarkers is commonly used in clinical trials to determine the optimal dose and the sensitivity of the drug. The gene expression analysis can reveal the action of the drug on target pathways (Mizuarai et al. 2009) and in toxicological pathways (Liebler and Guengerich 2005). The pipeline exposed here can help to prioritize the possible targets for drug discovery and depict the drug action at molecular level. Altogether, these results support the utility of gene expression as a tool in personalized medicine (Kamel and Al-Amodi 2017).

4.4.2. MIBC project

The statistical pipeline was tested with data from MIBC, whose aim was to identify a clinical-molecular signature that predict the efficacy of neoadjuvant chemotherapeutic treatment. The signature selected was compounded of three non-genomic variables (morphology, presence of hydronephrosis and taxonomic classification or cluster) and nine genes (*RAD51, CXCL9, PARP, 53BP1, HERC2, ERCC2, CHEK1, Ku80* and *RNF168*), resulting in an AUC of 0.663. The integrative model including genomic and non-genomic covariates resulted in better prognostic performance than the non-genomic model alone (AUC= 0.52 or 0.58 with the taxonomic variable).

Through the increasing predictive ability analysis, we can observe that with the two first selected genes (*RAD51* and *CXCL9*), the AUC get the higher value (AUC=0.671), also obtained including the next two genes (*PARP* and *53BP1*). Our results suggest that the overexpression of *RAD51*, *CXCL9* and *PARP*, and the lower expression of 53BP1 were associated with better response to the treatment. Several studies have been done about these biomarkers that support these results (Zhang et al. 2006). For example, the overexpression of RAD51 has been previously associated with increased resistance to platinum in several tumors including bladder cancer (Mullane et al. 2016).

Survival analysis of PFS, OS, DSS and EFS was performed between samples with favorable and unfavorable genetic signature to treatment response, over the increasing gene models. Results showed a higher distinction in the survival of the two groups of samples, as more genes were added in the genetic signature. MIBC patients with favorable gene expression profile for the nine genes selected by LASSO have higher survival than patients with unfavorable gene expression signature (PFS $p = 0.059$, OS $p = 0.013$, DSS $p = 0.026$, EFS $p = 0.019$). However, the signature including the clinical/pathological/taxonomical variables and three first genes selected by LASSO (*RAD51, CXCL9* and *PARP*) reached the statistical significance for OS, DSS and EFS ($p < 0.05$), and was close to the significance for PFS ($p = 0.059$).

Thanks to the developed statistical pipeline, we can conclude that our data suggest that the best signature to predict the response to cisplatin treatment is composed of three non-genomic variables (morphology, presence of hydronephrosis and taxonomic classification or cluster) and 9 genes (*RAD51, CXCL9, PARP, 53BP1, HERC2, ERCC2, CHEK1, Ku80* and *RNF168*). However, the signature including the non-genomic variables and three or four genes (*RAD51, CXCL9, PARP, 53BP1* and *HERC2)* might be powerful enough to predict the treatment response. Therefore, the signature should be validated in another cohort of patients with MIBC. Further investigation are required to achieve the final genomic and non-genomic signature.

4.4.3. Conditional integrative strategy

López de Maturana and colleges describe the challenges of the integration of omics and non-omics data sets and consider the conditional integrative approach as one of the most suitable in this context (López de Maturana et al. 2019; Bazzoli and Lambert-Lacroix 2018; Bøvelstad, Nygård, and Borgan 2009). In contrast with our study, Bazzoli research resulted in better prediction performance for only

gene expression dataset than the model that include genomics and non-genomics data (Bazzoli and Lambert-Lacroix 2018). However, the second omics dataset studied in that work (somatic copy number alterations) obtained higher prediction performance when was tested together with the non-genomic data. Bovelstad et al (Bøvelstad, Nygård, and Borgan 2009) studied three gene expression datasets of different tumors, and two of them obtained the higher prediction performance with the integrative model. In general, several studies confirm that models that integrate genomic and non-genomic data have better prognostic performance than one of them alone (Jayawardana et al. 2015; Thompson, Christensen, and Marsit 2018; van Vliet et al. 2012; Zhu et al. 2017), independently of the integrative model approach. However, some studies have reported no improvement in predictive ability when genomic data and non-genomic data is integrated (Chaudhary et al. 2018; de Maturana et al. 2014).

A complex trait is as a trait that does not follow Mendelian inheritance laws and is likely the consequence of the effect of multiple genes. However, the molecular part is not the only factor to be considered when studying these traits, since other non-biological factors that are probably influencing the phenotype (i.e. environmental). The integration of omics and non-omics data can guide to find new risk factors in these traits, design better predictive models, or discriminate patients with favorable response to treatment. The pipeline developed here can help to all these purposes and therefore, boosting biomedical research towards personalized medicine.

FINAL DISCUSSION

General discussion of all projects

This industrial doctorate has accomplished the four objectives of the company Whole Genix: (1) acquire better know-how of Next Generation Sequencing (NGS) protocols; (2) cover different types of cancer and expand to other genetic diseases; (3) develop new bioinformatic and telemedicine tools for analysis of different omics data; and (4) progress towards personalized medicine. This has been possible thanks to the collaboration with academia (Josep Carreras Research Institute and the University of Vic - Central University of Catalonia) that provided the theoretical and practical guidance in hematologic and bioinformatic research, respectively.

1. Acquire better expertise in Next Generation Sequencing (NGS).

The design and development of the HHD panel described in chapter 1 has increased the knowledge about the NGS technology, not only from the bioinformatics point of view, but also from others such as the design of targeting NGS panels, or optimization of wet lab protocols. The study of four design improvements of the HHD panel, has helped to understand how faster the genomic knowledge is increasing in short period of time for all genetic fields, including rare diseases and cancer (Fernandez-Marmiesse, Gouveia, and Couce 2018; Sabour, Sabour, and Ghorbian 2017). In the last years, the number of identified novel genetic alteration associated with diseases has grown exponentially due to the emergence of NGS-based studies. In four years (from 2015 to 2019), the number of genes identified to be associated to HHD has increased from 58 genes (panel version 13) to 107 (panel version 16), only considering the same groups of diseases and not the new diseases. From panel version 15 to version 16, the number of genes increased from 153 to 203. This number will be continually increasing since more genetic cases are continuously diagnosticated and studied. Similar to these rare diseases, the number of genes associated with oncogenesis, progression and metastasis of cancer have increased since the emerging of NGS technology. This fact forces to the genetic diagnostic laboratories to continuously search and be updated for new causative genes of these disorders.

Furthermore, WG has gain experience in designing targeting NGS panels. Targeted sequencing may be more appropriate when there is previous knowledge of the associated-disease genes, since it increases the sensibility to detect variants due to the deep coverage augmentation, while reducing significatively the time and cost (Shefer Averbuch et al. 2018; Suwinski et al. 2019). The development of custom pipeline to extract the coordinates of all isoforms of our selected genes, increases the chance of success to detect mutations that are only in particular isoforms. Also, the use of targeting sequencing reduces considerably the number of incidental findings (Buermans and den Dunnen 2014). However, it is a reality that the WES is gaining more space in the genetic labs due to the drop in price and the increase in significant findings. The price of targeted sequencing gene panel per samples is on average €1455 (US$1609, range: US$488-US$3443), and although the average price for WES per samples is still considerable higher (€5534), the minimum has fallen to €432 (£382) (Schwarze et al. 2018). The cost-effectiveness is higher in targeted sequencing panel and WES compared to WGS. Additionally to the exposed in this research, this industrial Ph.D has contributing in enabling WES and WGS services in WG company for clinical and research purposes.

Whole Genix was, and still is, a company specialized in the bioinformatic analysis of NGS data both in WES, WGS and targeted sequencing. The immersion of WG in the industrial PhD with the collaboration of BloodGenetics company and the development of the HHD panel, has brought the company new

awareness about NGS technology. In order to speed up the analysis process of HHD with the NGS gene panel, our own database which variants associated with HHD has been created. Those variants have been identified from previously reported publications, databases for human variations (i.e. ClinVar and HGMD), our own results, and from collaborations with other hematological laboratories. The development of this database helps to identify reported pathogenic variants, filter variants that seem to have more significant pathogenic effect, and prioritize the functional analysis for those VUS variants that resemble to other pathogenic variant in position or function.

Besides, it is a reality that the third-generation sequencing (TGS) is approaching for clinical diagnosis (Au et al. 2019). As it was commented in chapter 1, the TGS technology will solve some of the difficulties that the NGS is facing; as 1) the detection of structural variants, 2) the proper sequencing of homologous genes and 3) the sequencing of problematic regions (i.e. high GC content) (van Dijk et al. 2018), that are implicated in both in Mendelian and somatic-oncological diseases (Weischenfeldt et al. 2013).

2. Cover different types of cancer and expand to other genetic diseases.

WG is specialized in neoplastic diseases. However, the company has obtained greater specialization in bladder cancer (chapter 4) and myelodysplastic syndromes (chapter 3) due to the collaboration in two external research projects: the DJCLS R 14/04 Myelodysplastic syndromes project, which Principal Investigators are Dr. Mayka Sanchez (Spain) and Dr. Nobert Gattermann (Germany); and the bladder cancer project (SOG-CVI-2014-05) which principal investigator is Francisco X. Real Arribas. In the Sanchez & Gatterman´s project, a gene expression signature of 23 genes was identified to be differentially expressed comparing MDS patients with and without ring sideroblasts development, and the healthy controls. These genes contribute to a better understanding of the molecular pathways and the pathogenesis of MDS and the classification in disease subtypes. In the collaboration with Real´s project, a genetic signature to predict the response to cisplatin treatment in patients with MIBC was studied. The genetic signature resulted to be in 9 genes and 3 non-genomic (clinical/pathological/taxonomic) variables, that might predict the response to cisplatin with a predictive ability of 66.3% (AUC=0.663). Although further analysis should be performed to corroborate the results, the gene expression signatures studied helped to better understand the molecular pathways behind the MDS and MIBC disorders.

WG has spread to other non-oncological diseases: the hereditary hematological diseases (HDD). The collaboration with BloodGenetics company and the development of the HHD targeted panel, have introduced a new spectrum of diseases, no related to neoplastic disorders. This places WG as a versatile and flexible company that provides services to analyze genomic and transcriptomic data from oncological and non-oncological diseases.

3. Develop new bioinformatic and telemedicine tools for analysis of different omics data.

WG has been introduced and expanded in other high-throughput technology from different omics data, the transcriptomics. The development of BEA program (chapter 3) and the statistical pipeline to identify gene expression signatures (chapter 4), have made that WG gets new techniques of analysis of high-throughput transcriptomics data from BioMark-Fluidigm and Nanostring technologies. As was mentioned in chapter 4, the analysis of transcriptomic data can be used for the construction of gene expression signatures. These signatures are mostly studied in oncology to study the diagnosis, the prognosis and the drug response of cancer (Itadani, Mizuarai, and Kotani 2008; Chibon 2013).

Addittionally, the machine learning techniques developed here have been also employed in other research projects of WG, such as the study of epilepsy.

On the other hand, CoDysAn is a new telemedicine tool for the management and diagnosis of patients with this congenital dyserythropoietic anemia. The web tool provides awareness of the disease, and this is relevant since some patients suffering from CDA are under-diagnosed. The telemedicine tool includes a diagnosis algorithm to facilitate the classification and diagnosis of CDA types. (Tornador et al. 2019). Computational science is the base of this biotech company, and the development of new computational tools and bioinformatics pipelines are crucial for the sustainability of WG.

4. Progress towards personalized medicine.

One of the definitions of personalized medicine is the set of factors (mainly genomics), that can influence in the effectiveness or the adverse effect observed during drug treatment. The presence of the DNA variants and the differential expression of determined genes can be related to drug response (pharmacogenomics). All the four projects undertaken here contribute in one way or another in the advancement to the future personalized medicine. Congenital anemias represent a heterogeneous group of anemias in which the identification of genetic variants is essential for suitable treatment and patient care (Iolascon, Andolfo, and Russo 2015). The emergence of NGS technologies helps to improve timely diagnosis of anemias, understand the RCB function in disease, and applicate the personalized avoiding possible harmful treatments. The development of the HHD gene panel, supports one of the key priorities of the European Anemia Research Roadmap: “use of new technology for a personalized diagnosis and therapy” (Engert et al. 2016). The correct genetic diagnosis and the continuous discovery of genes associated with HHD increase the possibilities of development of gene therapy. Preclinical gene therapy studies for some HHD, such as hemolytic anemia due to Pyruvate Kinase Deficiency (Garcia-Gomez et al. 2016), have been successfully described.

The CoDysAn web tool is also closely linked with personalized medicine (Tornador et al. 2019). The diagnostic algorithm not only presents the clinical suspicion of CDA subtype, but also provide a link to connect to the NCBI Genetic Testing Registry (GTR) to perform a complete genetic test, if required. The importance of a correct and early genetic diagnosis for the subsequent implementation of proper treatment, avoiding detrimental or inappropriate treatments is crucial in CDA diseases.

As was described in chapter 4, the objective of transcriptomics analysis is to identify a gene expression profile that is associated to a pathological condition in terms of pathogenesis, course of the disease or response to a treatment. The creation of BEA program and the statistical pipeline to construct gene expression signatures with integrative models encourage the identification of those gene expression profiles, that can better diagnose a particular disease, improve classification of disorder subtypes that can course with worse prognosis, or predict the response to a specific treatment.

WG wants to endorse genetics and genomics in research and integrate this knowledge into clinical medicine to help accomplish the promise of personalized medicine.

CONCLUSIONS

CONCLUSIONS

CONCLUSION CHAPTER 1: Hereditary Hematologic Diseases Next Generation Sequencing Panel

- The implementation of the NGS methodology in clinical practice for the diagnosis of hereditary hematological diseases allows the inclusion of the study of multiple genes and a rapid and effective diagnosis of these cases.

- The use of the new targeted-NGS panels have allowed to reach a final diagnosis in 42.9% of analyzed cases.

- The use of the panel has allowed to identify clinically positive but genetic negative cases that were subsequently studied by WES (analysis on progress) that may allow for the future identification of new genes involved in HHD.

- As knowledgment is increasing in the field, the HHD gene panel must be frequently updated due to the continuous discovery of new genes associated with HHD.

- L-ferritin disease is a clear case of Mendelian disease-related genes that are associated with multiple diseases, including Hereditary Hyperferritinemia Cataract Syndrome (HHCS), benign hyperferritinemia, neuroferritinopathy (NBIA3), and L-ferritin deficiency with autosomal recessive and autosomal dominant inheritance. We have reported three variants: two novel mutations in L-ferritin deficiency with autosomal dominant inheritance and HHCS diseases (Publication: Cadenas et al. page 193), and the second case reported for L-ferritin deficiency due to p.Met1Val (Cremonesi et al. 2004).

CONCLUSION CHAPTER 2: CoDysAn: A telemedicine tool to improve awareness and diagnosis for patients with Congenital Dyserythropoietic Anemia (Publication: Tornador et al. 2019)

- CoDysAn is a freely accessible web site that provides awareness of congenital dyserythropoietic anemia (CDA) where general public, patients and medical doctors can better understand and learn more about this disease, and use the diagnosis algorithm tool to ease the classification and diagnostic of CDA types.

- The diagnostic algorithm of CoDysAn has been validated, resulting in a high specificity (89.5%) and sensitivity (78.5%).

- CoDysAn algorithm is connected to the NCBI Genetic Testing Registry (GTR) in a way to inform medical doctors about the existence of these accredited diagnostic centers to perform a complete genetic test, if required.

- CoDysAn includes a questionnaire to evaluate user opinion and degree of satisfaction with the website. This information will lead us to future improvements and updates for the website.

CONCLUSION CHAPTER 3: BEA (BioMark Expression Analysis), a web-based tool for gene expression Analysis from BioMark-Fluidigm Dynamic Arrays.

- BEA is a ready-to-use desktop program based in R/Shiny language that is easy to install and deploy, whose purpose is to provide an interactive statistic tools for the analysis of Biomark-Fluidigm data in an intuitive step-by-step process.

- BEA is oriented to users with limited knowledge in bioinformatics and advanced statistics, and allows the user to benefit from the power of R/Bioconductor packages without having programming experience.

- BEA has been designed to analyze different IFC array formats, multiple conditions and replicates. Analysis include relative quantification using ΔΔCT normalization and appropriate statistical testing to identify differential expressed genes between tested groups, considering test of normality and number of groups.

- BEA program helps the user to select suitable reference gene for reliable normalization and a qPCR data analysis, by providing three essential parameters: mean Ct value, deviation of the data, and gene-stability measure value.

- The analysis pipeline, the results and plots generated in each step of the analysis are incorporated in a final report, that can be exported in a PDF file format. Other plots and tables including the complete heatmap, the normalized data table and results table, can be exported in a PDF and CSV format.

- The use of BEA in a setting of biological sample of patients with MDS (with and without RS) and control samples has detected 23 genes differentially expressed (in a set of 96 genes selected as potential interactors with the iron regulatory proteins or IRPs). Further studies are needed to decipher the biological mean of these changes in expression profiles and their potential use as biomarkers in MDS.

CONCLUSION CHAPTER 4: Pipeline for construction of gene expression signatures for prediction and prognosis

- The statistical pipeline developed here, helps to build gene expression signatures for prediction and prognosis purposes by integrating genomic and non-genomic data.

- The pipeline is based on an integrative conditional modeling approach, in which the non-genomic model is built first using stepwise regression, and a second model is estimated by including the genomic variables to the first model and performing variable selection by LASSO in a cross validation procedure.

- The pipeline estimates the best genomic and non-genomic signature for the tested dataset, but also provides the flexibility to value the reduction of the number of genetic variables in the signature without compromising the predictive ability of the model.

- The pipeline has been tested using data from MIBC patients. The predictive signature that includes the tumor morphology type, the presence of hydronephrosis and the taxonomic

classification variables (non-genomic covariates) and 9 genes (*RAD51, CXCL9, PARP, 53BP1, HERC2, ERCC2, CHEK1, Ku80* and *RNF168*) obtained the highest predictive ability to predict cisplatin response (AUC = 0.663). However, results suggest that the non-genomic variables and the first three or four genes selected by LASSO (*RAD51, CXCL9, PARP* and *53BP1*) may have enough predictive ability to predict response to the treatment (AUC =0.671) and higher survival (Log-rank test p.value < 0.05). Nevertheless, further analysis must be performed.

- The statistical pipeline can help to better understand complex traits and find new risk factors, design better predictive models, or discriminate patients with favorable response to treatment, and therefore, to move closer to the future personalized medicine.

BIBLIOGRAPHY

BIBLIOGRAPHY

A

Allen, Katrina J., Lyle C. Gurrin, Clare C. Constantine, Nicholas J. Osborne, Martin B. Delatycki, Amanda J. Nicoll, Christine E. McLaren, et al. 2009. "Iron-Overload–Related Disease in HFE Hereditary Hemochromatosis." Research-article. https://doi.org/10.1056/NEJMoa073286.

Allerson, C. R., M. Cazzola, and T. A. Rouault. 1999. "Clinical Severity and Thermodynamic Effects of Iron-Responsive Element Mutations in Hereditary Hyperferritinemia-Cataract Syndrome." *J Biol Chem* 274 (37): 26439–47.

Allikmets, Rando, Wendy H. Raskind, Amy Hutchinson, Nichole D. Schueck, Michael Dean, and David M. Koeller. 1999. "Mutation of a Putative Mitochondrial Iron Transporter Gene (ABC7) in X-Linked Sideroblastic Anemia and Ataxia (XLSA/A)." *Human Molecular Genetics* 8 (5): 743–49. https://doi.org/10.1093/hmg/8.5.743.

Alloisio, N., P. Texier, L. Denoroy, C. Berger, E. Miraglia del Giudice, S. Perrotta, A. Iolascon, F. Gilsanz, G. Berger, and J. Guichard. 1996. "The Cisternae Decorating the Red Blood Cell Membrane in Congenital Dyserythropoietic Anemia (Type II) Originate from the Endoplasmic Reticulum." *Blood* 87 (10): 4433–39.

Altes, A., M. J. Perez-Lucena, and M. Bruguera. 2014. "[Systematic approach to the diagnosis of hyperferritinemia]." *Med Clin (Barc)* 142 (9): 412–17. https://doi.org/10.1016/j.medcli.2013.06.010.

Alvarez-Coca-Gonzalez J, Moreno-Carralero MI, Martinez-Perez J, Mendez M, and Moran-Jimenez MJ. 2010. "The Hereditary Hyperferritinemia-Cataract Syndrome: A Family Study." *European Journal of Pediatrics* 169: 1553–55.

Andolfo, Immacolata, Roberta Russo, Francesco Manna, Boris E. Shmukler, Antonella Gambale, Giuseppina Vitiello, Gianluca De Rosa, et al. 2015. "Novel Gardos Channel Mutations Linked to Dehydrated Hereditary Stomatocytosis (Xerocytosis)." *American Journal of Hematology* 90 (10): 921–26. https://doi.org/10.1002/ajh.24117.

Andolfo, Immacolata, Roberta Russo, Barbara Eleni Rosato, Francesco Manna, Antonella Gambale, Carlo Brugnara, and Achille Iolascon. 2018. "Genotype-Phenotype Correlation and Risk Stratification in a Cohort of 123 Hereditary Stomatocytosis Patients." *American Journal of Hematology* 93 (12): 1509–17. https://doi.org/10.1002/ajh.25276.

Ang, Sonny O, Hua Chen, Victor R Gordeuk, Adelina I Sergueeva, Lydia A Polyakova, Galina Y Miasnikova, Robert Kralovics, David W Stockton, and Josef T Prchal. 2002. "Endemic Polycythemia in Russia: Mutation in the VHL Gene." *Blood Cells, Molecules, and Diseases* 28 (1): 57–62. https://doi.org/10.1006/bcmd.2002.0488.

Arber, Daniel A., Attilio Orazi, Robert Hasserjian, Jürgen Thiele, Michael J. Borowitz, Michelle M. Le Beau, Clara D. Bloomfield, Mario Cazzola, and James W. Vardiman. 2016. "The 2016 Revision to the World Health Organization Classification of Myeloid Neoplasms and Acute Leukemia." *Blood* 127 (20): 2391. https://doi.org/10.1182/blood-2016-03-643544.

Arnaud, Lionel, Carole Saison, Virginie Helias, Nicole Lucien, Dominique Steschenko, Marie-Catherine Giarratana, Claude Prehu, et al. 2010. "A Dominant Mutation in the Gene Encoding the Erythroid Transcription Factor KLF1 Causes a Congenital Dyserythropoietic Anemia." *American Journal of Human Genetics* 87 (5): 721–27. https://doi.org/10.1016/j.ajhg.2010.10.010.

Ask, Katrine, Zuzana Jasencakova, Patrice Menard, Yunpeng Feng, Geneviève Almouzni, and Anja Groth. 2012. "Codanin-1, Mutated in the Anaemic Disease CDAI, Regulates Asf1 Function in S-Phase Histone Supply." *The EMBO Journal* 31 (8): 2013–23. https://doi.org/10.1038/emboj.2012.55.

Athiyarath, Rekha, Neeraj Arora, Francisco Fuster, Robert Schwarzenbacher, Rayaz Ahmed, Biju George, Mammen Chandy, et al. 2013. "Two Novel Missense Mutations in Iron Transport Protein Transferrin Causing Hypochromic Microcytic Anaemia and Haemosiderosis: Molecular Characterization and Structural Implications." *British Journal of Haematology* 163 (3): 404–7. https://doi.org/10.1111/bjh.12487.

Au, Chun Hang, Dona N. Ho, Beca B. K. Ip, Thomas S. K. Wan, Margaret H. L. Ng, Edmond K. W. Chiu, Tsun Leung Chan, and Edmond S. K. Ma. 2019. “Rapid Detection of Chromosomal Translocation and Precise Breakpoint Characterization in Acute Myeloid Leukemia by Nanopore Long-Read Sequencing.” *Cancer Genetics* 239 (August): 22–25. https://doi.org/10.1016/j.cancergen.2019.08.005.

B

Babbs, Christian, Nigel A. Roberts, Luis Sanchez-Pulido, Simon J. McGowan, Momin R. Ahmed, Jill M. Brown, Mohamed A. Sabry, et al. 2013. “Homozygous Mutations in a Predicted Endonuclease Are a Novel Cause of Congenital Dyserythropoietic Anemia Type I.” *Haematologica* 98 (9): 1383–87. https://doi.org/10.3324/haematol.2013.089490.

Bagby, Grover. 2018. “Recent Advances in Understanding Hematopoiesis in Fanconi Anemia.” *F1000Research* 7 (January). https://doi.org/10.12688/f1000research.13213.1.

Bai, Jane P. F., Alexander V. Alekseyenko, Alexander Statnikov, I-Ming Wang, and Peggy H. Wong. 2013. “Strategic Applications of Gene Expression: From Drug Discovery/Development to Bedside.” *The AAPS Journal* 15 (2): 427–37. https://doi.org/10.1208/s12248-012-9447-1.

Balas, A., M. J. Aviles, F. Garcia-Sanchez, and J. L. Vicario. 1999. “Description of a New Mutation in the L-Ferrin Iron-Responsive Element Associated with Hereditary Hyperferritinemia-Cataract Syndrome in a Spanish Family.” *Blood* 93 (11): 4020–21.

Baraibar, Martin A., Ana G. Barbeito, Barry B. Muhoberac, and Ruben Vidal. 2008. “Iron-Mediated Aggregation and a Localized Structural Change Characterize Ferritin from a Mutant Light Chain Polypeptide That Causes Neurodegeneration.” *The Journal of Biological Chemistry* 283 (46): 31679–89. https://doi.org/10.1074/jbc.M805532200.

Baraibar, Martin A., Ana G. Barbeito, Barry B. Muhoberac, and Ruben Vidal.. 2012. “A Mutant Light Chain Ferritin That Causes Neurodegeneration Has Enhanced Propensity toward Oxidative Damage.” *Free Radical Biology & Medicine* 52 (9): 1692–97. https://doi.org/10.1016/j.freeradbiomed.2012.02.015.

Baraibar, Martin A., Barry B. Muhoberac, Holly J. Garringer, Thomas D. Hurley, and Ruben Vidal. 2010. “Unraveling of the E-Helices and Disruption of 4-Fold Pores Are Associated with Iron Mishandling in a Mutant Ferritin Causing Neurodegeneration.” *The Journal of Biological Chemistry* 285 (3): 1950–56. https://doi.org/10.1074/jbc.M109.042986.

Bazzoli, Caroline, and Sophie Lambert-Lacroix. 2018. “Classification Based on Extensions of LS-PLS Using Logistic Regression: Application to Clinical and Multiple Genomic Data.” *BMC Bioinformatics* 19 (1): 314. https://doi.org/10.1186/s12859-018-2311-2.

Beaumont, C., P. Leneuve, I. Devaux, J. Y. Scoazec, M. Berthier, M. N. Loiseau, B. Grandchamp, and D. Bonneau. 1995. “Mutation in the Iron Responsive Element of the L Ferritin MRNA in a Family with Dominant Hyperferritinaemia and Cataract.” *Nat Genet* 11 (4): 444–46. https://doi.org/10.1038/ng1295-444.

Bento, Celeste, Melanie J. Percy, Betty Gardie, Tabita Magalhães Maia, Richard van Wijk, Silverio Perrotta, Fulvio Della Ragione, et al. 2014. “Genetic Basis of Congenital Erythrocytosis: Mutation Update and Online Databases.” *Human Mutation* 35 (1): 15–26. https://doi.org/10.1002/humu.22448.

Bergmann, Anke K., Dean R. Campagna, Erin M. McLoughlin, Suneet Agarwal, Mark D. Fleming, Sylvia S. Bottomley, and Ellis J. Neufeld. 2009. “Systematic Molecular Genetic Analysis of Congenital Sideroblastic Anemia: Evidence for Genetic Heterogeneity and Identification of Novel Mutations.” *Pediatric Blood & Cancer*, n/a-n/a. https://doi.org/10.1002/pbc.22244.

Berhe, Simon, Matthew M. Heeney, Dean R. Campagna, John F. Thompson, Eric J. White, Tristen Ross, Roy W. A. Peake, et al. 2018. “Recurrent Heteroplasmy for the MT-ATP6 p.Ser148Asn (m.8969G>A) Mutation in Patients with Syndromic Congenital Sideroblastic Anemia of Variable Clinical Severity.” *Haematologica* 103 (12): e561–63. https://doi.org/10.3324/haematol.2018.199109.

Beshlawi, Ismail, Shoaib Al Zadjali, Wafa Bashir, Mohamed Elshinawy, Abdulhakim Alrawas, and Yasser Wali. 2014. “Thiamine Responsive Megaloblastic Anemia: The Puzzling Phenotype.” *Pediatric Blood & Cancer* 61 (3): 528–31. https://doi.org/10.1002/pbc.24849.

Beutler, Ernest. 2008. “Glucose-6-Phosphate Dehydrogenase Deficiency: A Historical Perspective.” *Blood* 111 (1): 16–24. https://doi.org/10.1182/blood-2007-04-077412.

Bhuva, Meha, Sambit Sen, Terence Elsey, Wale Atoyebi, Helene Dreau, Charlotte Bradbury, Rosalynd Johnston, Patricia Bignell, and William Griffiths. 2018. “Sequence Analysis of Exon 1 of the Ferritin Light Chain (FTL) Gene Can Reveal the Rare Disorder ‘Hereditary Hyperferritinaemia without Cataracts.’” *British Journal of Haematology*, May. https://doi.org/10.1111/bjh.15231.

Björkstén, B., G. Holmgren, G. Roos, and R. Stenling. 1978. “Congenital Dyserythropoietic Anaemia Type III: An Electron Microscopic Study.” *British Journal of Haematology* 38 (1): 37–42. https://doi.org/10.1111/j.1365-2141.1978.tb07106.x.

Bolton-Maggs, Paula H. B., Jacob C. Langer, Achille Iolascon, Paul Tittensor, May-Jean King, and General Haematology Task Force of the British Committee for Standards in Haematology. 2012. “Guidelines for the Diagnosis and Management of Hereditary Spherocytosis--2011 Update.” *British Journal of Haematology* 156 (1): 37–49. https://doi.org/10.1111/j.1365-2141.2011.08921.x.

Bosio, Sandra, Alessandro Campanella, Enrico Gramaglia, Paolo Porporato, Filomena Longo, Laura Cremonesi, Sonia Levi, and Clara Camaschella. 2004. “C29G in the Iron-Responsive Element of L-Ferritin: A New Mutation Associated with Hyperferritinemia-Cataract.” *Blood Cells, Molecules & Diseases* 33 (1): 31–34. https://doi.org/10.1016/j.bcmd.2004.04.010.

Bottomley, Sylvia S. 2006. “Congenital Sideroblastic Anemias.” *Current Hematology Reports* 5 (1): 41–49.

Bottomley, Sylvia S., and Mark D. Fleming. 2014. “Sideroblastic Anemia: Diagnosis and Management.” *Hematology/Oncology Clinics of North America* 28 (4): 653–70, v. https://doi.org/10.1016/j.hoc.2014.04.008.

Bøvelstad, Hege M, Ståle Nygård, and Ørnulf Borgan. 2009. “Survival Prediction from Clinico-Genomic Models - a Comparative Study.” *BMC Bioinformatics* 10 (December): 413. https://doi.org/10.1186/1471-2105-10-413.

Brancaleoni, V., M. Balwani, F. Granata, G. Graziadei, P. Missineo, V. Fiorentino, S. Fustinoni, et al. 2016. “X-Chromosomal Inactivation Directly Influences the Phenotypic Manifestation of X-Linked Protoporphyria.” *Clinical Genetics* 89 (1): 20–26. https://doi.org/10.1111/cge.12562.

Bray, Freddie, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, and Ahmedin Jemal. 2018. “Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.” *CA: A Cancer Journal for Clinicians* 68 (6): 394–424. https://doi.org/10.3322/caac.21492.

Brown, Jessica S., Brent O’Carrigan, Stephen P. Jackson, and Timothy A. Yap. 2017. “Targeting DNA Repair in Cancer: Beyond PARP Inhibitors.” *Cancer Discovery* 7 (1): 20–37. https://doi.org/10.1158/2159-8290.CD-16-0860.

Buermans, H.P.J., and J.T. den Dunnen. 2014. “Next Generation Sequencing Technology: Advances and Applications.” *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1842 (10): 1932–41. https://doi.org/10.1016/j.bbadis.2014.06.015.

Bumgarner, Roger. 2013. “DNA Microarrays: Types, Applications and Their Future.” *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]* 0 22 (January): Unit-22.1. https://doi.org/10.1002/0471142727.mb2201s101.

Burdon, K. P., S. Sharma, C. S. Chen, D. P. Dimasi, D. A. Mackey, and J. E. Craig. 2007. “A Novel Deletion in the FTL Gene Causes Hereditary Hyperferritinemia Cataract Syndrome (HHCS) by Alteration of the Transcription Start Site.” *Hum Mutat* 28 (7): 742. https://doi.org/10.1002/humu.9501.

Burrage, Lindsay C., Sha Tang, Jing Wang, Taraka R. Donti, Magdalena Walkiewicz, J. Michael Luchak, Li-Chieh Chen, et al. 2014. “Mitochondrial Myopathy, Lactic Acidosis, and Sideroblastic Anemia (MLASA) plus Associated with a Novel de Novo Mutation (m.8969G>A) in the Mitochondrial Encoded ATP6 Gene.” *Molecular Genetics and Metabolism* 113 (3): 207–12. https://doi.org/10.1016/j.ymgme.2014.06.004.

Buttigliero, Consuelo, Marcello Tucci, Francesca Vignani, Giorgio V. Scagliotti, and Massimo Di Maio. 2017. “Molecular Biomarkers to Predict Response to Neoadjuvant Chemotherapy for Bladder Cancer.” *Cancer Treatment Reviews* 54 (March): 1–9. https://doi.org/10.1016/j.ctrv.2017.01.002.

C

Cadenas, Beatriz, Josep Fita-Torró, Mar Bermúdez-Cortés, Inés Hernandez-Rodriguez, José Luis Fuster, María Esther Llinares, Ana María Galera, et al. 2019. "L-Ferritin: One Gene, Five Diseases; from Hereditary Hyperferritinemia to Hypoferritinemia—Report of New Cases." *Pharmaceuticals* 12 (1). https://doi.org/10.3390/ph12010017.

Camaschella, C., A. Campanella, L. De Falco, L. Boschetto, R. Merlini, L. Silvestri, S. Levi, and A. Iolascon. 2007. "The Human Counterpart of Zebrafish Shiraz Shows Sideroblastic-like Microcytic Anemia and Iron Overload." *Blood* 110 (4): 1353–58. https://doi.org/10.1182/blood-2007-02-072520.

Camaschella, C., A. Roetto, A. Calì, M. De Gobbi, G. Garozzo, M. Carella, N. Majorano, A. Totaro, and P. Gasparini. 2000. "The Gene TFR2 Is Mutated in a New Type of Haemochromatosis Mapping to 7q22." *Nature Genetics* 25 (1): 14–15. https://doi.org/10.1038/75534.

Camaschella, C., G. Zecchina, G. Lockitch, A. Roetto, A. Campanella, P. Arosio, and S. Levi. 2000. "A New Mutation (G51C) in the Iron-Responsive Element (IRE) of L-Ferritin Associated with Hyperferritinaemia-Cataract Syndrome Decreases the Binding Affinity of the Mutated IRE for Iron-Regulatory Proteins." *British Journal of Haematology* 108 (3): 480–82.

Campillos, Monica, Ildefonso Cases, Matthias W. Hentze, and Mayka Sanchez. 2010. "SIREs: Searching for Iron-Responsive Elements." *Nucleic Acids Research* 38 (Web Server issue): W360–67. https://doi.org/10.1093/nar/gkq371.

Castiglioni, E., N. Soriani, D. Girelli, C. Camaschella, I. Spiga, M. G. Della Porta, M. Ferrari, and L. Cremonesi. 2010. "High Resolution Melting for the Identification of Mutations in the Iron Responsive Element of the Ferritin Light Chain Gene." *Clin Chem Lab Med* 48 (10): 1415–18. https://doi.org/10.1515/CCLM.2010.281.

Cazzola, M., G. Bergamaschi, L. Tonon, E. Arbustini, M. Grasso, E. Vercesi, G. Barosi, P. E. Bianchi, G. Cairo, and P. Arosio. 1997. "Hereditary Hyperferritinemia-Cataract Syndrome: Relationship between Phenotypes and Specific Mutations in the Iron-Responsive Element of Ferritin Light-Chain MRNA." *Blood* 90 (2): 814–21.

Cazzola, M., B. Foglieni, G. Bergamaschi, S. Levi, M. Lazzarino, and P. Arosio. 2002. "A Novel Deletion of the L-Ferritin Iron-Responsive Element Responsible for Severe Hereditary Hyperferritinaemia-Cataract Syndrome." *Br J Haematol* 116 (3): 667–70.

Chakraborty, P. K., K. Schmitz-Abe, E. K. Kennedy, H. Mamady, T. Naas, D. Durie, D. R. Campagna, et al. 2014. "Mutations in TRNT1 Cause Congenital Sideroblastic Anemia with Immunodeficiency, Fevers, and Developmental Delay (SIFD)." *Blood* 124 (18): 2867–71. https://doi.org/10.1182/blood-2014-08-591370.

Chapelle, A. de la, P. Sistonen, H. Lehväslaiho, E. Ikkala, and E. Juvonen. 1993. "Familial Erythrocytosis Genetically Linked to Erythropoietin Receptor Gene." *The Lancet* 341 (8837): 82–84. https://doi.org/10.1016/0140-6736(93)92558-B.

Charache, S., D. J. Weatherall, and J. B. Clegg. 1966. "Polycythemia Associated with a Hemoglobinopathy." *The Journal of Clinical Investigation* 45 (6): 813–22. https://doi.org/10.1172/JCI105397.

Chaudhary, Kumardeep, Olivier B. Poirion, Liangqun Lu, and Lana X. Garmire. 2018. "Deep Learning Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer." *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research* 24 (6): 1248–59. https://doi.org/10.1158/1078-0432.CCR-17-0853.

Chen, Neng. 2013. "Incorporate Gene Signature Profiling into Routine Molecular Testing." *Applied & Translational Genomics* 2 (April): 28–33. https://doi.org/10.1016/j.atg.2013.03.002.

Chibon, Frederic. 2013. "Cancer Gene Expression Signatures – The Rise and Fall?" *European Journal of Cancer* 49 (8): 2000–2009. https://doi.org/10.1016/j.ejca.2013.02.021.

Chobot, Agata, Ewa Rusak, Janet Wenzlau, Howard Davidson, Piotr Adamczyk, Agnieszka Krzywicka, Bogdan Mazur, Joanna Polańska, and Marian Rewers. 2018. "ATP4A Autoimmunity in Pediatric Patients with Type 1 Diabetes and Its Relationship to Blood Count, Iron Metabolism, and Vitamin B12." *Pediatric Diabetes* 19 (1): 80–84. https://doi.org/10.1111/pedi.12528.

Choi, Kwangmin, and Nancy Ratner. 2019. “IGEAK: An Interactive Gene Expression Analysis Kit for Seamless Workflow Using the R/Shiny Platform.” *BMC Genomics* 20 (1): 177. https://doi.org/10.1186/s12864-019-5548-x.

Choi, Woonyoung, Sima Porten, Seungchan Kim, Daniel Willis, Elizabeth R. Plimack, Jean Hoffman-Censits, Beat Roth, et al. 2014. “Identification of Distinct Basal and Luminal Subtypes of Muscle-Invasive Bladder Cancer with Different Sensitivities to Frontline Chemotherapy.” *Cancer Cell* 25 (2): 152–65. https://doi.org/10.1016/j.ccr.2014.01.009.

Chu, Yongjun, and David R. Corey. 2012. “RNA Sequencing: Platform Selection, Experimental Design, and Data Interpretation.” *Nucleic Acid Therapeutics* 22 (4): 271–74. https://doi.org/10.1089/nat.2012.0367.

Colombo, Pierre-Emmanuel, Fernanda Milanezi, Britta Weigelt, and Jorge S Reis-Filho. 2011. “Microarrays in the 2010s: The Contribution of Microarray-Based Gene Expression Profiling to Breast Cancer Classification, Prognostication and Prediction.” *Breast Cancer Research : BCR* 13 (3): 212. https://doi.org/10.1186/bcr2890.

Conte, Simona, Shintaro Katayama, Liselotte Vesterlund, Mohsen Karimi, Marios Dimitriou, Monika Jansson, Teresa Mortera-Blanco, et al. 2015. “Aberrant Splicing of Genes Involved in Haemoglobin Synthesis and Impaired Terminal Erythroid Maturation in SF3B1 Mutated Refractory Anaemia with Ring Sideroblasts.” *British Journal of Haematology* 171 (4): 478–90. https://doi.org/10.1111/bjh.13610.

Cozzi, Anna, Paolo Santambrogio, Daniela Privitera, Vania Broccoli, Luisa Ida Rotundo, Barbara Garavaglia, Rudolf Benz, et al. 2013. “Human L-Ferritin Deficiency Is Characterized by Idiopathic Generalized Seizures and Atypical Restless Leg Syndrome.” *The Journal of Experimental Medicine* 210 (9): 1779–91. https://doi.org/10.1084/jem.20130315.

Cremonesi, L., A. Cozzi, D. Girelli, F. Ferrari, I. Fermo, B. Foglieni, S. Levi, et al. 2004. “Case Report: A Subject with a Mutation in the ATG Start Codon of L-Ferritin Has No Haematological or Neurological Symptoms.” *Journal of Medical Genetics* 41 (6): e81.

Cremonesi, L., A. Fumagalli, N. Soriani, M. Ferrari, S. Levi, S. Belloli, G. Ruggeri, and P. Arosio. 2001. “Double-Gradient Denaturing Gradient Gel Electrophoresis Assay for Identification of L-Ferritin Iron-Responsive Element Mutations Responsible for Hereditary Hyperferritinemia-Cataract Syndrome: Identification of the New Mutation C14G.” *Clin Chem* 47 (3): 491–97.

Cremonesi, L., R. Paroni, B. Foglieni, S. Galbiati, I. Fermo, N. Soriani, S. Belloli, et al. 2003. “Scanning Mutations of the 5’UTR Regulatory Sequence of L-Ferritin by Denaturing High-Performance Liquid Chromatography: Identification of New Mutations.” *Br J Haematol* 121 (1): 173–79.

Crispin, Andrew, Paul Schmidt, Dean Campagna, Chang Cao, Daniel Lichtenstein, Anoop Sendamarai, Chaoshe Guo, et al. 2017. “Hscb</Em>, a Mitochondrial Iron-Sulfur Cluster Assembly Co-Chaperone, Is a Novel Candidate Gene for Congenital Sideroblastic Anemia.” *Blood* 130 (Suppl 1): 79.

Crownover, Brian K, Carlton J Covey, Nellis Family Medicine Residency, and Nellis Air Force Base. 2013. “Hereditary Hemochromatosis” 87 (3): 8.

Curtis, A. R., C. Fey, C. M. Morris, L. A. Bindoff, P. G. Ince, P. F. Chinnery, A. Coulthard, et al. 2001. “Mutation in the Gene Encoding Ferritin Light Polypeptide Causes Dominant Adult-Onset Basal Ganglia Disease.” *Nature Genetics* 28 (4): 350–54. https://doi.org/10.1038/ng571.

D

Da Costa, Lydie, Marie-Françoise O’Donohue, Birgit van Dooijeweert, Katarzyna Albrecht, Sule Unal, Ugo Ramenghi, Thierry Leblanc, et al. 2018. “Molecular Approaches to Diagnose Diamond-Blackfan Anemia: The EuroDBA Experience.” *European Journal of Medical Genetics* 61 (11): 664–73. https://doi.org/10.1016/j.ejmg.2017.10.017.

Dai, Tian-Mei, Zhi-Chuang Lü, Wan-Xue Liu, and Fang-Hao Wan. 2017. “Selection and Validation of Reference Genes for QRT-PCR Analysis during Biological Invasions: The Thermal Adaptability of Bemisia Tabaci MED.” Edited by Xiaoming Pang. *PLOS ONE* 12 (3): e0173821. https://doi.org/10.1371/journal.pone.0173821.

De Falco, Luigia, Mayka Sanchez, Laura Silvestri, Caroline Kannengiesser, Martina U. Muckenthaler, Achille Iolascon, Laurent Gouya, Clara Camaschella, and Carole Beaumont. 2013. “Iron Refractory Iron Deficiency Anemia.” *Haematologica* 98 (6): 845–53. https://doi.org/10.3324/haematol.2012.075515.

De Falco, Luigia, Laura Silvestri, Caroline Kannengiesser, Erica Morán, Claire Oudin, Marco Rausa, Mariasole Bruno, et al. 2014. “Functional and Clinical Impact of Novel TMPRSS6 Variants in Iron-Refractory Iron-Deficiency Anemia Patients and Genotype-Phenotype Studies.” *Human Mutation* 35 (11): 1321–29. https://doi.org/10.1002/humu.22632.

Devos, David, P. Jissendi Tchofo, Isabelle Vuillaume, Alain Destée, Stephanie Batey, John Burn, and Patrick F. Chinnery. 2009. “Clinical Features and Natural History of Neuroferritinopathy Caused by the 458dupA FTL Mutation.” *Brain* 132 (6): e109. https://doi.org/10.1093/brain/awn274.

Dgany, Orly, Nili Avidan, Jean Delaunay, Tatyana Krasnov, Lea Shalmon, Hanna Shalev, Tal Eidelitz-Markus, et al. 2002. “Congenital Dyserythropoietic Anemia Type I Is Caused by Mutations in Codanin-1.” *American Journal of Human Genetics* 71 (6): 1467–74.

Dheda, Keertan, Jim F. Huggett, Stephen A. Bustin, Margaret A. Johnson, Graham Rook, and Alimuddin Zumla. 2004. “Validation of Housekeeping Genes for Normalizing RNA Expression in Real-Time PCR.” *BioTechniques* 37 (1): 112–14, 116, 118–19. https://doi.org/10.2144/04371RR03.

Di Resta, Chiara, Silvia Galbiati, Paola Carrera, and Maurizio Ferrari. 2018. “Next-Generation Sequencing Approach for the Diagnosis of Human Diseases: Open Challenges and New Opportunities.” *EJIFCC* 29 (1): 4–14.

Dijk, Erwin L. van, Yan Jaszczyszyn, Delphine Naquin, and Claude Thermes. 2018. “The Third Revolution in Sequencing Technology.” *Trends in Genetics* 34 (9): 666–81. https://doi.org/10.1016/j.tig.2018.05.008.

Dinesen, Birthe, Brandie Nonnecke, David Lindeman, Egon Toft, Kristian Kidholm, Kamal Jethwani, Heather M. Young, et al. 2016. “Personalized Telehealth in the Future: A Global Research Agenda.” *Journal of Medical Internet Research* 18 (3): e53. https://doi.org/10.2196/jmir.5257.

Ding, Ye, Chi Yu Chan, and Charles E. Lawrence. 2004. “Sfold Web Server for Statistical Folding and Rational Design of Nucleic Acids.” *Nucleic Acids Research* 32 (suppl_2): W135–41. https://doi.org/10.1093/nar/gkh449.

Doerken, Sam, Marta Avalos, Emmanuel Lagarde, and Martin Schumacher. 2019. “Penalized Logistic Regression with Low Prevalence Exposures beyond High Dimensional Settings.” *PLoS ONE* 14 (5). https://doi.org/10.1371/journal.pone.0217057.

Dokal, Inderjeet. 2011. “Dyskeratosis Congenita.” *Hematology. American Society of Hematology. Education Program* 2011: 480–86. https://doi.org/10.1182/asheducation-2011.1.480.

Dolatshad, H, A Pellagatti, M Fernandez-Mercado, B H Yip, L Malcovati, M Attwood, B Przychodzen, et al. 2015. “Disruption of SF3B1 Results in Deregulated Expression and Splicing of Key Genes and Pathways in Myelodysplastic Syndrome Hematopoietic Stem and Progenitor Cells.” *Leukemia* 29 (5): 1092–1103. https://doi.org/10.1038/leu.2014.331.

Ducamp, Sarah, and Mark D. Fleming. 2019. “The Molecular Genetics of Sideroblastic Anemia.” *Blood* 133 (1): 59. https://doi.org/10.1182/blood-2018-08-815951.

Durupt, S., I. Durieu, B. Salles, C. Beaumont, A. Hot, H. Rousset, and D. Vital Durand. 2001. “[A potential etiology of elevated ferritin: hyperferritinemia-cataract syndrome].” *Rev Med Interne* 22 (1): 83–85.

E

Engert, Andreas, Carlo Balduini, Anneke Brand, Bertrand Coiffier, Catherine Cordonnier, Hartmut Döhner, Thom Duyvené de Wit, et al. 2016. “The European Hematology Association Roadmap for European Hematology Research: A Consensus Document.” *Haematologica* 101 (2): 115–208. https://doi.org/10.3324/haematol.2015.136739.

Eszlinger, Markus, Knut Krohn, Aleksandra Kukulska, Barbara Jarząb, and Ralf Paschke. 2007. “Perspectives and Limitations of Microarray-Based Gene Expression Profiling of Thyroid Tumors.” *Endocrine Reviews* 28 (3): 322–38. https://doi.org/10.1210/er.2006-0047.

European Association for the Study of the Liver. 2010. “EASL Clinical Practice Guidelines for HFE Hemochromatosis.” *Journal of Hepatology* 53 (1): 3–22. https://doi.org/10.1016/j.jhep.2010.03.001.

F

Fan, Dong-Mei, Dong-Mei Fan, Xu Yang, Xu Yang, Li-Min Huang, Li-Min Huang, Guo-Jun Ouyang, et al. 2019. "Simultaneous Detection of Target CNVs and SNVs of Thalassemia by Multiplex PCR and Next-generation Sequencing." *Molecular Medicine Reports* 19 (4): 2837–48. https://doi.org/10.3892/mmr.2019.9896.

Faniello, Maria Concetta, Maddalena Di Sanzo, Barbara Quaresima, Antonia Nisticò, Annalisa Fregola, Michela Grosso, Giovanni Cuda, and Francesco Costanzo. 2009. "Bilateral Cataract in a Subject Carrying a C to A Transition in the L Ferritin Promoter Region." *Clinical Biochemistry* 42 (9): 911–14. https://doi.org/10.1016/j.clinbiochem.2009.02.013.

Fassbinder-Orth, Carol A. 2014. "Methods for Quantifying Gene Expression in Ecoimmunology: From QPCR to RNA-Seq." *Integrative and Comparative Biology* 54 (3): 396–406. https://doi.org/10.1093/icb/icu023.

Fernández García, M Soledad, and Julie Teruya-Feldstein. 2014. "The Diagnosis and Treatment of Dyskeratosis Congenita: A Review." *Journal of Blood Medicine* 5 (August): 157–67. https://doi.org/10.2147/JBM.S47437.

Fernandez-Marmiesse, Ana, Sofia Gouveia, and Maria L. Couce. 2018. "NGS Technologies as a Turning Point in Rare Disease Research , Diagnosis and Treatment." *Current Medicinal Chemistry* 25 (3). https://doi.org/10.2174/0929867324666170718101946.

Ferrari, Francesca, Barbara Foglieni, Paolo Arosio, Clara Camaschella, Filomena Daraio, Sonia Levi, José Antonio García Erce, et al. 2006. "Microelectronic DNA Chip for Hereditary Hyperferritinemia Cataract Syndrome, a Model for Large-Scale Analysis of Disorders of Iron Metabolism." *Human Mutation* 27 (2): 201–8. https://doi.org/10.1002/humu.20294.

Font, A., M. Taron, J. L. Gago, C. Costa, J. J. Sánchez, C. Carrato, M. Mora, et al. 2011. "BRCA1 MRNA Expression and Outcome to Neoadjuvant Cisplatin-Based Chemotherapy in Bladder Cancer." *Annals of Oncology: Official Journal of the European Society for Medical Oncology* 22 (1): 139–44. https://doi.org/10.1093/annonc/mdq333.

Fontenay, M., S. Cathelin, M. Amiot, E. Gyan, and E. Solary. 2006. "Mitochondria in Hematopoiesis and Hematological Diseases." *Oncogene* 25 (34): 4757–67. https://doi.org/10.1038/sj.onc.1209606.

Freson, K., K. Devriendt, G. Matthijs, A. Van Hoof, R. De Vos, C. Thys, K. Minner, M. F. Hoylaerts, J. Vermylen, and C. Van Geet. 2001. "Platelet Characteristics in Patients with X-Linked Macrothrombocytopenia Because of a Novel GATA1 Mutation." *Blood* 98 (1): 85–92. https://doi.org/10.1182/blood.v98.1.85.

Freson, Kathleen, Gert Matthijs, Chantal Thys, Paul Mariën, Marc F. Hoylaerts, Jos Vermylen, and Chris Van Geet. 2002. "Different Substitutions at Residue D218 of the X-Linked Transcription Factor GATA1 Lead to Altered Clinical Severity of Macrothrombocytopenia and Anemia and Are Associated with Variable Skewed X Inactivation." *Human Molecular Genetics* 11 (2): 147–52. https://doi.org/10.1093/hmg/11.2.147.

G

Gallagher, Patrick G. 2015. "Diagnosis and Management of Rare Congenital Nonimmune Hemolytic Disease." *Hematology. American Society of Hematology. Education Program* 2015: 392–99. https://doi.org/10.1182/asheducation-2015.1.392.

Gambale, Antonella, Achille Iolascon, Immacolata Andolfo, and Roberta Russo. 2016. "Diagnosis and Management of Congenital Dyserythropoietic Anemias." *Expert Review of Hematology* 9 (3): 283–96. https://doi.org/10.1586/17474086.2016.1131608.

Garber, Ian, and Morris Pudek. 2014. "A Novel Deletion in the Iron-Response Element of the L-Ferritin Gene, Causing Hyperferritinaemia Cataract Syndrome." *Annals of Clinical Biochemistry* 51 (Pt 6): 710–13. https://doi.org/10.1177/0004563214542289.

Garcia Erce, J. A., T. Cortes, L. Cremonesi, M. Cazzola, G. Perez-Lungmus, and M. Giralt. 2006. "[Hyperferritinemia-cataract syndrome associated to the HFE gene mutation. Two new Spanish families and a new mutation (A37T: 'Zaragoza')]." *Med Clin (Barc)* 127 (2): 55–58.

Garcia-Closas, M., N. Rothman, J. D. Figueroa, L. Prokunina-Olsson, S. S. Han, D. Baris, E. J. Jacobs, et al. 2013. "Common Genetic Polymorphisms Modify the Effect of Smoking on Absolute Risk of Bladder Cancer." *Cancer Research* 73 (7): 2211–20. https://doi.org/10.1158/0008-5472.CAN-12-2388.

Garcia-Gomez, Maria, Andrea Calabria, Maria Garcia-Bravo, Fabrizio Benedicenti, Penelope Kosinski, Sergio López-Manzaneda, Collin Hill, et al. 2016. "Safe and Efficient Gene Therapy for Pyruvate Kinase Deficiency."

Molecular Therapy: The Journal of the American Society of Gene Therapy 24 (7): 1187–98. https://doi.org/10.1038/mt.2016.87.

Garderet, L., B. Hermelin, N. C. Gorin, and O. Rosmorduc. 2004. "Hereditary Hyperferritinemia-Cataract Syndrome: A Novel Mutation in the Iron-Responsive Element of the L-Ferritin Gene in a French Family." *Am J Med* 117 (2): 138–39. https://doi.org/10.1016/j.amjmed.2004.02.033.

Garringer, Holly J., Jose M. Irimia, Wei Li, Charles B. Goodwin, Briana Richine, Anthony Acton, Rebecca J. Chan, et al. 2016. "Effect of Systemic Iron Overload and a Chelation Therapy in a Mouse Model of the Neurodegenerative Disease Hereditary Ferritinopathy." *PLoS ONE* 11 (8). https://doi.org/10.1371/journal.pone.0161341.

Giansily, M., C. Beaumont, C. Desveaux, G. Hetet, J. F. Schved, and P. Aguilar-Martinez. 2001. "Denaturing Gradient Gel Electrophoresis Screening for Mutations in the Hereditary Hyperferritinaemia Cataract Syndrome." *Br J Haematol* 112 (1): 51–54.

Girelli, D., R. Corrocher, L. Bisceglia, O. Olivieri, L. Zelante, G. Panozzo, and P. Gasparini. 1997. "Hereditary Hyperferritinemia-Cataract Syndrome Caused by a 29-Base Pair Deletion in the Iron Responsive Element of Ferritin L-Subunit Gene." *Blood* 90 (5): 2084–88.

Girelli, D., O. Olivieri, L. De Franceschi, R. Corrocher, G. Bergamaschi, and M. Cazzola. 1995. "A Linkage between Hereditary Hyperferritinaemia Not Related to Iron Overload and Autosomal Dominant Congenital Cataract." *Br J Haematol* 90 (4): 931–34.

González Fernández, F. A., A. Villegas, P. Ropero, M. D. Carreño, E. Anguita, M. Polo, A. Pascual, and A. Henández. 2009. "Haemoglobinopathies with High Oxygen Affinity. Experience of Erythropathology Cooperative Spanish Group." *Annals of Hematology* 88 (3): 235–38. https://doi.org/10.1007/s00277-008-0581-x.

Gonzalez-Huerta, L., V. Ramirez-Sanchez, M. Rivera-Vega, O. Messina-Baas, and S. Cuevas-Covarrubias. 2008. "A Family with Hereditary Hyperferritinaemia Cataract Syndrome: Evidence of Incomplete Penetrance and Clinical Heterogeneity." *Br J Haematol* 143 (4): 596–98. https://doi.org/10.1111/j.1365-2141.2008.07345.x.

Gordis L. 2013. *Epidemiology.* Philadelphia, PA: Elsevier.

Grandchamp, Bernard, Gilles Hetet, Caroline Kannengiesser, Claire Oudin, Carole Beaumont, Sylvie Rodrigues-Ferreira, Robert Amson, et al. 2011. "A Novel Type of Congenital Hypochromic Anemia Associated with a Nonsense Mutation in the STEAP3/TSAP6</Em> Gene." *Blood* 118 (25): 6660. https://doi.org/10.1182/blood-2011-01-329011.

Green, Robert C., Jonathan S. Berg, Wayne W. Grody, Sarah S. Kalia, Bruce R. Korf, Christa L. Martin, Amy McGuire, et al. 2013. "ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing." *Genetics in Medicine : Official Journal of the American College of Medical Genetics* 15 (7): 565–74. https://doi.org/10.1038/gim.2013.73.

Griffiths, G, R. Hall, R Sylvester, D Raghavan, and MK Parmar. 2011. "International Phase III Trial Assessing Neoadjuvant Cisplatin, Methotrexate, and Vinblastine Chemotherapy for Muscle-Invasive Bladder Cancer: Long-Term Results of the BA06 30894 Trial." *Journal of Clinical Oncology* 29 (16): 2171–77. https://doi.org/10.1200/JCO.2010.32.3139.

Grossman, H. Barton, Ronald B. Natale, Catherine M. Tangen, V.O. Speights, Nicholas J. Vogelzang, Donald L. Trump, Ralph W. deVere White, et al. 2003. "Neoadjuvant Chemotherapy plus Cystectomy Compared with Cystectomy Alone for Locally Advanced Bladder Cancer." *New England Journal of Medicine* 349 (9): 859–66. https://doi.org/10.1056/NEJMoa022148.

H

Haley, Kristina. 2017. "Congenital Hemolytic Anemia." *The Medical Clinics of North America* 101 (2): 361–74. https://doi.org/10.1016/j.mcna.2016.09.008.

Harigae, Hideo, and Kazumichi Furuyama. 2010. "Hereditary Sideroblastic Anemia: Pathophysiology and Gene Mutations." *International Journal of Hematology* 92 (3): 425–31. https://doi.org/10.1007/s12185-010-0688-4.

Heimpel, Hermann, Kerstin Kellermann, Nadine Neuschwander, Josef Högel, and Klaus Schwarz. 2010. "The Morphological Diagnosis of Congenital Dyserythropoietic Anemia: Results of a Quantitative Analysis of

Peripheral Blood and Bone Marrow Cells." *Haematologica* 95 (6): 1034–36. https://doi.org/10.3324/haematol.2009.014563.

Heimpel, Hermann, Andreas Matuschek, Momin Ahmed, Brigitte Bader-Meunier, Adriana Colita, Jean Delaunay, Loic Garcon, et al. 2010. "Frequency of Congenital Dyserythropoietic Anemias in Europe." *European Journal of Haematology* 85 (1): 20–25.

Heimpel, Hermann, Klaus Schwarz, Monika Ebnöther, Jeroen S. Goede, Detlev Heydrich, Torsten Kamp, Lothar Plaumann, et al. 2006. "Congenital Dyserythropoietic Anemia Type I (CDA I): Molecular Genetics, Clinical Appearance, and Prognosis Based on Long-Term Observation." *Blood* 107 (1): 334–40. https://doi.org/10.1182/blood-2005-01-0421.

Heiss, N. S., S. W. Knight, T. J. Vulliamy, S. M. Klauck, S. Wiemann, P. J. Mason, A. Poustka, and I. Dokal. 1998. "X-Linked Dyskeratosis Congenita Is Caused by Mutations in a Highly Conserved Gene with Putative Nucleolar Functions." *Nature Genetics* 19 (1): 32–38. https://doi.org/10.1038/ng0598-32.

Hensley, Patrick J., Natasha Kyprianou, Matthew S. Purdom, Daheng He, Vincent DiCarlo, Chi Wang, and Andrew C. James. 2019. "Predictive Value of Phenotypic Signatures of Bladder Cancer Response to Cisplatin-Based Neoadjuvant Chemotherapy." *Urologic Oncology*, July. https://doi.org/10.1016/j.urolonc.2019.06.020.

Hetet, G., I. Devaux, N. Soufir, B. Grandchamp, and C. Beaumont. 2003. "Molecular Analyses of Patients with Hyperferritinemia and Normal Serum Iron Values Reveal Both L Ferritin IRE and 3 New Ferroportin (Slc11A3) Mutations." *Blood* 102 (5): 1904–10. https://doi.org/10.1182/blood-2003-02-0439.

Hoerl, Arthur E., and Robert W. Kennard. 1970. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics* 12 (1): 55. https://doi.org/10.2307/1267351.

Hoffman, Ronald, Edward J Benz, Leslie E Silberstein, Helen Heslop, Jeffrey Weitz, and John L. Hopper. 2018. *Hematology: Basic Principles and Practice,*. 7th ed. Elsevier Saunders.

Hollanda, Luciana M., Carmen S. P. Lima, Anderson F. Cunha, Dulcinéia M. Albuquerque, José Vassallo, Margareth C. Ozelo, Paulo P. Joazeiro, Sara T. O. Saad, and Fernando F. Costa. 2006. "An Inherited Mutation Leading to Production of Only the Short Isoform of GATA-1 Is Associated with Impaired Erythropoiesis." *Nature Genetics* 38 (7): 807–12. https://doi.org/10.1038/ng1825.

Hosmer DW, Lemeshow S, Sturdivant RX. 2013. *Applied Logistic Regression.* John Wiley & Sons;

I

Igci, Mehri, Mehmet Baysan, Remzi Yigiter, Mustafa Ulasli, Sirma Geyik, Recep Bayraktar, İbrahim Bozgeyik, Esra Bozgeyik, Ali Bayram, and Ecir Ali Cakmak. 2016. "Gene Expression Profiles of Autophagy-Related Genes in Multiple Sclerosis." *Gene* 588 (1): 38–46. https://doi.org/10.1016/j.gene.2016.04.042.

Iolascon, Achille, Immacolata Andolfo, and Roberta Russo. 2015. "Red Cells in Post-Genomic Era: Impact of Personalized Medicine in the Treatment of Anemias." *Haematologica* 100 (1): 3–6. https://doi.org/10.3324/haematol.2014.120733.

Iolascon, Achille, Maria d'Apolito, Veronica Servedio, Flora Cimmino, Antonio Piga, and Clara Camaschella. 2006. "Microcytic Anemia and Hepatic Iron Overload in a Child with Compound Heterozygous Mutations in DMT1 (SCL11A2)</Em>." *Blood* 107 (1): 349. https://doi.org/10.1182/blood-2005-06-2477.

Iolascon, Achille, Maria Rosaria Esposito, and Roberta Russo. 2012. "Clinical Aspects and Pathogenesis of Congenital Dyserythropoietic Anemias: From Morphology to Molecular Approach." *Haematologica* 97 (12): 1786–94. https://doi.org/10.3324/haematol.2012.072207.

Iolascon, Achille, Hermann Heimpel, Anders Wahlin, and Hannah Tamary. 2013. "Congenital Dyserythropoietic Anemias: Molecular Insights and Diagnostic Approach." *Blood* 122 (13): 2162–66. https://doi.org/10.1182/blood-2013-05-468223.

Iolascon, Achille, Roberta Russo, Maria Rosaria Esposito, Roberta Asci, Carmelo Piscopo, Silverio Perrotta, Madeleine Fénéant-Thibault, Loïc Garçon, and Jean Delaunay. 2010. "Molecular Analysis of 42 Patients with Congenital Dyserythropoietic Anemia Type II: New Mutations in the SEC23B Gene and a Search for a Genotype-Phenotype Relationship." *Haematologica* 95 (5): 708–15. https://doi.org/10.3324/haematol.2009.014985.

Ishizawa, Jo, Kenji Nakamaru, Takahiko Seki, Koichi Tazaki, Kensuke Kojima, Dhruv Chachad, Ran Zhao, et al. 2018. “Predictive Gene Signatures Determine Tumor Sensitivity to MDM2 Inhibition.” *Cancer Research* 78 (10): 2721–31. https://doi.org/10.1158/0008-5472.CAN-17-0949.

Itadani, Hiraku, Shinji Mizuarai, and Hidehito Kotani. 2008. “Can Systems Biology Understand Pathway Activation? Gene Expression Signatures as Surrogate Markers for Understanding the Complexity of Pathway Activation.” *Current Genomics* 9 (5): 349–60. https://doi.org/10.2174/138920208785133235.

J

Jaffray, Julie A., W. Beau Mitchell, Merlin Nithya Gnanapragasam, Surya V. Seshan, Xinhuo Guo, Connie M. Westhoff, James J. Bieker, and Deepa Manwani. 2013. “Erythroid Transcription Factor EKLF/KLF1 Mutation Causing Congenital Dyserythropoietic Anemia Type IV in a Patient of Taiwanese Origin: Review of All Reported Cases and Development of a Clinical Diagnostic Paradigm.” *Blood Cells, Molecules & Diseases* 51 (2): 71–75. https://doi.org/10.1016/j.bcmd.2013.02.006.

Jang, Jin Sung, Vernadette A Simon, Rod M Feddersen, Fariborz Rakhshan, Debra A Schultz, Michael A Zschunke, Wilma L Lingle, Christopher P Kolbert, and Jin Jen. 2011. “Quantitative MiRNA Expression Analysis Using Fluidigm Microfluidics Dynamic Arrays.” *BMC Genomics* 12 (March): 144. https://doi.org/10.1186/1471-2164-12-144.

Jayawardana, Kaushala, Sarah-Jane Schramm, Lauren Haydu, John F. Thompson, Richard A. Scolyer, Graham J. Mann, Samuel Müller, and Jean Yee Hwa Yang. 2015. “Determination of Prognosis in Metastatic Melanoma through Integration of Clinico-Pathologic, Mutation, MRNA, MicroRNA, and Protein Information.” *International Journal of Cancer* 136 (4): 863–74. https://doi.org/10.1002/ijc.29047.

Juvonen, E., E. Ikkala, F. Fyhrquist, and T. Ruutu. 1991. “Autosomal Dominant Erythrocytosis Caused by Increased Sensitivity to Erythropoietin.” *Blood* 78 (11): 3066–69.

K

Kallio, M. Aleksi, Jarno T. Tuimala, Taavi Hupponen, Petri Klemelä, Massimiliano Gentile, Ilari Scheinin, Mikko Koski, Janne Käki, and Eija I. Korpelainen. 2011. “Chipster: User-Friendly Analysis Software for Microarray and Other High-Throughput Data.” *BMC Genomics* 12 (October): 507. https://doi.org/10.1186/1471-2164-12-507.

Kamel, Hala Fawzy Mohamed, and Hiba Saeed A. Bagader Al-Amodi. 2017. “Exploitation of Gene Expression and Cancer Biomarkers in Paving the Path to Era of Personalized Medicine.” *Genomics, Proteomics & Bioinformatics* 15 (4): 220–35. https://doi.org/10.1016/j.gpb.2016.11.005.

Kannengiesser, C., A. M. Jouanolle, G. Hetet, A. Mosser, F. Muzeau, D. Henry, E. Bardou-Jacquet, et al. 2009. “A New Missense Mutation in the L Ferritin Coding Sequence Associated with Elevated Levels of Glycosylated Ferritin in Serum and Absence of Iron Overload.” *Haematologica* 94 (3): 335–39. https://doi.org/10.3324/haematol.2008.000125.

Kannengiesser, Caroline, Anne-Marie Jouanolle, Gilles Hetet, Annick Mosser, Françoise Muzeau, Dominique Henry, Edouard Bardou-Jacquet, et al. 2009. “A New Missense Mutation in the L Ferritin Coding Sequence Associated with Elevated Levels of Glycosylated Ferritin in Serum and Absence of Iron Overload.” *Haematologica* 94 (3): 335–39. https://doi.org/10.3324/haematol.2008.000125.

Kannengiesser, Caroline, Mayka Sanchez, Marion Sweeney, Gilles Hetet, Briedgeen Kerr, Erica Moran, Jose L. Fuster Soler, et al. 2011. “Missense SLC25A38 Variations Play an Important Role in Autosomal Recessive Inherited Sideroblastic Anemia.” *Haematologica* 96 (6): 808–13. https://doi.org/10.3324/haematol.2010.039164.

Kato, (first), and Casella, F. 1999. “L-Ferritin Baltimore-1: A Novel Mutation in the Iron Respinsive Element (C32G) as a Cause of Hyperferritinemia-Cataract Syndrome.” *Blood* 94: 407a.

Kato, J., K. Fujikawa, M. Kanda, N. Fukuda, K. Sasaki, T. Takayama, M. Kobune, et al. 2001. “A Mutation, in the Iron-Responsive Element of H Ferritin MRNA, Causing Autosomal Dominant Iron Overload.” *Am J Hum Genet* 69 (1): 191–97.

Kebschull, Moritz, Melanie Julia Fittler, Ryan T. Demmer, and Panos N. Papapanou. 2017. “Differential Expression and Functional Analysis of High-Throughput -Omics Data Using Open Source Tools.” *Methods in Molecular Biology (Clifton, N.J.)* 1537: 327–45. https://doi.org/10.1007/978-1-4939-6685-1_19.

Kelmenson, Daniel A., and Michael Hanley. 2017. “Dyskeratosis Congenita.” *The New England Journal of Medicine* 376 (15): 1460. https://doi.org/10.1056/NEJMicm1613081.

Kieke, Michele C., Jacob Klemm, Arthur Rech Tondin, Victor Alencar, Nathan Johnson, Ashley M. Driver, Thomas Lentz, Gregory J. Fischer, Diane A. Caporale, and Luke J. Drury. 2019. “Characterization of a Novel Pathogenic Variant in the FECH Gene Associated with Erythropoietic Protoporphyria.” *Molecular Genetics and Metabolism Reports* 20 (September): 100481. https://doi.org/10.1016/j.ymgmr.2019.100481.

Kiemeney, Lambertus A, Steinunn Thorlacius, Patrick Sulem, Frank Geller, Katja K H Aben, Simon N Stacey, Julius Gudmundsson, et al. 2008. “Sequence Variant on 8q24 Confers Susceptibility to Urinary Bladder Cancer.” *Nature Genetics* 40 (11): 1307–12. https://doi.org/10.1038/ng.229.

Knies, Kerstin, Shojiro Inano, María J. Ramírez, Masamichi Ishiai, Jordi Surrallés, Minoru Takata, and Detlev Schindler. 2017. “Biallelic Mutations in the Ubiquitin Ligase RFWD3 Cause Fanconi Anemia.” *The Journal of Clinical Investigation* 127 (8): 3013–27. https://doi.org/10.1172/JCI92069.

Koeppen, Katja, Bruce A Stanton, and Thomas H Hampton. 2017. “ScanGEO: Parallel Mining of High-Throughput Gene Expression Data.” Edited by Jonathan Wren. *Bioinformatics* 33 (21): 3500–3501. https://doi.org/10.1093/bioinformatics/btx452.

Kono, Satoshi. 2013. “Aceruloplasminemia: An Update.” *International Review of Neurobiology* 110: 125–51. https://doi.org/10.1016/B978-0-12-410502-7.00007-7.

Koralkova, P., W. W. van Solinge, and R. van Wijk. 2014. “Rare Hereditary Red Blood Cell Enzymopathies Associated with Hemolytic Anemia - Pathophysiology, Clinical Aspects, and Laboratory Diagnosis.” *International Journal of Laboratory Hematology* 36 (3): 388–97. https://doi.org/10.1111/ijlh.12223.

Korenková, Vlasta, Justin Scott, Vendula Novosadová, Marie Jindřichová, Lucie Langerová, David Švec, Monika Šídová, and Robert Sjöback. 2015. “Pre-Amplification in the Context of High-Throughput QPCR Gene Expression Experiment.” *BMC Molecular Biology* 16 (March). https://doi.org/10.1186/s12867-015-0033-9.

Kozera, Bartłomiej, and Marcin Rapacz. 2013. “Reference Genes in Real-Time PCR.” *Journal of Applied Genetics* 54 (4): 391–406. https://doi.org/10.1007/s13353-013-0173-x.

Kralovics, Robert, Karel Indrak, Tomas Stopka, Brian W. Berman, Jaroslav F. Prchal, and Josef T. Prchal. 1997. “Two New EPO Receptor Mutations: Truncated EPO Receptors Are Most Frequently Associated With Primary Familial and Congenital Polycythemias.” *Blood* 90 (5): 2057–61.

Kubota, Akatsuki, Ayumi Hida, Yaeko Ichikawa, Yoshio Momose, Jun Goto, Yukifusa Igeta, Hideji Hashida, et al. 2009. “A Novel Ferritin Light Chain Gene Mutation in a Japanese Family with Neuroferritinopathy: Description of Clinical Features and Implications for Genotype-Phenotype Correlations.” *Movement Disorders: Official Journal of the Movement Disorder Society* 24 (3): 441–45. https://doi.org/10.1002/mds.22435.

L

Labay, Valentina, Tal Raz, Dana Baron, Hanna Mandel, Hawys Williams, Timothy Barrett, Raymonde Szargel, et al. 1999. “Mutations in SLC19A2 Cause Thiamine-Responsive Megaloblastic Anaemia Associated with Diabetes Mellitus and Deafness.” *Nature Genetics* 22 (3): 300–304. https://doi.org/10.1038/10372.

Ladroue, Charline, Romain Carcenac, Michel Leporrier, Sophie Gad, Claire Le Hello, Françoise Galateau-Salle, Jean Feunteun, Jacques Pouysségur, Stéphane Richard, and Betty Gardie. 2008. “PHD2 Mutation and Congenital Erythrocytosis with Paraganglioma.” *The New England Journal of Medicine* 359 (25): 2685–92. https://doi.org/10.1056/NEJMoa0806277.

Lang, Stefan, Amol Ugale, Eva Erlandsson, Göran Karlsson, David Bryder, and Shamit Soneji. 2015. “SCExV: A Webtool for the Analysis and Visualisation of Single Cell QRT-PCR Data.” *BMC Bioinformatics* 16 (October). https://doi.org/10.1186/s12859-015-0757-z.

Lazarini, Mariana, João Agostinho Machado-Neto, Leticia Fröhlich Archangelo, Bruna Fernandes Mendes-Silva, Carolina Louzão Bigarella, Fabiola Traina, and Sara Teresinha Olalla Saad. 2013. “PTK2 and PTPN11 Expression in Myelodysplastic Syndromes.” *Clinics* 68 (10): 1371–75. https://doi.org/10.6061/clinics/2013(10)13.

Lee, Julia M., John R. Roche, Danny J. Donaghy, Anthony Thrush, and Puthigae Sathish. 2010. “Validation of Reference Genes for Quantitative RT-PCR Studies of Gene Expression in Perennial Ryegrass (Lolium Perenne L.).” *BMC Molecular Biology* 11 (January): 8. https://doi.org/10.1186/1471-2199-11-8.

Lemarchandel, V., V. Joulin, C. Valentin, R. Rosa, F. Galacteros, J. Rosa, and M. Cohen- Solal. 1992. “Compound Heterozygosity in a Complete Erythrocyte Bisphosphoglycerate Mutase Deficiency.” *Blood* 80 (10): 2643–49.

Lenzhofer, M., F. Schroedl, A. Trost, A. Kaser-Eichberger, H. Wiedemann, C. Strohmaier, M. Hohensinn, et al. 2015. “Aqueous Humor Ferritin in Hereditary Hyperferritinemia Cataract Syndrome.” *Optom Vis Sci* 92 (4 Suppl 1): S40-7. https://doi.org/10.1097/OPX.0000000000000544.

Lerner, Seth P., Mark P. Schoenberg, and Cora N. Sternberg. 2015. *Bladder Cancer: Diagnosis and Clinical Management*. John Wiley & Sons.

Levi, Sonia, and Valeria Tiranti. 2019. “Neurodegeneration with Brain Iron Accumulation Disorders: Valuable Models Aimed at Understanding the Pathogenesis of Iron Deposition.” *Pharmaceuticals* 12 (1). https://doi.org/10.3390/ph12010027.

Li, Junchang, Zhixin Jiao, Ruishi He, Yulong Sun, Qiaoqiao Xu, Jing Zhang, Yumei Jiang, Qiaoyun Li, and Jishan Niu. 2019. “Gene Expression Profiles and MicroRNA Regulation Networks in Tiller Primordia, Stem Tips, and Young Spikes of Wheat Guomai 301.” *Genes* 10 (9). https://doi.org/10.3390/genes10090686.

Lichtenstein, Daniel A, Andrew W Crispin, Anoop K Sendamarai, Dean R Campagna, Klaus Schmitz-Abe, Cristovao M Sousa, Martin D Kafina, et al. 2016. “A Recurring Mutation in the Respiratory Complex 1 Protein NDUFB11 Is Responsible for a Novel Form of X-Linked Sideroblastic Anemia.” *Blood* 128 (15): 1913–17. https://doi.org/10.1182/blood-2016-05-719062.

Liebler, Daniel C., and F. Peter Guengerich. 2005. “Elucidating Mechanisms of Drug-Induced Toxicity.” *Nature Reviews Drug Discovery* 4 (5): 410–20. https://doi.org/10.1038/nrd1720.

Liljeholm, Maria, Andrew F. Irvine, Ann-Louise Vikberg, Anna Norberg, Stacy Month, Herbert Sandström, Anders Wahlin, Masanori Mishima, and Irina Golovleva. 2013. “Congenital Dyserythropoietic Anemia Type III (CDA III) Is Caused by a Mutation in Kinesin Family Member, KIF23</Em>.” *Blood* 121 (23): 4791. https://doi.org/10.1182/blood-2012-10-461392.

Liu, G., S. Guo, G. J. Anderson, C. Camaschella, B. Han, and G. Nie. 2014. “Heterozygous Missense Mutations in the GLRX5 Gene Cause Sideroblastic Anemia in a Chinese Patient.” *Blood* 124 (17): 2750–51. https://doi.org/10.1182/blood-2014-08-598508.

Liu, Qi, and Richard I. Gregory. 2019. “RNAmod: An Integrated System for the Annotation of MRNA Modifications.” *Nucleic Acids Research* 47 (W1): W548–55. https://doi.org/10.1093/nar/gkz479.

Livak, K. J., and T. D. Schmittgen. 2001. “Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2(-Delta Delta C(T)) Method.” *Methods (San Diego, Calif.)* 25 (4): 402–8. https://doi.org/10.1006/meth.2001.1262.

Long, Zhangbiao, Hongmin Li, Yali Du, and Bing Han. 2018. “Congenital Sideroblastic Anemia: Advances in Gene Mutations and Pathophysiology.” *Gene* 668 (August): 182–89. https://doi.org/10.1016/j.gene.2018.05.074.

López de Maturana, Evangelina, Lola Alonso, Pablo Alarcón, Isabel Adoración Martín-Antoniano, Silvia Pineda, Lucas Piorno, M. Luz Calle, and Núria Malats. 2019. “Challenges in the Integration of Omics and Non-Omics Data.” *Genes* 10 (3): 238. https://doi.org/10.3390/genes10030238.

Luscieti, S., P. Santambrogio, B. Langlois d’Estaintot, T. Granier, A. Cozzi, M. Poli, B. Gallois, et al. 2010. “Mutant Ferritin L-Chains That Cause Neurodegeneration Act in a Dominant-Negative Manner to Reduce Ferritin Iron Incorporation.” *J Biol Chem* 285 (16): 11948–57. https://doi.org/10.1074/jbc.M109.096404.

Luscieti, Sara, Gabriele Tolle, Jessica Aranda, Carmen Benet Campos, Frank Risse, Érica Morán, Martina U Muckenthaler, and Mayka Sánchez. 2013. “Novel Mutations in the Ferritin-L Iron-Responsive Element That Only Mildly Impair IRP Binding Cause Hereditary Hyperferritinaemia Cataract Syndrome.” *Orphanet Journal of Rare Diseases* 8 (February): 30. https://doi.org/10.1186/1750-1172-8-30.

M

Ma, Rui, Sheng Xu, Yucheng Zhao, Bing Xia, and Ren Wang. 2016. “Selection and Validation of Appropriate Reference Genes for Quantitative Real-Time PCR Analysis of Gene Expression in Lycoris Aurea.” *Frontiers in Plant Science* 7 (April). https://doi.org/10.3389/fpls.2016.00536.

Macgregor, Pascale F., and Jeremy A. Squire. 2002. “Application of Microarrays to the Analysis of Gene Expression in Cancer.” *Clinical Chemistry* 48 (8): 1170–77.

Maciel, P., V.T. Cruz, M. Constante, I. Iniesta, M.C. Costa, S. Gallati, N. Sousa, J. Sequeiros, P. Coutinho, and M.M. Santos. 2005. “Neuroferritinopathy: Missense Mutation in FTL Causing Early-Onset Bilateral Pallidal Involvement.” *Neurology* 65 (4): 603–5. https://doi.org/10.1212/01.wnl.0000178224.81169.c2.

Malcovati, Luca, and Mario Cazzola. 2016. “Recent Advances in the Understanding of Myelodysplastic Syndromes with Ring Sideroblasts.” *British Journal of Haematology* 174 (6): 847–58. https://doi.org/10.1111/bjh.14215.

Mallona, Izaskun, Anna Díez-Villanueva, Berta Martín, and Miguel A Peinado. 2017. “Chainy: An Universal Tool for Standardized Relative Quantification in Real-Time PCR.” *Bioinformatics* 33 (9): 1411–13. https://doi.org/10.1093/bioinformatics/btw839.

Mancuso, Michelangelo, Guido Davidzon, Roger M. Kurlan, Rabi Tawil, Eduardo Bonilla, Salvatore Di Mauro, and James M. Powers. 2005. “Hereditary Ferritinopathy: A Novel Mutation, Its Cellular Pathology, and Pathogenetic Insights.” *Journal of Neuropathology and Experimental Neurology* 64 (4): 280–94.

Martin, M. E., S. Fargion, P. Brissot, B. Pellat, and C. Beaumont. 1998. “A Point Mutation in the Bulge of the Iron-Responsive Element of the L Ferritin Gene in Two Families with the Hereditary Hyperferritinemia-Cataract Syndrome.” *Blood* 91 (1): 319–23.

Mattila, R. M., A. Sainio, M. Jarvelainen, J. Pursiheimo, and H. Jarvelainen. 2018. “A Novel Double Nucleotide Variant in the Ferritin-L Iron-Responsive Element in a Finnish Patient with Hereditary Hyperferritinaemia-Cataract Syndrome.” *Acta Ophthalmol* 96 (1): 95–99. https://doi.org/10.1111/aos.13492.

Maturana, Evangelina López de, Stephen J. Chanok, Antoni C. Picornell, Nathaniel Rothman, Jesús Herranz, M. Luz Calle, Montserrat García-Closas, et al. 2014. “Whole Genome Prediction of Bladder Cancer Risk with the Bayesian LASSO.” *Genetic Epidemiology* 38 (5): 467–76. https://doi.org/10.1002/gepi.21809.

McDonnell, S. M., B. L. Preston, S. A. Jewell, J. C. Barton, C. Q. Edwards, P. C. Adams, and R. Yip. 1999. “A Survey of 2,851 Patients with Hemochromatosis: Symptoms and Response to Treatment.” *The American Journal of Medicine* 106 (6): 619–24.

McLeod, J. L., J. Craig, S. Gumley, S. Roberts, and M. A. Kirkland. 2002. “Mutation Spectrum in Australian Pedigrees with Hereditary Hyperferritinaemia-Cataract Syndrome Reveals Novel and de Novo Mutations.” *Br J Haematol* 118 (4): 1179–82.

Mehaffey, M. G., A. L. Newton, M. J. Gandhi, M. Crossley, and J. G. Drachman. 2001. “X-Linked Thrombocytopenia Caused by a Novel Mutation of GATA-1.” *Blood* 98 (9): 2681–88. https://doi.org/10.1182/blood.v98.9.2681.

Meneses, F. G. A., B. Schnabel, I. D. C. G. Silva, F. L. Alberto, L. Toma, H. B. Nader, and C. C. Lopes. 2011. “Identification of the Mutations Associated with Hereditary Hyperferritinemia Cataract Syndrome and Hemochromatosis in a Brazilian Family.” *Clinical Genetics* 79 (2): 189–92. https://doi.org/10.1111/j.1399-0004.2010.01517.x.

Merryweather-Clarke, Alison T., Jennifer J. Pointon, Anne Marie Jouanolle, Jacques Rochette, and Kathryn J. H. Robson. 2000. “Geography of HFE C282Y and H63D Mutations.” *Genetic Testing* 4 (2): 183–98. https://doi.org/10.1089/10906570050114902.

Messa, E., R. M. Pellegrino, A. Palmieri, S. Carturan, D. Cilloni, G. Saglio, and A. Roetto. 2009. “Identification of a Novel Mutation in the L Ferritin Iron-Responsive Element Causing Hereditary Hyperferritinemia-Cataract Syndrome.” *Acta Haematol* 122 (4): 223–25. https://doi.org/10.1159/000253031.

Mizuarai, Shinji, Kazunori Yamanaka, Hiraku Itadani, Tsuyoshi Arai, Toshihide Nishibata, Hiroshi Hirai, and Hidehito Kotani. 2009. “Discovery of Gene Expression-Based Pharmacodynamic Biomarker for a P53 Context-Specific Anti-Tumor Drug Wee1 Inhibitor.” *Molecular Cancer* 8 (1): 34. https://doi.org/10.1186/1476-4598-8-34.

Moradkhani, Kamran, Claude Préhu, John Old, Shirley Henderson, Vera Balamitsa, Hong-Yuan Luo, Man-Chiu Poon, David H. K. Chui, Henri Wajcman, and George P. Patrinos. 2009. “Mutations in the Paralogous Human α-

Globin Genes Yielding Identical Hemoglobin Variants." *Annals of Hematology* 88 (6): 535–43. https://doi.org/10.1007/s00277-008-0624-3.

Moutton, Sébastien, Patricia Fergelot, Jean-Marc Trocello, Violaine Plante-Bordeneuve, Nada Houcinat, Emilie Wenisch, Vincent Larue, et al. 2014. "A Novel FTL Mutation Responsible for Neuroferritinopathy with Asymmetric Clinical Features and Brain Anomalies." *Parkinsonism & Related Disorders* 20 (8): 935–37. https://doi.org/10.1016/j.parkreldis.2014.04.026.

Muckenthaler, Martina U., Stefano Rivella, Matthias W. Hentze, and Bruno Galy. 2017. "A Red Carpet for Iron Metabolism." *Cell* 168 (3): 344–61. https://doi.org/10.1016/j.cell.2016.12.034.

Muhoberac, Barry B., Martin A. Baraibar, and Ruben Vidal. 2011. "Iron Loading-Induced Aggregation and Reduction of Iron Incorporation in Heteropolymeric Ferritin Containing a Mutant Light Chain That Causes Neurodegeneration." *Biochimica et Biophysica Acta* 1812 (4): 544–48. https://doi.org/10.1016/j.bbadis.2010.10.010.

Muhoberac, Barry B., and Ruben Vidal. 2013. "Abnormal Iron Homeostasis and Neurodegeneration." *Frontiers in Aging Neuroscience* 5. https://doi.org/10.3389/fnagi.2013.00032.

Mullane, Stephanie A., Lillian Werner, Elizabeth A. Guancial, Rosina T. Lis, Edward C. Stack, Massimo Loda, Philip W. Kantoff, Toni K. Choueiri, Jonathan Rosenberg, and Joaquim Bellmunt. 2016. "Expression Levels of DNA Damage Repair Proteins Are Associated with Overall Survival in Platinum Treated Advanced Urothelial Carcinoma." *Clinical Genitourinary Cancer* 14 (4): 352–59. https://doi.org/10.1016/j.clgc.2015.12.029.

Mumford, A. D., I. A. Cree, J. D. Arnold, M. C. Hagan, K. C. Rixon, and J. J. Harding. 2000. "The Lens in Hereditary Hyperferritinaemia Cataract Syndrome Contains Crystalline Deposits of L-Ferritin." *Br J Ophthalmol* 84 (7): 697–700.

Mumford, A. D., T. Vulliamy, J. Lindsay, and A. Watson. 1998. "Hereditary Hyperferritinemia-Cataract Syndrome: Two Novel Mutations in the L-Ferritin Iron-Responsive Element." *Blood* 91 (1): 367–68.

Muñoz-Muñoz, J., N. Cuadrado-Grande, M.-I. Moreno-Carralero, B. Hoyos-Sanabria, A. Manubes-Guarch, A.-F. González, P. Tejada-Palacios, A. Del-Castillo-Rueda, and M.-J. Morán-Jiménez. 2013. "Hereditary Hyperferritinemia Cataract Syndrome in Four Patients with Mutations in the IRE of the FTL Gene." *Clinical Genetics* 83 (5): 491–93. https://doi.org/10.1111/j.1399-0004.2012.01934.x.

Muramatsu, Hideki, Yusuke Okuno, Kenichi Yoshida, Yuichi Shiraishi, Sayoko Doisaki, Atsushi Narita, Hirotoshi Sakaguchi, et al. 2017. "Clinical Utility of Next-Generation Sequencing for Inherited Bone Marrow Failure Syndromes." *Genetics in Medicine* 19 (7): 796–802. https://doi.org/10.1038/gim.2016.197.

Murta-Nascimento, Cristiane, Debra T. Silverman, Manolis Kogevinas, Montserrat García-Closas, Nathaniel Rothman, Adonina Tardón, Reina García-Closas, et al. 2007. "Risk of Bladder Cancer Associated with Family History of Cancer: Do Low-Penetrance Polymorphisms Account for the Increase in Risk?" *Cancer Epidemiology Biomarkers & Prevention* 16 (8): 1595. https://doi.org/10.1158/1055-9965.EPI-06-0743.

N

Napier, I. 2005. "Iron Trafficking in the Mitochondrion: Novel Pathways Revealed by Disease." *Blood* 105 (5): 1867–74. https://doi.org/10.1182/blood-2004-10-3856.

Nguyen, Hao G., Christopher J. Welty, and Matthew R. Cooperberg. 2015. "Diagnostic Associations of Gene Expression Signatures in Prostate Cancer Tissue." *Current Opinion in Urology* 25 (1): 65–70. https://doi.org/10.1097/MOU.0000000000000131.

Ni, Wang, Hong-Fu Li, Yi-Cen Zheng, and Zhi-Ying Wu. 2016. "FTL Mutation in a Chinese Pedigree with Neuroferritinopathy." *Neurology. Genetics* 2 (3): e74. https://doi.org/10.1212/NXG.0000000000000074.

Nichols, Kim E., John D. Crispino, Mortimer Poncz, James G. White, Stuart H. Orkin, John M. Maris, and Mitchell J. Weiss. 2000. "Familial Dyserythropoietic Anaemia and Thrombocytopenia Due to an Inherited Mutation in GATA1." *Nature Genetics* 24 (3): 266–70. https://doi.org/10.1038/73480.

Nishida, Katsuya, Holly J. Garringer, Naonobu Futamura, Itaru Funakawa, Kenji Jinnai, Ruben Vidal, and Masaki Takao. 2014. "A Novel Ferritin Light Chain Mutation in Neuroferritinopathy with an Atypical Presentation." *Journal of the Neurological Sciences* 342 (1–2): 173–77. https://doi.org/10.1016/j.jns.2014.03.060.

Njajou, O. T., N. Vaessen, M. Joosse, B. Berghuis, J. W. van Dongen, M. H. Breuning, P. J. Snijders, et al. 2001. "A Mutation in SLC11A3 Is Associated with Autosomal Dominant Hemochromatosis." *Nature Genetics* 28 (3): 213–14. https://doi.org/10.1038/90038.

Noy-Lotan, Sharon, Orly Dgany, Roxane Lahmi, Nathaly Marcoux, Tanya Krasnov, Nissan Yissachar, Doron Ginsberg, et al. 2009. "Codanin-1, the Protein Encoded by the Gene Mutated in Congenital Dyserythropoietic Anemia Type I (CDAN1), Is Cell Cycle-Regulated." *Haematologica* 94 (5): 629–37. https://doi.org/10.3324/haematol.2008.003327.

O

Ogino, Shuji, Margaret L. Gulley, Johan T. den Dunnen, and Robert B. Wilson. 2007. "Standard Mutation Nomenclature in Molecular Diagnostics." *The Journal of Molecular Diagnostics : JMD* 9 (1): 1–6. https://doi.org/10.2353/jmoldx.2007.060081.

Ogutu, Joseph O, Torben Schulz-Streeck, and Hans-Peter Piepho. 2012. "Genomic Selection Using Regularized Linear Regression Models: Ridge Regression, Lasso, Elastic Net and Their Extensions." *BMC Proceedings* 6 (S2). https://doi.org/10.1186/1753-6561-6-S2-S10.

Ohta, E., T. Nagasaka, K. Shindo, S. Toma, K. Nagasaka, K. Ohta, and Z. Shiozawa. 2008. "Neuroferritinopathy in a Japanese Family with a Duplication in the Ferritin Light Chain Gene." *Neurology* 70 (16 Pt 2): 1493–94. https://doi.org/10.1212/01.wnl.0000310428.74624.95.

Olwagen, Courtney P., Peter V. Adrian, and Shabir A. Madhi. 2019. "Performance of the Biomark HD Real-Time QPCR System (Fluidigm) for the Detection of Nasopharyngeal Bacterial Pathogens and Streptococcus Pneumoniae Typing." *Scientific Reports* 9 (1). https://doi.org/10.1038/s41598-019-42846-y.

P

Pagani, Alessia, Silvia Colucci, Renata Bocciardi, Marta Bertamino, Carlo Dufour, Roberto Ravazzolo, Laura Silvestri, and Clara Camaschella. 2017. "A New Form of IRIDA Due to Combined Heterozygous Mutations of TMPRSS6 and ACVR1A Encoding the BMP Receptor ALK2." *Blood* 129 (25): 3392–95. https://doi.org/10.1182/blood-2017-03-773481.

Palmer, William C., Prakash Vishnu, William Sanchez, Bashar Aqel, Doug Riegert-Johnson, Leigh Ann Kenda Seaman, Andrew W. Bowman, and Candido E. Rivera. 2018. "Diagnosis and Management of Genetic Iron Overload Disorders." *Journal of General Internal Medicine* 33 (12): 2230–36. https://doi.org/10.1007/s11606-018-4669-2.

Papaemmanuil, E., M. Cazzola, J. Boultwood, L. Malcovati, P. Vyas, D. Bowen, A. Pellagatti, et al. 2011. "Somatic SF3B1 Mutation in Myelodysplasia with Ring Sideroblasts." *The New England Journal of Medicine* 365 (15): 1384–95. https://doi.org/10.1056/NEJMoa1103283.

Papanikolaou, George, Mark E Samuels, Erwin H Ludwig, Marcia L E MacDonald, Patrick L Franchini, Marie-Pierre Dubé, Lisa Andres, et al. 2004. "Mutations in HFE2 Cause Iron Overload in Chromosome 1q–Linked Juvenile Hemochromatosis." *Nature Genetics* 36 (1): 77–82. https://doi.org/10.1038/ng1274.

Patnaik, Mrinal M., and Ayalew Tefferi. 2017. "Refractory Anemia with Ring Sideroblasts (RARS) and RARS with Thrombocytosis (RARS-T) – '2017 Update on Diagnosis, Risk-Stratification, and Management.'" *American Journal of Hematology* 92 (3): 297–310. https://doi.org/10.1002/ajh.24637.

Pearson, Howard A., Jeffrey S. Lobel, Samuel A. Kocoshis, J. Lawrence Naiman, Joan Windmiller, Ahti T. Lammi, Ronald Hoffman, and John C. Marsh. 1979. "A New Syndrome of Refractory Sideroblastic Anemia with Vacuolization of Marrow Precursors and Exocrine Pancreatic Dysfunction." *The Journal of Pediatrics* 95 (6): 976–84. https://doi.org/10.1016/S0022-3476(79)80286-3.

Pellagatti, Andrea, Mario Cazzola, Aristoteles A. N. Giagounidis, Luca Malcovati, Matteo G. Della Porta, Sally Killick, Lisa J. Campbell, et al. 2006. "Gene Expression Profiles of CD34+ Cells in Myelodysplastic Syndromes: Involvement of Interferon-Stimulated Genes and Correlation to FAB Subtype and Karyotype." *Blood* 108 (1): 337–45. https://doi.org/10.1182/blood-2005-12-4769.

Percy, Melanie J., Paul W. Furlow, Guy S. Lucas, Xiping Li, Terence R. J. Lappin, Mary Frances McMullin, and Frank S. Lee. 2008. "A Gain-of-Function Mutation in the HIF2A Gene in Familial Erythrocytosis." *The New England Journal of Medicine* 358 (2): 162–68. https://doi.org/10.1056/NEJMoa073123.

Phillips, J. D., C. A. Warby, and J. P. Kushner. 2005. “Identification of a Novel Mutation in the L-Ferritin IRE Leading to Hereditary Hyperferritinemia-Cataract Syndrome.” *Am J Med Genet A* 134A (1): 77–79. https://doi.org/10.1002/ajmg.a.30425.

Pierro, Elena Di, Roberta Russo, Zeynep Karakas, Valentina Brancaleoni, Antonella Gambale, Ismail Kurt, S. Stuart Winter, et al. 2015. “Congenital Erythropoietic Porphyria Linked to GATA1-R216W Mutation: Challenges for Diagnosis.” *European Journal of Haematology* 94 (6): 491–97. https://doi.org/10.1111/ejh.12452.

Pitroda, Sean P., Itai M. Pashtan, Hillary L. Logan, Brian Budke, Thomas E. Darga, Ralph R. Weichselbaum, and Philip P. Connell. 2014. “DNA Repair Pathway Gene Expression Score Correlates with Repair Proficiency and Tumor Sensitivity to Chemotherapy.” *Science Translational Medicine* 6 (229): 229ra42. https://doi.org/10.1126/scitranslmed.3008291.

Prchal, Josef T. 2005. “Polycythemia Vera and Other Primary Polycythemias.” *Current Opinion in Hematology* 12 (2): 112–16.

Prchal, Josef T., and Xylina T. Gregg. 2005. “Red Cell Enzymes.” *Hematology. American Society of Hematology. Education Program*, 19–23. https://doi.org/10.1182/asheducation-2005.1.19.

R

Rabinovitch, A. 1990. “Hematology Reference Ranges.” *Archives of Pathology & Laboratory Medicine* 114 (12): 1189.

Radonić, Aleksandar, Stefanie Thulke, Ian M. Mackay, Olfert Landt, Wolfgang Siegert, and Andreas Nitsche. 2004. “Guideline to Reference Gene Selection for Quantitative Real-Time PCR.” *Biochemical and Biophysical Research Communications* 313 (4): 856–62. https://doi.org/10.1016/j.bbrc.2003.11.177.

Ramirez, R N, K Bedirian, S M Gray, and A Diallo. 2019. “DNA Rchitect: An R Based Visualizer for Network Analysis of Chromatin Interaction Data.” *Bioinformatics*, August. https://doi.org/10.1093/bioinformatics/btz608.

Ramirez-Herrick, Ashley M., Shannon E. Mullican, Andrea M. Sheehan, and Orla M. Conneely. 2011. “Reduced NR4A Gene Dosage Leads to Mixed Myelodysplastic/Myeloproliferative Neoplasms in Mice.” *Blood* 117 (9): 2681–90. https://doi.org/10.1182/blood-2010-02-267906.

Rao, Mohan S., Terry R. Van Vleet, Rita Ciurlionis, Wayne R. Buck, Scott W. Mittelstadt, Eric A. G. Blomme, and Michael J. Liguori. 2019. “Comparison of RNA-Seq and Microarray Gene Expression Platforms for the Toxicogenomic Evaluation of Liver From Short-Term Rat Toxicity Studies.” *Frontiers in Genetics* 9 (January). https://doi.org/10.3389/fgene.2018.00636.

Ravasi, Giulia, Sara Pelucchi, Raffaella Mariani, Marco Casati, Federico Greni, Cristina Arosio, Irene Pelloni, et al. 2017. “Unexplained Isolated Hyperferritinemia without Iron Overload.” *American Journal of Hematology* 92 (4): 338–43. https://doi.org/10.1002/ajh.24641.

Ravindranath, Yaddanapudi, Gerald Goyette, Steven Buck, Manisha Gadgeel, Alan Dombkowski, Laurence A. Boxer, Patrick G. Gallagher, and Robert M. Johnson. 2011. “A New Case of KLF1 G973A Mutation and Congenital Dyserythropoeitic Anemia (CDA)- Further Definition of Emerging New Syndrome and Possible Association with Gonadal Dysgenesis.” *Blood* 118 (21): 2101–2101.

Rehm, Heidi L. 2013. “Disease-Targeted Sequencing: A Cornerstone in the Clinic.” *Nature Reviews Genetics* 14 (4): 295–300. https://doi.org/10.1038/nrg3463.

Rey, Mónica del, Rocío Benito, Celia Fontanillo, Francisco J. Campos-Laborie, Kamila Janusz, Talía Velasco-Hernández, María Abáigar, et al. 2015. “Deregulation of Genes Related to Iron and Mitochondrial Metabolism in Refractory Anemia with Ring Sideroblasts.” *PloS One* 10 (5): e0126555. https://doi.org/10.1371/journal.pone.0126555.

Richards, Sue, Nazneen Aziz, Sherri Bale, David Bick, Soma Das, Julie Gastier-Foster, Wayne W. Grody, et al. 2015. “Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology.” *Genetics in Medicine : Official Journal of the American College of Medical Genetics* 17 (5): 405–24. https://doi.org/10.1038/gim.2015.30.

Ricketts, Christopher J., Jayne A. Minton, Jacob Samuel, Indra Ariyawansa, Jerry K. Wales, Ivan F. Lo, and Timothy G. Barrett. 2006. "Thiamine-Responsive Megaloblastic Anaemia Syndrome: Long-Term Follow-up and Mutation Analysis of Seven Families." *Acta Paediatrica (Oslo, Norway : 1992)* 95 (1): 99–104. https://doi.org/10.1080/08035250500323715.

Riley, Lisa G, Minal J Menezes, Joëlle Rudinger-Thirion, Rachael Duff, Pascale de Lonlay, Agnes Rotig, Michel C Tchan, Mark Davis, Sandra T Cooper, and John Christodoulou. 2013. "Phenotypic Variability and Identification of Novel YARS2 Mutations in YARS2 Mitochondrial Myopathy, Lactic Acidosis and Sideroblastic Anaemia." *Orphanet Journal of Rare Diseases* 8 (December): 193. https://doi.org/10.1186/1750-1172-8-193.

Riley, Lisa G., Joëlle Rudinger-Thirion, Klaus Schmitz-Abe, David R. Thorburn, Ryan L. Davis, Juliana Teo, Susan Arbuckle, et al. 2015. "LARS2 Variants Associated with Hydrops, Lactic Acidosis, Sideroblastic Anemia, and Multisystem Failure." In *JIMD Reports, Volume 28*, edited by Eva Morava, Matthias Baumgartner, Marc Patterson, Shamima Rahman, Johannes Zschocke, and Verena Peters, 28:49–57. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/8904_2015_515.

Roetto, Antonella, George Papanikolaou, Marianna Politou, Federica Alberti, Domenico Girelli, John Christakis, Dimitris Loukopoulos, and Clara Camaschella. 2003. "Mutant Antimicrobial Peptide Hepcidin Is Associated with Severe Juvenile Hemochromatosis." *Nature Genetics* 33 (1): 21–22. https://doi.org/10.1038/ng1053.

Rothman KJ, Greenland S, Lash TL. 2012. *Modern Epidemiology*. Philadelphia, PA: Lippincott Williams & Wilkins.

Rothman, Nathaniel, Montserrat Garcia-Closas, Nilanjan Chatterjee, Nuria Malats, Xifeng Wu, Jonine D Figueroa, Francisco X Real, et al. 2010. "A Multi-Stage Genome-Wide Association Study of Bladder Cancer Identifies Multiple Susceptibility Loci." *Nature Genetics* 42 (11): 978–84. https://doi.org/10.1038/ng.687.

Rotig, A., M. Colonna, J. P. Bonnefont, S. Blanche, A. Fischer, J. M. Saudubray, and A. Munnich. 1989. "Mitochondrial DNA Deletion in Pearson's Marrow/Pancreas Syndrome." *Lancet (London, England)* 1 (8643): 902–3. https://doi.org/10.1016/s0140-6736(89)92897-3.

Rubinstein, Wendy S., Donna R. Maglott, Jennifer M. Lee, Brandi L. Kattman, Adriana J. Malheiro, Michael Ovetsky, Vichet Hem, et al. 2013. "The NIH Genetic Testing Registry: A New, Centralized Database of Genetic Tests to Enable Access to Comprehensive Information and Improve Transparency." *Nucleic Acids Research* 41 (D1): D925–35. https://doi.org/10.1093/nar/gks1173.

Rufer, A., J. P. Howell, A. P. Lange, R. Yamamoto, J. Heuscher, M. Gregor, and W. A. Wuillemin. 2011. "Hereditary Hyperferritinemia-Cataract Syndrome (HHCS) Presenting with Iron Deficiency Anemia Associated with a New Mutation in the Iron Responsive Element of the L Ferritin Gene in a Swiss Family." *Eur J Haematol*, March. https://doi.org/10.1111/j.1600-0609.2011.01607.x.

Russo, Roberta, Immacolata Andolfo, Francesco Manna, Antonella Gambale, Roberta Marra, Barbara Eleni Rosato, Paola Caforio, et al. 2018. "Multi-Gene Panel Testing Improves Diagnosis and Management of Patients with Hereditary Anemias." *American Journal of Hematology* 93 (5): 672–82. https://doi.org/10.1002/ajh.25058.

Russo, Roberta, Immacolata Andolfo, Francesco Manna, Gianluca De Rosa, Luigia De Falco, Antonella Gambale, Mariasole Bruno, et al. 2016. "Increased Levels of ERFE-Encoding FAM132B in Patients with Congenital Dyserythropoietic Anemia Type II." *Blood* 128 (14): 1899–1902. https://doi.org/10.1182/blood-2016-06-724328.

Russo, Roberta, Antonella Gambale, Maria Rosaria Esposito, Maria Luisa Serra, Annaelena Troiano, Ilaria De Maggio, Mario Capasso, et al. 2011. "Two Founder Mutations in the SEC23B Gene Account for the Relatively High Frequency of CDA II in the Italian Population." *American Journal of Hematology* 86 (9): 727–32. https://doi.org/10.1002/ajh.22096.

Russo, Roberta, Antonella Gambale, Concetta Langella, Immacolata Andolfo, Sule Unal, and Achille Iolascon. 2014. "Retrospective Cohort Study of 205 Cases with Congenital Dyserythropoietic Anemia Type II: Definition of Clinical and Molecular Spectrum and Identification of New Diagnostic Scores." *American Journal of Hematology* 89 (10): E169–75. https://doi.org/10.1002/ajh.23800.

Russo, Roberta, Concetta Langella, Maria Rosaria Esposito, Antonella Gambale, Francesco Vitiello, Fara Vallefuoco, Torben Ek, Elizabeth Yang, and Achille Iolascon. 2013. "Hypomorphic Mutations of SEC23B Gene

Account for Mild Phenotypes of Congenital Dyserythropoietic Anemia Type II." *Blood Cells, Molecules, and Diseases* 51 (1): 17–21. https://doi.org/10.1016/j.bcmd.2013.02.003.

S

Sabour, Leila, Maryam Sabour, and Saeid Ghorbian. 2017. "Clinical Applications of Next-Generation Sequencing in Cancer Diagnosis." *Pathology & Oncology Research* 23 (2): 225–34. https://doi.org/10.1007/s12253-016-0124-z.

Saint-Jacques, Nathalie, Louise Parker, Patrick Brown, and Trevor Jb Dummer. 2014. "Arsenic in Drinking Water and Urinary Tract Cancers: A Systematic Review of 30 Years of Epidemiological Evidence." *Environmental Health: A Global Access Science Source* 13 (June): 44. https://doi.org/10.1186/1476-069X-13-44.

Samorodnitsky, Eric, Jharna Datta, Benjamin M. Jewell, Raffi Hagopian, Jharna Miya, Michele R. Wing, Senthilkumar Damodaran, et al. 2015. "Comparison of Custom Capture for Targeted Next-Generation DNA Sequencing." *The Journal of Molecular Diagnostics* 17 (1): 64–75. https://doi.org/10.1016/j.jmoldx.2014.09.009.

Samorodnitsky, Eric, Benjamin M. Jewell, Raffi Hagopian, Jharna Miya, Michele R. Wing, Ezra Lyon, Senthilkumar Damodaran, et al. 2015. "Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing." *Human Mutation* 36 (9): 903–14. https://doi.org/10.1002/humu.22825.

Sanchez, Mayka, and Albert Altés. 2017. "Eritropatología." In *Eritropatología*, 569–82. Ambos Marketing.

Sanchez, Mayka, Bruno Galy, Bjoern Schwanhaeusser, Jonathon Blake, Tomi Bähr-Ivacevic, Vladimir Benes, Matthias Selbach, Martina U. Muckenthaler, and Matthias W. Hentze. 2011. "Iron Regulatory Protein-1 and -2: Transcriptome-Wide Definition of Binding MRNAs and Shaping of the Cellular Proteome by Iron Regulatory Proteins." *Blood* 118 (22): e168–79. https://doi.org/10.1182/blood-2011-04-343541.

Sandström, H., and A. Wahlin. 2000. "Congenital Dyserythropoietic Anemia Type III." *Haematologica* 85 (7): 753–57.

Schmitz-Abe, Klaus, Szymon J Ciesielski, Paul J Schmidt, Dean R Campagna, Fedik Rahimov, Brenda A Schilke, Marloes Cuijpers, et al. 2015. "Congenital Sideroblastic Anemia Due to Mutations in the Mitochondrial HSP70 Homologue HSPA9." *Blood* 126 (25): 2734–38. https://doi.org/10.1182/blood-2015-09-659854.

Schwarz, Klaus, Achille Iolascon, Fatima Verissimo, Nikolaus S. Trede, Wyatt Horsley, Wen Chen, Barry H. Paw, et al. 2009. "Mutations Affecting the Secretory COPII Coat Component SEC23B Cause Congenital Dyserythropoietic Anemia Type II." *Nature Genetics* 41 (8): 936–40. https://doi.org/10.1038/ng.405.

Schwarze, Katharina, James Buchanan, Jenny C. Taylor, and Sarah Wordsworth. 2018. "Are Whole-Exome and Whole-Genome Sequencing Approaches Cost-Effective? A Systematic Review of the Literature." *Genetics in Medicine : Official Journal of the American College of Medical Genetics* 20 (10): 1122–30. https://doi.org/10.1038/gim.2017.247.

Setoodeh, Aria, Amirreza Haghighi, Nasrollah Saleh-Gohari, Sian Ellard, and Alireza Haghighi. 2013. "Identification of a SLC19A2 Nonsense Mutation in Persian Families with Thiamine-Responsive Megaloblastic Anemia." *Gene* 519 (2): 295–97. https://doi.org/10.1016/j.gene.2013.02.008.

Shefer Averbuch, Noa, Orna Steinberg-Shemer, Orly Dgany, Tanya Krasnov, Sharon Noy-Lotan, Joanne Yacobovich, Amir A. Kuperman, et al. 2018. "Targeted next Generation Sequencing for the Diagnosis of Patients with Rare Congenital Anemias." *European Journal of Haematology* 101 (3): 297–304. https://doi.org/10.1111/ejh.13097.

Skop, Ahna R., Hongbin Liu, John Yates, Barbara J. Meyer, and Rebecca Heald. 2004. "Dissection of the Mammalian Midbody Proteome Reveals Conserved Cytokinesis Mechanisms." *Science (New York, N.Y.)* 305 (5680): 61–66. https://doi.org/10.1126/science.1097931.

Solomon, Ponnumony John, Priya Margaret, Ramya Rajendran, Revathy Ramalingam, Godfred A Menezes, Alph S Shirley, Seung Jun Lee, et al. 2015. "A Case Report and Literature Review of Fanconi Anemia (FA) Diagnosed by Genetic Testing." *Italian Journal of Pediatrics* 41 (May). https://doi.org/10.1186/s13052-015-0142-6.

Sompele, Stijn Van de, Lucie Pécheux, Jorge Couso, Audrey Meunier, Mayka Sanchez, and Elfride De Baere. 2017. "Functional Characterization of a Novel Non - Coding Mutation 'Ghent +49A > G' in the Iron-Responsive

Element of L-Ferritin Causing Hereditary Hyperferritinaemia-Cataract Syndrome." *Scientific Reports* 7 (1): 18025. https://doi.org/10.1038/s41598-017-18326-6.

Song, Jianning, Zhigang Bai, Wei Han, Jun Zhang, Hua Meng, Jintao Bi, Xuemei Ma, Shiwei Han, and Zhongtao Zhang. 2012. "Identification of Suitable Reference Genes for QPCR Analysis of Serum MicroRNA in Gastric Cancer Patients." *Digestive Diseases and Sciences* 57 (4): 897–904. https://doi.org/10.1007/s10620-011-1981-7.

Spurgeon, Sandra L., Robert C. Jones, and Ramesh Ramakrishnan. 2008. "High Throughput Gene Expression Measurement with Real Time PCR in a Microfluidic Dynamic Array." Edited by Cathal Seoighe. *PLoS ONE* 3 (2): e1662. https://doi.org/10.1371/journal.pone.0001662.

Suwinski, Pawel, ChuangKee Ong, Maurice H. T. Ling, Yang Ming Poh, Asif M. Khan, and Hui San Ong. 2019. "Advancing Personalized Medicine Through the Application of Whole Exome Sequencing and Big Data Analytics." *Frontiers in Genetics* 10 (February). https://doi.org/10.3389/fgene.2019.00049.

T

Tabur, S., S. Oztuzcu, I. V. Duzen, A. Eraydin, S. Eroglu, M. Ozkaya, and A. T. Demiryürek. 2015. "Role of the Transient Receptor Potential (TRP) Channel Gene Expressions and TRP Melastatin (TRPM) Channel Gene Polymorphisms in Obesity-Related Metabolic Syndrome." *European Review for Medical and Pharmacological Sciences* 19 (8): 1388–97.

Tefferi, Ayalew, and Tiziano Barbui. 2017. "Polycythemia Vera and Essential Thrombocythemia: 2017 Update on Diagnosis, Risk-Stratification, and Management." *American Journal of Hematology* 92 (1): 94–108. https://doi.org/10.1002/ajh.24607.

Thellin, O., W. Zorzi, B. Lakaye, B. De Borman, B. Coumans, G. Hennen, T. Grisar, A. Igout, and E. Heinen. 1999. "Housekeeping Genes as Internal Standards: Use and Limits." *Journal of Biotechnology* 75 (2–3): 291–95.

Thomas, G R, E A Roberts, J M Walshe, and D W Cox. 1995. "Haplotypes and Mutations in Wilson Disease." *American Journal of Human Genetics* 56 (6): 1315–19.

Thompson, Jeffrey A., Brock C. Christensen, and Carmen J. Marsit. 2018. "Methylation-to-Expression Feature Models of Breast Cancer Accurately Predict Overall Survival, Distant-Recurrence Free Survival, and Pathologic Complete Response in Multiple Cohorts." *Scientific Reports* 8 (1): 5190. https://doi.org/10.1038/s41598-018-23494-0.

Thurlow, Vanessa, Bipin Vadher, Adrian Bomford, Corinne DeLord, Caroline Kannengiesser, Carole Beaumont, and Bernard Grandchamp. 2012. "Two Novel Mutations in the L Ferritin Coding Sequence Associated with Benign Hyperferritinaemia Unmasked by Glycosylated Ferritin Assay." *Annals of Clinical Biochemistry* 49 (Pt 3): 302–5. https://doi.org/10.1258/acb.2011.011229.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society. Series B (Methodological)* 58 (1): 267–88.

Tonella, Luca, Marco Giannoccaro, Salvatore Alfieri, Silvana Canevari, and Loris De Cecco. 2017. "Gene Expression Signatures for Head and Neck Cancer Patient Stratification: Are Results Ready for Clinical Application?" *Current Treatment Options in Oncology* 18 (5): 32. https://doi.org/10.1007/s11864-017-0472-2.

Tornador, Cristian, Edgar Sánchez-Prados, Beatriz Cadenas, Roberta Russo, Veronica Venturi, Immacolata Andolfo, Ines Hernández-Rodriguez, Achille Iolascon, and Mayka Sánchez. 2019. "CoDysAn: A Telemedicine Tool to Improve Awareness and Diagnosis for Patients With Congenital Dyserythropoietic Anemia." *Frontiers in Physiology* 10. https://doi.org/10.3389/fphys.2019.01063.

U

Ulirsch, Jacob C., Jeffrey M. Verboon, Shideh Kazerounian, Michael H. Guo, Daniel Yuan, Leif S. Ludwig, Robert E. Handsaker, et al. 2018. "The Genetic Landscape of Diamond-Blackfan Anemia." *The American Journal of Human Genetics* 103 (6): 930–47. https://doi.org/10.1016/j.ajhg.2018.10.027.

Unal, Sule, Roberta Russo, Fatma Gumruk, Baris Kuskonmaz, Mualla Cetin, Tulin Sayli, Betul Tavil, Concetta Langella, Achille Iolascon, and Duygu Uckan Cetinkaya. 2014. "Successful Hematopoietic Stem Cell Transplantation in a Patient with Congenital Dyserythropoietic Anemia Type II." *Pediatric Transplantation* 18 (4): E130–33. https://doi.org/10.1111/petr.12254.

V

Van de Sompele, S., L. Pecheux, J. Couso, A. Meunier, M. Sanchez, and E. De Baere. 2017. “Functional Characterization of a Novel Non-Coding Mutation ‘Ghent +49A > G’ in the Iron-Responsive Element of L-Ferritin Causing Hereditary Hyperferritinaemia-Cataract Syndrome.” *Sci Rep* 7 (1): 18025. https://doi.org/10.1038/s41598-017-18326-6.

Vandesompele, Jo, Katleen De Preter, Filip Pattyn, Bruce Poppe, Nadine Van Roy, Anne De Paepe, and Frank Speleman. 2002. “Accurate Normalization of Real-Time Quantitative RT-PCR Data by Geometric Averaging of Multiple Internal Control Genes.” *Genome Biology* 3 (7): research0034.1-research0034.11.

Veldman-Jones, M. H., Z. Lai, M. Wappett, C. G. Harbron, J. C. Barrett, E. A. Harrington, and K. S. Thress. 2015. “Reproducible, Quantitative, and Flexible Molecular Subtyping of Clinical DLBCL Samples Using the NanoString NCounter System.” *Clinical Cancer Research* 21 (10): 2367–78. https://doi.org/10.1158/1078-0432.CCR-14-0357.

Veldman-Jones, Margaret H., Roz Brant, Claire Rooney, Catherine Geh, Hollie Emery, Chris G. Harbron, Mark Wappett, et al. 2015. “Evaluating Robustness and Sensitivity of the NanoString Technologies NCounter Platform to Enable Multiplexed Gene Expression Analysis of Clinical Samples.” *Cancer Research* 75 (13): 2587–93. https://doi.org/10.1158/0008-5472.CAN-15-0262.

Vidal, R., B. Ghetti, M. Takao, C. Brefel-Courbon, E. Uro-Coste, B. S. Glazier, V. Siani, et al. 2004. “Intracellular Ferritin Accumulation in Neural and Extraneural Tissue Characterizes a Neurodegenerative Disease Associated with a Mutation in the Ferritin Light Polypeptide Gene.” *Journal of Neuropathology and Experimental Neurology* 63 (4): 363–80.

Vidal, Ruben, Leticia Miravalle, Xiaoying Gao, Ana G. Barbeito, Martin A. Baraibar, Shahryar K. Hekmatyar, Mario Widel, Navin Bansal, Marie B. Delisle, and Bernardino Ghetti. 2008. “Expression of a Mutant Form of the Ferritin Light Chain Gene Induces Neurodegeneration and Iron Overload in Transgenic Mice.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 28 (1): 60–67. https://doi.org/10.1523/JNEUROSCI.3962-07.2008.

Vliet, Martin H. van, Hugo M. Horlings, Marc J. van de Vijver, Marcel J. T. Reinders, and Lodewyk F. A. Wessels. 2012. “Integration of Clinical and Gene Expression Data Has a Synergetic Effect on Predicting Breast Cancer Outcome.” *PloS One* 7 (7): e40358. https://doi.org/10.1371/journal.pone.0040358.

Vrijenhoek, Terry, Ken Kraaijeveld, Martin Elferink, Joep de Ligt, Elcke Kranendonk, Gijs Santen, Isaac J Nijman, et al. 2015. “Next-Generation Sequencing-Based Genome Diagnostics across Clinical Genetics Centers: Implementation Choices and Their Effects.” *European Journal of Human Genetics* 23 (9): 1142–50. https://doi.org/10.1038/ejhg.2014.279.

W

Wakeman, L., S. Al-Ismail, A. Benton, A. Beddall, A. Gibbs, S. Hartnell, K. Morris, and R. Munro. 2007. “Robust, Routine Haematology Reference Ranges for Healthy Adults.” *International Journal of Laboratory Hematology* 29 (4): 279–83. https://doi.org/10.1111/j.1365-2257.2006.00883.x.

Wang, Zhong, Mark Gerstein, and Michael Snyder. 2009. “RNA-Seq: A Revolutionary Tool for Transcriptomics.” *Nature Reviews. Genetics* 10 (1): 57–63. https://doi.org/10.1038/nrg2484.

Weischenfeldt, Joachim, Orsolya Symmons, François Spitz, and Jan O. Korbel. 2013. “Phenotypic Impact of Genomic Structural Variation: Insights from and for Human Disease.” *Nature Reviews Genetics* 14 (2): 125–38. https://doi.org/10.1038/nrg3373.

Whatley, Sharon D., Sarah Ducamp, Laurent Gouya, Bernard Grandchamp, Carole Beaumont, Michael N. Badminton, George H. Elder, et al. 2008. “C-Terminal Deletions in the ALAS2 Gene Lead to Gain of Function and Cause X-Linked Dominant Protoporphyria without Anemia or Iron Overload.” *The American Journal of Human Genetics* 83 (3): 408–14. https://doi.org/10.1016/j.ajhg.2008.08.003.

Wickramasinghe, S. N., N. Illum, and P. D. Wimberley. 1991. “Congenital Dyserythropoietic Anaemia with Novel Intra-Erythroblastic and Intra-Erythrocytic Inclusions.” *British Journal of Haematology* 79 (2): 322–30. https://doi.org/10.1111/j.1365-2141.1991.tb04541.x.

Williams, Emma L., Eleanor A. L. Bagg, Michael Mueller, Jana Vandrovcova, Timothy J. Aitman, and Gill Rumsby. 2015. “Performance Evaluation of Sanger Sequencing for the Diagnosis of Primary Hyperoxaluria and Comparison with Targeted next Generation Sequencing.” *Molecular Genetics & Genomic Medicine* 3 (1): 69–78. https://doi.org/10.1002/mgg3.118.

Wiseman, D. H., A. May, S. Jolles, P. Connor, C. Powell, M. M. Heeney, P. J. Giardina, et al. 2013. “A Novel Syndrome of Congenital Sideroblastic Anemia, B-Cell Immunodeficiency, Periodic Fevers, and Developmental Delay (SIFD).” *Blood* 122 (1): 112–23. https://doi.org/10.1182/blood-2012-08-439083.

Wong, Martin C. S., Franklin D. H. Fung, Colette Leung, Wilson W. L. Cheung, William B. Goggins, and C. F. Ng. 2018. “The Global Epidemiology of Bladder Cancer: A Joinpoint Regression Analysis of Its Incidence and Mortality Trends and Projection.” *Scientific Reports* 8 (1). https://doi.org/10.1038/s41598-018-19199-z.

X

Xu, Weiming, Kenneth H. Astrin, and Robert J. Desnick. 1996. “Molecular Basis of Congenital Erythropoietic Porphyria: Mutations in the Human Uroporphyrinogen III Synthase Gene.” *Human Mutation* 7 (3): 187–92. https://doi.org/10.1002/(SICI)1098-1004(1996)7:3<187::AID-HUMU1>3.0.CO;2-8.

Y

Yang, Chengqing, Guoqin Hu, Zezhi Li, Qingzhong Wang, Xuemei Wang, Chengmei Yuan, Zuowei Wang, et al. 2017. “Differential Gene Expression in Patients with Subsyndromal Symptomatic Depression and Major Depressive Disorder.” *PloS One* 12 (3): e0172692. https://doi.org/10.1371/journal.pone.0172692.

Yien, Yvette Y., Sarah Ducamp, Lisa N. van der Vorm, Julia R. Kardon, Hana Manceau, Caroline Kannengiesser, Hector A. Bergonia, et al. 2017. “Mutation in Human CLPX Elevates Levels of δ-Aminolevulinate Synthase and Protoporphyrin IX to Promote Erythropoietic Protoporphyria.” *Proceedings of the National Academy of Sciences* 114 (38): E8045–52. https://doi.org/10.1073/pnas.1700632114.

Yohe, Sophia, and Bharat Thyagarajan. 2017. “Review of Clinical Next-Generation Sequencing.” *Archives of Pathology & Laboratory Medicine* 141 (11): 1544–57. https://doi.org/10.5858/arpa.2016-0501-RA.

Yoshida, Takahiro, Max Kates, Kazutoshi Fujita, Trinity J. Bivalacqua, and David J. McConkey. 2019. “Predictive Biomarkers for Drug Response in Bladder Cancer.” *International Journal of Urology: Official Journal of the Japanese Urological Association*, August. https://doi.org/10.1111/iju.14082.

Z

Zanella, Alberto, Elisa Fermo, Paola Bianchi, and Giovanna Valentini. 2005. “Red Cell Pyruvate Kinase Deficiency: Molecular and Clinical Aspects.” *British Journal of Haematology* 130 (1): 11–25. https://doi.org/10.1111/j.1365-2141.2005.05527.x.

Zhang, R., L. Tian, L-J Chen, F. Xiao, J-M Hou, X. Zhao, G. Li, et al. 2006. “Combination of MIG (CXCL9) Chemokine Gene Therapy with Low-Dose Cisplatin Improves Therapeutic Efficacy against Murine Carcinoma.” *Gene Therapy* 13 (17): 1263–71. https://doi.org/10.1038/sj.gt.3302756.

Zhu, Bin, Nan Song, Ronglai Shen, Arshi Arora, Mitchell J. Machiela, Lei Song, Maria Teresa Landi, et al. 2017. “Integrating Clinical and Multiple Omics Data for Prognostic Assessment across Human Cancers.” *Scientific Reports* 7 (1): 16954. https://doi.org/10.1038/s41598-017-17031-8.

Zmajkovic, Jakub, Pontus Lundberg, Ronny Nienhold, Maria L. Torgersen, Anders Sundan, Anders Waage, and Radek C. Skoda. 2018. “A Gain-of-Function Mutation in EPO in Familial Erythrocytosis.” *The New England Journal o f Medicine* 378 (10): 924–30. https://doi.org/10.1056/NEJMoa1709064.

Zou, Hui, and Trevor Hastie. 2005. “Regularization and Variable Selection via the Elastic Net.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (2): 301–20. https://doi.org/10.1111/j.1467-9868.2005.00503.x.

APPENDIX

APPENDIX

1. Supplementary Material

Tables

Supplementary Material Table S1. Summary of HHD covered by panel 15. (Chapter I)

Supplementary Material Table S2. Summary of FTL identified mutations. (Chapter I)

Supplementary Material Table S3. DNA sequence for FTL RNA fold predictions. (Chapter I)

Supplementary Material Table S4. Biomarkers candidates to predict response to cisplatin treatment (Chapter IV).

Supplementary Material Table S5. Univariant analysis (simple logistic regression) of all genes (Chapter IV).

Figures

Supplementary Material Figure S1. Algorithm for differential diagnosis of Congenital Dyserythropoietic Anemia (CDA). (Chapter II)

Supplementary Material Table S1. Summary of hereditary hematological disorders covered by HHD panel version 15. (Chapter I)

Panel	Disease group	Disease	Acronym	Gen	Inheritance	OMIM
10010	Hereditary hemochromatosis and Hyperferritinemia / hypoferritinemia	Hereditary hemochromatosis Type 1	HFE1	*HFE*	AR	235200
		Hereditary hemochromatosis Type 2	HFE2A	*HFE2*	AR	602390
		Hereditary hemochromatosis Type 2	HFE2B	*HAMP*	AR	613313
		Hereditary hemochromatosis Type 3	HFE3	*TFR2*	AR	604250
		Hereditary hemochromatosis Type 4	HFE4	*SLC40A1*	AD	606069
		Hereditary hemochromatosis Type 5	HFE5	*FTH1*	AD	615517
		Hereditary hemochromatosis		*BMP6*		
		HFE modifier gene		*GNPAT*		
		L-ferritin deficiency (due to *ATP4A*)		*ATP4A*		
		Hyperferritinemia with Cataract Syndrome	HHCS	*FTL*	AD	600886
		Benign Hyperferritinemia		*FTL*	AD	600886
		L-Ferritin deficiency		*FTL*	AD,AR	615604
		Neuroferritinopathy	NCBI3	*FTL*	AD	606159
10020	Congenital sideroblastic anemias and acquired sideroblastic anemia	X-linked sideroblastic anemia due to ALAS2 mutation	XLSA	*ALAS2*	X-linked	300751
		SLC25A38 deficiency	SA	*SLC25A38*	AR	205950
		Thiamine-responsive megaloblastic anemia due to SLC19A2 mutation	TRMA	*SLC19A2*	AR	249270
		X-linked SA with ataxia due to defects in ABCB7	XLSA/A	*ABCB7*	X-linked	301310
		Glutaredoxin 5 deficiency	SA	*GLRX5*	AR	616860
		HSPA9 deficiency	SA	*HSPA9*	AR/D?	600548
		Mitochondrial myopathy with lactic acidosis and sideroblastic 1	MLASA	*PUS1*	AR	600462
		Mitochondrial myopathy with lactic acidosis and sideroblastic 2	MLASA	*YARS2*	AR	613561
		MLASA-like	MLASA	*mt-ATP6*	Maternal or sporadic	500011
		Sideroblastic anemia, B cell immunodeficiency, fevers and developmental	SIFD	*TRNT1*	AR	616084
		LARS2 deficiency	SA	*LARS2*	AR	
		CSA due to defects in NDUFB11	XLSA	*NDUFB11*	X-linked	

		CSA due to mutation in STEAP3	SA	*STEAP3*	AD	
		Acquired sideroblastic anemia, myelodysplastic syndrome with ring sideroblasts		*SF3B1*	Somatic	
10030	Iron and Copper related Anemia	Atransferrinemia		*TF*	AR	209300
		Aceruloplasminemia		*CP*	AR	604290
		Microcytic and hypochromic anemia with iron overload	AHMIO1	*SLC11A2*	AR	206100
		Microcytic and hypochromic anemia with iron overload	AHMIO1	*STEAP3*		
		Iron-refractory Iron-Deficiency Anemia	IRIDA	*TMPRSS6*	AR	206200
		New form of IRIDA	IRIDA	*ACVR1*	AR	
		Wilson disease		*ATP7B*		
10040	Congenital dyserythropoietic anemia	Congenital dyserythropoietic anemia Type I	CDA I	*CDAN1*	AR	224120
		Congenital dyserythropoietic anemia Type I	CDA I	*C15ORF41*	AR	224120
		Congenital dyserythropoietic anemia Type II	CDA II	*SEC23B*	AR	224100
		Congenital dyserythropoietic anemia Type III	CDA III	*KIF23*	AD	105600
		Congenital dyserythropoietic anemia Type IV	CDA IV	*KLF1*	AD	613673
		X-linked thrombocytopenia and dyserythropoietic anemia	XLTDA	*GATA1*	X-linked	300367
10050	Congenital Erythrocytosis/Familiar Polycythemia	Type 1 familial erythrocytosis	ECYT1	*JAK2; SH2B3; EPOR*	AD	133100
		Type 2 familial erythrocytosis	ECYT2	*VHL*	AR	263400
		Type 3 familial erythrocytosis	ECYT3	*EGLN1*	AD	609820
		Type 4 familial erythrocytosis	ECYT4	*EPAS1*	AD	611783
		Type 8 familial erythrocytosis	ECYT8	*BPGM*	AR	222800
10061	Hereditary hemolytic anemias due to RBC enzymopathies or glycogen storage disease	Pyruvate kinase deficiency		*PKLR*	AR	266200
		Hemolytic anemia, G6PD deficient		*G6PD*	XLD	300908
		Hemolytic anemia due to hexokinase deficiency		*HK1*	AR	235700
		Hemolytic anemia, nonspherocytic, due to glucose phosphate isomerase deficiency		*GPI*	AR	613470
		Diphosphoglycerate mutase deficiency of erythrocyte, DPGM Deficiency (Erythrocytosis, familial, 8)		*BPGM*	AR	222800

		Hemolytic anemia due to glutathione reductase deficiency		*GSR*	A codominant	
		Hemolytic anemia due to glutathione synthetase deficiency		*GSS*	AR	231900
		Hemolytic anemia due to glutathione peroxidase deficiency		*GPX1*	AR	614164
		Hemolytic anemia due to gamma-glutamylcysteine synthetase deficiency		*GCLC*	AR	230450
		Anemia, hemolytic, due to UMPH1 deficiency		*NT5C3A*	AR	266120
		Hemolytic anemia due to adenylate kinase deficiency		*AK1*	AR	612631
		Phosphoglycerate kinase 1 deficiency		*PGK1*	XLR	300653
		Hemolytic anemia due to triosephosphate isomerase deficiency		*TPI1*	AR	615512
		Glycogen storage disease VII		*PFKM*	AR	232800
		Glycogen storage disease XII		*ALDOA*	AR	611881
		Hemolytic anemia due to elevated adenosine deaminase		*ADA*		102700
		Methemoglobinemia, type I and type II		*CYB5R3*	AR	613213
10062	Hereditary hemolytic anemias Membranopathies (spherocytosis, elliptocytosis, pyropoikilocytosis, ovalocytosis, stomatocytosis) and Gilbert syndrome	Hereditary spherocytosis Type 1		*ANK1*	AD	182900
		Hereditary spherocytosis Type 2		*SPTB*	AD	616649
		Hereditary spherocytosis Type 3		*SPAT1*	AR	270970
		Hereditary spherocytosis Type 4		*SLC4A1*	AD	612653
		Hereditary spherocytosis Type 5		*EPB42*	AR	612690
		Hereditary elliptocytosis Type 1		*EPB41*	AR,AD	611804
		Hereditary elliptocytosis Type 2		*SPAT1*	AD	130600
		Hereditary elliptocytosis Type 3		*SPTB*	AD	617948
		Hereditary elliptocytosis Type 4 / Ovalocytosis, SA type		*SLC4A1*	AD	166900
		Gilbert syndrome		*UGT1A1*	AR	143500
		Hereditary elliptocytosis by GYPC		*GYPC*	AR, AD	616089
		Dehydrated hereditary stomatocytosis with or without pseudohyperkalemia and/or perinatal edema		*PIEZO1*	AD	194380
		Dehydrated hereditary stomatocytosis 2		*KCNN4*	AD	616689
		Stomatin-deficient cryohydrocytosis with neurologic defects	SDCHCN	*SLC2A1*	AD	608885

		Pseudohyperkalemia, familial, 2, due to red cell leak	PSHK2	*ABCB6*	AD	609153
		Overhydrated hereditary stomatocytosis	OHST	*RHAG*	AD	185000
		Sitosterolemia		*ABCG5*	AR	210250
		Sitosterolemia		*ABCG8*	AR	210250
		McLeod syndrome with or without chronic granulomatous disease		*XK*	X-linked	300842
		Pyropoikilocytosis		*SPTA1*	AR	266140
		Cryohydrocytosis		*SLC4A1*	AD	185020
10080	Erythropoietic Protoporphyria and Congenital Erythropoietic Porphyria	Protoporphyria, erythropoietic, 1	EPP1	*FECH*	AR	177000
		Protoporphyria, erythropoietic, X-linked	XLEPP	*ALAS2*	X-Linked	300752
		Protoporphyria, erythropoietic, 2	EEP2	*CLPX*	AD	618015
		Porphyria, congenital erythropoietic	CEO	*UROS*	AR	263700
10090	Fanconi Anemia	Fanconi anemia, complementation group A	FANCA	*FANCA*	AR	227650
		Fanconi anemia, complementation group B	FANCB	*FANCB*	XL	300514
		Fanconi anemia, complementation group C	FANCC	*FANCC*	AR	227645
		Fanconi anemia, complementation group D1	FANCD1	*BRCA2*	AR	605724
		Fanconi anemia, complementation group D2	FANCD2	*FANCD2*	AR	227646
		Fanconi anemia, complementation group E	FANCE	*FANCE*	AR	600901
		Fanconi anemia, complementation group F	FANCF	*FANCF*	AR	603467
		Fanconi anemia, complementation group G	FANCG	*XRCC9*	AR	614082
		Fanconi anemia, complementation group I	FANCI	*FANCI*	AR	609053
		Fanconi anemia, complementation group J	FANCJ	*BRIP1*	AR	609054
		Fanconi anemia, complementation group L	FANCL	*FANCL*	AR	608111
		Fanconi anemia, complementation group M	FANCM	*FANCM*	AR	
		Fanconi anemia, complementation group N	FANCN	*PALB2*	AR	610832
		Fanconi anemia, complementation group O	FANCO	*RAD51C*	AR	613390
		Fanconi anemia, complementation group P	FANCP	*SLX4*	AR	613951
		Fanconi anemia, complementation group Q	FANCQ	*ERCC4*	AR	615272
		Fanconi anemia, complementation group R	FANCR	*RAD51*	AD	617244

		Fanconi anemia, complementation group S	FANCS	*BRCA1*	AR	617883
		Fanconi anemia, complementation group T	FANCT	*UBE2T*	AR	616435
		Fanconi anemia, complementation group U	FANCU	*XRCC2*	AR	617247
		Fanconi anemia, complementation group V	FANCV	*MAD2L2*	AR	617243
		Fanconi anemia, complementation group W	FANCW	*RFWD3*	AR	617784
10100	Diamond- Blackfan-Anemia	Diamond-Blackfan anemia 1	DBA1	*RPS19*	AD	105650
		Diamond-Blackfan anemia 3	DBA3	*RPS24*	AD	610629
		Diamond-Blackfan anemia 4	DBA4	*RPS17*	AD	612527
		Diamond-Blackfan anemia 5	DBA5	*RPL35A*	AD	612528
		Diamond-Blackfan anemia 6	DBA6	*RPL5*	AD	612561
		Diamond-Blackfan anemia 7	DBA7	*RPL11*	AD	612562
		Diamond-Blackfan anemia 8	DBA8	*RPS7*	AD	612563
		Diamond-Blackfan anemia 9	DBA9	*RPS10*	AD	613308
		Diamond-Blackfan anemia 10	DBA10	*RPS26*	AD	613309
		Diamond-Blackfan anemia 11	DBA11	*RPL26*	AD	614900
		Diamond-Blackfan anemia 12	DBA12	*RPL15*	AD	615550
		Diamond-Blackfan anemia 13	DBA13	*RPS29*	AD	615909
		Diamond-Blackfan anemia 14	DBA14	*TSR2*	XLR	300946
		Diamond-Blackfan anemia 15	DBA15	*RPS28*	AD	606164
		Diamond-Blackfan anemia 16	DBA16	*RPL27*	AD	617408
		Diamond-Blackfan anemia 17	DBA17	*RPS27*	AD	617409
		Diamond-Blackfan anemia		*GATA1*	XLR	
		Diamond-Blackfan anemia		*RPL19*		
		Diamond-Blackfan anemia		*RPL31*	AD	
		Diamond-Blackfan anemia		*RPS15*		
		Diamond-Blackfan anemia		*RPS27A*		
		Diamond-Blackfan anemia		*RPL36*		
		Diamond-Blackfan anemia		*RPL9*		

10110	Dyskeratosis congenita	Dyskeratosis congenita, X-linked	DKCX	*DKC1*	XLR	305000
		Dyskeratosis congenita, autosomal dominant 1	DKCA1	*TERC*	AD	127550
		Dyskeratosis congenita, autosomal dominant 2	DKCA2	*TERT*	AD, AR	613989
		Dyskeratosis congenita, autosomal recessive 4		*TERT*	AD, AR	613989
		Dyskeratosis congenita, autosomal recessive 6		*PARN*	AR	616353
		Dyskeratosis congenita, autosomal recessive 2		*NOLA2*	AR	613987
		Dyskeratosis congenita, autosomal dominant 3		*TINF2*	AD	613990
		Dyskeratosis congenita, autosomal dominant 6		*ACD*	AD, AR	616553
		Dyskeratosis congenita, autosomal recessive 7		*ACD*	AD, AR	616553
		Dyskeratosis congenita, autosomal recessive 1		*NOLA3*	AR	224230
		Dyskeratosis congenita, autosomal recessive 3		*WRAP53*	AR	613988
		Dyskeratosis congenita, autosomal dominant 4	DKCA4	*RTEL1*	AD, AR	615190
		Dyskeratosis congenita, autosomal recessive 5	DKCB5	*RTEL1*	AD, AR	615190
		Dyskeratosis congenita		*USB1*	AR	
		Dyskeratosis congenita		*CTC1*	AR	
		Pantothenate Kinase-associated neurodegeneration (PKAN)		*PANK2*	AR	234200
		PLA2G6-associated neurodegeneration (PLAN)		*PLA2G6*	AR	256600
10120	NBIA, Neurodegeneration with brain iron accumulation	Neuroferritinopathy (NF)	NBIA1	*FTL*	AD	606159
		Mitochondrial membrane protein-associated neurodegeneration (MPAN)	NBIA2	*C19orf12*	AR	614298
		B-propeller-associated neurodegeneration (BPAN)	NBIA3	*WDR45*	X-Linked	300894
		COASY protein-associated neurodegeneration (CoPAN)	NBIA4	*COASY*	AR	615643
		Aceruplasminemia	NBIA5	*CP*	AR	
		Fatty acid hydroxylase-associated neurodegeneration	NBIA6	*FA2H*	AR	
		Kufor-Rakeb disease (KRS)		*ATP13A2*	AR	
		Woodhouse-Sakati syndrome (WSS)		*DCAF17*	AR	

Supplementary Material Table S2. Summary of FTL identified mutations. (Chapter I)

Table summarizing all mutations in *FTL* described up to now in the literature that may cause one of the five diseases. The table shows for each mutation, the conventional nomenclature according to HGVS (corresponding to [NCBI:NM_000146.3] reference sequence), the traditional nomenclature, the position in the IRE structure, and the corresponding published report. NA (not available).

N	Disease	HGVS nomenclature	Mutation type	Mutation position	First Publication	Old Nomenclature
1	**HHCS**	c.-216C>A	IRE regulatory	Promoter *FTL*	(Faniello et al. 2009)	NA
2	HHCS	c.-193C>G + c.-160A>G	IRE regulatory	lower stem + hexanucleotide loop	(Castiglioni et al. 2010)	+7C>G & +40A>G
3	HHCS	c.-190C>T	IRE regulatory	lower stem	(Cremonesi et al. 2003)	+10C>U
4	HHCS	c.-186C>G	IRE regulatory	lower stem	(Cremonesi et al. 2001)	+14C>G
5	HHCS	c.-184C>T	IRE regulatory	lower stem	(Cremonesi et al. 2003)	+16C>U
6	HHCS	c.-182C>T + c.-178T>G	IRE regulatory	lower stem	(Cazzola et al. 1997)	Paiva-2 + 18C>U & 22U>G
7	HHCS	c.-176T>C	IRE regulatory	lower stem	(Rufer et al. 2011)	+24U>C
8	HHCS	c.-171C>G	IRE regulatory	lower stem	(Bosio et al. 2004)	Torino +29C>G
9	HHCS	c.-168G>A	IRE regulatory	lower stem	(Cazzola et al. 1997)	Pavia-1 +32G>A
10	HHCS	c.-168G>C	IRE regulatory	lower stem	((first) Kato and Casella, F 1999)	Baltimore-1 +32G>C
11	HHCS	c.-168G>T	IRE regulatory	lower stem	(Martin et al. 1998)	Paris-2 or Milano-1 +32G>U
12	HHCS	c.-167C>A	IRE regulatory	C bulge	(Durupt et al. 2001)	Paris +33C>A
13	HHCS	c.-167C>T	IRE regulatory	C bulge	(Balas et al. 1999)	Madrid or Philadelphia +33C>U
14	HHCS	c.-166T>C	IRE regulatory	upper stem	(Hetet et al. 2003)	Paris +34U>C
15	HHCS	c.-164C>A	IRE regulatory	upper stem	(Mumford et al. 1998)	London-2 +36C>A
16	HHCS	c.-164C>G	IRE regulatory	upper stem	(Cremonesi et al. 2003)	Milano +36C>G
17	HHCS	c.-164C>T	IRE regulatory	upper stem	(Sara Luscieti et al. 2013)	Badalona +36C>U

18	HHCS	c.-163A>C	IRE regulatory	upper stem	(Ferrari et al. 2006)	Pavia +37A>C
19	HHCS	c.-163A>G	IRE regulatory	upper stem	(Cremonesi et al. 2003)	Milano +37A>G
20	HHCS	c.-163A>T	IRE regulatory	upper stem	(Garcia Erce et al. 2006)	Zaragoza +37A>U
21	HHCS	c.-161C>A	IRE regulatory	hexanucleotide loop	(McLeod et al. 2002)	Geelong +39C>A
22	HHCS	c.-161C>G	IRE regulatory	hexanucleotide loop	(Garderet et al. 2004)	Paris +39C>G
23	HHCS	c.-161C>T	IRE regulatory	hexanucleotide loop	(Mumford et al. 1998)	London-1 +39C>U
24	HHCS	c.-160A>G	IRE regulatory	hexanucleotide loop	(Beaumont et al. 1995)	Paris-1 or Montpellier-1 +40A>G
25	HHCS	c.-160A>G + c.-159G>C	IRE regulatory	hexanucleotide loop	(Cremonesi et al. 2001)	Paris-1 or Montpellier-1 +40A>G & Verona-1 +41G>C
26	HHCS	c.-159G>C	IRE regulatory	hexanucleotide loop	(Meneses et al. 2011)	Verona-1 +41G>C
27	HHCS	c.-157G>A	IRE regulatory	hexanucleotide loop	(Phillips, Warby, and Kushner 2005)	Salt Lake City +43G>A
28	HHCS	c.-154T>G	IRE regulatory	upper stem	(Messa et al. 2009)	+46U>G
29	HHCS	c.-153G>A	IRE regulatory	upper stem	(Hetet et al. 2003)	Paris +47G>A
30	HHCS	c.-151A>C	IRE regulatory	lower stem	(Castiglioni et al. 2010)	+49A>C
31	HHCS	c.-151A>G	IRE regulatory	lower stem	(Sompele et al. 2017)	Ghent +49A > G
32	HHCS	c.-150C>A	IRE regulatory	lower stem	(Gonzalez-Huerta et al. 2008)	+50C>A
33	HHCS	c.-149G>C	IRE regulatory	lower stem	(Camaschella, Zecchina, et al. 2000)	Torino +51G>C
34	HHCS	c.-148G>C	IRE regulatory	lower stem	(Sara Luscieti et al. 2013)	Heidelberg +52G>C
35	HHCS	c.-144A>T	IRE regulatory	lower stem	(Ferrari et al. 2006)	Paris +56A>U
36	HHCS	c.-110C>T	IRE regulatory	5’ UTR	(Cremonesi et al. 2003)	+90C>U

37	HHCS	c.-220_-196del25	IRE regulatory	new transcription starting site (resulting IRE lacks nt 1-24)	(Burdon et al. 2007)	NA
38	HHCS	c.-190-162del29	IRE regulatory	eliminating IRE	(Girelli et al. 1997)	Verona-2 +10_38del29
39	HHCS	c.-182_-174delCGGGTCTGTinsAGGGGCCGG [$]	IRE regulatory	eliminating part of lower stem	(Lenzhofer et al. 2015)	+18_+26 delCGGGTCTGTinsAGGGGCCGG
40	HHCS	c.-178_-173del6	IRE regulatory	eliminating part of lower stem	(Cazzola et al. 2002)	+22_27del6
41	HHCS	c.-168_-165delGCTT	IRE regulatory	eliminating C bulge	(Garber and Pudek 2014)	+32_35delGCTT
42	HHCS	c.-164_158del7	IRE regulatory	eliminating part of hexanucleotide loop	**(Cadenas et al. 2019)**	Esplugues +36_42del7
43	HHCS	c.-161delC	IRE regulatory	eliminating IRE	(Muñoz-Muñoz et al. 2013)	+39delC
44	HHCS	c.-162_-161delCA	IRE regulatory	eliminating part of hexanucleotide loop	(Giansily et al. 2001)	+38_39del AC
45	HHCS	c.-158_-143del16	IRE regulatory	eliminating part of hexanucleotide loop	(Giansily et al. 2001)	+42_57del16
46	HHCS	c.-153_-152delGGinsCT	IRE regulatory	eliminating part of upper stem	(Mattila et al. 2018)	+47_48delGGinsCT
47	HHCS	c.-44delT	IRE regulatory	eliminating IRE	(Cremonesi et al. 2003)	
1	**Neuroferritinopathy**	c.[474G>A]; p.(Ala96Thr)	Missense	exon3	(Maciel et al. 2005)	
2	Neuroferritinopathy	c.641_642 4bp_dup	Frameshift	exon 4	(Kubota et al. 2009)	
3	Neuroferritinopathy	c.646InsC	Frameshift	exon 4	(Mancuso et al. 2005)	
4	Neuroferritinopathy	c.458dupA	Frameshift	exon 4	(Devos et al. 2009)	
5	Neuroferritinopathy	c.460InsA	Frameshift	exon 4	(Curtis et al. 2001)	
6	Neuroferritinopathy	c.468_483 dup16nt	Frameshift	exon 4	(Nishida et al. 2014)	
7	Neuroferritinopathy	c.469_484 dup16nt	Frameshift	exon 4	(Ohta et al. 2008)	
8	Neuroferritinopathy	c.498InsTC	Frameshift	exon 4	(R. Vidal et al. 2004)	
9	Neuroferritinopathy	c.468dupT	Frameshift	exon 4	(Moutton et al. 2014)	

10	Neuroferritinopathy	c.467_470dupGTGG	Frameshift	exon 4	(Ni et al. 2016)
1	**Benign Hyperferritinemia**	c.[77A<T]; p.(Gln26Leu)	Missense	exon 1	(Thurlow et al. 2012)
2	Benign Hyperferritinemia	c.[80C>T]; p.(Ala27Val)	Missense	exon 1	(Thurlow et al. 2012)
3	Benign Hyperferritinemia	c.[89C>T]; p.Thr30Ile	Missense	exon 1	(C. Kannengiesser et al. 2009)
1	**L-ferritin deficiency Dominant**	c[1A>G]; p.(M1V)	Missense	start codon	(Cremonesi et al. 2004)
2	L-ferritin deficiency Dominant	c.375+2T>A	Splicing	intronic	**(Cadenas et al. 2019)**
1	**L-ferritin deficiency Recessive**	c.[310G>T];p. (E104X]	Nonsense	exon3	(Cozzi et al. 2013)

NOTES:

[$]Previously reported by HGMD as c.-182_-176delCGGGTCTinsAGGGGCC, correct: c.-182_-174delCGGGTCTGTinsAGGGGCCGG

Supplementary Material Table S3. DNA sequence for FTL RNA fold predictions. (Chapter I)

WT-IRE	5’-GCAGTTCGGCGGTCCCGCGGGTCTGTCTCTTGCTTCAACAGTGTTTGGACGGAACAGATCC GGGGACTCTCTTCC-3’
Mut-IRE	5’-GCAGTTCGGCGGTCCCGCGGGTCTGTCTCTTGCTTGTTTGGACGGAACAGATCCGGGGACT CTCTTCC-3’

Supplementary Material Table S4. Biomarkers candidates to predict response to cisplatin treatment (Chapter IV).

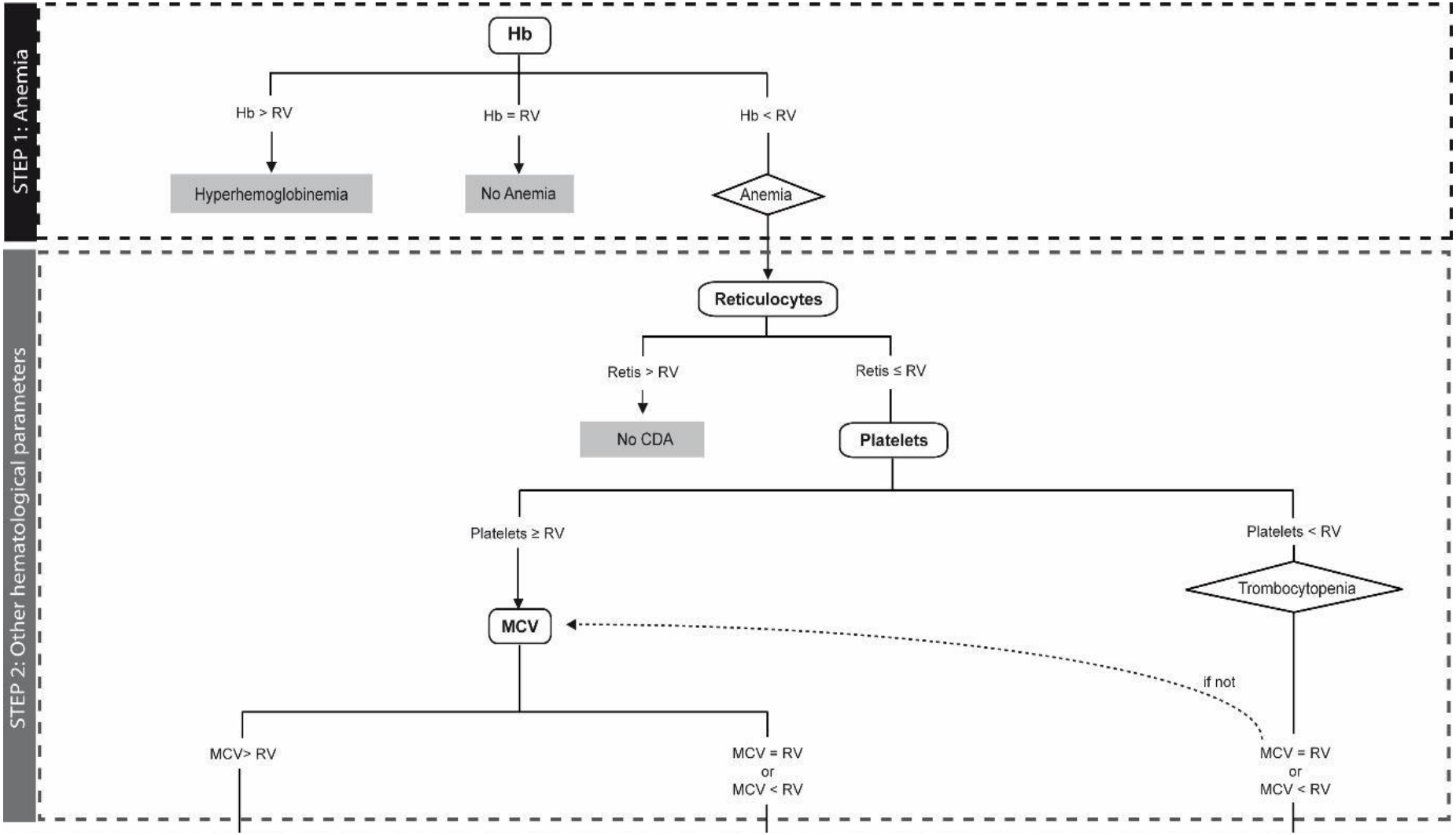
Pathway/mechanism		Gene
DNA repair pathways	Nucleotide excision repair (NER)	*ERCC1*
		ERCC2
		ERCC5
		ERCC6
		eIF3a
	Non-homologous end joining (NHEJ)	*Ku80*
		CHEK1
		CHEK2
		REV7/ MAD2L2
		PARP
		Rif1
	Fanconi Anemia	*FANCC*
		FANCA
		FANCD2
	Homologous recombination	*BRCA1*
		BRCA2
		NBN
		PALB2
		RAD50
		RAD51
		POLϑ
		PTIP
		RNF8
		RNF168
		HERC2
		53BP1
		ATR
		ATM
Genes of therapeutic targets and chromatin remodeling	Tyrosine kinase receptor	*Axl*
		c-MET
		RON
	chromatin remodeling	*TERT*
Taxonomic classification genes	Luminal	*FOXA1*
		GATA3
	Basal	*KRT 5/6*
		KRT14
Immune response pathway	Apoptosis	*PD1*
		PDL1
	chemokines	*CXCL9*
	INF pathway	*INFG*
	Inhibitors receptors / ligands	*LAG3*

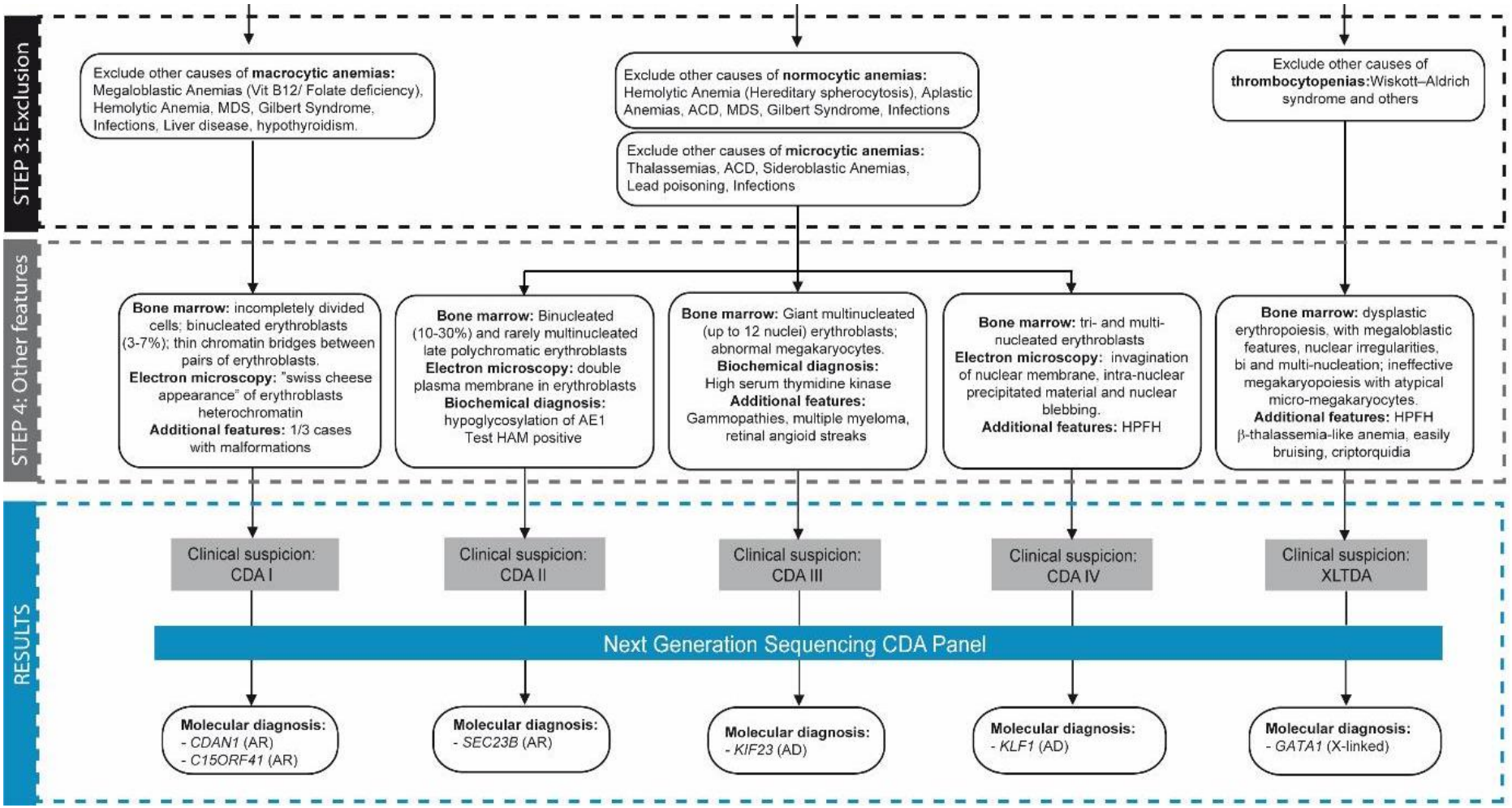
Supplementary Material Table S5. Univariant analysis (simple logistic regression) of all genes (Chapter IV).

Gene	β Cofactor	Std. Error	z value	Pr(>|z|)
53BP1	0.274	0.172	1.596	0.111
ATM	0.045	0.122	0.370	0.711
ATR	0.325	0.316	1.028	0.304
Axl	-0.098	0.142	-0.686	0.493
BRCA1	-0.083	0.127	-0.651	0.515
BRCA2	-0.079	0.123	-0.641	0.522
CHEK1	-0.121	0.128	-0.943	0.346
CHEK2	-0.016	0.121	-0.135	0.892
CXCL9	-0.209	0.093	-2.254	0.024
ERCC1	0.212	0.248	0.856	0.392
ERCC2	0.177	0.129	1.375	0.169
ERCC5	0.057	0.133	0.425	0.671
ERCC6	-0.035	0.142	-0.247	0.805
FANCA	-0.033	0.131	-0.252	0.801
FANCD2	0.032	0.119	0.272	0.786
FOXA1	0.105	0.080	1.310	0.190
GATA3	-0.007	0.093	-0.080	0.936
HERC2	0.359	0.236	1.520	0.129
KRT14	-0.037	0.056	-0.667	0.505
KRT5/6	0.016	0.060	0.259	0.795
Ku80	-0.170	0.399	-0.426	0.670
LAG3	-0.090	0.107	-0.848	0.396
NBN	0.000	0.109	-0.004	0.997
PALB2	-0.027	0.124	-0.220	0.826
PARP	-0.440	0.319	-1.381	0.167
PDL1	-0.096	0.146	-0.660	0.509
POL theta	-0.027	0.112	-0.242	0.809
PTIP	0.004	0.117	0.034	0.973
RAD50	0.074	0.109	0.675	0.500
RAD51	-0.343	0.185	-1.858	0.063
REV7 (MAD2L2)	0.105	0.112	0.937	0.349
RNF168	0.146	0.129	1.126	0.260
RNF8	-0.073	0.161	-0.454	0.650
RON	0.052	0.100	0.514	0.607
Rif1	0.050	0.151	0.334	0.739
TERT	-0.186	0.150	-1.240	0.215
c-Met	0.000	0.137	0.002	0.998
eIF3a	0.079	0.154	0.514	0.607

Supplementary Material Figure S1. Algorithm for differential diagnosis of Congenital Dyserythropoietic Anemia (CDA). (Chapter II)

This figure is also showed in chapter II (Figure 2.2). Here, the imagen has been zoomed in and divided in two pages for better visualization. The algorithm is structured in five steps. (1) Anemia, the hemoglobin levels are considered for the presence of anemia. (2) Other hematological parameters are evaluated. (3) Exclusion of other diseases (4) Other clinical features are examined. (5) The clinical suspicious of CDA is shown, including the gene that is involved in the CDA subtype.




STEP 3: Exclusion
Exclude other causes of **macrocytic anemias:** Megaloblastic Anemias (Vit B12/ Folate deficiency), Hemolytic Anemia, MDS, Gilbert Syndrome, Infections, Liver disease, hypothyroidism.
Exclude other causes of **normocytic anemias:** Hemolytic Anemia (Hereditary spherocytosis), Aplastic Anemias, ACD, MDS, Gilbert Syndrome, Infections
Exclude other causes of **microcytic anemias:** Thalassemias, ACD, Sideroblastic Anemias, Lead poisoning, Infections
Exclude other causes of **thrombocytopenias:** Wiskott–Aldrich syndrome and others
STEP 4: Other features
Bone marrow: incompletely divided cells; binucleated erythroblasts (3-7%); thin chromatin bridges between pairs of erythroblasts. **Electron microscopy:** "swiss cheese appearance" of erythroblasts heterochromatin **Additional features:** 1/3 cases with malformations
Bone marrow: Binucleated (10-30%) and rarely multinucleated late polychromatic erythroblasts **Electron microscopy:** double plasma membrane in erythroblasts **Biochemical diagnosis:** hypoglycosylation of AE1 Test HAM positive
Bone marrow: Giant multinucleated (up to 12 nuclei) erythroblasts; abnormal megakaryocytes. **Biochemical diagnosis:** High serum thymidine kinase **Additional features:** Gammopathies, multiple myeloma, retinal angioid streaks
Bone marrow: tri- and multi-nucleated erythroblasts **Electron microscopy:** invagination of nuclear membrane, intra-nuclear precipitated material and nuclear blebbing. **Additional features:** HPFH
Bone marrow: dysplastic erythropoiesis, with megaloblastic features, nuclear irregularities, bi and multi-nucleation; ineffective megakaryopoiesis with atypical micro-megakaryocytes. **Additional features:** HPFH β-thalassemia-like anemia, easily bruising, criptorquidia
RESULTS
Clinical suspicion: CDA I
Clinical suspicion: CDA II
Clinical suspicion: CDA III
Clinical suspicion: CDA IV
Clinical suspicion: XLTDA
Next Generation Sequencing CDA Panel
Molecular diagnosis: - *CDAN1* (AR) - *C15ORF41* (AR)
Molecular diagnosis: - *SEC23B* (AR)
Molecular diagnosis: - *KIF23* (AD)
Molecular diagnosis: - *KLF1* (AD)
Molecular diagnosis: - *GATA1* (X-linked)

2. National and International congress

2.1. Congress presentations

First Author:

"Clinical cases solved by NGS"
Beatriz Cadenas; Josep Fita; Alex Negro; Mayka Sánchez
EHA-SWG Scientific Meeting on Anemias: Diagnosis and Treatment in the Omics Era, Barcelona, Spain 3 February 2017. Oral presentation by Mayka Sanchez.

"Mutations in YARS2 cause congenital sideroblastic anaemia without showing evidences of myopathy and lactic acidosis"
Beatriz Cadenas; Josep Fita; Leonor Arenillas; Sara Montesdeoca; Laia Mártinez; Carme Pedro; Mayka Sánchez.
22nd Congress European Hematology Association (congress), Madrid, Spain,22-25 June 2017. Poster presentation.

"BEA: a web tool for BioMark gene expression analysis"
Cadenas, Beatriz; Cebrián, Isaac; Sanchez, Edgar; Calle, M. Luz; Tornador, Cristian; Sanchez, Mayka.
European Human Genetics Conference 2018, Milan, Lombardia, Italia 16-19 June 2018. Poster presentation.

"BEA: a web tool for BioMark gene expression analysis"
Cadenas, Beatriz; Cebrián, Isaac; Sanchez, Edgar; Calle, M. Luz; Tornador, Cristian; Sanchez, Mayka.
VI Jornada de Bioinformàtica i Genòmica, Spain 20 December 2018. Poster presentation.

L-Ferritin: One Gene, Five Diseases; from Hereditary Hyperferritinemia to Hypoferritinemia - Report of New Cases
Beatriz Cadenas, Ferran Celma, Josep Fita-Torro, Mar Bermúdez-Cortés, Inés Hernandez-Rodriguez , José L Fuster, Mª Esther Llinares, Ana Mª Galera, Julia L Romero, Santiago Pérez-Montero, Cristian Tornador, Mayka Sánchez.
European Human Genetics Conference 2019, Gothenburg, Sweden, 15-18 June 2018. Poster presentation.

Collaboration:

"DIAGNÓSTICOS EN ENFERMEDADES HEMATOLÓGICAS HEREDITARIAS MEDIANTE PANELES DE GENES DE NEXT GENERATION SEQUENCING (NGS)"
Mayka Sánchez-Fernandez; Josep Fita-Torró; **Beatriz Cadenas-Sevilla;** Clara Esteban-Jurado; David Beneitez-Pastor; Jose Luis Dapena; Adoración Blanco-Alvarez; Victor Marco Betes; José Luis Fuster; Mar Bermúdez-Cortés; Leonor Arenillas; Laila Martínez; Sara Montesdeoca; Carme Pedro; Santiago Pérez-Montero; Inés Hernández-Rodriguez; Cristian Esteban-Jurado.
LIX Congreso Nacional de la SEHH y XXXIII Congreso Nacional de la SETH, Malaga, Andalucía, España 26-28 October 2017. Oral presentation by Mayka Sanchez.

Clinical diagnosis by next generation sequencing of hereditary haematological diseases
Sònia Sabaté-Soler, Josep Fita-Torró, **Beatriz Cadenas-Sevilla**, Clara Esteban-Jurado, David Beneitez-Pastor, Jose Luis Dapena, Adoración Blanco-Alvarez, Eugènia Rivero-Arango, José Luis Fuster, Mar Bermúdez-Cortés, Leonor Arenillas, Sara Montesdeoca, Laia Martínez, Carme Pedro, Salvador Payan-Perna, Isabel Badell-Serra, Angel Remacha, Santiago Pérez-Montero, Inés Hernández-Rodriguez, Cristian Tornador-Antolín, Mayka Sanchez.
European Iron Club, Zurich, 8-11 February 2018. Poster presentation by Mayka Sánchez

"Diagnosis of hereditary Haematological diseases using next generation sequencing panels."
Laura Zalba; Josep Fita; **Beatriz Cadenas**; David Beneitez; Jose Luis Dapena; Adoración Blanco; Eugènia Rivero; Mar Bermúdez; Leonor Arenillas; Sara Montesdeoca; Laia Martínez; Carme Pedro; Salvador Payan; Jose Luis Fuster; Isabel Badell; Angel Remacha; Santiago Pérez; Inés Hernández; Cristian Tornador; Mayka Sanchez.
23nd Annual Congress of the European Hematology Association (EHA) Stockholm, Norway 14-17 June 2018. Poster presentation by Mayka Sanchez.

"L-Ferritin: One Gene, Five Diseases; from Hereditary Hyperferritinemia to Hypoferritinemia - Report of New Cases"
Ferran Celma, **Beatriz Cadenas**, Josep Fita-Torro, Mar Bermúdez-Cortés, Inés Hernandez-Rodriguez, José L Fuster, Mª Esther Llinares, Ana Mª Galera, Julia L Romero, Santiago Pérez-Montero, Cristian Tornador, Mayka Sánchez
8th Congress of the International BioIron Society, EMBL Heidelberg, Germany 5 - 10 May 2019.
Poster presented by Ferran Celma

2.2. Attendance to Courses and congress

Com presentar el Pla d'Investigació (PI) en salut (course)
9 February 2017
UVIC, Vic, Spain

The EHA-SWG Scientific Meeting on Anemias: Diagnosis and Treatment in the Omics Era (congress)
2-4 February 2017
Barcelona, Spain

OPEN MULTISCALE SYSTEMS MEDICINE (OpenMultiMed)Training School (course)
21-23 February 2017
Porto, Portugal.

Illumina Training for MiniSeq Systen (course)
10-11 April 2017
Esplugues de Llobregat, Spain

22nd Congress European Hematology Association (congress)
22-25 June 2017
Madrid, Spain

NGS Course; next-generation sequencing in a diagnostic setting (course)
25 -27th September 2017
Ljubljana, Slovenia

Developing skills in scientific writing (course)
4-5 October 2017
Institute Josep Carreras, Badalona 2017

Cross-disciplinary competences of the Plan of INDUSTRIAL DOCTORS 2017 (course)
20 October – 24 November 2017
Barcelona, Spain

V JORNADA DE BIOINFORMÀTICA I GENÒMICA (congress)
20 December 2017
Barcelona, Spain

3rd Open MultiMed Training School. Multi-scale and Multi-level Modelling Methodologies in Biomedicine (course)
21-23 February 2018
Erlangen, Germany

Participation in the QuantStudio12K Flex Real-Time PCR System and Open Array Training (course)
27-28 March 2018
Esplugues de Llobregat, Spain

European Human Genetics Conference Milan 2018 European Society of Human Genetics ESHG congress)
16-19 June 2018
Milan, Italy

VI JORNADA DE BIOINFORMÀTICA I GENÒMICA (congress)
20 December 2018
Barcelona, Spain

II Jornadas de Investigación Traslacional en Tumores Urológicos: del laboratorio a la práctica clínica
31 Januatry – 1 February 2019
Badalona, Spain

European Society of Human Genetics Conference (congress)
15-18 June 2019
Gothenburg, Sweden

3. Publications

Results presented in this dissertation (chapter 1 and 2) are part of the following published manuscripts. Some text passages of those manuscripts have been adapted in this dissertation.

From chapter 1:

L-Ferritin: One Gene, Five Diseases; from Hereditary Hyperferritinemia to Hypoferritinemia-Report of New Cases.
Cadenas B, Fita-Torró J, Bermúdez-Cortés M, Hernandez-Rodriguez I, Fuster JL, Llinares ME, Galera AM, Romero JL, Pérez-Montero S, Tornador C, Sanchez M.
Pharmaceuticals 2019, 12, 17; doi:10.3390/ph12010017
ISSN:1424-8247
CiteScore 2017: 4.20
SJR 2017: 1.293
SNIP 2017: 1.370
(see page 193)

From chapter 2:

CoDysAn: A telemedicine tool to improve awareness and diagnosis for patients with Congenital Dyserythropoietic Anemia
Cristian Tornador, Edgar Sánchez-Prados, **Beatriz Cadenas**, Roberta Russo, Veronica Venturi, Immacolata Andolfo, Ines Hernández-Rodriguez, Achille Iolascon, Mayka Sánchez*
Frontiers in Physiology - Red Blood Cell Physiology section; doi: 10.3389/fphys.2019.01063
Accepted: 02 August 2019
Impact factor: 3.201
CiteScore: 3.22
(see page 209)

pharmaceuticals

MDPI

Article

L-Ferritin: One Gene, Five Diseases; from Hereditary Hyperferritinemia to Hypoferritinemia—Report of New Cases

Beatriz Cadenas [1,2,3], Josep Fita-Torr ó [4], Mar Berm údez-Cort és [5], In és Hernandez-Rodriguez [6], José Luis Fuster [5], Mar ía Esther Llinares [5], Ana Mar ía Galera [5], Julia Lee Romero [7], Santiago P érez-Montero [4], Cristian Tornador [1,4] and Mayka Sanchez [4,8,9,*]

1 Whole Genix SL., 08021 Barcelona, Spain; Beatriz.cadenas@wholegenix.com (B.C.); cristian.tornador@wholegenix.com (C.T.)
2 Iron Metabolism: Regulation and Diseases Group, Josep Carreras Leukemia Research Institute (IJC), Campus Can Ruti, Badalona, 08916 Barcelona, Spain
3 Experimental Sciences and Technology Department, Universitat de Vic-Universitat Central de Catalunya, 08500 Vic, Spain
4 BloodGenetics SL, Esplugues de Llobregat, 08950 Barcelona, Spain; josepfita@hotmail.com (J.F.-T.); sperez@bloodgenetics.com (S.P.-M.)
5 Pediatric OncoHematology Service, Clinic University Hospital Virgen de la Arrixaca, Instituto Murciano de Investigaci ón Biosanitaria (IMIB), 30120 Murcia, Spain; mariam.bermudez2@carm.es (M.B.-C.); josel.fuster@carm.es (J.L.F.); mariae.llinares@carm.es (M.E.L.); anam.galera@carm.es (A.M.G.)
6 Hematology Service, University Hospital Germans Trias i Pujol (HGTiP), Institut Catal à d'Oncologia (ICO), Badalona, 08916 Barcelona, Spain; agnesrh@iconcologia.net
7 Biomedical Engineering Department, University of Texas at Austin, Austin, TX 78712, USA; julialromero@utexas.edu
8 Program of Predictive and Personalised Medicine of Cancer (PMPPC), Institut d'Investigaci ó Germans Trias i Pujol (IGTP), Campus Can Ruti, Badalona, 08916 Barcelona, Spain
9 Iron Metabolism: Regulation and Diseases Group, Faculty of Medicine and Health Sciences, Universitat Internacional de Catalunya (UIC), 08195 Barcelona, Spain
* Correspondence: msanchezfe@uic.es; Tel.: +34-935-042-000

Received: 23 December 2018; Accepted: 19 January 2019; Published: 23 January 2019

Abstract: Ferritin is a multimeric protein composed ofl ight (L-ferritin) and heavy (H-ferritin) subunits that binds and stores iron inside the cell. A variety of mutations have been reported in the L-ferritin subunit gene (FTL gene) that cause the following five diseases: (1) hereditary hyperferritinemia with cataract syndrome (HHCS), (2) neuroferritinopathy, a subtype of neurodegeneration with brain iron accumulation (NBIA), (3) benign hyperferritinemia, (4) L-ferritin deficiency with autosomal dominant inheritance, and (5) L-ferritin deficiency with autosomal recessive inheritance. Defects in the FTL gene lead to abnormally high levels of serum ferritin (hyperferritinemia) in HHCS and benign hyperferritinemia, while low levels (hypoferritinemia) are present in neuroferritinopathy and in autosomal dominant and recessive L-ferritin deficiency. Iron disturbances as well as neuromuscular and cognitive deficits are present in some, but not all, of these diseases. Here, we identified two novel FTL variants that cause dominant L-ferritin deficiency and HHCS (c.375+2T > A and 36_42delCAACAGT, respectively), and one previously reported variant (Met1Val) that causes dominant L-ferritin deficiency. Globally, genetic changes in the FTL gene are responsible for multiple phenotypes and an accurate diagnosis is useful for appropriate treatment. To help in this goal, we included a diagnostic algorithm for the detection of diseases caused by defects in FTL gene.

Keywords: ferritin; hereditary hyperferritinemia; hereditary hypoferritinemia; iron metabolism; cataracts syndrome; neurodegenerative disease

1. Introduction

Ferritin is an iron-binding protein that stores and releases iron and thus contributes to maintaining and controlling iron homeostasis. Iron is stored in ferritin in the Fe^{+3} form and released in the Fe^{+2} form. Tissue ferritin is a multimeric protein formed from the assembly of 24 peptide subunits, known as the light (L) and heavy (H) ferritin subunits. The H subunit has ferroxidase activity and converts iron from Fe^{+2} to Fe^{+3}, which enables iron storage; it also serves to regulate pH and increase concentrations off ree radicals in the body, which can be extremely damaging to cellular structures and proteins [1]. The L-ferritin subunit helps with electron transport in and out of the ferritin core protein and plays a role in iron release, as Fe^{+2}, from ferritin. Unlike tissue ferritin, serum ferritin is partially glycosylated and nearly completely iron-free [2,3] and is mainly composed of L-ferritin subunits [4–6].

Focusing on the L-ferritin (FTL) gene, five diseases have been identified as directly resulting from mutations in this gene. These diseases include hereditary hyperferritinemia with cataract syndrome (HHCS), neuroferritinopathy, benign hyperferritinemia (or hyperferritinemia without iron overload), autosomal dominant L-ferritin deficiency and autosomal recessive L-ferritin deficiency.

HHCS (OMIM#600886, ORPHA163) is associated with mutations located in the iron responsive element (IRE) at the 5′ untranslated region (UTR) of the L-ferritin mRNA, which results in the disruption of binding with iron regulatory proteins (IRP1 and 2), this is known as the IRP-IRE post-transcriptional regulatory system [7–9]. Mutations in this RNA motif result in the loss of ferritin translation repression and excess ferritin production, even though iron levels remain normal. Ferritin overproduction leads to deposits in the lens of the eye, resulting in the development of cataracts [10]. Up to now, there are 47 known mutations associated with HHCS [11].

In 2001, Curtis and collaborators [12] described for the first time neuroferritinopathy (OMIM#606159, ORPHA:157846), an autosomal dominant condition characterized by normal to low serum ferritin levels, progressive chorea or dystonia, and subtle cognitive deficits. Neuroferritinopathy is classified as a member of the group of diseases known as neurodegeneration with brain iron accumulation (NBIA). So far, there have been ten reported mutations causing this condition, mostly located at the C-terminal region of the FLT gene [13]. FTL mutations diminish ferritin's ability to store iron so, in an attempt to control free iron levels, neurons produce more ferritin, resulting in iron and ferritin accumulation in the basal ganglia of the brain and leading to movement and cognitive disabilities [14].

Benign hyperferritinemia or genetic hyperferritinemia without iron overload (OMIM#600886, ORPHA:254704) is another FTL mutated disorder where patients have high (greater than 90%) glycosylated serum ferritin levels. There are three known mutations in the N-terminal region of the FTL gene that cause benign hyperferritinemia [15]. Despite serum ferritin hyperglycosylation, no other harmful effects have been detected in patients with this disorder [16].

Finally, two variants in FTL have been reported causing L-ferritin deficiency, i.e., hypoferritinemia (OMIM#615604, ORPHA:440731). Mutation Glu104Ter was described in a single patient with inheritance in autosomal recessive mode and consists of a G > C nucleotide substitution in exon 3 (c.310G > T). This mutation causes a complete lack of translation of the FTL gene with subsequently undetectable levels of serum ferritin. This patient presented with seizures and restless leg syndrome [17]. The FTL mutation Met1Val, resulting in a change at the start codon (c.-1A > G), has also been described in a single case; however, this mutation was inherited in an autosomal dominant manner. The patient presented with decreased levels of serum ferritin, but no history ofi ron deficiency anemia or neurologic dysfunction [18].

In this study, we report the identification of two novel mutations in the FTL gene detected by gene sequencing. One mutation is associated with a diagnosis of HHCS and the other with a diagnosis of dominant L-ferritin deficiency. We also describe an additional dominant L-ferritin deficiency case with a previously described (Met1Val) mutation in the FTL gene. Moreover, we have performed an extensive review of all reported variants in the FTL gene linked with the previously described five conditions to help in the understanding of the phenotypes.

2. Results

We completely sequenced the entire coding region, intron-exon boundaries and 5′ and 3′ regulatory regions for the FTL gene either by Sanger sequencing or by next generation sequencing (NGS).

2.1. Case Studies

2.1.1. Family 1—A Case with Autosomal Dominant L-Ferritin Deficiency

Proband II.1 from family 1 (Figure 1A and Table 1) is a four-year-old female of Spanish origin referred to the department of Pediatric OncoHematology Service of the Clinic University Hospital Virgen de la Arrixaca because of refractory hypoferritinemia (serum ferritin 4–9 ng/mL) unresponsive to oral iron supplementation, without any accompanying sign or symptoms. Physical examination was normal with normal weight and size for age. At the age of six, the proband complained of recurrent severe headaches. Cerebral CT and MRI were normal except for a small (non-specific) subcortical area of gliosis in the right frontal lobe. This was considered an incidental finding by pediatric neurologists who established a provisional diagnosis of primary headache and recommended treatment with flunarizine.

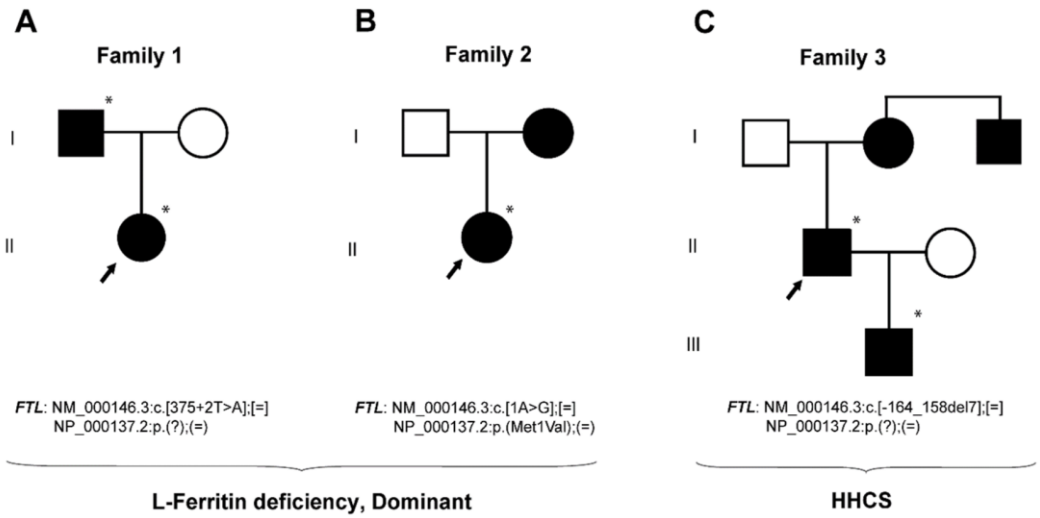


Figure 1. Pedigree trees from three studied families affected from dominant L-ferritin deficiency and HHCS. Squares indicate males and circles females. Probands are pointed with an arrow. Filled symbols indicate affected members and asterisks indicate subjects with genetic studies done at BloodGenetics SL. Mutations are named according the HGVS nomenclature.

Mutation analysis revealed the presence of T > A change in the intron 3 of the FTL gene, at position 375 + 2 (NM_000146.3:c.[375 + 2T > A];[=], HGSV nomenclature). This variant was also found in a heterozygous state in the father of the proband (I.1). Human Splicing Finder software predicted that this variant alters the wild type splicing donor site, affecting mRNA splicing. This mutation is novel and has not been previously reported either in the literature or in any public database (ENSEMBL, NCBI, 1000Genomes, public HGMD). However, another variant without a known clinical significance exists at the same position, c.375 + 2T > C, reported in the SNP database as rs1371561306 and with a very low allele frequency (MAF = 0.000008 reported in TOPMED database).

Table 1. Genetic and clinical features of the probands.

Case	Family 1	Family 2	Family 3	Reference Values
Patient	II.1	II.1	II.1	-
Gender	F	F	M	-
Age at diagnosis (years)	4	2	67	-
Hb (g/dL)	12.2–13.3	13.1–13.7	14.0	13.5–17.5 (M). 12.1–15.1 (F)
MCV (fL)	78–84	80	90.2	80–95
Ferritin (ng/mL)	4–9	2–7	3037	12–300 (M), 12–200 (F)
Transferrin sat (%)	12.9	17.2–26.2	22.0–41.0	25–50
Iron (µL/dL)	n/a	61.95	46	49–226
Mutation	c.375 + 2T > A	p.Met1Val	c.-164_158del7	-
	Novel	Previously reported [18]	Novel	-
Disease	L-ferritin deficiency	L-ferritin deficiency	HHCS	-
Inheritance	AD	AD	AD	-

The following abbreviations were used: HB, hemoglobin; MCV, mean corpuscular volume; TF sat, transferrin saturation; F, female; M, male; n/a, not available. The mutation nomenclature used follows the HGVS guidelines.

2.1.2. Family 2—A Case with Autosomal Dominant L-Ferritin Deficiency

Proband II.1 from family 2 (Figure 1B and Table 1) is an asymptomatic two-year-old girl evaluated at the department of Pediatric OncoHematology Service of the Clinic University Hospital Virgen de la Arrixaca for further investigation of mild neutropenia and eosinophilia. The patient had been previously diagnosed with a small ventricular septal defect. During her follow-up, we did not find neutropenia or other hematological anomalies, but rather marked hypoferritinemia without anemia. Hypoferritinemia was unresponsive to oral iron supplementation. She occasionally complained of mild asthenia and an occasional mild headache, which were found to be tension headaches after a full evaluation by a pediatric neurologist. The proband ´s mother had low serum ferritin levels (<6 ng/mL), low transferrin saturation (9.6%) with normal levels of transferrin.

Sequencing analysis of this proband (Figure 1B, II.1) showed an A > G substitution at position 1 in the heterozygous state, causing the start codon methionine to change into valine. This mutation was previously reported in 2004 [18] and was described in the SNP database as rs139732572 with a very low allele frequency (MAF = 0.000008 reported in the ExAc database). This variant has been classified in ClinVar database as pathogenic (Variation ID 96689), causing L-ferritin deficiency in dominant inheritance mode. Here we report the second case of a patient with hypoferritinemia and this same mutation in the FTL gene (NM_000146.3:c[1A > G];[=], NP_000137.2:p(Met1Val);(=), HGSV nomenclature).

2.1.3. Family 3—A Case with HHCS

The proband I.1 in family 3 (Figure 1C and Table 1) is a 65-year-old man with a history of enolism and dyslipidemia, showing high levels of serum ferritin (>3000 ng/mL), motive for what he was referred to the Hematology Service at the University Hospital Germans Trias i Pujol (HGTiP). At age of 45, he underwent cataract surgery. Initially, he underwent three therapeutic phlebotomies, but they were suspended due to the development of anemia. Magnetic resonance (MR) imaging showed normal deposits ofl iver iron (30 µmol/g). The family history suggests the presence of HHCS due to the existence of additional cases of hyperferritinemia (proband ´s son and uncle) and cataracts (proband ´s mother). The proband's son (II.1) is a 39-year-old male with history of stage 1 orchiectomized and disease-free seminoma and no surgical removed cataracts. He was contacted by the same hematology service (HGTiP) under suspicion of HHCS because of hyperferritinemia (>2000 ng/mL) and cataracts. The hematological evaluation was normal except for the high ferritin levels, and liver magnetic resonance showed normal levels of hepatic iron (20 µmol/g).

The genetic studies performed on family 3 showed the presence a heterozygous deletion c.[-164_-158del7] located in the 5 ⊠FTL IRE in the proband (I.1) and his son (II.1), both affected with hereditary hyperferritinemia with cataracts syndrome (Figure 1). Genetic analyses were not available for the mother and the uncle of the proband. This variant consists of a deletion of seven nucleotides (CAACAGT), excising part of the hexanucleotide loop and upper stem of the FTL IRE (Figure 2). Following the traditional nomenclature for FTL IRE mutations, we refer to these mutations as Esplugues +36_42del7 mutation (HGVS nomenclature as NM_000146.3:c.[-164_158del7];[=]). This deletion is predicted to impair the IRE structure. RNA secondary structure modelling of WT and mutated FTL 5⊠ IRE sequences was performed using the Sfold web server, which predicted that -164_-158del7 mutation, located at the hexanucleotide loop, is likely to disturb the WT IRE conformation (Figure 3). In addition, the SIREs web server prediction [19] indicate loss of the IRE structure, as the mutated query returned no results. This mutation has not been previously described in the literature, but other similar IRE deletions have been previously demonstrated to be pathogenic for HHCS [11]. The location and severity of this mutation, together with the clinical manifestations of HHCS present in the affected individuals of this family, indicates that this variant is most probably the genetic cause of disease in this family.

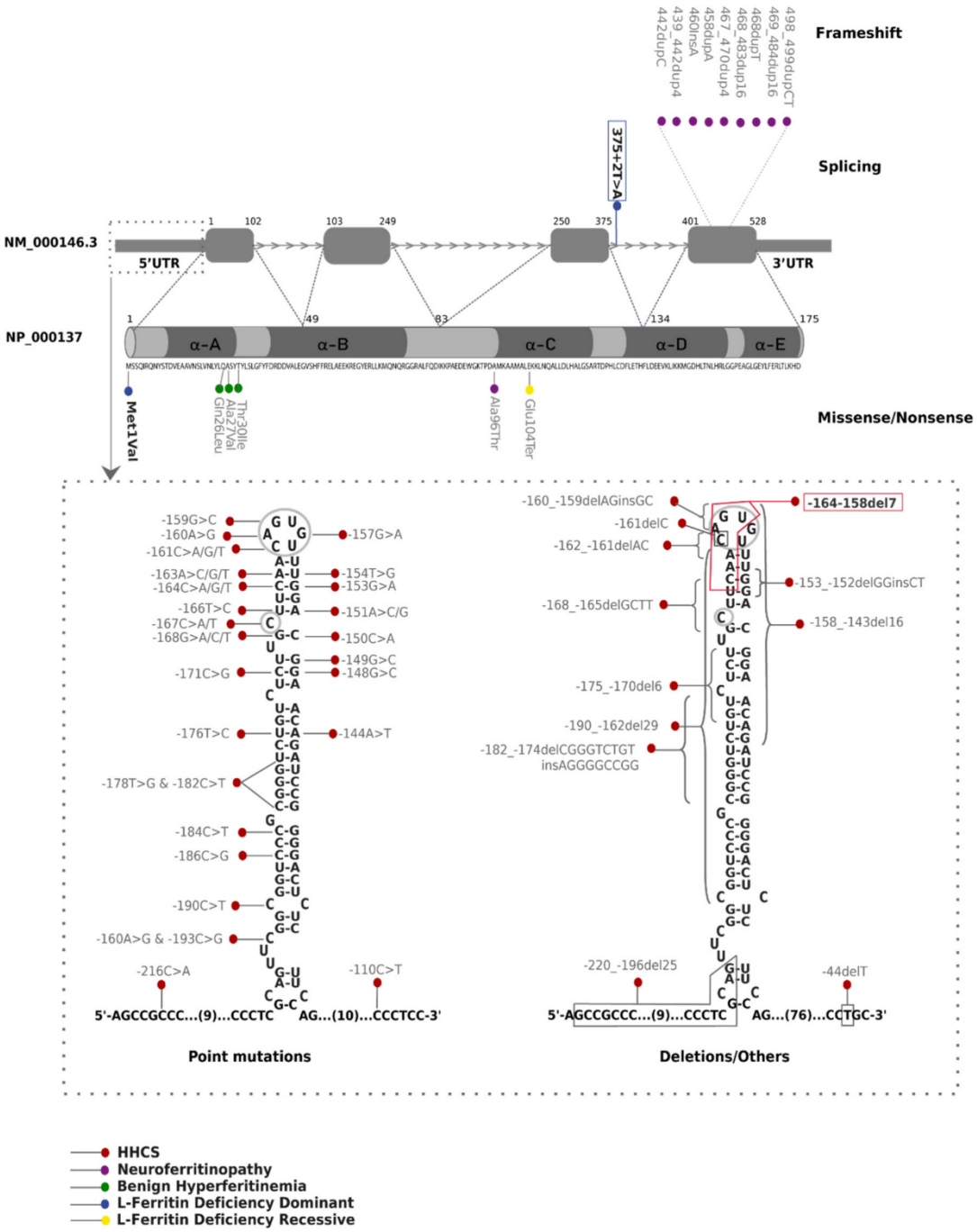


Figure 2. Schematic localization ofl iterature reported and new FTL mutations. Mutations described in this work are in bold and new mutations are boxed. The domains of the five alpha helices (A to E) are represented in the protein (NP_000137.2). Mutations are classified as nonsense, frameshift, missense, or splicing. Here we report FTL protein changes using the three-letter amino acid code.

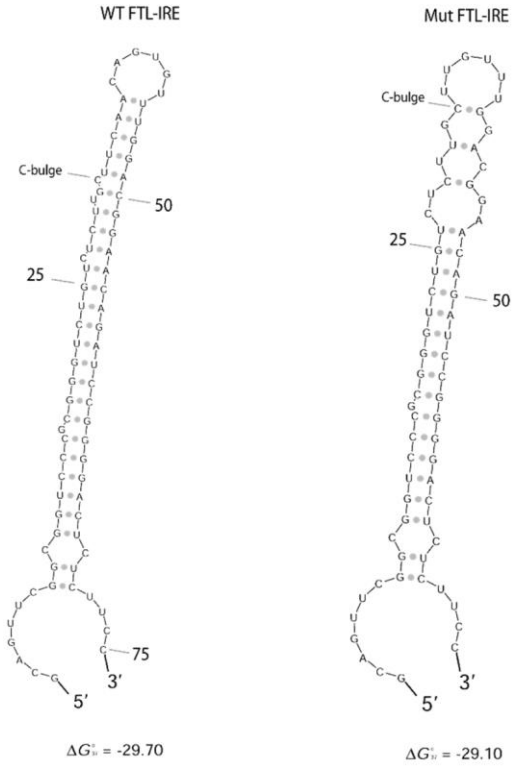


Figure 3. WT and mutated FTL -IRE fold prediction. Predicted secondary structure of WT and mutated FTL -IRE using Sfold web server [20]. Deletion in hexanucleotide loop (c.-164_158del7) is expected to disturb completely the IRE structure. Nucleotides are numbered from the transcription starting site. Free energy (ΔG) is detailed.

2.2. Update on L-Ferritin Mutations and Diseases

Mutations in L-ferritin gene causes five different phenotypic diseases, as summarized in Table 2. After an extensive literature search, we have collected all 63 different mutations reported so far for these five diseases, including the two novel mutations reported in the present work (Figure 2 and Supplementary Material: Table S1).

Most of the FTL mutations described correspond to hereditary hyperferritinemia with cataract syndrome (HHCS), including 36 point mutations, 9 deletions, and 2 insertion-deletions. All the mutations for HHCS are located at the 5 ´UTR of the FTL gene (chr19:49468566-49468764), affecting the primary sequence and structure of FTL -IRE (Figure 2). HHCS is an autosomal dominant disorder, and all reported variants are in heterozygous state except for three cases [11,21,22] where homozygous mutations have been described. Patients with HHCS do not present clinical manifestations other than high serum ferritin levels and congenial bilateral nuclear cataracts.

Neuroferritinopathy is the only NBIA disorder with an autosomal dominant inheritance. It is caused by mutation in the FTL gene. Up to now, nine insertions from one to 16 nucleotides located in the exon 4 of FTL gene have been reported to cause this neurodegenerative disorder (Figure 2). These insertions alter the reading frame of the C-terminal region, generating a longer protein with additional aberrant amino acids. In addition, a missense Ala96Thr mutation (rs104894685) has been also described to cause neuroferritinopathy; amino acid 96 is predicted to be situated at the same tertiary structure region as the pathogenic insertions [13]. Mutations in the C-terminal region of FTL disrupt α-helix D or E, which seems to be essential for the stability of the peptide [23,24]. Clinical and biochemical manifestations of this disease include low serum ferritin levels, iron accumulation in the basal ganglia and progressive and severe neurological dysfunctions with subtle cognitive deficits in some cases.

Table 2. Summary of diseases caused by mutations in FTL gene.

Disease	Hereditary Hyperferritinemia Cataract Syndrome	Benign Hyperferritinemia	Neurodegeneration with Brain Iron Accumulation 3	L-Ferritin Deficiency, Dominant	L-Ferritin Deficiency, Recessive
First publication	[7,8]	[15]	[12]	[18]	[17]
Inheritance	Autosomal dominant	Autosomal dominant	Autosomal dominant	Autosomal dominant	Autosomal recessive
Mechanism	LOST OF IRP REGULATION	(DO NOT PROCEED)	DOMINANT NEGATIVE EFFECT	HAPLOINSUFICIENCY	TOTAL LOSS OF FTL
Mutation/s	Many in the 5 IRE	Missense in exon 1	Frameshift in exon 4	p.(M1V; =)	p.(E104X; E104X)
Type	5UTR	Affects the A α-helix near the N-terminus	Predicted to cause loss of the C-terminal secondary structure	Start loss	Nonsense
Hematological features	High serum ferritin Normal serum iron Normal transferrin saturation Normal red cell counts Normal hematologic parameters	High serum ferritin Serum ferritin hyperglycosylation Normal hematologic parameters	Low serum ferritin	Low serum ferritin Low transferrin saturation (17%) Normal serum iron Normal hematologic parameters	Undetectable serum ferritin Normal Transferrin saturation Normal hematologic parameters
Other features	Congenital bilateral nuclear cataract		Severe neurological dysfunction Gastrointestinal dysphagia		Idiopathic generalized seizures Atypical RLS Progressive hair loss

The following abbreviations were used: IRP, iron regulatory protein; IRE, iron responsive element; UTR, untranslated region; RLS, restless leg syndrome; A α-helix, first domain alpha helix of FTL protein.

Three heterozygous mutations in FTL exon 1 have been associated with hyperferritinemia without iron overload where cataracts were absent; this condition has also been named benign hyperferritinemia. The Thr30Ile mutation (rs397514540, ExAc MAF = 0.000008) has been identified in French and British families [15,25]. Two further pathogenic mutations—p.(Ala27-Val) and p.(Gln26Ile)—have been reported in two additional patients [26]. Mutations in exon 1 of FTL are associated with higher than normal serum ferritin glycosylation. These three variants alter the A α-helix near the N terminus of L-ferritin; it has been hypothesized that the aberrant peptide extends the length of the hydrophobic cluster of amino acids at the N terminus, increasing the secretion of L-ferritin [15]. However, the reason for the development of hyperferritinemia and hyperglycosylation associated with these mutant ferritin forms is still not fully elucidated [25].

We have described in this report a novel intronic splicing mutation in the FTL gene (NM_000146.3:c.[375 + 2T > A];[=]) associated also with autosomal dominant L-ferritin deficiency. Including this novel splicing mutation, two heterozygous mutations have been described to cause L-ferritin deficiency with autosomal dominant transmission. Cremonesi and collaborators identified in 2004 a heterozygous A > G substitution in the first nucleotide of FTL, which change the ATG start codon (methionine) into a valine [18]. This mutation is predicted to encode a non-functional and unstable protein. Despite the low serum levels of L-ferritin, the proband presenting this mutation does not show either serious neurological problems (other than headaches) indicating that the molecular mechanism of L-ferritin deficiency with autosomal dominant inheritance is haploinsufficiency.

Finally, only one Italian case has been reported to cause L-ferritin deficiency with autosomal recessive inheritance; a homozygous substitution at nucleotide 310 G to T that produces a premature stop codon (E104X) [17]. This amino acid change is predicted to be located in the middle of the α-helix C domain, a critical region for the stability of the protein. In silico analysis predicted that a stop codon at this position produces a truncated protein unable to fold into a functional peptide and, therefore, leads to the generation an L-ferritin subunit with a complete loss off unction. This homozygous mutation is associated with a more aggressive phenotype, which is characterized by undetectable ferritin levels, idiopathic generalized seizures and atypical restless leg syndrome.

3. Discussion

L-ferritin disease is a clear case of Mendelian disease-related genes that are associated with multiple diseases, including hyperferritinemia linked with cataracts in HHCS, isolated hyperferritinemia (benign hyperferritinemia), hypoferritinemia associated with other symptoms (neurological symptoms as in neuroferritinopathy, or muscular symptoms in autosomal recessive L-ferritin deficiency), or isolated hypoferritinemia (autosomal dominant L-ferritin deficiency). For improving the diagnosis of these different diseases, we have created a clinical diagnostic algorithm of diseases caused by mutations in FTL gene (Figure 4).

The underlying mechanisms for these disease patterns are not fully elucidated for all these diseases. In HHCS, it is known that the pathological molecular mechanism is linked to the disruption of the IRP-IRE post-transcriptional regulatory system, with the de-repression of L-ferritin mRNA and the subsequent overproduction of L-ferritin protein that precipitates and deposits in the lens of the eyes, producing cataracts. In HHCS, authors have argued that an association exists between the clinical severity of the disease and the location of the mutation in FTL -IRE [11,27,28]. Higher serum ferritin levels are mostly associated with mutations in the apical loop or the C-bulge area, compared with mutations in the upper or lower stems. These findings are consistent with our cases in family 3, i.e., a father and son with a seven-nucleotide deletion in the IRE of the FTL gene (5′-CAACAGU3′). The deletion partially affects the hexanucleotide loop of the IRE, and patients show considerably high serum ferritin levels (2071–3037 ng/mL) and develop early bilateral cataracts (Table 1).

Mutations in exon 4 of FTL gene that alter the C-terminal sequence and the length of the protein have been reported to cause neuroferritinopathy. These mutations affect the D- or E α-helix (see Figure 2), leading to significant disruption of the tertiary and quaternary structure of the FTL protein

and producing an unstable and leaky ferritin. Studies have shown that mutant ferritin maintains the normal spherical shell structure and size [29]. However, mutant FTL C-termini, rich in amino acids with iron-binding properties, may be extended and unravel out of the shell. These mutant structures could initiate aggregation, forming ferritin inclusion bodies/precipitates [29,30]. Previous in vitro functional analysis and in vivo mouse models have revealed that these molecular-level defects have two main consequences: (1) the loss of normal protein function, reducing iron incorporation; (2) and the acquisition of toxic function through radical production, ferritin aggregation, and reactive oxygen species (ROS) generation [14,31–33]. It has been demonstrated that mutations in the C terminus have a dominant-negative effect, which explains the dominant transmission of the disorder [34].

In benign hyperferritinemia, missense variations in exon 1 of FTL increase the hydrophobicity of the ferritin N-terminal site. The A- α-helix of the protein is altered and hyperglycosylated, although the precise molecular mechanism is not clear yet.

L-ferritin deficiency is characterized by haploinsufficiency of the FTL protein in autosomal dominant L-ferritin deficiency, or complete loss off unction in autosomal recessive L-ferritin deficiency. Mutations associated with this condition are predicted to affect protein expression. The inactivation of one allele of FTL , which occurs in the two described heterozygous cases of autosomal dominant L-ferritin deficiency, produces a significant reduction of L-ferritin in serum. The phenotype of this dominant condition is normal apart from low levels of serum ferritin and low transferrin saturation. However, the homozygous mutation Glu104Ter causes a total loss off unction of L-ferritin and leads to ferritin missing the FTL chain. Previous mutational studies have suggested that the presence of H homopolymer ferritin in fibroblasts is associated with reduced cellular iron availability and increased ROS production, which triggers cellular damage. These findings were also found in neurons derived from patient fibroblasts and correlate with the neurological phenotype of this more severe condition [17].

In this study, we have identified two new cases of autosomal dominant L-ferritin deficiency, one due to a new mutation in intron 3 of the FTL gene that most probably affects splicing. The T > G substitution in nucleotide 375 + 2 (genomic coordinates g.49469665 in Assembly GRCh37.p13) modifies a highly conserved dinucleotide GT donor site, a key sequence recognized by the spliceosome during splicing. Mutations in splice site junctions are likely to lead to exon skipping or total or partial intron retention in the mRNA transcript in most cases [32]. Our in silico analysis using Human Splicing Finder suggests that our variant would activate a new alternative donor site onwards (375 + 5G and +6T), leading to the insertion off our additional nucleotides in the mRNA sequence between exon 3 and 4 and altering the reading frame of the transcript [33]. Therefore, this change would lead to a premature stop codon formation and the putative generation of a truncated FTL protein of 128 amino acids, if the aberrant mRNA is not detected and degraded by the nonsense mediated decay (NMD) system. According to our bioinformatics prediction, the c.375 + 2T > A mutation will not generate a dominant-negative version of FTL protein, but a truncated FTL protein completely missing the E- α-helix domain and partially lacking the D- α-helix domain. Supported by the clinical manifestations found in this family (low serum ferritin levels, low transferrin saturation and lack of serious neurological or movement abnormalities), the molecular mechanism in this case is most probably due to the loss off unction of FTL and it will be not expected that the disease derives in neuroferritinopathy. However, we cannot totally and completely exclude the later development in life of brain iron overload and neuroferritinopathy in these patients by a yet unknown and novel mechanism.

Patients with hereditary hyperferritinemia could be misdiagnosed as patients suffering from hereditary hemochromatosis, liver dysfunction or inflammation. Some patients have received unnecessary invasive diagnostic techniques, such as liver biopsy, and are inappropriately treated with venesections and phlebotomies that can cause severe iron-deficiency anemia. On the other hand, clinical manifestations of neuroferritinopathy including tremor, parkinsonism, psychiatric problems, and abnormal involuntary movements, the presence off erritin-iron precipitation in glia cells

and neurons [30], are often misdiagnosed and treated as Huntington's or Parkinson's disease. Potential treatment targets for neuroferritinopathy may include an optimized iron chelator to induce the re-solubilization ofi ron aggregations, in combination with radical scavengers to prevent oxidative ferritin damage [35]. Patients with benign hyperferritinemia could be misdiagnosed as patients with HHCS due to the presence of high ferritin levels and the possible late-onset appearance of cataracts. Apart from the surgical removal of cataracts in HHCS, HHCS and L-ferritin deficiency have no specific therapy.

These facts emphasize the importance of a correct and early genetic diagnosis for the subsequent implementation of proper treatment, avoiding detrimental or inappropriate treatments. Following clinical algorithms, such as the one included in this publication (Figure 4) and also HIGHFERRITIN Web Server (http://highferritin.imppc.org/) [36] will surely help in this goal.

4. Materials and Methods

4.1. Patients

All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee on 10th July 2015.

4.2. DNA Extraction, PCR Amplification, and DNA Sequencing

Genetic studies were performed with minor differences for all the pedigrees. Genomic DNA was extracted from peripheral blood using the FlexiGene DNA kit (Qiagen) according to manufacturer's instructions.

Ferritin gene regions (exonic, intron–exon boundaries, and untranslated regions) were sequenced using the Sanger method or next generation sequencing (NGS).

For the Sanger method, the FTL gene was amplified using 50 ng of genomic DNA. Primer sequences and PCR conditions are available upon request. The resulting amplification products were verified on a 2% ethidium bromide agarose gel. The purified PCR products were sequenced using a conventional Sanger method. Sequencing results were analyzed using Mutation Surveyor software (SoftGenetics LLC, PA, USA).

For NGS methods, patients were analyzed using a targeted NGS gene panel (v14) for hereditary hemochromatosis and hyper/hypoferritinemia that included the following nine genes: HFE , HFE2 , HAMP , TFR2 , SLC40A1 , BMP6 , FTL , FTH1 , and GNPAT .

Briefly, the capture of genomic regions was conducted starting from 225 ng of gDNA using a custom design HaloPlex ^{TM} Target Enrichment 1–500 kb kit (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's instructions. Library quality was determined using the Agilent 4200 TapeStation System. The quantity was measured using a QubitdsDNA assay kit with a Qubitfluorometer (Life Technologies, Carlsbad, CA, USA) and fluorescence was detected using a SpectraMax Gemini EM microplate reader (Molecular Devices). Libraries were sequenced using MiSeq reagent kit v2 (300 cycles) (Illumina, San Diego, CA, USA) on an Illumina or a MiSeq or MiniSeq sequencer (Illumina, San Diego, CA, USA), generating 150-bp paired-end reads. Samples were aligned with the reference human genome GRCh37/hg19 and data analysis was performed using our algorithms. Mutations detected by NGS were confirmed by conventional Sanger sequencing.

Genetic variants are reported following the official Human Genome Variation Sequence (HGVS) nomenclature and refer to NM_ 000146.3 for the Homo sapiens FTL transcript variant and NP_ 000137.2 for the Homo sapiensFTL protein.

Reported mutations in this study have been submitted to the Leiden Open Variation Database (http://www.lovd.nl) or to ClinVar (http://www.ncbi.nlm.nih.gov/clinvar).

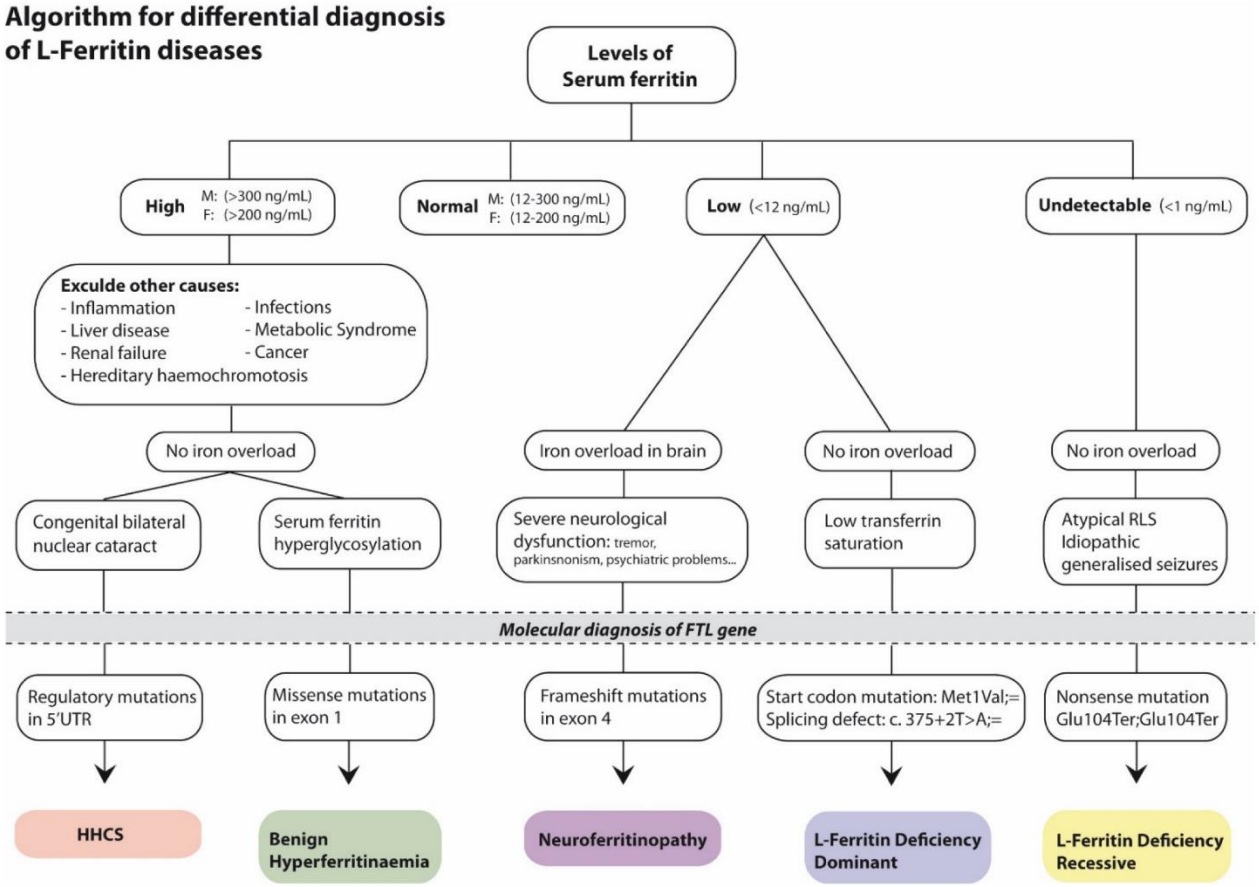


Figure 4. Algorithm for diagnosis of diseases caused by defects in FTL gene. The following abbreviations were used: F, female; M, male. The mutation nomenclature used follows the HGVS guidelines.

4.3. FTL RNA Fold Predictions

RNA folding analysis was carried out to predict the IRE structure of wild-type (WT) and mutated FTL -IRE using the Sfold web server (http://sfold.wadsworth.org/) [20]. DNA sequences used for folding predictions are shown in Supplementary Material Table S2. The SIREs web server tool was also used to predict iron-responsive elements [19].

Supplementary Materials: The following are available online at http://www.mdpi.com/1424-8247/12/1/17/s1 , Table S1: Table summarizing all FTL mutations identified up now in the literature, including the two novel mutations described here. Table S2: DNA sequence for FTL RNA fold predictions.

Author Contributions: Conceptualization, M.S. and B.C.; Methodology, C.T., B.C., and M.S.; Software, C.T.; Validation, J.F.-T.; Formal Analysis, J.F.-T.; Investigation, J.F.-T. and B.C.; Resources, S.P.-M., I.H.-R., J.L.F., M.E.L., A.M.G., and M.B.-C.; Data Curation, J.F.-T.; Writing—Original Draft Preparation, B.C., J.L.R., and M.S.; Writing—Review and Editing, M.S., I.H.-R., J.L.F., and M.B.-C.; Visualization, B.C. and M.S.; Supervision, M.S.; Project Administration, M.S.; Funding Acquisition, M.S.

Funding: This research was partially supported by grant SAF2015-70412-R from Spanish Secretary of Research, Development and Innovation (MINECO) Spain to M.S.

Acknowledgments: The authors would like to thank all patients and their family members for their participation in the study.

Conflicts of Interest: The authors declare no conflict ofi nterest.

References

1. Arosio, P.; Ingrassia, R.; Cavadini, P. Ferritins: A family of molecules for iron storage, antioxidation and more. Biochim. Biophys. Acta 2009, 1790, 589–599. [CrossRef] [PubMed]
2. Arosio, P.; Yokota, M.; Drysdale, J.W. Characterization of serum ferritin in iron overload: Possible identity to natural apoferritin. Br. J. Haematol. 1977, 36, 199–207. [CrossRef] [PubMed]
3. Fardet, L.; Coppo, P.; Kettaneh, A.; Dehoux, M.; Cabane, J.; Lambotte, O. Low glycosylated ferritin, a good marker for the diagnosis of hemophagocytic syndrome. Arthritis Rheum. 2008, 58, 1521–1527. [CrossRef] [PubMed]
4. Harrison, P.M.; Arosio, P. The ferritins: Molecular properties, iron storage function and cellular regulation. Biochim. Biophys. Acta 1996, 1275, 161–203. [CrossRef]
5. Santambrogio, P.; Cozzi, A.; Levi, S.; Arosio, P. Human serum ferritin G-peptide is recognized by anti-L ferritin subunit antibodies and concanavalin-A. Br. J. Haematol. 1987, 65, 235–237. [CrossRef] [PubMed]
6. Cohen, L.A.; Gutierrez, L.; Weiss, A.; Leichtmann-Bardoogo, Y.; Zhang, D.L.; Crooks, D.R.; Sougrat, R.; Morgenstern, A.; Galy, B.; Hentze, M.W.; et al. Serum ferritin is derived primarily from macrophages through a nonclassical secretory pathway. Blood 2010, 116, 1574–1584. [CrossRef]
7. Beaumont, C.; Leneuve, P.; Devaux, I.; Scoazec, J.Y.; Berthier, M.; Loiseau, M.N.; Grandchamp, B.; Bonneau, D. Mutation in the iron responsive element of the L ferritin mRNA in a family with dominant hyperferritinaemia and cataract. Nat. Genet. 1995, 11, 444–446. [CrossRef]
8. Girelli, D.; Olivieri, O.; De Franceschi, L.; Corrocher, R.; Bergamaschi, G.; Cazzola, M. A linkage between hereditary hyperferritinaemia not related to iron overload and autosomal dominant congenital cataract. Br. J. Haematol. 1995, 90, 931–934. [CrossRef]
9. Muckenthaler, M.U.; Rivella, S.; Hentze, M.W.; Galy, B. A Red Carpet for Iron Metabolism. Cell 2017, 168, 344–361. [CrossRef]
10. Mumford, A.D.; Cree, I.A.; Arnold, J.D.; Hagan, M.C.; Rixon, K.C.; Harding, J.J. The lens in hereditary hyperferritinaemia cataract syndrome contains crystalline deposits of L-ferritin. Br. J. Ophthalmol. 2000, 84, 697–700. [CrossRef]
11. Luscieti, S.; Tolle, G.; Aranda, J.; Campos, C.B.; Risse, F.; Mor án, É.; Muckenthaler, M.U.; S ánchez, M. Novel mutations in the ferritin-L iron-responsive element that only mildly impair IRP binding cause hereditary hyperferritinaemia cataract syndrome. Orphanet J. Rare Dis. 2013, 8, 30. [CrossRef] [PubMed]
12. Curtis, A.R.; Fey, C.; Morris, C.M.; Bindoff, L.A.; Ince, P.G.; Chinnery, P.F.; Coulthard, A.; Jackson, M.J.; Jackson, A.P.; McHale, D.P.; et al. Mutation in the gene encoding ferritin light polypeptide causes dominant adult-onset basal ganglia disease. Nat. Genet. 2001, 28, 350–354. [CrossRef] [PubMed]

13. Maciel, P.; Cruz, V.T.; Constante, M.; Iniesta, I.; Costa, M.C.; Gallati, S.; Sousa, N.; Sequeiros, J.; Coutinho, P.; Santos, M.M. Neuroferritinopathy: Missense mutation in FTL causing early-onset bilateral pallidal involvement. Neurology 2005, 65, 603–605. [CrossRef] [PubMed]
14. Muhoberac, B.B.; Vidal, R. Abnormal iron homeostasis and neurodegeneration. Front. Aging Neurosci. 2013, 5, 32. [CrossRef]
15. Kannengiesser, C.; Jouanolle, A.-M.; Hetet, G.; Mosser, A.; Muzeau, F.; Henry, D.; Bardou-Jacquet, E.; Mornet, M.; Brissot, P.; Deugnier, Y.; et al. A new missense mutation in the L ferritin coding sequence associated with elevated levels of glycosylated ferritin in serum and absence ofi ron overload. Haematologica2009, 94, 335–339. [CrossRef]
16. Ravasi, G.; Pelucchi, S.; Mariani, R.; Casati, M.; Greni, F.; Arosio, C.; Pelloni, I.; Majore, S.; Santambrogio, P.; Levi, S.; et al. Unexplained isolated hyperferritinemia without iron overload. Am. J. Hematol. 2017, 92, 338–343. [CrossRef]
17. Cozzi, A.; Santambrogio, P.; Privitera, D.; Broccoli, V.; Rotundo, L.I.; Garavaglia, B.; Benz, R.; Altamura, S.; Goede, J.S.; Muckenthaler, M.U.; et al. Human L-ferritin deficiency is characterized by idiopathic generalized seizures and atypical restless leg syndrome. J. Exp. Med. 2013, 210, 1779–1791. [CrossRef]
18. Cremonesi, L.; Cozzi, A.; Girelli, D.; Ferrari, F.; Fermo, I.; Foglieni, B.; Levi, S.; Bozzini, C.; Camparini, M.; Ferrari, M.; et al. Case report: A subject with a mutation in the ATG start codon of L-ferritin has no haematological or neurological symptoms. J. Med. Genet. 2004, 41, e81. [CrossRef] [PubMed]
19. Campillos, M.; Cases, I.; Hentze, M.W.; Sanchez, M. SIREs: Searching for iron-responsive elements. Nucleic Acids Res. 2010, 38, W360–W367. [CrossRef] [PubMed]
20. Ding, Y.; Chan, C.Y.; Lawrence, C.E. Sfold web server for statistical folding and rational design of nucleic acids. Nucleic Acids Res. 2004, 32, W135–W141. [CrossRef]
21. Alvarez-Coca-Gonzalez, J.; Moreno-Carralero, M.I.; Martinez-Perez, J.; Mendez, M.; Moran-Jimenez, M.J. The hereditary hyperferritinemia-cataract syndrome: A family study. Eur. J. Pediatr. 2010, 169, 1553–1555. [CrossRef] [PubMed]
22. Van de Sompele, S.; Pecheux, L.; Couso, J.; Meunier, A.; Sanchez, M.; De Baere, E. Functional characterization of a novel non-coding mutation “Ghent + 49A > G” in the iron-responsive element of L-ferritin causing hereditary hyperferritinaemia-cataract syndrome. Sci. Rep. 2017, 7, 18025. [CrossRef] [PubMed]
23. Kubota, A.; Hida, A.; Ichikawa, Y.; Momose, Y.; Goto, J.; Igeta, Y.; Hashida, H.; Yoshida, K.; Ikeda, S.-I.; Kanazawa, I.; et al. A novel ferritin light chain gene mutation in a Japanese family with neuroferritinopathy: Description of clinical features and implications for genotype-phenotype correlations. Mov. Disord. 2009, 24, 441–445. [CrossRef] [PubMed]
24. Baraibar, M.A.; Muhoberac, B.B.; Garringer, H.J.; Hurley, T.D.; Vidal, R. Unraveling of the E-helices and Disruption of 4-Fold Pores Are Associated with Iron Mishandling in a Mutant Ferritin Causing Neurodegeneration. J. Biol. Chem. 2010, 285, 1950–1956. [CrossRef] [PubMed]
25. Bhuva, M.; Sen, S.; Elsey, T.; Atoyebi, W.; Dreau, H.; Bradbury, C.; Johnston, R.; Bignell, P.; Griffiths, W. Sequence analysis of exon 1 of the ferritin light chain (FTL) gene can reveal the rare disorder “hereditary hyperferritinaemia without cataracts”. Br. J. Haematol. 2018. [CrossRef] [PubMed]
26. Thurlow, V.; Vadher, B.; Bomford, A.; DeLord, C.; Kannengiesser, C.; Beaumont, C.; Grandchamp, B. Two novel mutations in the L ferritin coding sequence associated with benign hyperferritinaemia unmasked by glycosylated ferritin assay. Ann. Clin. Biochem. 2012, 49, 302–305. [CrossRef] [PubMed]
27. Faniello, M.C.; Di Sanzo, M.; Quaresima, B.; Nistic ò, A.; Fregola, A.; Grosso, M.; Cuda, G.; Costanzo, F. Bilateral cataract in a subject carrying a C to A transition in the L ferritin promoter region. Clin. Biochem. 2009, 42, 911–914. [CrossRef]
28. Allerson, C.R.; Cazzola, M.; Rouault, T.A. Clinical severity and thermodynamic effects ofi ron-responsive element mutations in hereditary hyperferritinemia-cataract syndrome. J. Biol. Chem. 1999, 274, 26439–26447. [CrossRef]
29. Baraibar, M.A.; Barbeito, A.G.; Muhoberac, B.B.; Vidal, R. Iron-mediated Aggregation and a Localized Structural Change Characterize Ferritin from a Mutant Light Chain Polypeptide That Causes Neurodegeneration. J. Biol. Chem. 2008, 283, 31679–31689. [CrossRef]
30. Vidal, R.; Ghetti, B.; Takao, M.; Brefel-Courbon, C.; Uro-Coste, E.; Glazier, B.S.; Siani, V.; Benson, M.D.; Calvas, P.; Miravalle, L.; et al. Intracellular ferritin accumulation in neural and extraneural tissue characterizes

a neurodegenerative disease associated with a mutation in the ferritin light polypeptide gene. J. Neuropathol. Exp. Neurol. 2004, 63, 363–380. [CrossRef]

31. Garringer, H.J.; Irimia, J.M.; Li, W.; Goodwin, C.B.; Richine, B.; Acton, A.; Chan, R.J.; Peacock, M.; Muhoberac, B.B.; Ghetti, B.; et al. Effect of Systemic Iron Overload and a Chelation Therapy in a Mouse Model of the Neurodegenerative Disease Hereditary Ferritinopathy. PLoS ONE 2016, 11, e0161341. [CrossRef]
32. Vidal, R.; Miravalle, L.; Gao, X.; Barbeito, A.G.; Baraibar, M.A.; Hekmatyar, S.K.; Widel, M.; Bansal, N.; Delisle, M.B.; Ghetti, B. Expression of a Mutant Form of the Ferritin Light Chain Gene Induces Neurodegeneration and Iron Overload in Transgenic Mice. J. Neurosci. 2008, 28, 60–67. [CrossRef]
33. Muhoberac, B.B.; Baraibar, M.A.; Vidal, R. Iron Loading-Induced Aggregation and Reduction of Iron Incorporation in Heteropolymeric Ferritin Containing a Mutant Light Chain that Causes Neurodegeneration. Biochim. Biophys. Acta 2011, 1812, 544–548. [CrossRef] [PubMed]
34. Luscieti, S.; Santambrogio, P.; Langlois d'Estaintot, B.; Granier, T.; Cozzi, A.; Poli, M.; Gallois, B.; Finazzi, D.; Cattaneo, A.; Levi, S.; et al. Mutant ferritin L-chains that cause neurodegeneration act in a dominant-negative manner to reduce ferritin iron incorporation. J. Biol. Chem. 2010, 285, 11948–11957. [CrossRef]
35. Baraibar, M.A.; Barbeito, A.G.; Muhoberac, B.B.; Vidal, R. A mutant light chain ferritin that causes neurodegeneration has enhanced propensity toward oxidative damage. Free Radic. Biol. Med. 2012, 52, 1692–1697. [CrossRef] [PubMed]
36. Altes, A.; Perez-Lucena, M.J.; Bruguera, M. [Systematic approach to the diagnosis of hyperferritinemia]. Med. Clin. (Barc.) 2014, 142, 412–417. [PubMed]

Tornador et al. 2019 Frontiers in Physiology

frontiers in Physiology

TECHNOLOGY AND CODE
published: 13 September 2019
doi: 10.3389/fphys.2019.01063

CoDysAn: A Telemedicine Tool to Improve Awareness and Diagnosis for Patients With Congenital Dyserythropoietic Anemia

Cristian Tornador[1,2], *Edgar Sánchez-Prados*[3], *Beatriz Cadenas*[4,5,6], *Roberta Russo*[7,8], *Veronica Venturi*[9], *Immacolata Andolfo*[7,8], *Ines Hernández-Rodriguez*[10], *Achille Iolascon*[7,8] *and Mayka Sánchez*[1,9]*

[1] BloodGenetics S.L., Barcelona, Spain, [2] Teresa Moreto Foundation, Barcelona, Spain, [3] Bioinformatics for Health Sciences Master Programme, Universitat Pompeu Fabra, Barcelona, Spain, [4] Whole Genix SL., Barcelona, Spain, [5] Universitat de Vic-Universitat Central de Catalunya, Vic, Spain, [6] Iron Metabolism: Regulation and Diseases Group, Josep Carreras Leukaemia Research Institute, Campus Can Ruti, Barcelona, Spain, [7] Department of Molecular Medicine and Medical Biotechnologies, University of Naples Federico II, Naples, Italy, [8] CEINGE–Biotecnologie Avanzate, Naples, Italy, [9] Iron Metabolism: Regulation and Diseases Group, Department of Basic Sciences, Faculty of Medicine and Health Sciences, Universitat Internacional de Catalunya, Barcelona, Spain, [10] Haematology Service, Hospital Germans Trias i Pujol University Hospital, Oncology Catalan Institute, Barcelona, Spain

OPEN ACCESS

Edited by:
Richard Van Wijk,
Utrecht University, Netherlands

Reviewed by:
Anna Rita Migliaccio,
Icahn School of Medicine at Mount Sinai, United States
Pedro Cabrales,
University of California, San Diego, United States

***Correspondence:**
Mayka Sánchez
msanchezfe@uic.es

Specialty section:
This article was submitted to Red Blood Cell Physiology, a section of the journal Frontiers in Physiology

Received: *28 March 2019*
Accepted: *02 August 2019*
Published: *13 September 2019*

Citation:
Tornador C, Sánchez-Prados E, Cadenas B, Russo R, Venturi V, Andolfo I, Hernández-Rodriguez I, Iolascon A and Sánchez M (2019) CoDysAn: A Telemedicine Tool to Improve Awareness and Diagnosis for Patients With Congenital Dyserythropoietic Anemia. Front. Physiol. 10:1063. doi: 10.3389/fphys.2019.01063

Congenital Dyserythropoietic Anemia (CDA) is a heterogeneous group of hematological disorders characterized by chronic hyporegenerative anemia and distinct morphological abnormalities of erythroid precursors in the bone marrow. In many cases, a final diagnosis is not achieved due to different levels of awareness for the diagnosis of CDAs and lack of use of advanced diagnostic procedures. Researchers have identified five major types of CDA: types I, II, III, IV, and X-linked dyserythropoietic anemia and thrombocytopenia (XLDAT). Proper management in CDA is still unsatisfactory, as the different subtypes of CDA have different genetic causes and different but overlapping patterns of signs and symptoms. For this reason, we developed a new telemedicine tool that will help doctors to achieve a faster diagnostic for this disease. Using open access code, we have created a responsive webpage named CoDysAn (**Co**ngenital **Dys**erythropoietic **An**emia) that includes practical information for CDA awareness and a step-by-step diagnostic tool based on a CDA algorithm. The site is currently available in four languages (Catalan, Spanish, Italian, and English). This telemedicine webpage is available at http://www.codysan.eu.

Keywords: telemedicine tool, congenital dyserythropoietic anemia, diagnosis, algorithm, hematological disease

INTRODUCTION

Congenital Dyserythropoietic Anemia (CDA) is a heterogeneous group of hematological disorders characterized by chronic hyporegenerative anemia and distinct morphological abnormalities of erythroid precursors in the bone marrow. Patients with CDA present congenital and chronic anemia of variable degree with a reticulocytosis not corresponding to the degree of anemia (ineffective erythropoiesis), jaundice and frequently splenomegaly and/or hepatomegaly (Iolascon et al., 2012, 2013).

Five classical types of CDAs (I–II–III–IV and XLTDA) have been defined based on bone marrow morphology. Among all types, CDA type II is the most common and well-known form. Genetically,

CDA type Ia (OMIN 224120) and CDA type Ib (OMIM 615631) are caused by mutations in codanin 1 (*CDAN1*) (chr 15q15.2) and *C15orf41* (chr15q14) genes, respectively (Dgany et al., 2002; Babbs et al., 2013). CDA type II (OMIM 224100) is due to pathogenic variants in Sec23 homolog B, coat complex II component (*SEC23B*) gene (chr20p11.23) (Bianchi et al., 2009; Schwarz et al., 2009). Few patients with CDA type III (OMIM 105600) have been described: they present the same mutation in the Kinesin Family Member 23 (*KIF23*) gene (chr15q23) (Liljeholm et al., 2013). CDA type IV (OMIM 613673) is due to mutations in the Kruppel Like Factor 1 (*KLF1*) gene (chr19p13.13) (Arnaud et al., 2010; Jaffray et al., 2013). Finally, X-linked dyserythropoietic anemia and thrombocytopenia (XLDAT) (OMIM 300367) is caused by mutations in transcription factor GATA Binding Protein 1 (*GATA1*) gene (chr Xp11.23) (Nichols et al., 2000; Del Vecchio et al., 2005). CDA types I and II are inherited in an autosomal recessive manner, CDA type III and IV present an autosomal dominant inheritance pattern and X-linked dyserythropoietic anemia with thrombocytopenia has an X-linked mode of inheritance.

Depending on the type of CDA, different treatments have been established. Allogenic bone marrow transplantation has been successfully employed in a few severe cases of CDAI and CDAII. CDA III patients may require a transfusion only during times of extreme anemia e.g., pregnancy or surgery. Treatment focuses on hemoglobin normalization with the administration of interferon (IFN) alpha is used with success in CDA I patients with *CDAN1* mutations; however, patients bearing a mutation in a different gene i.e., *C15ORF41* were unresponsive to this same treatment. Severe cases of fetal anemia associated with CDAI, CDAII, and XLTDA may require intrauterine transfusions. Blood iron levels should be closely monitored in CDA I, CDAII, and other CDA patients undergoing regular transfusions. In these cases, morbidity may be severe due to iron overload complications that can be fatal if left untreated (Gambale et al., 2016; Palmer et al., 2018); therefore, it is imperative to monitor iron overload and induce iron depletion, when needed, by iron chelation. This working classification of CDA is still in use in clinical practice; however, the identification of the mutated genes involved in the majority of CDA subgroups will improve the diagnostic possibilities and allow a better classification of CDA patients. At present, in many cases, a final diagnosis is not achieved due to different levels of awareness for the diagnosis of CDAs and lack of use of advanced diagnostic procedures. In addition, there are families that fulfill the general definition of CDAs, but do not conform to any of the classical CDA variants. Therefore, it is very plausible that new forms of CDA may exist. These new forms will be possibly identified if a proper diagnosed is achieved in each patient suspected with CDA. Toward this goal, we have developed a new telemedicine tool named CoDysAn (**Co**ngenital **Dys**erythropoietic **An**emia) for the management and diagnosis of patients with this disease.

The aim of CoDysAn webpage is to provide a freely accessible website where general public, patients and medical doctors can better understand and learn more about this disease. Moreover, CoDysAn web page includes a diagnosis algorithm tool to ease the classification and diagnostic of CDA types.

METHODS

Patients and Validation

CoDysAn web algorithm has been developed with a set of 24 patients genetically diagnosed of different types of CDA (18 CDA type II, 1 CDA type Ib, 4 CDA type Ia, and 1 XLTDA) and with a set of 19 additional patients genetically diagnosed of non-CDA hereditable anemias including eight hereditary spherocytosis, four patients with pyruvate kinase defects, one patient with pyruvate kinase defect and a beta thalassemia trait, one patient with defects in hemolytic anemia due to adenylate kinase deficiency (AK1) gene, one patient with X-linked sideroblastic anemia, three patients with dehydrated hereditary stomatocytosis type 1 (DHS1) and one patient with dehydrated hereditary stomatocytosis type 2 (DHS2). A different set of 23 CDAII patients was utilized to independently validate the algorithm. Patients were previously reported (Iolascon et al., 2009; Schwarz et al., 2009; Russo et al., 2010, 2011, 2013, 2014, 2016, 2018; Unal et al., 2014; Andolfo et al., 2015, 2018; Di Pierro et al., 2015) and diagnosed at the Medical Genetics Unit of A.O.U. Federico II, CEINGE–Biotecnologie Avanzate (Napoli).

Design of Web Server

CoDysAn is implemented in PHP, HTML5, CSS, and Javascript. The web server is executed in a XAMPP. Network visualization and interactive exploration modules are based on several open-source projects: Bootstrap, jQuery and Filezilla. The source code of the diagnostic tool algorithm is implemented in php at http://www.codysan.eu/diagnostics-tool.html. It is integrated within this web page between lines 661 and 1,204 in four steps corresponding to the four steps of the form. The code can be checked by typing in a browser: "view-source: http://www.codysan.eu/diagnostics-tool.html."

Implementation

CoDysAn algorithm is based on the diagnostic workflow previously proposed (Iolascon et al., 2012; Gambale et al., 2016). This algorithm is based on hematological parameters depending on age and gender (**Table 1**). Age is split in three groups: from 0 to 6 months old; from 6 months to 12 years old; and older than 12 years. Hematological tested parameters include: hemoglobin levels, mean corpuscular volume (MCV), reticulocytes count and platelets count. Exclusion of other possible causes of anemia is also considered in the final step of the algorithm. References values for hematological data are adapted from general hematological reference books (Rabinovitch, 1990; Wakeman et al., 2007; Hoffman et al., 2018).

RESULTS

CoDySan Scope

Following previous experience of the group in developing telemedicine tools for management and diagnosis of patients (HIGHFERRITIN Web server http://highferritin.imppc.org/

TABLE 1 | Parameter thresholds used by the diagnostic CoDysAn algorithm.

Parameter		0–6 months	6 months to 12 years	>12 years	Units
Hemoglobin	M	9.5–18	11–15.5	13–17.5	g/dL
	F	9.5–18	11–15.5	12–16	
MCV*	M	77.5–111.5	74–89.5	80–100	fL
	F	77.5–111.5	74–89.5	80–100	
Reticulocytes	M	61–134	24–114	29–95	$\times 10^9$/L
	F	67–142	40–162	27–91	
Platelets	M	145–450	145–450	145–450	$\times 10^9$/L
	F	145–450	145–450	145–450	

MCV stands for Mean Corpuscular Volume. M stands for male and F stands for female.

tool) (Altes et al., 2014), we have developed CoDySan web tool. CoDySan is a user-friendly webpage for a better awareness on the rare hereditary hematological diseases, congenital dyserythropoietic anemias (CDAs). CoDySan webpage includes a step-by-step diagnostic algorithm based on **Figure 1**. The site is freely available at URL http://www.codysan.eu in four languages (Catalan, Spanish, Italian, and English).

Webpage Structure and Design

The CoDysAn website is currently containing seven sections (see **Supplementary Figure 1** to visualize several screenshots of different sections of CoDysAn webpage): A home or main webpage section including links to other sections of the site to make information more accessible; The CoDysAn section, where one can found information about congenital dyserythropoietic anemia (CDA) disease, the CoDysAn project as a whole research project, the privacy policy, cookies policy and medical disclaimer reminding that, as any other telemedicine project, CoDysAn diagnostic tool is a preliminary diagnostic test and expert medical doctors should be contacted for a conclusive diagnosis; The diagnostic section, including the CDA algorithm flowchart (**Figure 1**) and a step-by-step diagnostic tool with specific instructions on how to use it; The collaborators section, including links to the contributors for the CoDysAn project, patient associations and links to similar web tools, such as HIGHFERRITIN web server; A resource section, including news on the CoDysAn project, bibliographical references and reference values used for the diagnostic algorithm (see also **Table 1**); An opinion section containing a Google form that allows users to express their opinion and degree of satisfaction with the website; A contact section, where users can directly contact CoDysAn developers to address any doubt regarding the webpage.

Diagnostic Telemedicine Tool

The diagnostic algorithm used for setting up the CoDysAn diagnostic tool is depicted in **Figure 1**. A step-by-step and user-friendly form will progressively ask relevant patient information; in the first stage, age, gender and hemoglobin levels should be provided to discern if the patient has hyperhemoglobinemia (high hemoglobin values in regards to the reference value), anemia or if the values are inside the normal range, in which case the web tool will return a text indicating that there is no anemia.

Due to fluctuations in the hematological parameters, the algorithm correlates to the reference values for the hematological provided data (hemoglobin level, reticulocytes, platelets, etc.) according to age and gender, see **Table 1**. To simplify the algorithm, we have only considered three different age ranges, from 0 to 6 months, from 6 to 12 years, and older than 12 years old.

If anemia is detected, i.e., the hemoglobin levels are below the normal values for the indicated gender and age range, the algorithm will ask for three additional hematological parameters: mean corpuscular volume (MCV), reticulocytes count, and platelets count.

Users can change the input units of the provided hematological parameter. These values are converted to the international system of reference units and the value is used to check if the parameters are within range for their given thresholds (see **Table 1**).

Depending on the data provided, a new form will appear asking to exclude specific possible causes of macrocytic, normocytic or microcytic anemia. At least one alternative cause of anemia should be excluded to proceed with the diagnostic tool. In the following step, the user is asked to select additional patient's clinical or biochemical factors, such as binucleated erythroblasts, malformations or electron microscopy features.

Finally, depending on the provided information, CoDySan tool will return a result of clinical suspicion (any of the CDA types) if it applies, or a brief explanation if there is no clinical suspicion of CDA.

If a clinical suspicion of CDA is indicated, the user has the option to search for world-wide genetic laboratories that provide clinical diagnostic tests for a particular CDA gene via the button "Search Lab." The list of world-wide genetic laboratories is taken from the NCBI's Genetic Testing Registry (GTR) webpage (Rubinstein et al., 2013). There is also the possibility to refresh the webpage and perform a new diagnostic test via the button "New diagnostic."

Validation

The CodysAn algorithm has been designed with 43 patients with hereditable anemia, including 24 patients genetically diagnosed with different types of CDA (18 CDA type II, 1 CDA type Ib, 4 CDA type IIa, and 1 XLTDA) and 19 additional patients genetically diagnosed with non-CDA hereditable anemias. The algorithm achieved a specificity of 89.5% and a sensibility of 87.5%. An additional set of 23 patients (all CDA II) was utilized to validate the algorithm, which returned a specificity of 87%.

DISCUSSION

Telemedicine webpages and tools are significantly changing the way medical doctors and patients approach health care and diagnosis (Dinesen et al., 2016). CoDysAn telemedicine tool is a webpage intended to increase awareness about the rare disease CDA as, currently, patients suffering from this disease are under-diagnosed (Russo et al., 2014). The content of the webpage serves as an informative and training resource for the general public, patients and medical doctors. The use of this

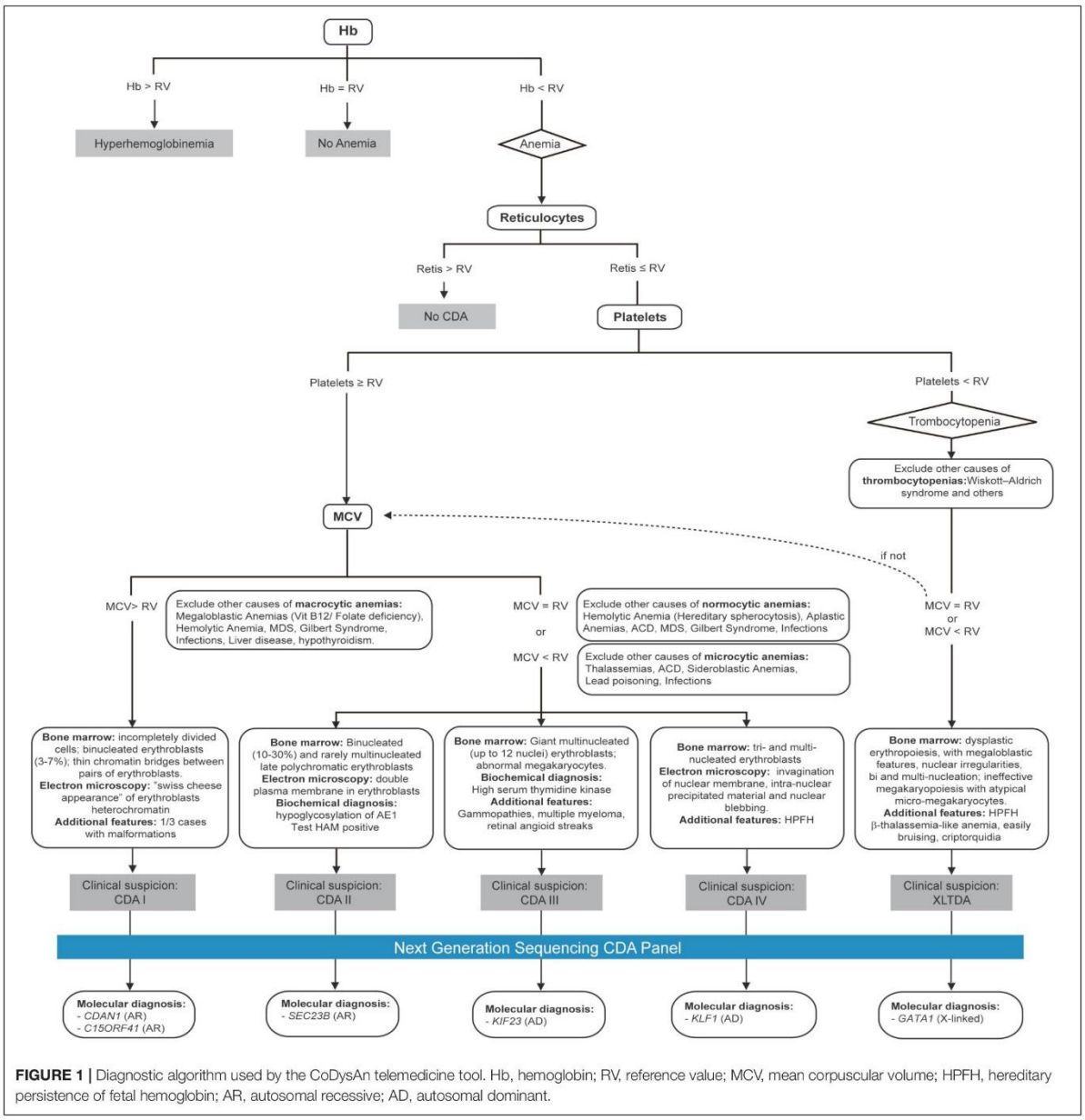


FIGURE 1 | Diagnostic algorithm used by the CoDysAn telemedicine tool. Hb, hemoglobin; RV, reference value; MCV, mean corpuscular volume; HPFH, hereditary persistence of fetal hemoglobin; AR, autosomal recessive; AD, autosomal dominant.

tool presents limits: patients should be considered as a whole entity and multiple biochemical determinations are needed due to daily parameters' variability within the same subject. Although hematological reference ranges are useful in results interpretation and in clinical decision-making, it should be borne in mind that variations within the population may affect some outcomes. CoDysAn incorporates a diagnostic algorithm that proved to be useful for a preliminary diagnostic. It will help medical doctors to know which molecular diagnostics they should request, reducing time and effort necessary for the diagnostic of CDA and allowing a direct implementation of a proper treatment once reached a definitive molecular diagnosis. Few reference centers are now offering genetic diagnostic panels screening the six known genes causing CDA. CoDySan algorithm is connected to the NCBI Genetic Testing Registry (GTR) in a way to inform medical doctors about the existence of these accredited diagnostic centers to perform a complete genetic test, if required. This telemedicine tool aims to inform the general public and aid in

the diagnosis of CDA. It is not intended as an attempt to practice medicine or provide specific medical advice and it should not be used to replace or overrule a qualified health care provider's judgment. Users should not rely upon this website for self-medication. We believe that CoDysAn webpage will positively contribute to improve medical and scientific communication on the anemia field.

DATA AVAILABILITY

All datasets generated for this study are included in the manuscript and/or the **Supplementary Files**.

AUTHOR CONTRIBUTIONS

MS designed the webpage, designed the study, and wrote the manuscript. ES-P, CT, and BC created the webpage and wrote the diagnostic algorithm. BC designed **Figure 1**. IA and RR translated the CoDysAn webpage to Italian. AI, IA, and RR provided the patients data to test CoDysAn algorithm. VV wrote and revised the manuscript. IH-R helped with reference values and **Table 1**. All authors read and approved the final version of the manuscript.

FUNDING

The CoDysAn project was supported by a research grant from Ramón Areces Foundation (Reference CIVP18A1857) to MS and AI, and by a collaborative agreement between the patient associations Pablo Ugarte Association (APU) and Patient and Relative Association of Congenital Dyserythropoietic Anemia (ADISCON).

ACKNOWLEDGMENTS

We would like to thank Nuria Centeno from UPF for Master Project advisor tasks and Francisco Fuster from Josep Carreras Leukaemia Research Institute for assistance with the opinion questionnaire. This work was performed in the context of the Experimental Sciences and Technology Ph.D. program of the University of Vic (UVic).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2019.01063/full#supplementary-material

REFERENCES

Altes, A., Perez-Lucena, M. J., Bruguera, M., and Grp Iberico, F. (2014). Systematic approach to the diagnosis of hyperferritinemia. *Med. Clin. (Barc).* 142, 412–417. doi: 10.1016/j.medcli.2013.06.010

Andolfo, I., Russo, R., Manna, F., Shmukler, B. E., Gambale, A., Vitiello, G., et al. (2015). Novel Gardos channel mutations linked to dehydrated hereditary stomatocytosis (xerocytosis). *Am. J. Hematol.* 90, 921–926. doi: 10.1002/ajh.24117

Andolfo, I., Russo, R., Rosato, B. E., Manna, F., Gambale, A., Brugnara, C., et al. (2018). Genotype-phenotype correlation and risk stratification in a cohort of 123 hereditary stomatocytosis patients. *Am. J. Hematol.* 93, 1509–1517. doi: 10.1002/ajh.25276

Arnaud, L., Saison, C., Helias, V., Lucien, N., Steschenko, D., Giarratana, M. C., et al. (2010). A Dominant mutation in the gene encoding the erythroid transcription factor KLF1 causes a congenital dyserythropoietic anemia. *Am. J. Hum. Genet.* 87, 721–727. doi: 10.1016/j.ajhg.2010.10.010

Babbs, C., Roberts, N. A., Sanchez-Pulido, L., McGowan, S. J., Ahmed, M. R., Brown, J. M., et al. (2013). Homozygous mutations in a predicted endonuclease are a novel cause of congenital dyserythropoietic anemia type I. *Haematologica* 98, 1383–1387. doi: 10.3324/haematol.2013.089490

Bianchi, P., Fermo, E., Vercellati, C., Boschetti, C., Barcellini, W., Iurlo, A., et al. (2009). Congenital dyserythropoietic anemia type II (CDAII) is caused by mutations in the SEC23B gene. *Hum Mutat.* 30, 1292–1298. doi: 10.1002/humu.21077

Del Vecchio, G. C., Giordani, L., De Santis, A., and De Mattia, D. (2005). Dyserythropoietic anemia and thrombocytopenia due to a novel mutation in GATA-1. *Acta Haematol.* 114, 113–116. doi: 10.1159/000086586

Dgany, O., Avidan, N., Delaunay, J., Krasnov, T., Shalmon, L., Shalev, H., et al. (2002). Congenital dyserythropoietic anemia type I is caused by mutations in codanin-1. *Am. J. Hum. Genet.* 71, 1467–1474. doi: 10.1086/344781

Di Pierro, E., Russo, R., Karakas, Z., Brancaleoni, V., Gambale, A., Kurt, I., et al. (2015). Congenital erythropoietic porphyria linked to GATA1-R216W mutation: challenges for diagnosis. *Eur. J. Haematol.* 94, 491–497. doi: 10.1111/ejh.12452

Dinesen, B., Nonnecke, B., Lindeman, D., Toft, E., Kidholm, K., Jethwani, K., et al. (2016). Personalized telehealth in the future: a global research agenda. *J. Med. Internet Res.* 18:e53. doi: 10.2196/jmir.5257

Gambale, A., Iolascon, A., Andolfo, I., and Russo, R. (2016). Diagnosis and management of congenital dyserythropoietic anemias. *Expert Rev. Hematol.* 9, 283–296. doi: 10.1586/17474086.2016.1131608

Hoffman, R., Benz, E. J. Jr., Silberstein, L. E., Heslop, H., Weitz, J., and Anastasi, J. (2018). *Hematology: Basic Principles and Practice*, *7th Edn*. Philadelphia, PA: Elsevier Saunders.

Iolascon, A., Esposito, M. R., and Russo, R. (2012). Clinical aspects and pathogenesis of congenital dyserythropoietic anemias: from morphology to molecular approach. *Haematologica* 97, 1786–1794. doi: 10.3324/haematol.2012.072207

Iolascon, A., Heimpel, H., Wahlin, A., and Tamary, H. (2013). Congenital dyserythropoietic anemias: molecular insights and diagnostic approach. *Blood* 122, 2162–2166. doi: 10.1182/blood-2013-05-468223

Iolascon, A., Russo, R., Esposito, M. R., Asci, R., Piscopo, C., Perrotta, S., et al. (2009). Molecular analysis of 42 patients with congenital dyserythropoietic anemia type II: new mutations in the SEC23B gene and a search for a genotype-phenotype relationship. *Haematologica* 95, 708–715. doi: 10.3324/haematol.2009.014985

Jaffray, J. A., Mitchell, W. B., Gnanapragasam, M. N., Seshan, S. V., Guo, X., Westhoff, C. M., et al. (2013). Erythroid transcription factor EKLF/KLF1 mutation causing congenital dyserythropoietic anemia type IV in a patient of Taiwanese origin: review of all reported cases and development of a clinical diagnostic paradigm. *Blood Cells Mol. Dis.* 51, 71–75. doi: 10.1016/j.bcmd.2013.02.006

Liljeholm, M., Irvine, A. F., Vikberg, A. L., Norberg, A., Month, S., Sandström, H., et al. (2013). Congenital dyserythropoietic anemia type III (CDA III) is caused by a mutation in kinesin family member, KIF23. *Blood* 121, 4791–4799. doi: 10.1182/blood-2012-10-461392

Nichols, K. E., Crispino, J. D., Poncz, M., White, J. G., Orkin, S. H., Maris, J. M., et al. (2000). Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nat. Genet.* 24, 266–270. doi: 10.1038/73480

Palmer, W. C., Vishnu, P., Sanchez, W., Aqel, B., Riegert-Johnson, D., Seaman, L. A. K., et al. (2018). Diagnosis and management of genetic iron overload disorders. *J. Gen. Intern. Med.* 33, 2230–2236. doi: 10.1007/s11606-018-4669-2

Rabinovitch, A. (1990). Hematology reference ranges. *Arch. Pathol. Lab. Med.* 114:1189.

Rubinstein, W. S., Maglott, D. R., Lee, J. M., Kattman, B. L., Malheiro, A. J., Ovetsky, M., et al. (2013). The NIH genetic testing registry: a new, centralized database of genetic tests to enable access to comprehensive information and improve transparency. *Nucleic Acids Res.* 41, D925–D935. doi: 10.1093/nar/gks1173

Russo, R., Andolfo, I., Manna, F., De Rosa, G., De Falco, L., Gambale, A., et al. (2016). Increased levels of ERFE-encoding FAM132B in patients with congenital dyserythropoietic anemia type II. *Blood* 128, 1899–1902. doi: 10.1182/blood-2016-06-724328

Russo, R., Andolfo, I., Manna, F., Gambale, A., Marra, R., Rosato, B. E., et al. (2018). Multi-gene panel testing improves diagnosis and management of patients with hereditary anemias. *Am. J. Hematol.* 93, 672–682. doi: 10.1002/ajh.25058

Russo, R., Esposito, M. R., Asci, R., Gambale, A., Perrotta, S., Ramenghi, U., et al. (2010). Mutational spectrum in congenital dyserythropoietic anemia type II: identification of 19 novel variants in SEC23B gene. *Am. J. Hematol.* 85, 915–920. doi: 10.1002/ajh.21866

Russo, R., Gambale, A., Esposito, M. R., Serra, M. L., Troiano, A., De Maggio, I., et al. (2011). Two founder mutations in the SEC23B gene account for the relatively high frequency of CDA II in the Italian population. *Am. J. Hematol.* 86, 727–732. doi: 10.1002/ajh.22096

Russo, R., Gambale, A., Langella, C., Andolfo, I., Unal, S., and Iolascon, A. (2014). Retrospective cohort study of 205 cases with congenital dyserythropoietic anemia type II: definition of clinical and molecular spectrum and identification of new diagnostic scores. *Am. J. Hematol.* 89, E169–E175. doi: 10.1002/ajh.23800

Russo, R., Langella, C., Esposito, M. R., Gambale, A., Vitiello, F., Vallefuoco, F., et al. (2013). Hypomorphic mutations of SEC23B gene account for mild phenotypes of congenital dyserythropoietic anemia type II. *Blood Cells Mol. Dis.* 51, 17–21. doi: 10.1016/j.bcmd.2013.02.003

Schwarz, K., Iolascon, A., Verissimo, F., Trede, N. S., Horsley, W., Chen, W., et al. (2009). Mutations affecting the secretory COPII coat component SEC23B cause congenital dyserythropoietic anemia type II. *Nat. Genet.* 41, 936–940. doi: 10.1038/ng.405

Unal, S., Russo, R., Gumruk, F., Kuskonmaz, B., Cetin, M., Sayli, T., et al. (2014). Successful hematopoietic stem cell transplantation in a patient with congenital dyserythropoietic anemia type II. *Pediatr. Transplant.* 18, E130–E133. doi: 10.1111/petr.12254

Wakeman, L., Al-Ismail, S., Benton, A., Beddall, A., Gibbs, A., Hartnell, S., et al. (2007). Robust, routine haematology reference ranges for healthy adults. *Int. J. Lab. Hematol.* 29, 279–283. doi: 10.1111/j.1365-2257.2006.00883.x

Conflict of Interest Statement: CT was employed by company BloodGenetics SL. BC was employed by company Whole Genix SL.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

4. Other documents

Certificate of participation in 2018 EMQN external quality assessment for NGS and Sanger sequencing. Both results were satisfactory (see page 216).

Example of Clinical report of BloodGenetics (see page 217)

Instruction of use of BEA program (see page 221)

Example of report performed by BEA program (see page 237)

Certificate of participation in 2018 EMQN external quality assessment for NGS and Sanger sequencing.

EMQN

The European Molecular Genetics Quality Network

Certificate of Participation

2018

This is to certify that

BLOODGENETICS S.L.

Participated in 2018 in the following European Molecular Genetics Quality Network (EMQN) external quality assessment schemes:

Scheme	Genotyping	Interpretation	Clerical	Result
DNA SEQUENCING - NGS (v Germline)				Satisfactory
DNA SEQUENCING - SANGER (Full version)	2.00	2.00		Satisfactory

The laboratory participated in 2 schemes and passed 2 of them. The detailed performance data are given on the Individual Laboratory Report (ILR). When viewing the certificates, you should ensure that 2 schemes are listed.

Key: > Scheme Mean | < Scheme Mean | Poor performance

NRS: No Results Submitted
WFS: Withdrew From Scheme

Signed by:

Prof. David Barton
EMQN Chairman

Dr. Simon Patton
EMQN Director

EMQN is a not for profit organisation which provides external quality assessment schemes for genetic testing laboratories. EMQN is a UKAS accredited provider of EQA services and is based at the Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester, UK.

Page 1 of 1

Example of Clinical report

BLOOD GENETICS
Diagnostics in Inherited Blood Diseases

INFORME DIAGNÓSTICO GENÉTICO (NGS)

Paciente:	Test Realizado:	Fecha Informe:
	NGS-10010v14	15/06/2017

Fecha de nacimiento:		**Médico/Petición:**	Dr. xxxx	**ID de muestra:**	P-00124-D2
Sexo:		**Hospital/Clínica:**	Hospital xxxx	**Tipo de muestra:**	Sangre total (ST)
nº Caso:	00289	**nº Historia Clínica:**	xxx	**Fecha recepción:**	
nº Paciente:	P-00124	**Documentación:**	Consentimiento informado + informe médico.	**Nº Solicitud:**	

RESUMEN DE RESULTADOS – 1 alteración/es encontrada/s*

1 variante/s patogénica/s o probablemente patogénica/s

0 variante/s de significación incierta

1 Terapia asociada con beneficio clínico

(*) Las variantes benignas o probablemente benignas no se incluyen en este informe, no obstante esta información está a la disposición del médico si se solicita.

Historia Clínica:
Caso con hipoferritinemia sin respuesta a terapia con hierro oral.

Diagnóstico Clínico:
Sospecha de Hipoferritinemia

Genes de estudio:
Secuenciación completa de los genes HFE, HFE2, HAMP, TFR2, SLC40A1, BMP6, FTL, FTH1, GNPAT mediante Panel NGS-10010v14

INFORMACIÓN GENERAL DE LA ENFERMEDAD

La **hemocromatosis hereditaria (HH)** es un grupo de enfermedades de origen genético caracterizadas por una acumulación excesiva de hierro en los tejidos. Existen 5 tipos de HH llamadas HH tipo 1 (gen HFE), HH tipo 2 (genes HJV y HAMP), HH tipo 3 (gen TFR2), HH tipo 4 (gen SLC40A1) y HH debida a mutaciones en el gen BMP6. La presencia de una variante en el gen GNPAT ha sido relacionada con una clínica más severa en HH tipo 1. Las formas 1, 2 y 3 son enfermedades con herencia autosómica recesiva y la forma 4 y la debida a mutaciones en gen BMP6 son dominantes. La HH tipo 1 es la forma más común de HH, siendo las otras formas enfermedades raras. En general las HH pueden causar fatiga crónica, pigmentación oscura de la piel y complicaciones clínicas como fibrosis o cirrosis hepática, carcinoma hepatocelular, diabetes mellitus, artropatía, osteoporosis, hipogonadismo hipogonadotrópico e insuficiencia cardíaca.

La **hiperferritinemia** es la presencia de niveles elevados de ferritina en sangre. La ferritina es la proteína encargada del almacenamiento y distribución intracelular del hierro. Existen dos genes ferritina H y ferritina L (FTH1 y FTL) que codifican para las dos subunidades de la ferritina. Todas las formas de hiperferritinemia familiar presentan una herencia autosómica dominante. Mutaciones en estos dos genes dan lugar a:

a) El **Síndrome de hiperferritinemia y cataratas congénitas (HHCS)** caracterizado por niveles elevados de ferritina sérica sin sobrecarga de hierro, cataratas congénitas bilaterales y debido a mutaciones en el elemento regulador "Iron-responsive element (IRE) en el 5'UTR del gen FTL.

b) **Hiperferritinemia benigna**, cursa sin sobrecarga de hierro y es debida a mutaciones en la secuencia codificante del gen FTL.

c) **Hiperferritinemia con sobrecarga de hierro**, está causada por una mutación en el IRE del gen FTH1 en 5'UTR y se ha descrito solo en una familia.

La **hipoferritinemia hereditaria por deficiencia de ferritina L** está causada por mutaciones en el gen FTL y ha sido descrita en dos pacientes, en uno como enfermedad autosómica dominante y en otro como enfermedad autosómica recesiva.

RESULTADOS MAS RELEVANTES E INTERPRETACIÓN DE LAS VARIANTES

A.Variantes encontradas con relevancia clínica y/o consecuencias fenotípicas conocidas:

FTL: NM_000146.3:c.[375+2T>A];[=]. Este cambio es una mutación de splicing.

Este cambio **no** se ha descrito previamente en las bases de datos consultadas (ENSEMBL, NCBI y 1000Genomes, HGMD público).

Los resultados genéticos indican que el paciente presenta una variación intrónica en el gen *FTL* en heterocigosis que debido a su posición (+2) afectaría al splicing del ARNm y por lo tanto a la proteína. Este cambio es por tanto considerado como patogénico y responsable de la clínica observada en el paciente.

Ver además Annexo II para otras variantes encontradas.

Diagnostics in Inherited Blood Diseases BLOODGENETICS S.L.
C/Verge de Guadalupe18, 08950 Esplugues de Llobregat, Barcelona, España. info@bloodgenetics.com

Paciente:	Fecha:	Panel:
P-00124	15/06/2017	NGS-10010v14

RECOMENDACIONES CLÍNICAS

En casos de Hemocromatosis e hiperferritinemia se recomienda un seguimiento de los niveles de ferritina sérica, saturación de transferrina periódicamente y una evaluación de la sobrecarga hepática mediante métodos no invasivos (resonancia magnética) en el paciente y análisis bioquímicos en familiares que puedan estar afectados.

En el caso de detectarse variantes patogénicas o probablemente patogénicas se recomienda extender el estudio genético a otros miembros de la familia. En el caso de detectarse variaciones de significación desconocida la extensión de los análisis genéticos a otros familiares podría ayudar a una mejor clasificación de la variante en patogénica o benigna.

Ponemos a su disposición la herramienta médica HIGHFERRITIN para la evaluación de pacientes con niveles elevados de ferritina. Por favor consulte: http://highferritin.imppc.org.

Además, informamos de la existencia de la Asociación Española de Hemocromatosis (aeh@hes.scs.es; www.hemocromatosis.es) a la que los pacientes diagnosticados de **Hemocromatosis Hereditaria** se pueden asociar para recibir asesoramiento e información directa de esta enfermedad.

Tratamiento recomendado:

En el caso de **Hemocromatosis Hereditarias o hiperferritinemias con sobrecarga de hierro** si los niveles de ferritina son elevados se recomienda realizar eritroaferesis controlada o tratamiento con flebotomías hasta la normalización de los niveles de ferritinas para disminuir la sobrecarga de hierro hepática. Normalmente se requiere de un alto número de flebotomías semanales para normalizar los valores de ferritina en casos con ferritina inicial muy elevada.

En casos de **Hemocromatosis hereditaria tipo 4** es posible que las flebotomías causen anemia en el paciente por lo tanto se recomienda realizar flebotomías de menor cantidad (300 ml) y más espaciadas en el tiempo.

En el caso de **hiperferritinemia con cataratas o hiperferritinemia benigna** NO se recomienda y es contraproducente realizar sangrías, el paciente anemizará y los niveles de ferritina no se normalizarán. En caso de HHCS los niveles de ferritina alta no comportan graves manifestaciones clínicas solo cataratas que pueden ser intervenidas quirúrgicamente.

En el caso de **hipoferritinemia hereditaria por deficiencia de ferritina L** los bajos o ausentes niveles de ferritina sérica se deben a mutaciones directas en el gen FTL y no reflejarían un déficit de hierro, la terapia con hierro no sería necesaria si no se demuestra una falta real de hierro por anemia (con valores bajos de hemoglobina) y niveles bajos de hierro sérico. La haploinsuficiencia de Ferritina L (mutación en heterocigosis y niveles bajos de ferritina sérica) no causa ningún defecto neurológico o hematológico. La falta total de Ferritina L (mutación en homocigosis y niveles indetectables de ferritina sérica) se ha asociado con pérdida progresiva de cabello, convulsiones generalizadas que pueden tratarse con medicamentos anti-convulsivos, leve trastorno neuropsicológico y el síndrome de piernas inquietas.

ANÁLISIS REALIZADO Y INTERPRETADO POR:

Josep Fita, técnico de laboratorio
Responsable genetista: Dra. Mayka Sánchez
Responsable análisis Bioinformáticos: Dr. Cristian Tornador

Atentamente,

Dra. Mayka Sánchez
Responsable genetista
Acreditada en Genética Humana por la Asociación Española de Genética Humana (AEGH)
Núm. De Colegiada: 21589-C (Colegio de Biólogos de Cataluña)

Dr. Cristian Tornador Antolín
Responsable de análisis bioinformáticos
Núm. de Colegiado: EIC-19211 (Asociación y Escuela Oficial del Colegio de Ingenieros de Cataluña)

Paciente:	Fecha:	Panel:
P-00124	15/06/2017	NGS-10010v14

Anexo I: MÉTODOS, PARÁMETROS DE CALIDAD Y LIMITACIONES

Metodología:

Tras la recepción de sangre del paciente se procede a la extracción de DNA para la realización del estudio genético. Las muestras de DNA son procesadas mediante la tecnología Haloplex (Agilent) donde son fragmentadas utilizando enzimas de restricción y desnaturalizadas. La librería de sondas (diseño propio realizado por SureDesing) es añadida e hibridizada a los fragmentos diana. Cada sonda es un oligonucleótido diseñado para hibridar ambos extremos de los fragmentos diana del DNA, guiando así los fragmentos específicos para formar moléculas de DNA circulares. Las sondas se encuentran biotiniladas y por lo tanto los fragmentos pueden recuperarse con esferas de estreptavidina magnéticas. Las moléculas circulares se cierran por ligación y se amplifican mediante PCR. Tras purificar los productos de PCR, las muestras están preparadas para secuenciar en equipos MiSeq/MiniSeq (Illumina) con una longitud de lectura de 2x150bp. Durante todo el proceso se utiliza un control interno para validar las diferentes etapas del proceso. Tras la secuenciación, los datos son analizados con el software SureCall (Agilent) donde las muestras son alineadas con el genoma humano de referencia GRCh37/hg19 y procesadas para posteriormente efectuar la búsqueda de variantes. Para finalizar, las variantes encontradas son filtradas utilizando nuestra propia base de datos y sistema de clasificación.

Parámetros de calidad:

El Panel para Anemias relacionadas con el metabolismo del hierro 10010v14 se compone de 9 genes involucrados en estas patologías:

HFE HFE2 HAMP TFR2 SLC40A1 BMP6 FTL FTH1 GNPAT

Las regiones codificantes y zonas de splicing de estos genes son secuenciados mediante NGS con una alta cobertura lo que permite detectar variaciones genéticas con precisión.

El diseño del panel v14 incluye 51 regiones que cubren 23187 bases con un 99,96% de cobertura.
En este análisis se han producido 1.132.922 reads con una cobertura media de 291X.
En total se ha/n encontrado 1 variante/s de tipo 1, 2 y 3 (patogénicas, posiblemente patogénicas o VUS) que pasa el filtrado de calidad.

Todos los cambios patogénicos, documentados o no, son confirmados por secuenciación Sanger independiente. Los cambios clasificados como probablemente benignos o benignos no son confirmados.

Limitaciones:

La interpretación de los resultados está limitada por la información disponible en la actualidad. En un futuro es posible una mejor interpretación a medida que aumentan los datos y el conocimiento sobre genética humana y enfermedades específicas.

En el test son secuenciados los exones de cada transcrito, 50bp de los extremos de cada exón y las regiones 3' y 5' de los genes. A pesar de esto, el informe no contiene información sobre las zonas no cubiertas de los genes como regiones intrónicas profundas o cualquier exon alternativo actualmente no caracterizado.

Diagnostics in Inherited Blood Diseases BLOODGENETICS S.L.
C/Verge de Guadalupe18, 08950 Esplugues de Llobregat, Barcelona, España. info@bloodgenetics.com

Paciente:	Fecha:	Panel:
P-00124	15/06/2017	NGS-10010v14

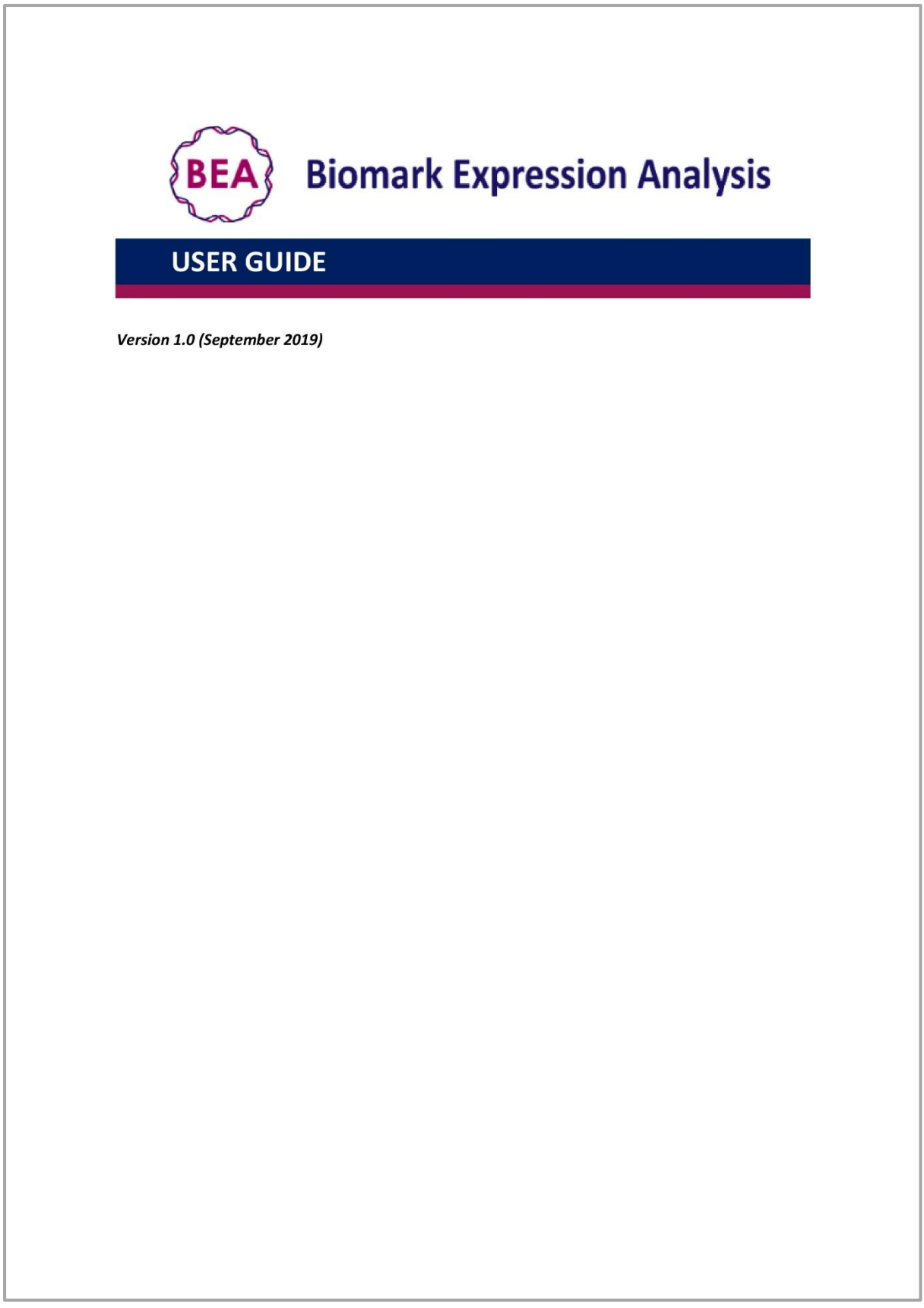
Anexo II: VARIANTES DE SIGNIFICACIÓN INCIERTA ENCONTRADAS EN EL ESTUDIO

	GEN	cDNA /PROTEÍNA	CIGOSIDAD	MAF	BASE DE DATOS	INTERPRETACIÓN Y OBSERVACIONES
Variantes en CDS						No se detectó ninguna
Variantes en región intrónica**						No se detectó ninguna
Variantes en 5'UTR						No se detectó ninguna
Variantes en 3'UTR***						No se detectó ninguna

· **La evidencia disponible actualmente es insuficiente para descartar un papel causante o contributivo en la enfermedad de estas variaciones. Sin ensayos funcionales y estudios genéticos familiares no es posible conocer la relevancia precisa de las VUS encontradas.**
· **No se reportan aquellas variantes conocidas como benignas o probablemente benignas, aquellas que poseen una frecuencia poblacional alta (MAF>3%), variantes sin cambio de aminoácidos (sinónimas) y las que se encuentran en regiones intrónicas profundas** (+/- 50bp) o regiones reguladores del 3'UTR más allá de 100bp*****.

Diagnostics in Inherited Blood Diseases BLOODGENETICS S.L.
C/Verge de Guadalupe18, 08950 Esplugues de Llobregat, Barcelona, España. info@bloodgenetics.com

Instruction of use of BEA program



BEFORE USING BEA:

The chip run folder contains the following folder and files. The ChipRun.bml file contains the results of the Fluidigm run, and it is the one we will use to analyze the data.

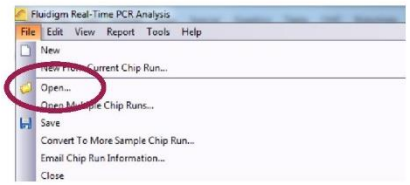
Type	Name	Description
Folder	Data	Folder with TIFF-images, used to extract RT-qPCR data
Folder	Cals	Folder with TIFF-images, used to set exposure times and focus
File	ChipRun.bml	File containing data from the run, used for analysis
File	ChipRunLog.txt	Log file over the chip run, used for troubleshooting and support

Before starting the BEA program for the gene expression analysis, export your Fluidigm results from the ChipRun.bml file (raw data) as a "Results table data" as .csv text file, using the Fluidigm Real-Time PCR Analysis software. To obtain this file:

0. Raw data preprocessing and extraction

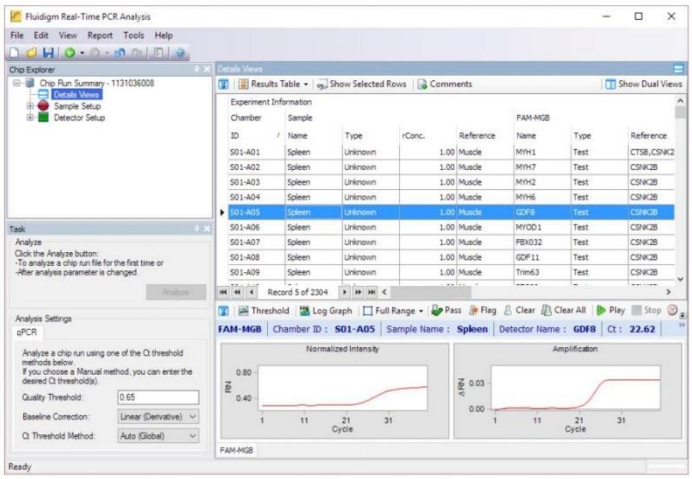
1- Open the Fluidigm **BioMark Real-Time PCR Analysis software**

2- Select **File > Open**, or click **Open a Chip Run**



3- Select the chip run file (.blm exention) you want to analyze and click **Open**

4- Click **Details Views**.



5- Click **Analyze**

IMPORTANT! Make sure that the **baseline correction** in the Fluidigm RT-PCR Analysis Software is set to **Linear** and that the **Ct Threshold Method** is set to **Auto (Global)** under "Analysis views". When changing this setting you need to click the Analyze button before exporting the data. Auto (Global) Threshold is used to avoid differences in dCq values between biomarkers depending on where the threshold is set for each run.

6- **OPTIONAL**: Setting Up a Sample Plate. If you want to set the sample types (Unknown, Reference, NTC, NAC, Blank, or Standard) , check manufacturer's instruction (www.fluidigm.com). Note that this step is not necessary, the reference samples will be also assigned in the sample information file.

7- **OPTIONAL**: Setting up a Detector (Assay) Plate. Select Reference Assays is recommended.

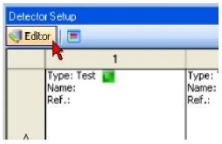
Click Detector **Setup > New**.

Choose the appropriate **Container type** and **Container format.**

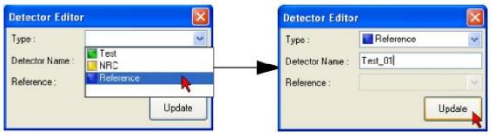
Click **OK** to open the **Detector Plate screen**.

Select cells containing the reference genes

Click **Editor**



Select the appropriate type (Reference) and click **update.**

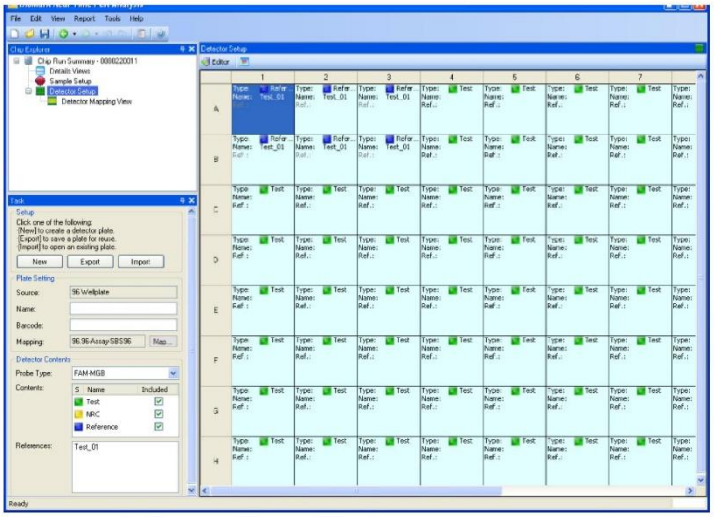


Close the Detector Editor

Click the **Open Mapping File icon**.

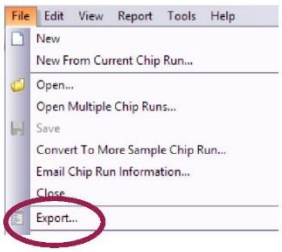
Double-click either left or right sample mapping file.

Your selection is displayed in light blue in the Mapping Viewer.

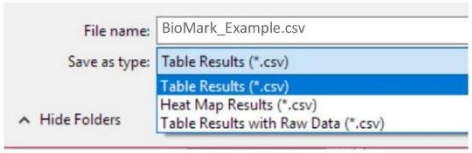


Your detector plate setup is complete.

8. Select **Export** in the **File** menu to export your data.



9. **Name** your data file in the File name box.



10. Select **Table results (*.csv)** in the Save as type menu.

11. Click **Save**. The exported ChipRun_Table_Resutls.csv file can now be used for analysis with the Bea program.

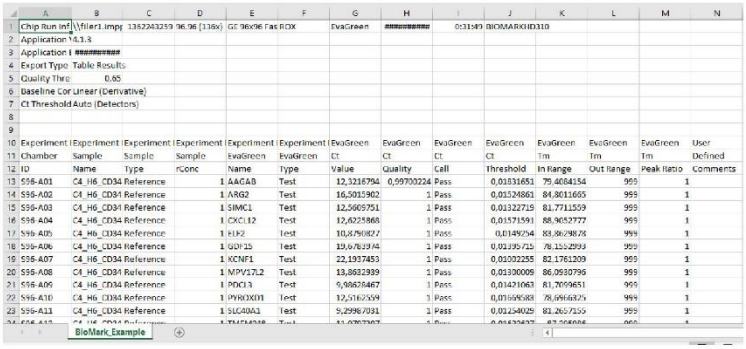
1. Prepare input files

Two input files are required before running Bea:

Input file #1: Ct-values results table (.csv)

This file is a coma separated value file containing the results from BioMark Fluidigm platform after preprocessing the raw data using the Real-Time PCR Analysis software (Fig. xx).

- Sample names must have no spaces, start with a letter, and can contain letter, numbers and certain special characters (dot or underscore).
- Replicates or triplicates must have exactly same sample name.
- The data must be separated by semicolon (;), and number are quoted.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Chip Run Inf	\\filer1.impg	1362243259	96.96 [136x]	GE 96x96 Fas	ROX	EvaGreen	##########	031549	BIOMARKHD310				
2	Application \	4.1.3												
3	Application I	##########												
4	Export Type	Table Results												
5	Quality Thre	0.65												
6	Baseline Cor	Linear (Derivative)												
7	Ct Threshold	Auto (Detectors)												
8														
9														
10	Experiment	Experiment	Experiment	Experiment	Experiment	Experiment	EvaGreen	EvaGreen	EvaGreen	EvaGreen	EvaGreen	EvaGreen	EvaGreen	User
11	Chamber	Sample	Sample	Sample	EvaGreen	EvaGreen	Ct	Ct	Ct	Ct	Tm	Tm	Tm	Defined
12	ID	Name	Type	nConc	Name	Type	Value	Quality	Call	Threshold	In Range	Out Range	Peak Ratio	Comments
13	S96-A01	C4_H6_CD34	Reference		1 AACAB	Test	12,3216794	0,99700224	Pass	0,01831651	79,4084154	999	1	
14	S96-A02	C4_H6_CD34	Reference		1 ARG2	Test	16,5015902		1 Pass	0,01524861	84,8011665	999	1	
15	S96-A03	C4_H6_CD34	Reference		1 SIMC1	Test	12,5609751		1 Pass	0,01322719	81,7711559	999	1	
16	S96-A04	C4_H6_CD34	Reference		1 CXCL12	Test	12,6225868		1 Pass	0,01571591	88,9052777	999	1	
17	S96-A05	C4_H6_CD34	Reference		1 EIF2	Test	10,8790827		1 Pass	0,0149754	83,8629878	999	1	
18	S96-A06	C4_H6_CD34	Reference		1 GDF15	Test	19,6761944		1 Pass	0,01330715	78,1552999	999	1	
19	S96-A07	C4_H6_CD34	Reference		1 KCNF1	Test	22,1937453		1 Pass	0,01002255	82,1761209	999	1	
20	S96-A08	C4_H6_CD34	Reference		1 MPV17L2	Test	13,8632939		1 Pass	0,01300009	86,0930796	999	1	
21	S96-A09	C4_H6_CD34	Reference		1 PDCL3	Test	9,98626467		1 Pass	0,01421063	81,7099651	999	1	
22	S96-A10	C4_H6_CD34	Reference		1 PYROXD1	Test	12,5162559		1 Pass	0,01699683	78,6666826	999	1	
23	S96-A11	C4_H6_CD34	Reference		1 SLC40A1	Test	5,29987031		1 Pass	0,01254029	81,2657159	999	1	

BioMark_Example

Input file #2: Sample Information file (.csv)

The SIF is a comma separated value file containing three columns:

- **SampleName**: Sample name used in the experiment. Remember that if you include replicates, those samples must be named identically. The program is letter case sensitive.
- **SampleType**: Sample Type as described in the results-table (Unknown, Reference, NTC, NAC, Blank, or Standard)
- **GROUP**: Group of analysis (e.g. “Case” vs. “Control”; “Healthy” vs.” Treated” vs.” No-treated”). Minimum two groups and maximum three groups can be assigned. Group names only can contain letters, numbers and some special characters (dot or underscore). First character must be a letter, and no spaces are permitted.

	A	B	C
1	SampleName	SampleType	GROUP
2	C4_H6_CD34	Reference	control
3	R9_ID11_RCMD	Unknown	nonRS
4	R24_ID31_RCMD	Unknown	nonRS
5	R1_ID1_RCMD	Unknown	nonRS
6	R16_ID19_RAEB	Unknown	nonRS
7	R25_ID32_RAEB	Unknown	nonRS
8	R8_ID9_RAEB	Unknown	nonRS
9	R17_ID20_RCMD	Unknown	nonRS
10	C3_H5_CD34	Reference	control
11	R10_ID13_RS	Unknown	RS
12	R23_ID30_RS	Unknown	RS
13	R2_ID2_RS	Unknown	RS
14	R15_ID18_RS	Unknown	RS
15	R26_ID33_RS	Unknown	RS
16	R7_ID8_RCMD	Unknown	nonRS
17	R18_ID22_RS	Unknown	RS
18	C2_H3_MNC	Reference	control
19	R11_ID14_RS	Unknown	RS
20	R22_ID29_RS	Unknown	RS
21	R3_ID4_RS	Unknown	RS
22	R14_ID17_RS	Unknown	RS
23	R27_ID34_RCMD	Unknown	nonRS

sample_information_file

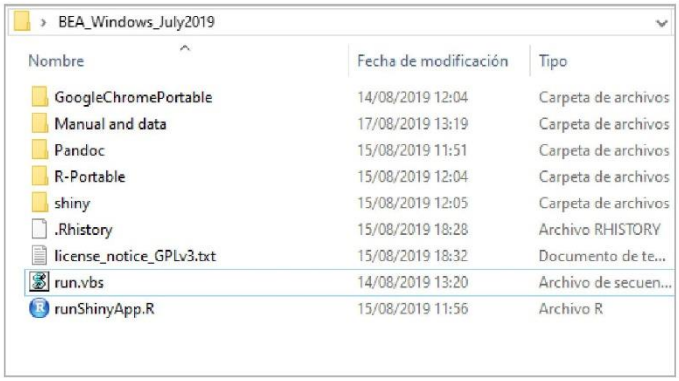
2. Launch the program (only for windows)

1. Download and unzip the compressed file into your desktop folder. **Do not change (sub)folder names.**

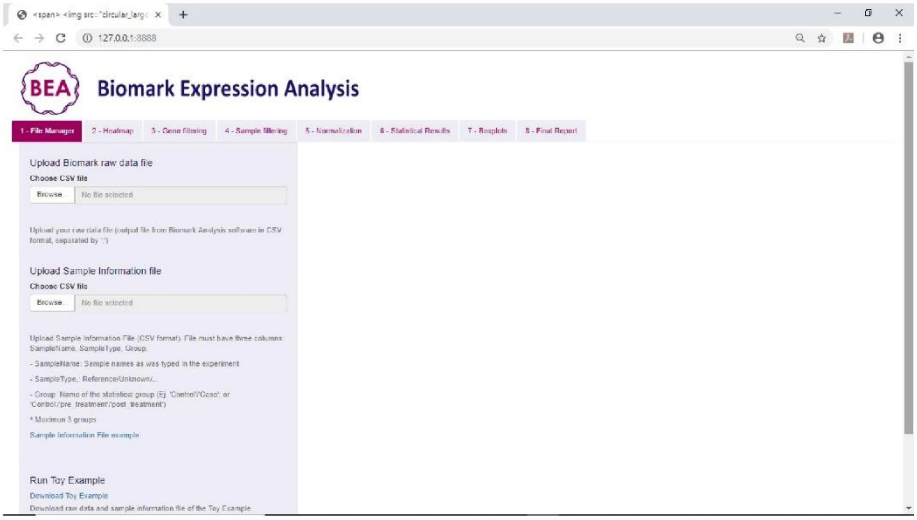
2. Open the "**BEA_Windosws_July2019**" folder.

3. Double-click the launcher icon (**run.vbs**).

IMPORTANT! The program takes ~15 seconds to initiate BEA. Please be patient and DO NOT initiate another job. If you launch BEA from an USB flash drive or an external HDD, it could take longer. Please be patient.

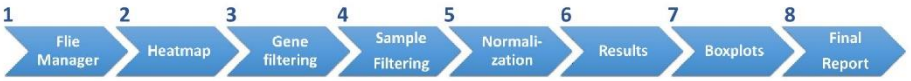


BEA is opened in a portal google chrome page. Internet is not required.



3. Analyze your data using BEA: Biomark Expression Analysis

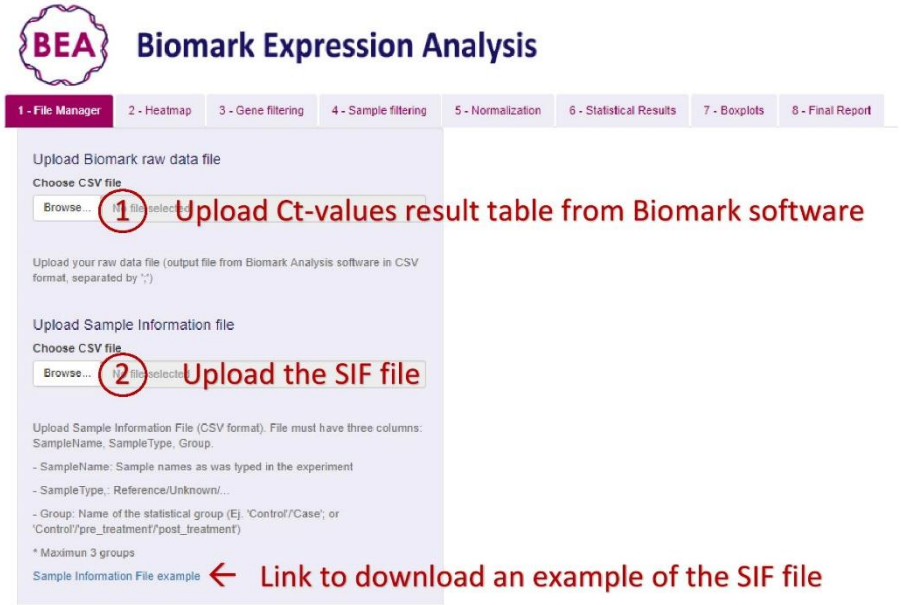
The program is structured in eight steps that are incorporated in eight tabs respectively



"1- File Manager" tab

The two required input files are uploaded to BEA using this interface.

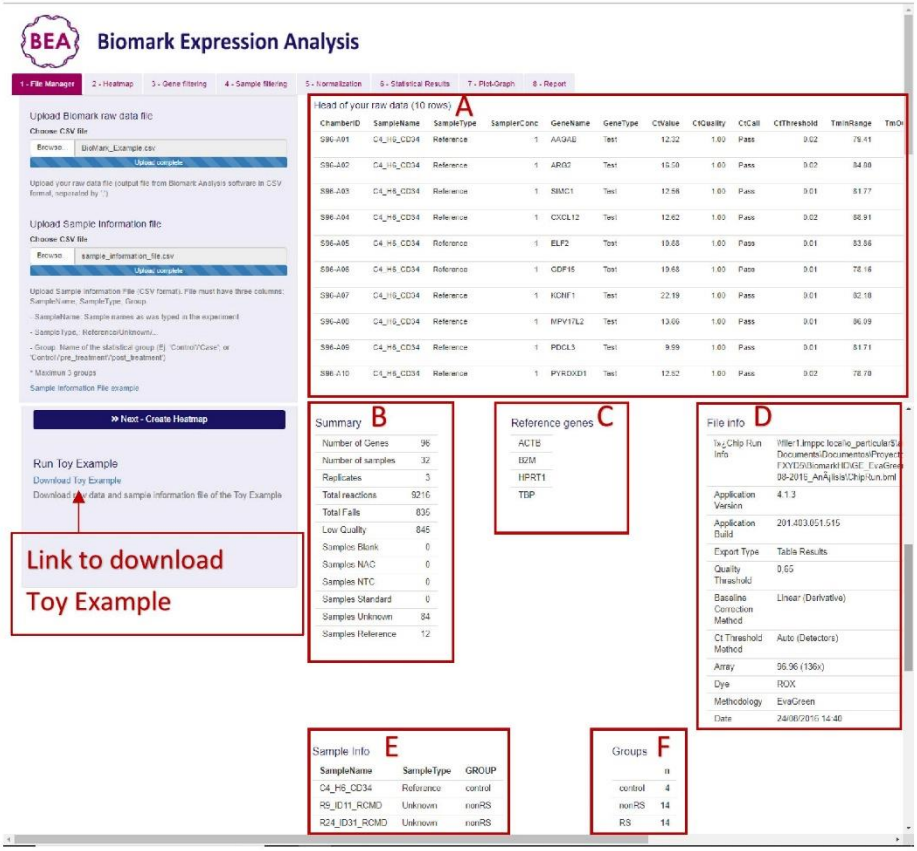
1- Insert the Ct-values results table (Biomark raw data).
2- Insert the SIF. An example of the sample information file can be downloaded through a link that redirects the user to another webpage



If the input files are uploaded properly, some tables appear automatically:

A- Head of the raw data input file from BioMark-Fluidigm preprocessed output;
B- Summary table of the experiment (samples, assays…)
C- Table with the reference genes included;
D- Table with information about the experiment and the chip;
E- Table showing the sample information file uploaded;
F- Table with groups names and number of samples per group.

From this page, users can also download a toy example to train themselves and understand the analysis. The toy example is a zip folder containing the two required input files (table results data, the sample information file), a standard protocol, a final report example and a README file with all information about the repository content.



ERROR MESSAGES! Some error messages can appear if:

Error message	Meaning
Wrong input file, please put a valid file.	The Ct-value results table is wrong (incorrect format…)
Wrong sample information file, please put a valid file.	The SIF file is wrong (incorrect number of columns or incorrect format)
No reference genes found	No reference genes have been found in the raw data. The user must select the reference genes from the list that appears in the left side of the page
Sample Names do NOT match, please check input files	The sample names from SIF file does not match with the sample names of the raw data file.

Only when the two input files are correct, the reference genes are set and the sample names from both inputs' files match, the next-step button ("Next – Create Heatmap") appears and can be clicked to go to the Heatmap tab.

"2- Heatmap" tab

BEA provides in the second step a general visualization of the experiment with a gene expression heatmap. The Ct value for all genes and samples (individual replicates) are plotted ranged from high expression or low Ct value (red) to low expression or high Ct value (bluish). A black square indicates no Ct value or a value outside of the spectrum range

- **A- Sort genes by:** Gene axe can be sorted alphabetically, or based on the expression (mean Ct by gene)
- **B- Sort sample by:** Sample axe can be sorted alphabetically or by groups (how is in the picture)
- **C- Download plot:** Button to download the heatmap plot in a pdf format.
- **D- Next:** Button to move to the next step

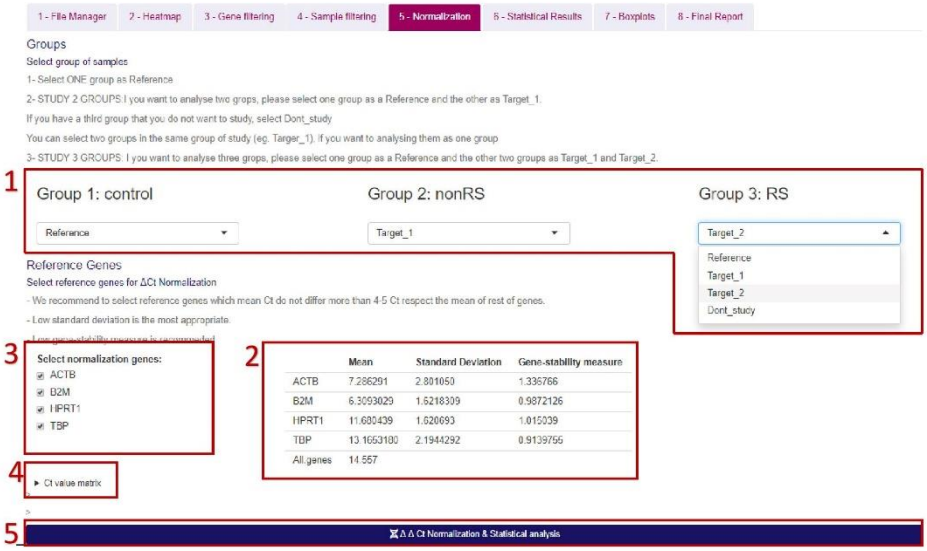
"3- Gene filtering" tab:

Some genes with low efficiency can be filtered out of the analysis to reduce noise in the analysis. The number of fail assays per gene is exposed in a barplot.

1- Select threshold to filter out those genes with higher number of fail than the selected cutoff (by default is 75%).
2- Barplot can be sorted alphabetically or numerically.
3- User can remove other genes from the analysis by selecting those genes from the "List of genes" or by clinking over the plot.
4- Genes discarded are displayed in this box
5- Click Next.

"5- Normalization" tab:

In this tab user select all the parameters required to perform delta delta Ct normalization (Livak).



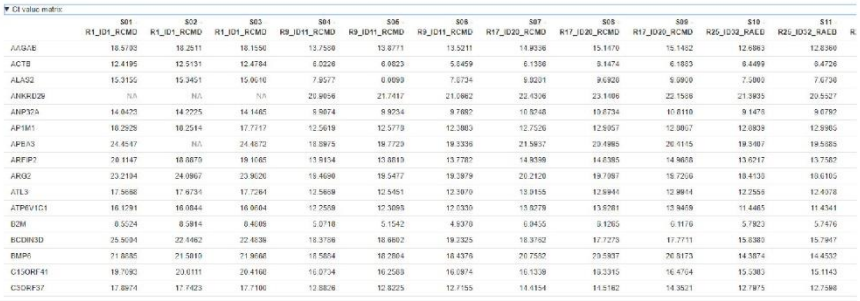
1- Set the group of samples you want to analyze

Tag	Meaning
"Reference"	Control group
"Target_1"	Group to test
"Target_2"	Second group to test
"Dont_study"	Don´t study this group and perform a two-groups analysis

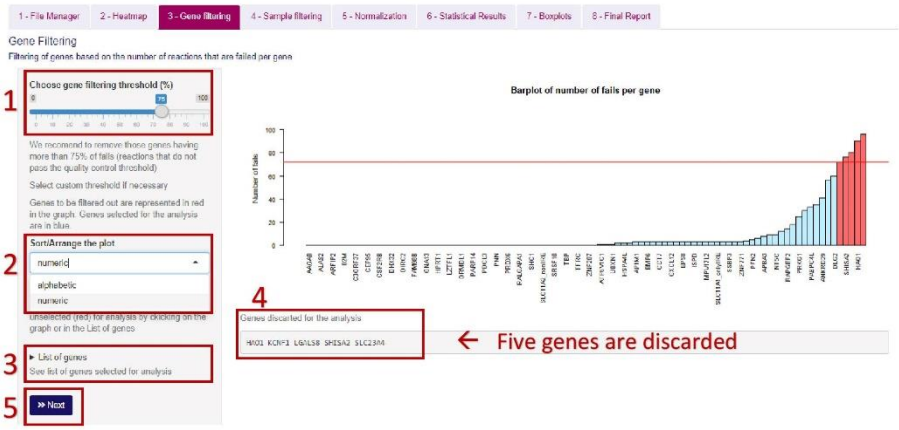
2- Check the reference genes table to choose appropriate normalization gene.

3- Select the normalization gene/s. More than one reference gene can be selected for normalization.

4- Ct value matrix can be displayed.



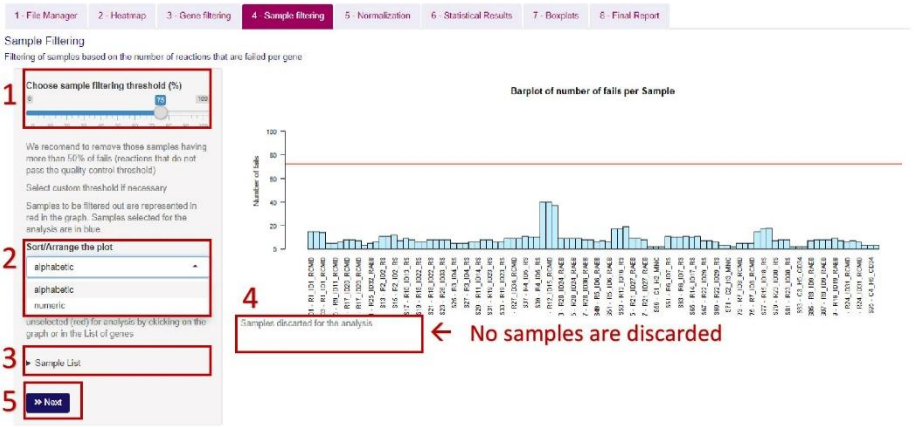
5- Click "ΔΔCt Normalization & Statistical analysis" to move to next step.



The barplot with the number of fails per gen can be rearranged by sorting gene axis alphabetically (A to Z) or numerically (low to highest abundance of fails). Discarded genes are colored in red on the barplot and typed under the plot.

"4- Sample filtering" tab:

The filtration of samples with high count of fails can be done in this tab. The filtering process is performed in the same way that in "Gene filtering" tab. A barplot is displayed, and a list of samples can be checked.



1- Select threshold to filter out those samples with higher number of fail than the selected cutoff (by default is 75%).
2- Barplot can be sorted alphabetically or numerically.
3- User can remove other samples from the analysis by selecting those sample names from the "List of genes" or by clinking over the plot.
4- Samples discarded are displayed in this box
5- Click Next.

"6- Statistical Results" tab

Genes are normalized (ΔΔCt Normalization) and analyzed following the statistical methodology described in the Statistics panel. Results tables are shown in the Results panel.

 # Statistical panel: The statistical testing and parameters can be customized:

1- Select statistical analysis method (as default is "Test Normality (Shapiro Test)").

Statistical method	Meaning
Test Normality (Shapiro Test)	Test normality and apply suitable statistical test for each gene: ANOVA/ T-test when the data follow a normal distribution; Kruskal-Wallis/ Wilcoxon when the data do not follow a normal distribution
Parametric Test (ANOVA/T-test)	Assume normality
Non-parametric Test (Kruskal-Wallis/ Wilcoxon)	Assume non-normality

2- Select multiple comparison correction method (as default is false discovery rate – "fdr").

Multiple comparison method	Meaning
fdr	False discovery rate
bonferroni	Bonferroni
none	Avoid multiple testing correction

3- Select Fold Change Threshold (as default is 2).
4- Select p value threshold (as default is 0.05).

 # Results panel: The results table contain the statistical results of the tested groups for each gene, including Shapiro-Wilk test, statistical test used, Fold Change, and adjusted p.value among others. Significant p.values are printed in bold. Genes that are differentially expressed in at least one group

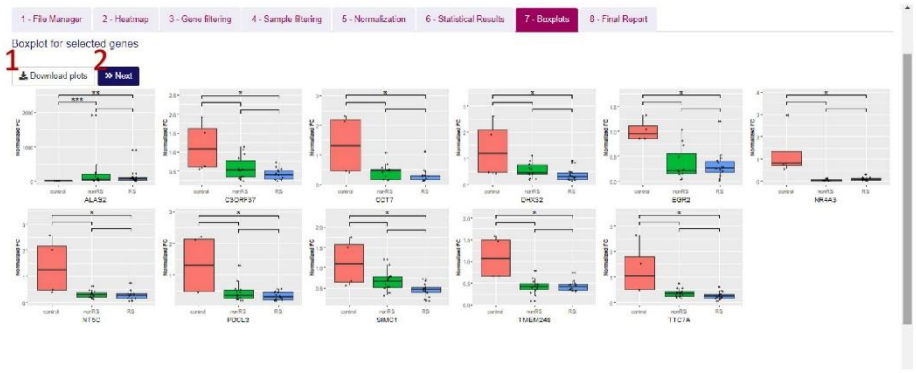
are printed in pink and marked with starts (*** p<0.005; **p<0.01; *p<0.05 or p< p-threshold). Genes are considered differentially expressed based on fold change and p.value.

5- Select on the table those genes that you want to visualize in a boxplot. User can select significant and non-significant genes. Selected genes are now colored in blue on the table.
6- Selected genes are displayed on the left
7- Click "Plot selected genes" button to move to next tab.
8- The results table can be downloaded from this button.
9- The normalized data, without statistics, can be also downloaded

"7- Boxplots" tab

Samples selected in previous tab are represented here by boxplots. One plot per gene is displayed, showing normalized fold change expression of the gene by tested groups. Boxplots can be downloaded in pdf format file.

1- Click "Download plots" to download the boxplots in a PDF format.
2- Click Next to move to final step

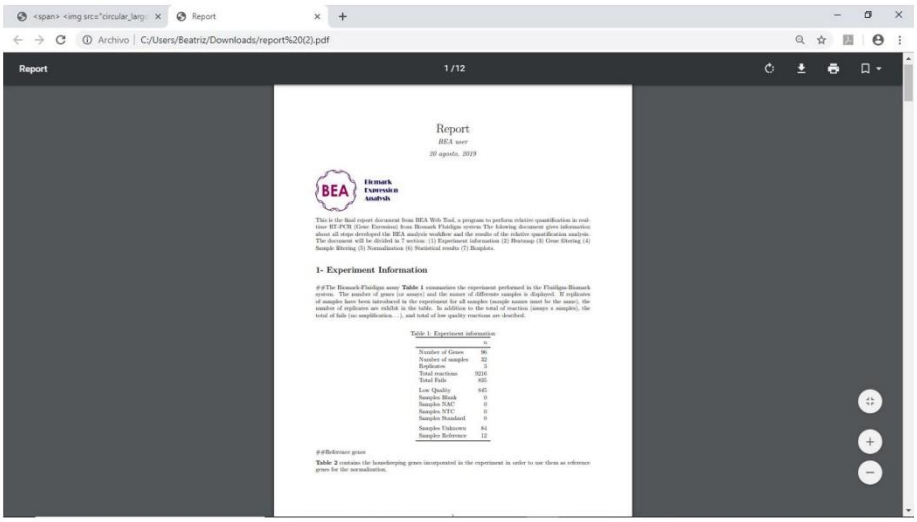


"8- Final report" tab

1- In this tab users have the option of downloading a final report in a pdf format. The report resumes the whole analysis process performed with BEA from all tabs, but the heatmap. Due to dimensionality problems, the heatmap representing the expression of all samples and all genes is not added in the final report. Instead, a heatmap of the significantly expressed genes is included.

2- A download bar gives to user information about how long the process will take to generate the report.

Example of report executed with the BEA program and the testing data:



Example of report performed by BEA program

Report

BEA user

22 septiembre, 2019

This is the final report document from BEA Web Tool, a program to perform relative quantification in real-time RT-PCR (Gene Exression) from Biomark Fluidigm system The folowing document gives information about all steps developed the BEA analysis workflow and the results of the relative quantification analysis. The document will be divided in 7 section: (1) Experiment information (2) Heatmap (3) Gene filtering (4) Sample filtering (5) Normalization (6) Statistical results (7) Boxplots.

1- Experiment Information

##The Biomark-Fluidigm assay **Table 1** summarizes the experiment performed in the Fluidigm-Biomark system. The number of genes (or assays) and the numer of differente samples is displayed. If replicates of samples have been introduced in the experiment for all samples (sample names must be the same), the number of replicates are exhibit in the table. In addition to the total of reaction (assays x samples), the total of fails (no amplification...), and total of low quality reactions are descibed.

Table 1: Experiment information

	n
Number of Genes	96
Number of samples	32
Replicates	3
Total reactions	9216
Total Fails	835
Low Quality	845
Samples Blank	0
Samples NAC	0
Samples NTC	0
Samples Standard	0
Samples Unknown	84
Samples Reference	12

##Reference genes

Table 2 contains the housekeeping genes incorporated in the experiment in order to use them as reference genes for the normalization.

1

Table 2: Reference genes

Ref genes
ACTB
B2M
HPRT1
TBP

##Groups Information from sample information file. **Table3** shows the different groups added in the sample information file, and the number of samples per group (n).

Table 3: Groups

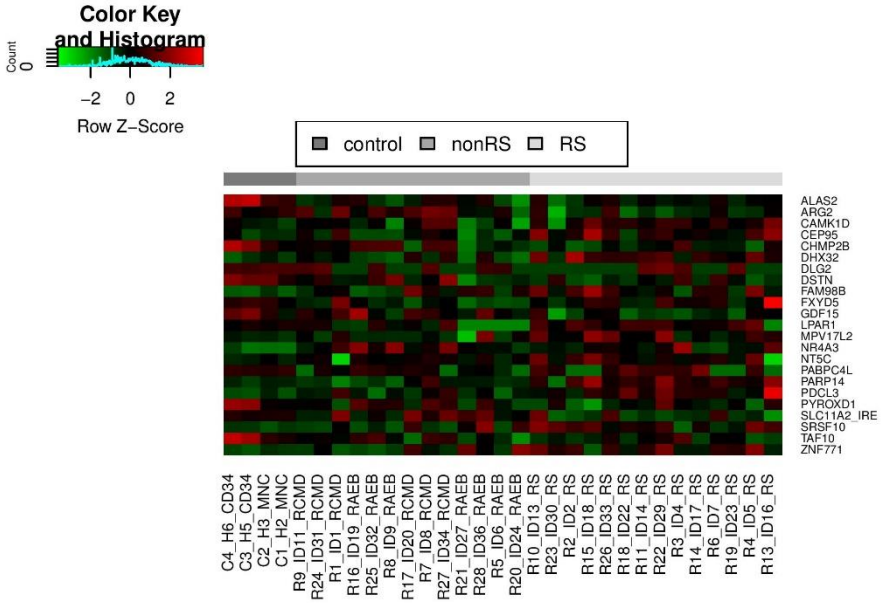
Group	n
control	4
nonRS	14
RS	14

2- Heatmap

##Heatmap

The heatmap plot for the differentially expressed genes are shown below. If replicates are included, the median of the replicates are used.The red, black and green colors represent overexpression, close to average expression and underexpression of the particular gene, respectively. Gene are in rows and samples in columns. Samples are ordered by groups (gray bar on top of the plot).

Warning! The heatmap plot with all samples are not plotted due to too large figure margins. If you want to access to the heatmap plot, please go back to BEA web tool program, select **2 - Heatmap** tab and press the **Download** button to obtain the heatmap in a PDF format. Remember that you can also sort the x- and y-axes of the plot (samples and genes respectively) at your convenience.



3 - Gene Filtering

Genes filtered:

Remove those genes classified as *Fail* by the Real-Time PCR Analysis program from Fluidigm. Remember that those fails are the reactions which amplification curve have low quality and may be problematic or suboptimal for the analysis.

Threshold used = 72

Number of genes removed = 5

Genes removed for the analysis:

```
## [1] "KCNF1"   "SHISA2"  "SLC23A4" "HAO1"    "LGALS8"
```

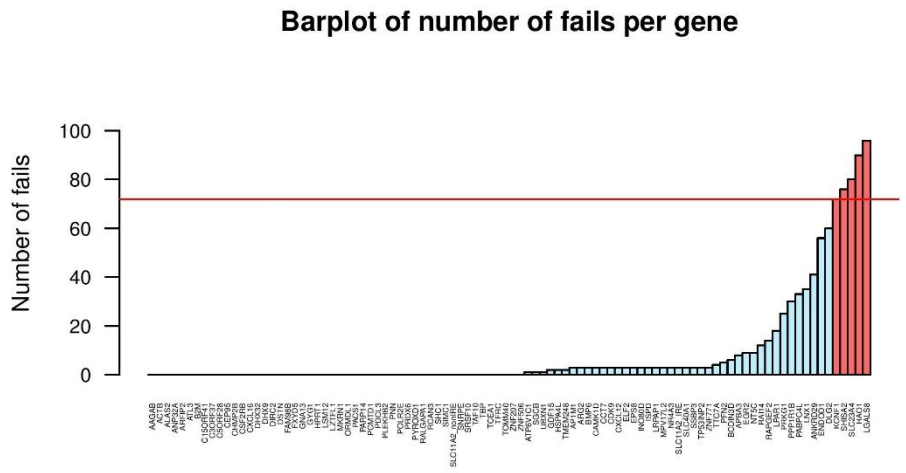


Figure 1. Barplot of number of fails per gene. The graph shows the number of *fails* (low amplification curve) per gene. All the studied genes are displayed on the x-axis. The horizonal red line is the threshold selected for the user. Those genes with more number of fails than the threshold are coloured in red and will be removed from the analysis.

4 - Sample Filtering

Samples filtered:

Remove those samples classified as *Fail* by the Real-Time PCR Analysis program from fluidigm. Remember that those fails are the reactions which amplification curve have low quality and may be problematic or suboptimal for the analysis.

Threshold used = 33.6

Number of samples removed = 3

Samples removed:

```
## [1] "S40 - R12_ID15_RCMD" "S41 - R12_ID15_RCMD" "S42 - R12_ID15_RCMD"
```

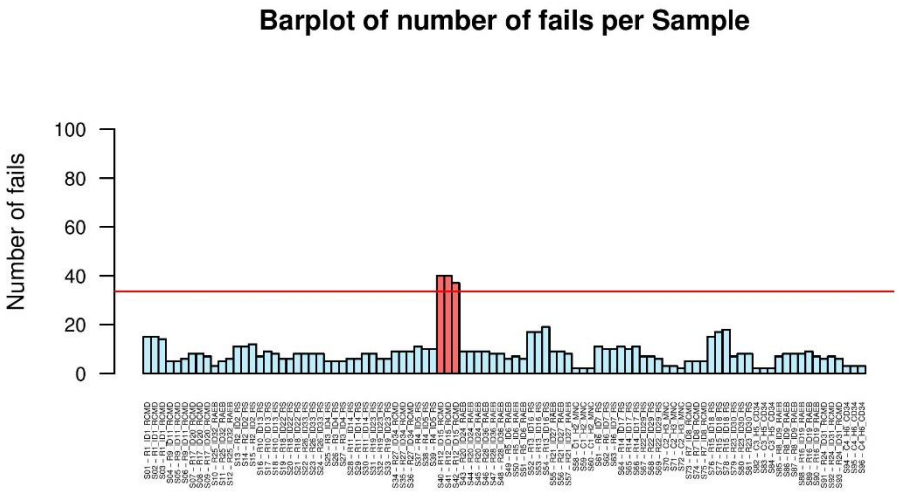


Figure 2. Barplot of number of fails per sample. The graph shows the number of *fails* (low amplification curve) per sample. All the studied samples are displayed on the x-axis. The horizonal red line is the threshold selected for the user. Those samples with more number of fails than the threshold are coloured in red and will be removed from the analysis.

5 - Normalization

Groups and housekeeping genes

The relative quantification is performed by using the $\Delta\Delta_{Ct}$ method (Livak and Schmittgen, 2001). To achieve the relative expression results, appropiate normalization must be performed to control for experimental error. To perform the normalization, at least one reference gene must be selected (more than one are accepted), and the goups of analysis must be defined. **Table 4 & 6** shows the groups of analysis and reference genes selected by the user. **Table 5** displays the genes predefined as *Reference*, includig the **Mean**, the **Standard Deviation** and the **Gene-stability measure (M)** (Vandesompele et al, 2002) for each gene. We recommend to select reference genes which mean Ct do not differ more than 4-5 Ct respect the mean of rest of genes, with low standard deviation and low gene-stability measure.

Table 4: Groups

Groups	Group_Name	Analysis
Group 1	control	Reference
Group 2	nonRS	Target_1
Group 3	RS	Target_2

Table 5: Reference genes

	ACTB	B2M	HPRT1	TBP	All.genes
Mean	6.941313	6.1038731	11.4914038	12.893182	14.4009
Standard Deviation	2.062040	1.1623020	1.2473722	1.603699	
Gene-stability measure	1.156331	0.8878971	0.8842936	0.817371	

Table 6: Reference genes selected

Normalization_genes
HPRT1
TBP

6 - Statistical results

Selected statiscits parameters

The **Table 7** shows the parameteres selected to perfmor statistic analysis.

Table 7: Statistics parameters

Statistics	Value
Statistic method	Test Normality (Shapiro Test)
Multiple comparison correction	none
Fold Change threshold	1.5
P.value threshold	0.05

Statistic methods:

- **Test Normality (Shapiro Test):** Test of normality Shapiro-Wilk test is performed for each gene data set. If that gene follow a normal distribution, a parametric test is performed (ANOVA + Post Hoc if 3 groups are analysed; or T-test for two groups). On the contrary, a non-parmetric test is performed (Kruskal-wallis or Wilcoxcon test) when the data set do not follow a normal distribution.
- **Assume normality (ANOVA / T-test):** Test of normality is not performed and all the genes are analysed with parametric test.
- **Assume non-normality (Kruskal-Wallis / Wilcoxon):** Test of normality is not performed and all genes are analysed with non-parametric test.

Multiple comparison correction:

- **fdr:** False Discovery Rate correction.
- **bonferroni:** Bonferroni correction
- **none:** No correction has been made.

Fold Change threshold

The Fold Change threshold is the minimun absolute value to consider any gene differential expressed.

P.value threshold

The P.value threshold is the maximun P.value to consider a gene **significatelly** and differentially expressed.

Number of significative genes:

```
## [1] "23"                " significant genes"
```

Table 8: Differentially expressed genes

				Ref vs Target 1			Ref vs Target 2			Target 1 vs Target 2			
Gene	Shapiro	Method	P.val	FC	P.val	*	FC	P.val	*	FC	P.val	*	Significative
ALAS2	0.228	One-Way ANOVA	0.000	128.198	0.000	***	90.030	0.000	***	-1.424	0.861		*
ARG2	0.444	One-Way ANOVA	0.012	-1.002	1.000		5.986	0.118		5.997	0.014	*	*
CAMK1D	0.020	Kruskal-Wallis	0.025	-1.030	0.871		-1.979	0.044	*	-1.922	0.044	*	*
CEP95	0.633	One-Way ANOVA	0.004	-1.018	0.995		-1.587	0.060		-1.559	0.006	**	*
CHMP2B	0.786	One-Way ANOVA	0.007	2.179	0.059		3.025	0.005	**	1.388	0.307		*
DHX32	0.959	One-Way ANOVA	0.000	-1.163	0.800		-2.229	0.005	**	-1.917	0.001	***	*
DLG2	0.683	One-Way ANOVA	0.016	-1.075	0.982		-3.700	0.028	*	-3.443	0.024	*	*
DSTN	0.304	One-Way ANOVA	0.010	1.844	0.040	*	2.186	0.007	**	1.186	0.545		*
FAM98B	0.858	One-Way ANOVA	0.014	-1.607	0.096		-1.976	0.011	*	-1.230	0.354		*
FXYD5	0.000	Kruskal-Wallis	0.014	2.524	0.034	*	1.169	0.277		-2.159	0.034		*
GDF15	0.945	One-Way ANOVA	0.003	2.345	0.652		17.468	0.015	*	7.450	0.012	*	*
LPAR1	0.443	One-Way ANOVA	0.029	1.541	0.762		-2.260	0.357		-3.482	0.025	*	*
MPV17L2	0.187	One-Way ANOVA	0.027	-1.062	0.979		-1.838	0.138		-1.730	0.037	*	*
NR4A3	0.010	Kruskal-Wallis	0.006	-19.086	0.004	***	-10.387	0.004	***	1.838	0.239		*
NT5C	0.116	One-Way ANOVA	0.019	-1.788	0.303		-2.988	0.021	*	-1.672	0.152		*
PABPC4L	0.954	One-Way ANOVA	0.003	3.113	0.192		-2.191	0.450		-6.819	0.002	***	*
PARP14	0.007	Kruskal-Wallis	0.001	1.025	0.956		-1.958	0.012	*	-2.006	0.001	***	*
PDCL3	0.008	Kruskal-Wallis	0.002	-1.367	0.102		-2.179	0.004	***	-1.594	0.016	*	*
PYROXD1	0.435	One-Way ANOVA	0.027	1.548	0.130		1.855	0.021	*	1.198	0.446		*
SLC11A2_IRE	0.047	Kruskal-Wallis	0.013	-1.013	0.785		1.786	0.152		1.809	0.012	*	*
SRSF10	0.930	One-Way ANOVA	0.018	-1.327	0.524		-1.973	0.033	*	-1.486	0.075		*
TAF10	0.856	One-Way ANOVA	0.003	2.183	0.005	**	2.333	0.002	***	1.069	0.902		*
ZNF771	0.209	One-Way ANOVA	0.033	-1.845	0.391		-3.253	0.040	*	-1.763	0.179		*

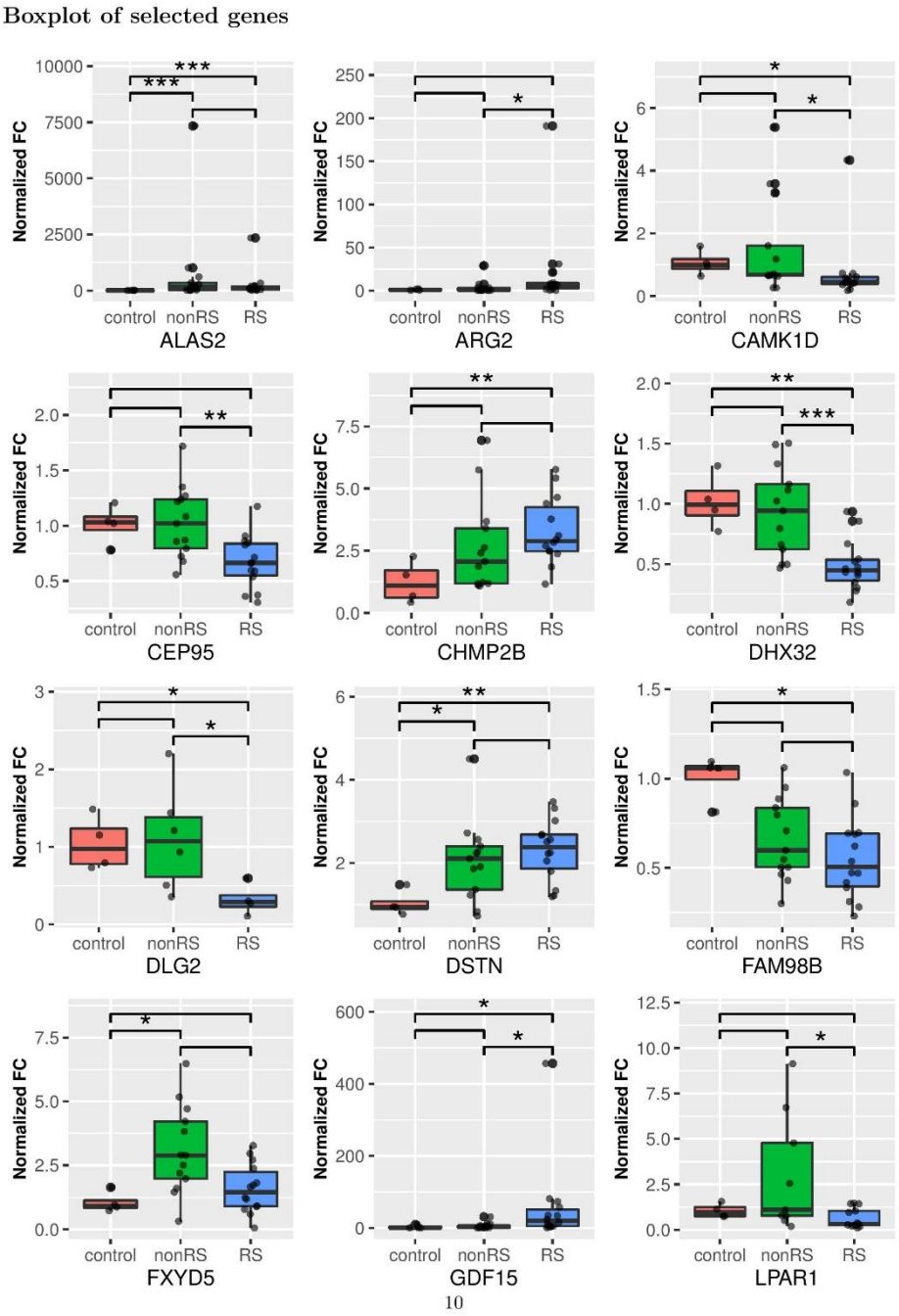
Note:
FC = Fold Change
*** p<0.005; **p<0.01; *p<0.05 or p< p-threshold

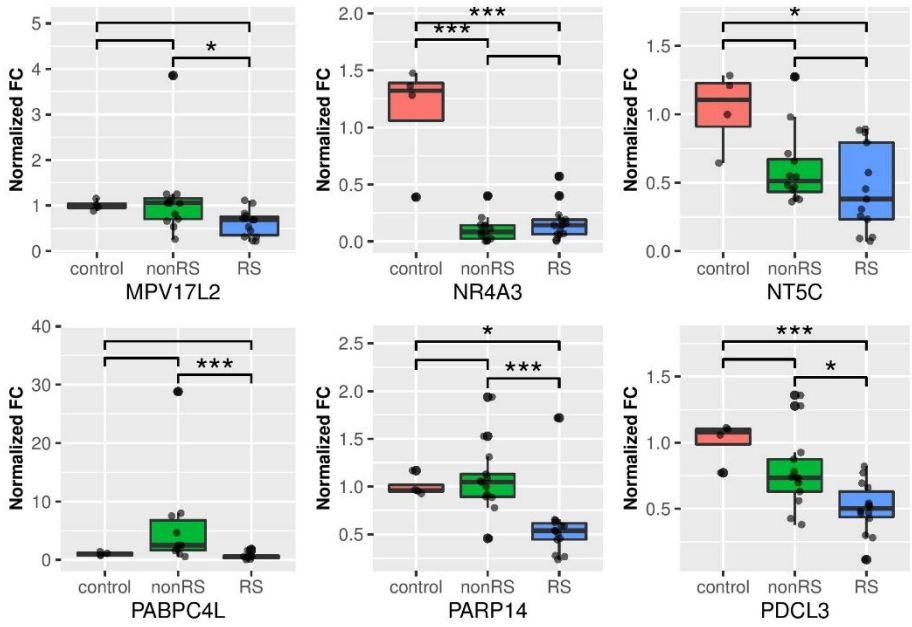
7- Boxplots

Genes selected for boxplot

Table 9: Genes selected

sel_genes
ALAS2
ARG2
CAMK1D
CEP95
CHMP2B
DHX32
DLG2
DSTN
FAM98B
FXYD5
GDF15
LPAR1
MPV17L2
NR4A3
NT5C
PABPC4L
PARP14
PDCL3
PYROXD1
SLC11A2_IRE
SRSF10
TAF10
ZNF771


Boxplot of selected genes
Normalized FC
control
nonRS
RS
ALAS2
ARG2
CAMK1D
CEP95
CHMP2B
DHX32
DLG2
DSTN
FAM98B
FXYD5
GDF15
LPAR1


Normalized FC
control
nonRS
RS
MPV17L2
NR4A3
NT5C
PABPC4L
PARP14
PDCL3

11

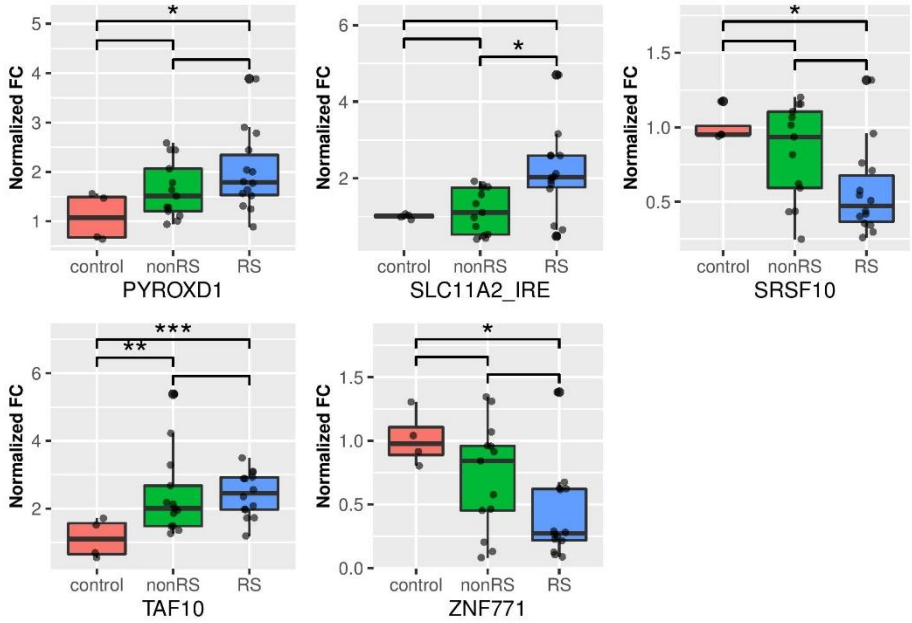


Figure 3. Boxplot of selected genes. Relative expression of selected genes for the studied groups.

ACKNOWLEDGMENTS

ACKNOLEDGEMENTS

Una de las famosas frases del jugador de baloncesto Michael Jordan es la siguiente: "*Talent wins games, but teamwork and intelligence wins championships*" (El talento gana partidos, pero el trabajo en equipo y la inteligencia ganan campeonatos). Y es que esta tesis no se hubiese conseguido sin el trabajo y el apoyo de muchas personas, tanto en el mundo laboral como en lo personal. Por eso, me gustaría dedicarles las siguientes palabras.

En primer lugar, a mis tres directores de tesis: Mayka Sanchez, Malu Calle y Cristian Tornador. **Mayka**, gracias por la fe y confianza que ha depositado en mi durante todo momento, por creer en mi trabajo y valorar siempre mis esfuerzos. Gracias por guiarme en este doctorado, por su inestimable aportación científica y su asesoramiento. Gracias por su paciencia, disponibilidad y diligencia, especialmente estos últimos meses.

Malu, gracias por ser una mujer científica digna de admirar. Gracias por su profesionalidad, su audacia y saber estar. Gracias por todos los conocimientos bioinformáticos y estadísticos que me ha aportado, y por su gran paciencia y constancia para que los asimilase. Gracias por su apoyo y amabilidad en todo momento.

Cristian, gracias por la fe, la confianza y admiración que ha mostrado siempre sobre mi. Gracias por su apoyo y su comprensión en todo momento. Pero sobre todo gracias por todos lo conocimientos sobre emprendimiento y dirección de empresas que he aprendido trabajando a su lado.

Agradecer a todos los **colaboradores** que han hecho posible el desarrollo de esta tesis:

A Dr. Francesc X. Real, a Nuria Malats y Albert Font por proporcionarme la oportunidad de participar en el proyecto de medicina predictiva personalizada de cáncer de vejiga.

Al Dr. Robeto Gatterman, por su colaboración el proyecto de Síndrome Mielódisplasico.

Al Plan de Doctorados industriales de la Generalitat de Catalunya, por el soporte financiero.

Al Instituto de Investigación Josep Carreras, la Universidad de Vic, a Whole Genix, a BloodGenetics, y a la Fundación Teresa Moretó por el soporte científico aportado.

Agradecer a todos los "Maykitos" y "Exmaykitos": Pep Fita, Jessica Aranda, Jorge Couso, Anna Barqué, Estrella Gómez, Sònia Sabaté, Clara Esteban, Laura Zalba, Marc Vila, Ferrán Celma, y Verónica Venturi. Gracias por vuestra ayuda, por la ciencia que hemos compartido y por la excelente atmósfera que creasteis en el laboratorio.

Pep (o Pepito), gracias por ser un gran científico en potencia, por ser mi enciclopedia del laboratorio, y por tenerte al lado durante gran parte del doctorado. Gracias por todos los buenos momentos, por los disfraces de princesa y de sandía, por las canciones de OT, por las comidas compartidas de Mercadona, y los cotilleos inventados en el Can Carmen. La lista de bueno momentos es interminable.

Gracias **Jessi**, por ser una mezcla de amiga, compañera, mamá, psicóloga y "payaseta". Todas estas facetas te hacen una mujer única. Pero sobre todo, gracias por la alegría y el humor que aportabas al grupo cada mañana.

Gracias **Laura**, por tu entusiasmo y amor por la ciencia. Por los murales sobre cristal y los pendientes de navidad, por darlo todo en los karaokes, y por los chocolates compartidos. Pero sobre todo por tu simpatía, amabilidad y por ser la mejor casera que he tenido.

Gracias **Jorge** por todos los consejos y recomendaciones sobre el doctorado, y las largas conversaciones sobre el futuro laboral. Gracias **Anna** por tu dulzura y tu compañerismo. Gracias **Clara** por los buenos momentos en el laboratorio (los baños de bromuro, los empujones a máquinas expendedoras…). Gracias **Sònia** por ser "tipo de incógnito" y hacerme reír sin parar. Gracias **Estrella** por los buenos momentos tanto dentro como fuera del lab. Gracias **Marc** por tus consejos y tu sabiduría. Gracias **Ferrán** por tu alegría y tu entusiasmo. Gracias **Vero** por tu colaboración estos últimos meses de doctorado.

Agradezco también a mis compañeros de "holding" empresarial: Sandra Martínez, Luca Badessi, Xenia Ferrer, Santiago Perez, Nuria Nabau y Oscar. A todos los vosotros, ya sabési que me tenéis para lo que queráis, y para ello solo hay que pronunciar las palabras mágicas, "que te ayude Bea".

Sandra, gracias por convertirte en una persona tan importante en mi vida. Gracias por hacerme ver que todavía existen personas buenas de verdad, personas de plena confianza y de buen corazón. Gracias por ser un gran apoyo y por estar siempre dispuesta a ayudar.

Luca, gracias por traer siempre esa alegría y entusiasmo al lab. Pero sobre todo, gracias por todos los desayunos en el can Carmen, por los viajes a "las fuera de Barcelona", y por convertirme en tu fiel consejera del amor. Grazie per essere un buon amico.

Xenia, gracias por toda tu ayuda en esta última etapa de la tesis, por tus consejos de post-doc y por aguantar y responder siempre con una sonrisa las millones de preguntas con las que te he bombardeado. Pero sobre todo gracias por tu humor, tu alegría y tu amabilidad. Me queda pendiente invitarte un café bien cargado de cafeína un día de estos.

Santi, gracias por tu gran ayuda en esta última etapa del doctorado. Gracias **Nuria** por hacerme partícipe de la Fundación Teresa Moretó, por aportar organización y estructura a la empresa, y por tu personalidad apacible y cercana. Gracias **Oscar**, por su aportación bioinformática.

Y como no, agradecer también a mi familia y seres más cercanos:

En primer lugar, **a mis padres**, José María Cadenas e Isabel Sevilla, por ser los mejores padres que una hija puede tener. Por dar todo, y cuando digo todo es TODO, por sus hijos. Gracias por los valores de esfuerzo, valentía y educación que nos habéis inculcado. Gracias por darme vuestro apoyo incondicional siempre, por creer en mí, y por darme todo lo que estaba en vuestro alcance para hacerme crecer como profesional y como mejor persona. A vosotros os debo todo.

A mi hermana **Isabel** (que ya se le está cayendo la lágrima), por ser mi guía y modelo a seguir desde muy pequeña. Por su inteligencia, su bondad, su cariño, su apoyo… La lista es interminable. Gracias por tener fe en mí. Tanta que siempre me dices que algún día ganaré el premio Nobel (ya estoy más cerca, jaja).

A mi hermano **Chema**, por querer a su "hermanita", por protegerla y adorarla. Gracias por todas esas luchas que me han hecho ser más fuerte, competitiva y sacar la rabia. Gracias por enseñarme a través del baloncesto lo que es el esfuerzo, el trabajo en equipo, y el compromiso.

Gracias a mis sobrinos, **Nicolás**, **Sofía** y **Jonás**, por con solo una mirada, hacerme olvidarme del resto del mundo.

Gracias a **Rebe** y **Rachel** por estar siempre cerca, aunque estemos en la distancia.

Gracias a mi familia de Salas de los Infantes, por vuestro apoyo, por celebrar los triunfos y por creer en mí.

Y por último, a la persona que más he de agradecer en estos tres años, a mi marido, mi mejor amigo y mi consultor de la vida. Gracias **David**, por todo. Y es que necesitaría una tesis nueva para explicar todo lo que te he de agradecer. Gracias por apoyarme en todo momento, por escuchar mis inquietudes y por ayudarme a tomar decisiones correctas. Gracias por empujarme en los momentos que lo necesitaba, por frenarme en los momentos de estrés y por hacerme ver otros puntos de vista cuando no veía salida. Gracias por creer en mí, por creer en mi potencial, por hacerme creer más en mí misma. Pero sobre todo, gracias por "hacerme risa". Gracias por potenciar mi competitividad con los concursos de triples, las competiciones de palas en la playa, o los cientos de juegos que compartimos (y nos picamos). Gracias por hacerme más fuertes con las aguadillas y con las peleas de churros. Gracias por los chistes frikis malos (tu eres LASSO, yo soy RIDGE, y juntos somos ELNET). Gracias David, por todo.

A todos los aquí nombrados, mi más sincera gratitud.