Alternative mechanisms of gene regulation during hematopoiesis

Sebastian Ullrich

TESI DOCTORAL UPF / year 2018

THESIS SUPERVISOR
Prof. Dr. Roderic Guigo
Department of Bioinformatics and Genomics
Center for Genomic Regulation (CRG), Barcelona

upf. Universitat
Pompeu Fabra
Barcelona

CRG
Centre
for Genomic
Regulation

Dem Leben

Acknowledgment

When I look back on the last years in which I did my Ph.D. I see a lot of things that happened to me and changed me, which comes usually down to people, environment, and experiences. In that regard I was glad to land a spot at CRG, a fantastic place do research in equilibrium with life. CRG is a place filled with exception scientists that are approachable and did not lose a sense of humor along the lines. One of these great scientists is Roderic, which I was glad to have as a supervisor throughout the learning period called Ph.D. Along with Rory, he had always good advice without being judgmental but great in patience. I further want to thank Emilio for technical advice and help whenever needed and Alessandra for the little gadgets she left on ISIS before she left. In general in want to thank anyone in the lab, not just for creating a pleasant working environment but also the methodological advice whenever needed. Here I want to especially thank Diego and Manu in regard to advice with computational problems and Silvia and Carme for experimental ones. Also, I greatly want to thank my companion Reza for discussing the essence of biology and life. Not to forget all the transient member of the lab Tomassso, Anna-laura, Bernardo and Tommy who despite being dispersed to different countries became friends. In addition, I want to thank Romina, Gloria and Imma from the administration side of CRG for making everything non-scientific work flawlessly.

Furthermore, I want to thank Hana (no 1. secret computational consultant), Andrea, Hannah and Martin for nightly energizing discussions and Linus for the relaxed habitation symbiosis that allowed to dedicate both of our energy unconditionally to science.

Having worked in the thesis on blood development I do not want to forget my blood. I especially want to thank my parents and my sister for supporting me all the way to where I am.

Last but not least I want to thank Circus Pizza, Pizza Sortidor and NAP for willingly accepting big chunks of my Ph.D. salary until late night in return to Italian energy.

Abstract

Gene regulation orchestrates the development of different cell types and organs from the same genetic blueprint. While the basic mode of gene regulation is driven by transcription factors, there are a variety of other mechanisms that determine the amount of RNA produced per genes. In this work we first investigate specifically intron retention as a mode of alternative splicing that alters the cellular transcriptomes. As a model, we use hematopoiesis. We compare intron retention in different stages of human and mouse B-cell development to granulocyte differentiation. We further explore expression and binding patterns of splicing regulatory factors. Second, we investigate the role of lncRNAs in the transdifferentiation of B-cell related lymphoma cells to macrophages. We specifically explore the role of a set of upregulated lncRNAs during this process. We deplete their expression during transdifferentiation with CRISPR/Cas9 to identify potential genes that retard or block the process and therefore are crucial for changing cell identity.

Resum

La regulació gènica determina el desenvolupament dels diferents tipus cel·lulars, teixits i òrgans. Tot i que el mode bàsic de regulació és dirigit per factors de transcripció, existeixen una gran varietat de mecanismes que contribueixen a determinar la quantitat de RNA produïda pels gens. En aquest treball, investiguem en primer lloc la retenció d'introns com un tipus d'splicing alternatiu que altera el transcriptome cel·lular. Com a model biològic, ens centrem en la hematopoesi. Comparem la retenció d'introns en diferents estadis del desenvolupament de limfòcits B en humà i ratolí amb la retenció durant la diferenciació del granulòcits. Estudiem també el patró d'expressió i d'unió (binding) dels factors de regulació de l'splicing. En segon lloc, investiguem el paper dels RNA llargs no codificants (long non coding RNAs, lncRNAs) en la transdiferenciació de limfòcits B a macròfags. En particular, el paper d'aquells lncRNAs que son regulats positivament durant aquest procés. Reduïm la seva expressió durant la transdiferenciació mitjançant la tècnica CRISPR/Cas9 amb l'objectiu d'identificar gens amb el potencial de retardar o de bloquejar el procés i que, en conseqüència, pugui jugar un paper crucial en el canvi de la identitat cel·lular.

List of publications during the thesis:

Internal transdifferentiation project

Sebastian Ullrich, Carme Arnan, Alexandre Esteban, Ramil Nurtdinov, Sílvia Pérez-Lluch, Rory Johnson, Roderic Guigó. Screening for novel regulators that affect speed and efficiency of transdifferentiation from B-cell like BLaER1 cells to macrophages with CRISPR/Cas9. (manuscript in preparation)

Ramil Nurtdinov, María Sanz, Alexandre Esteban, Amaya Abad, **Sebastian Ullrich**, Carme Arnan, Rory Johnson, Thomas Graf, Sílvia Pérez-Lluch, Roderic Guigó (2018) Comparative transcriptomics of B cell transdifferentiation in human and mouse. (submitted)

Blueprint

Sebastian Ullrich, Sílvia Pérez-Lluch, Roderic Guigó. Intron retention is tightly associated with regulation of splicing factors and proliferative activity during B-cell development. (manuscript in preparation)

Beekman R, Chapaprieta V, Russiñol N, Vilarrasa-Blasi R, Verdaguer-Dot N, Martens JHA, Duran-Ferrer M, Kulis M, Serra F, Javierre BM, Wingett SW, Clot G, Queirós AC, Castellano G, Blanc J, Gut M, Merkel A, Heath S, Vlasova A, **Ullrich S**, Palumbo E, Enjuanes A, Martín-García D, Beà S, Pinyol M, Aymerich M, Royo R, Puiggros M, Torrents D, Datta A, Lowy E, Kostadima M, Roller M, Clarke L, Flicek P, Agirre X, Prosper F, Baumann T, Delgado J, López-Guillermo A, Fraser P, Yaspo ML, Guigó R, Siebert R, Martí-Renom MA, Puente XS, López-Otín C, Gut I, Stunnenberg HG, Campo E, Martin-Subero JI. (2018) The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. Nat Med. 2018 Jun;24(6):868-880. doi: 10.1038/s41591-018-0028-4.

Luigi Grassi*, Farzin Pourfarzad*, **Sebastian Ullrich**, Angelika Merkel, Felipe Were, Enrique Carrillo de Santa Pau, Guoqiang Yi, Ida H Hiemstra, Anton TJ Tool, Erik Mul, Juliane Perner, Eva Janssen-Megens, Kim Berentsen, Hinri Kerstens, Ehsan Habibi, Marta Gut, Marie Laure Yaspo, Matthias Linser, Ernesto

Lowy, Avik Datta, Laura Clarke, Paul Flicek, Martin Vingron, Dirk Roos, Timo K van den Berg, Simon Heath, Daniel Rico, Mattia Frontini, Myrto Kostadima, Ivo Gut, Alfonso Valencia, Willem H Ouwehand, Hendrik G Stunnenberg, Joost HA Martens, Taco W Kuijpers (2018) Dynamics of transcription regulation in human bone marrow myeloid differentiation to mature blood neutrophils. Cell Reports Volume 24, Issue 10, P2784-2794, doi: 10.1016/j.celrep.2018.08.018

IHEC

Emilio Palumbo*, **Sebastian Ullrich***, Roderic Guigo. Grape2: A modular pipeline for parralellized processing of transcriptomic sequencing data. (manuscript in preparation)

Stunnenberg HG, International Human Epigenome Consortium, Hirst M (2016) The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. Cell. 2016 Nov 17;167(5):1145-1149. doi: 10.1016/j.cell.2016.11.007.

Contents

Chapter 1

INTRODUCTION

1.1 From heritable traits towards genomics

The way how complex multicellular eukaryotic organisms develop and maintain the integrity of their organs and tissues puzzled generations of scientists in the past. The quest for the underlying blueprint was tightly bound to the development of methods to study small structures not visible to the eye. The first prove of discrete heritable traits and along with that the foundation of modern genetics was set by Mendel with his experiments on peas (Mendel, 1866). At the end of the 19th century, it was clear that chromosomes carry those heritable traits but their molecular structure remained elusive. While ribonucleic acid (RNA) was already known as ”nuclein” and suspected to be the template for proteins synthesis (Caspersson and Schultz, 1939), it took until the middle of the 20th century until the desoxyribonucleic acid (DNA) was found to be the molecular underpinning of the chromosomes that carry those heritable traits (Avery et al., 1944, Hershey and Chase, 1952). With the development of crystallographic methods, the double stranded sugar backbone structure of DNA has been revealed and a model for replication was postulated (Watson and Crick, 1953). All pieces were put together as the central dogma of molecular biology that postulated an information flow from DNA, as information storage, to RNA, as a messenger, to proteins, as the building blocks of all living structures (Crick, 1958). From there on, molecular genetics took over with its rapid development and improvement of methods like gel electrophoresis to separate charged RNA and DNA molecules by size, chain-termination sequencing to determine the sequence of the DNA bases (Sanger et al., 1977) and polymer chain reaction to amplify nucleic acid sequences (Mullis et al., 1986).

With this toolbox on hand, scientists subsequently discovered the mechanisms and enzymes involved in DNA replication, RNA transcription and maturation as well as the translation into proteins. The way how these processes are timed for a specific gene were partially known for some well-studied genes or even simple signalling cascades but in its enormous complexity seemed overwhelming. While the nucleotide sequence for an increasing number of protein coding genes was determined, scientists believed knowing the entire sequence of a complex genome, as the human one, would solve the puzzle of how e.g. transcriptions factors find their target gene to recruit the transcription machinery to the locus. With the beginning of the 1990s, the human genome project was funded as a worldwide collaborative effort to sequence the entire human genome and finished in 2003 with the first human genome published (Lander et al., 2001, Venter et al., 2001, IHGSC, 2004). This started a new era of genomic sciences, where genes could be studied globally taking their surrounding genomic sequences in consideration.

1.2 New sequencing techniques pushed the boundaries

Against the expectations that having the whole sequence of human DNA would answer the key question of how genotypes unfold into phenotypes and how that happens in a temporal controlled manner during development, it was just the bare beginning of answering that question. Individual differences between the sequences of two individuals as short as single nucleotide polymorphisms (SNP) matter as the sequence units of DNA expressed at a given time in a given tissue in a given individual do. Microarrays spotted with oligonucleotides designed to bind previously selected genes (Taub et al., 1983) helped to scale expression analysis for genomes but failed to find genes that were not described before. Furthermore, SNP arrays were used for known mutations but could not widen the spectrum of observed SNPs. On the other hand, Sanger sequencing, used for the first human reference genome, required over a decade and employing sequencing centers around the world. It was not cost and time efficient enough to continue to produce more genomes for other human individuals or other species. New highly parallelized sequencing techniques were closing this gap by providing millions of short reads from a single sequencing run of one machine in less than 24 hours. From such short but overlapping reads, genomes and transcripts could be assembled with newly developed computational tools to provide new reference genomes in a more rapid way and to complement microarray approaches by allowing to find new transcripts and sequence alterations (Margulies et al., 2005, Bentley et al., 2008, Schuster, 2008).
In order to understand the information encoded in the DNA with the basic sequence known and new high throughput techniques on hand, the Encyclopedia of DNA Elements (ENCODE) project found about 80% of the genome to be biochemically active (The ENCODE Project Consortium, 2012). The before speculated number of human genes could be refined with every release of the GENCODE annotation with the latest version (GENCODE v28) containing 58,381 genes of which 19,901 are protein coding, down from speculations of up to 100k

protein coding genes in the past (Fields et al., 1994, Liang et al., 2000). Besides those advances in the annotation of genetic elements, whole genome sequencing, used to determine the DNA sequence of an individual, came down to 1,000 USD (Illumina 10x sequencer) to allow for cost-efficient personalized treatment of patients. Furthermore, combining high throughput short read sequencing immunoprecipitation and bead pull down protocols allowed to study epigenetic modifications and chromatin conformation (Dekker et al., 2002, Johnson et al, 2007, Lieberman-Aiden et al., 2009, de Wit and de Laat, 2012).

1.3 Sequential layers of gene expression regulation

While the first model of gene regulation, by other gene products, was proposed by François Jacob and Jacques Monod in the 1960s, multiple regulatory levels that affect timing and amount of a given gene product were discovered.

At the level of DNA organization in the nucleus, sequence stretches were found to stay in particular regions named chromosome territories where active genes reside on the surface of the territories, whereas suppressed genes are sequestrated on the inside (de Wit and de Laat, 2012, Dostie and Bickmore, 2012, Ethier et al., 2012, Vaquerizas et al., 2012). Genes with high expression have been found to colocalize in foci called transcription factories and thereby are coregulated (Razin et al., 2011, Dai and Dai, 2012).

On top of that, the occupation of DNA with histone proteins and their side tail modification matter for transcriptional activity. While actively transcribed genes were found to have a nucleosome free promoter region (Yuan et al., 2005, Lai and Pugh, 2017), the histone side tail modifications, known as the histone code, follow specific patterns along the transcript. For example, while H3K27ac is found on active promoters, H3K9m3 is found along repressed genes (Allfrey et al., 1964, Jenuwein and Allis, 2001).

About 2600 known human transcription factors act as a further layer of regulation by either binding to enhancer or promoter regions in relative proximity of the gene (Babu et al., 2004). Acting in various combinations with each other allows specific regulation for each gene (Brivanlou and Darnell, 2002). Besides recruiting the transcription machinery, they were found to interact with the histone code by bringing histone acetyltransferases or histone deacetylase on site (Narlikar et al., 2002). Besides modulation of their synthesis, they are regulated by localisation and activation, e.g. phosphorylation (Whiteside and Goodbourn, 1993, Weigel and Moore, 2007).

Further downstream and by far the best-studied process of gene regulation is the transcription initiation and elongation process with the RNA polymerase II as the core molecule and a multitude of co-regulators. Already during the synthesis of the RNA, maturation starts along the nascent strand for the majority of transcripts (Beyer and Osheim, 1988, Kotovic et al., 2003, Lacadie and Rosbash, 2005, Listerman et al., 2006, Dye et al., 2006, Pandya-Jones and Black, 2009, Ameur et al., 2011, Khodor et al., 2011, Vargas et al., 2011, Khodor et al., 2012, Tilgner et al., 2012). During the process of removing introns from the immature transcript, alterations in the exon composition can be introduced, referred to as alternative splicing (Wang and Burge, 2008). This part of the regulation will be detailed on in the next section. Furthermore, the regulation of transcripts by non-coding RNAs will be addressed in an upcoming section.

When the protein coding mRNA is capped and polyadenylated, it is ready for export into the cytoplasm, through the nuclear core complex. Besides nuclear export factors being involved the regulation of export, it has been shown that RNA modifications like ubiquitylation play a role in translocation to the cytoplasm (Durairaj et al., 2009). Finally, besides the rate of production, the stability of a transcript also impacts the steady state abundance at a given time. Besides non-coding RNAs, a complex interplay of RNA binding proteins, e.g. binding AU rich elements, determines the half-life of a transcript by stabilizing it or labelling it for degradation (Wu and Brewer, 2012).

1.4 Alternative splicing

All structures and tissues of complex multicellular organisms are built from proteins in a time controlled manner. For development and maintenance of structures, transcript maturation is a crucial regulatory step, where introns are removed and exons are joined together. By alternative splicing (AS) multiple transcripts can be derived from one gene for about 95% of all mammalian protein coding genes (Johnson et al., 2003, Pan et al., 2008, Barash et al., 2010). The functional spectrum of alternative exon composition can be as dramatic as for the FAS receptor, where membrane anchored and decoy receptors have opposing functions (Cascino et al., 1995). In general, alternative exons were found to be coiled structures on the protein surface that do not disrupt its overall structure (Wang et al., 2005, Romero et al., 2006). The finding that AS is more abundant in immune and neuronal related genes, suggests that it contributes to temporal and spatially complex regulation patterns (Modrek and Lee, 2002).

The underlying mechanism for splicing, which makes it robust and reproducible for the majority of the events during normal maturation of an mRNA transcript, relies on a well-conserved machinery among eukaryotic organisms. The spliceosome is a conglomerate of the five sub-complexes, consisting of the snRNAs U1, U2, U4, U5, U6 and associated proteins. Additional regulatory proteins bind transiently. All steps from detection of the splice sites to exon-exon joining are fulfilled by the spliceosome with its transition states (Nilsen, 2003, Jurica and Moore, 2003, Wang and Burge, 2008, Wahl et al., 2009, Hegele et al., 2012). For initiation of a splicing event, exon-intron boundaries are recognized as the potential splice sites. As initial step U1 snRNP binds to the 5' splice site by base pairing. Then, U2AF pairs with the polypyrimidine tract of the 3' splice site and non-snRNP factors like SF1/mBBP bind at the branchpoint (Madhani and Guthrie, 1994). In two consecutive nucleophilic attacks exons are joined together and the intron lariat is released (Hang et al., 2015). In eukaryotes there are two principal modes of splicing, exon and intron definition (Robberson et al., 1990, Berget, 1995, Fox-Walsh et al., 2005).

Exon definition is the preferential method, when introns are longer than approximately 250 bp, while intron definition is used if introns are short. From a steric standpoint it allows the complex to form faster and be more reproducible as ends to be joined are more proximate to each other (De Conti et al., 2013). In higher eukaryotes exon definition is dominant as introns are usually much longer than exons (Zhang 1998, Sakharkar et al., 2005). In contrast, in lower eukaryotes like yeast, where introns are usually below 100 bp, but also *Drosophila melanogaster* with about 50% of all introns below 100 bp, intron definition is the dominant mode of splicing (Lang and Spritz, 1983, Berget, 1995).

Regarding the core splicing machinery, one crucial aspect that influences alternative splicing is the strength of the splice site, which means how much the exon-intron junction sequence diverges from the consensus sequence. Strong splice sites with perfect sequence conservation usually lead to constitutive splicing. Due to lower affinity for junction binding, splice-component recruitment of the splicing machinery is lower for weak splice sites and therefore the usage of the splice site is lower as well. However, the processing outcome is very dependent on the cellular context (Kornblihtt et al., 2013).

For weak splice sites, trans-acting factors play a crucial role in making the decision whether the exon is included or excluded. The two major groups of them are Serine/Arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoprotein (hnRNP) proteins. Furthermore, there are tissue-specific factors like PTB15, NOVA16 and FOX (Jelen et al., 2007, Lee et al., 2009, Kafasla et al., 2012). Independent of the protein family to which they belong; these factors can activate or inhibit a certain splicing decisions. Most of them actually fulfil both functions, depending on the location where they bind and the interplay with other factors (Ule et al., 2006). In an approach to characterize the binding patterns of those transient splicing factors, cis-regulatory sequences were determined. Those sequences can lie either in exons or introns. They are classified into enhancers or silencers, which results in the following nomenclature: exonic splicing enhancer (ESE), exonic splicing silencer (ESS), intronic splicing enhancer (ISE), intronic splicing silencer (ISS) (Kornblihtt et al., 2013) (Figure 1.1).

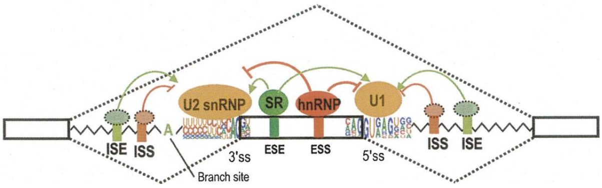


Figure 1.1: Splicing regulatory elements can be found either in exons or introns. Exonic splicing enhancers (ESE) and exonic splicing silencers (ESS) are found in exons, while intronic splicing enhancers (ISE) and intronic splicing silencers (ISS) are found in introns. (adapted from Wang and Burge, 2008)

Systematic analysis of splicing events has revealed five major types of alternative splicing events in cassette exons. Either the whole exon is skipped, in contrast to the canonical inclusion, or alternative 5' or 3' splice sites are used for the exon (Figure 1.2). Additionally, exons can be included mutually exclusive and introns can be retained in the mature mRNA (Black, 2003, Matlin et al., 2005, Sammeth et al., 2008, Pan et al., 2008). Two further options to generate alternative transcripts are the usage of an alternative transcription start or polyadenylation site.

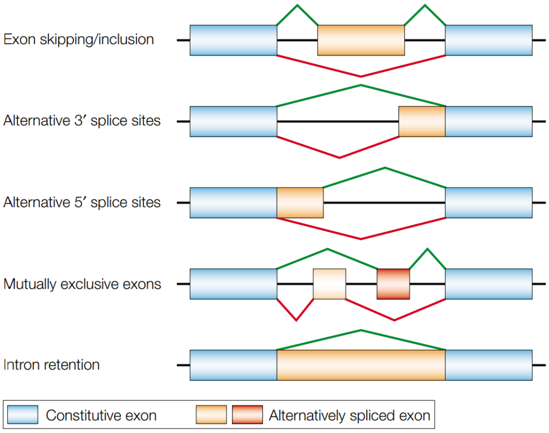

Exon skipping/inclusion
Alternative 3′ splice sites
Alternative 5′ splice sites
Mutually exclusive exons
Intron retention
Constitutive exon
Alternatively spliced exon

Figure 1.2: Types of alternative splicing events in cassette exons. (adapted from Cartegni et al., 2002)

While the occurrence of alternative splicing was initially studied mechanistically, over time more and more examples appeared where AS is assumed to be a major contribution to cellular development and the occurrence of disease. However, due to technical limitations, those examples were mostly observations of a small number of genes. With technological improvements in sequencing (next generation sequencing), genome wide methods became available and affordable. Due to its sensitivity, next generation sequencing allowed a detection of a multitude of AS events across higher vertebrates. Taking the conclusion from those studies together, they can be summed up as the following: Dynamic changes in alternative splicing unfold in development and cell differentiation in many genes at the same time in a coordinated fashion. Specific RNA binding proteins orchestrate those splicing programs. Interestingly, some genes affected by AS maintain stable expression profiles, while the relative isoform contribution changes. Coordinated splicing networks are cell type- and region-specific, e.g. differences have been observed between cerebral cortex and hippocampus in the brain or cardiomyocytes and cardiac fibroblasts in the heart. Tissues with repetitively observed AS patterns encompass neuronal tissues, muscle tissue and blood among others (Baralle and Giudice, 2017).

1.5 Intron retention

1.5.1 Mechanistic aspects of IR

As a subtype of alternative splicing, intron retention (IR) was documented to happen in a 2-5% fraction of human genes in the beginning of the 2000s (Clark and Thanaraj, 2002, Kan et al., 2002). Besides that, it was also described early in plants, especially *arabidopsis thaliana* (Ner-Gaon et al., 2004). In contrast to other types of alternative splicing, IR has initially been expected to be a form of mis-splicing that needs to be handled by nuclear sequestration or nonsense mediated decay in order to prevent the production of potentially harmful proteins (Jaillon et al., 2008, Roy and Irimia, 2008, Gudipati et al., 2012). However, later studies revealed that IR can actually self-regulate the transcript abundance of a given gene, the export of the transcript to the cytoplasm as well as the maturation of mRNA on demand (Lareau et al., 2007, Moran et al., 2008, Cuenca-Bono et al., 2011, Wong et al., 2013, Palazzo et al., 2013, Boothby et al., 2013).
In general, introns are considered to be retained if they are, unlike other introns within the transcript, not removed from the pre-mRNA but remain in the sequence of the mature mRNA. Those introns were described to be shorter, have a higher GC content and a lower splice site strength than other introns not affected by IR (Braunschweig et al., 2014). Furthermore, retained introns frequently contain one or more PTCs in their sequence that trigger degradation by nonsense mediated decay (Maquat, 2004). In addition, associations between IR and reduced expression levels of splicing factors, RNA polymerase II occupancies and epigenetic changes were observed (Wong et al., 2013, Braunschweig et al., 2014, Guo et al., 2014, Gascard et al., 2015, Wong et al., 2017, Middleton et al., 2017).
A comparative analysis of 2,567 mRNA sequencing datasets found SR protein binding sites to be enriched in introns with high IR values, suggesting SR proteins to be involved in the recruitment of spliceosomal components. While there were approximately 15,000 introns found to be retained in at least one dataset,

retention for each of them was limited to fewer than 7% of all samples (Middleton et al., 2017).

1.5.2 IR in animals

In the animal kingdom IR was predominantly described in higher vertebrates affecting a variety of functions. Autoregulation of gene expression is one of the most recurring, affecting even splicing-related genes themselves. SR proteins were found to contain ultraconserved regions between human and mouse. For SRSF1 and SRSF2, they fall into intronic 3' untranslated region that are retained and thereby target the messenger RNAs for degradation (Lareau et al., 2007). In human and mouse granulocyte development, the nuclear lamina protein LMNB1 is affected by IR. As a consequence of decreasing expression, the lamina structure has less reinforcement and the nucleus folds into the granulocyte typical lobes in return (Wong et al., 2013). In neuronal tissue presynaptic proteins are transcribed in neurons and non-neuronal cells. Binding of PTBP1 to retained 3' terminal introns however prevents export of these transcripts to the cytoplasm and instead leads to nuclear degradation of the transcripts in non-neuronal cells (Yap et al., 2012). Similarly, PABPN1 auto-regulates its expression levels to maintain homeostasis in human cells. PABPN1 binds an adenosine rich region in its 3' untranslated region that causes the retention of the 3'-terminal intron, which in turn induces degradation by the nuclear exosome (Bergeron et al., 2015). In neuronal differentiation, IR retaining transcripts were suspected to lower the abundance of proteins enriched for non-physiologically relevant processes (Braunschweig et al., 2014).
Dosage compensation is another process where IR was found to contribute to in *drosophila melanogaster*. While complete splicing of Male-specific lethal 2 (Msl-2) in male flies promotes physiologically normal expression of the X-chromosome, IR in Msl-2 prevents its own translation in female flies (Zhou et al., 1995, Bashaw and Baker, 1995). Another interesting function of IR in neuronal cells is to provide localization information for the mRNA. Intron retaining mRNAs were found to contain ID elements (a class of SINE retrotransposon) guiding them for dendritic localization (Buckley et al., 2011). Besides the aforementioned SR rich proteins that regulate their own expression and thereby impact splicing in general, the heterogeneous nuclear ribonucleoprotein hnRNPLL was described to affect alternative splicing during T-cell activation (Cho et al., 2014).
Enrichment of protein diversity, often associated with alternative splicing, has been linked to IR recently as well. For the calcium-activated big potassium (BKCa) channel, removal of an intron retaining variant altered localisation and intrinsic firing properties of hippocampal neurons (Bell et al., 2008, Bell et al., 2010). Furthermore, in *drosophila melanogaster* a retained intron causes a readthrough into non-coding DNA and thereby produces the novel protein Noble

from the rieske iron sulphur protein locus (Gontijo et al., 2011).

1.5.3 IR in plants

Contrary to animals, IR is the most frequent form of alternative splicing in plants (Wong et al., 2015). In the model species *Arabidopsis thaliana* a dominance of IR of up to 64% of all splicing events was observed (Ner-Gaon et al., 2004, Wang and Brendel, 2006, Kalyna et al., 2012). Similar to animals, retained introns are short and rich in GC, however, introns are shorter in plants in general (Ner-Gaon et al., 2004, Galante et al., 2004, Sakabe and de Souza, 2007, Braunschweig et al., 2014). Like in animals the relatively small lengths of those introns suggests intron definition as dominant mode of splice site recognition (Lim and Burge, 2001, Amit et al., 2012, Reddy et al., 2012). Steric hindrance was one hypothesized reason for the high abundance of IR in plants (Wang and Brendel, 2006). Experimental findings that animal introns are not accurately spliced in plants and vice versa indicates that the mechanisms are not conserved between the two kingdoms (Reddy et al., 2012). As the major plant model organism *Arabidopsis thaliana* is probably the best studied plant, thus, IR was described to regulate several physiological processes in *Arabidopsis thaliana*. At low temperatures, starch accumulation is promoted by IR dependent expression regulation of IDD14 (Seo et al., 2011). Furthermore, an intron retaining variant of the circadian rhythm related gene CCA1 links circadian rhythms to temperature adaptation (Seo et al., 2012). In *Arabidopsis thaliana* roots, intron-retaining transcripts of ZIF2 are more efficiently translated and thereby increase tolerance to higher abundance of zinc in the soil (Remy et al., 2014). Interestingly, in *Marsilea vestita*, a heterosporous fern, transcripts with retained introns were sequestrated in the nucleus during gametophyte development. In the sense of priming, those transcripts could be spliced for rapid production of proteins upon physiological requirements. Depletion of those transcripts with siRNA interference demonstrated that no alteration of gametophyte development was observed unless those intron-retaining transcripts were needed during progression of development (Boothby et al., 2013).

Direct comparison of described functions of IR in plants and animals revealed that in both kingdoms IR is used to produce transcript isoforms with new functions (Bell et al., 2008, Bell et al., 2010, Gontijo et al., 2011, Seo et al., 2011, Seo et al., 2012, Rocchi et al., 2012, Khaladkar et al., 2013, Remy et al., 2014) as well as it does regulate genes by sequestration or induced degradation of intron retaining transcripts (Lareau et al., 2007, Yap et al., 2012, Filichkin and Mockler, 2012, Wong et al., 2013, Braunschweig et al., 2014, Bergeron et al., 2015) and primed expression of transcripts that are spliced on demand (Moran et al., 2008, Boothby et al., 2013). However, functions in dosages compensation of sex chromosomes (Zhou et al., 1995, Bashaw and Baker, 1995), tags for localisation of transcripts (Buckley et al., 2011) and association with the surrounding of exons, prone to exclusion (Cho et al., 2014), were solely described for the animal kingdom.

Besides animals and plants, IR was also reported in unicellular organisms such as *Saccharomyces cerevisiae*, *Plasmodium*, *Capsaspora owczarzaki*, and *Tetrahymena thermophila* (Hossain et al., 2011, Xiong et al., 2012, Sebé-Pedrós et al., 2013, Lunghi et al., 2015).

1.5.4 IR in disease

Like alternative splicing in general, IR is also associated with disease, most importantly with cancer. In a variety of cancer types including bladder, colon, endometrium, head and neck, kidney, liver, lung, prostate, stomach, rectum and thyroid tumors as well as acute myeloid leukemia, IR was reported to be higher than in healthy tissue (Lu et al., 2003, Solomon et al., 2003, Comstock et al., 2009, Ren et al., 2012, Masood et al., 2012, Eswaran et al., 2013, Zhang et al., 2014, Simon et al., 2014, Dvinge and Bradley, 2015, Jung et al., 2015). In breast and lung cancer, IR is even one of the dominant forms of alternative splicing with 2,038 genes affected in breast cancer and 2,340 in lung cancer (Eswaran et al., 2013, Zhang et al., 2014). In agreement with IR features in healthy tissue, IR was associated with a 3' bias, weak splice sites and high GC content in cancer (Zhang et al., 2014, Dvinge and Bradley, 2015). Contrary to normal cells, where splicing factors were reported to be expressed at lower levels, splicing factor expression was not altered in lung cancer samples compared to healthy lung tissue (Zhang et al., 2014). In breast cancers splicing factor expression was even found to be upregulated (Shapiro et al., 2011).

Besides cancer, IR was also found in other disease like Xeroderma pigmentosum (Saredi et al., 2012), Late infantile neuronal ceroid lipofuscinosis (Cartault et al., 2011), Autoimmune polyendocrine syndrome type 1 (Maselli et al., 2014), Netherton Syndrome (Zhang et al., 2013), Amyotrophic lateral sclerosis (Lacroix et al., 2012), Inflammatory bowel disease (Flomen and Makoff, 2011) and Myotonic Dystrophy type 2 (Häsler et al., 2011). However, if IR in those disease is a side effect or a driver of disease progression remains to be investigated.

1.6 Long non-coding RNAs

While the abundance of genes was estimated to be about 100,000 in the mid 1980s, based on the size of a typical gene and the human genome, only a fraction of protein coding genes, relative to this number, could be confirmed by the human genome project. First estimates were ranging around 31,000 protein coding genes in 2001 and were reduced to 22,287 protein coding genes in 2004 (Lander et al., 2001, International Human Genome Sequencing Consortium, 2004) and further down to 19,901 in the most current release of the GENCODE annotation (v28, www.gencodegenes.org). This counterintuitive reduction of complexity in the coding genome was opposed by the noncoding. In the same way as the estimates of protein coding genes decrease, the number of genes that are transcribed but not translated increased. The Functional Annotation of Mammalian cDNA (FANTOM) project found approximately 10,000 non coding transcripts from distinct loci in its 3rd phase (Carninci et al., 2005), while in phase 5 the number increased to 27,919 long noncoding RNAs (lncRNAs) (Hon et al., 2008). Those findings were confirmed by the Encyclopedia of DNA Elements (ENCODE) project that found about 80% of the genome to be biochemically active (The ENCODE Project Consortium, 2012).

1.6.1 Classification of lncRNAs

In an attempt to distinguish lncRNAs from the multitude of short RNAs such as microRNAs (miRNAs), small interfering RNAs (siRNAs), Piwi-interacting RNAs (piRNAs) and small nucleolar RNAs (snoRNAs) a minimum length cutoff was set to 200 bp (Rinn and Chang, 2012, Ma et al., 2013, Knoll et al., 2015). While structural features as 5’ capping, splicing and polyadenylation emphasise similarities to mRNAs, lncRNAs distinguish from protein coding genes by short ($<$100 codons) or entirely lacking open reading frames (ORF) (Carninci et al., 2005, Morris and Mattick, 2014).

A further subclassification of lncRNAs was based on their relative position towards neighboring protein coding genes and divided them into 5 categories (Figure 1.3) (Carninci et al., 2005, Morris and Mattick, 2014, Devaux et al., 2015, Lorenzen and Thum, 2016). Sense lncRNAs are transcribed from the same stand as a protein coding gene, they structurally overlap with a fraction of introns and/or exons. In contrast, antisense lncRNAs overlap with protein coding genes as well but are transcribed from the opposite strand. Bidirectional lncRNAs are transcribed from the opposite strand of the nearby protein coding gene, like antisense lncRNAs, but do not overlap. They, however, remain in the proximity of $<$1 kb from the nearest protein coding gene. Intronic lncRNAs are entirely transcribed from within an intron of a protein coding gene whereas the most abundant form of lncRNAs, intergenic lncRNAs, are spaced between protein coding genes.

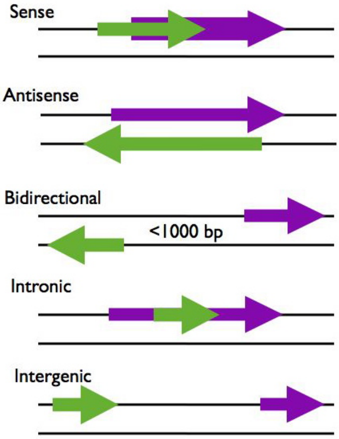


Figure 1.3: Different types of lncRNA classified by their relative position to nearby protein coding genes (adapted from https://mcmanuslab.ucsf.edu/node/251)

1.6.2 Discovery and properties of lncRNAs

Detection of lncRNAs is more challenging than it is for protein coding genes, as on average lncRNAs have 10-fold lower expression levels (Ravasi et al., 2006, Cabili et al., 2011). In addition, the majority of lncRNAs (80%) have a more tissue restricted expression relative to protein coding genes (20%) (Cabili et al., 2011). A further challenge for discovery of novel lncRNAs is the restricted expression in specific developmental stages (Yan et al., 2013). This specificity suggests that lncRNAs are important in shaping cell type and developmental specific transcriptomes (Knoll et al., 2015). Therefore, only lncRNAs with higher and broader expression levels are known since long. The first discovered lncRNA, H19, was thought to be protein coding at the time of discovery in the mid 1980s (Pachnis et al., 1984). Xist, a very prominent lncRNA important for sex determination that is also highly expressed, was discovered in 1991 (Borsani et al., 1991, Brown et al., 1991, Brockdorff et al., 1991). Hybridization to tailing arrays allowed to study lncRNAs in a broader, parallelized fashion, but was limited to prior knowledge regarding the oligo design (Martin and Wang, 2011). The true blooming of the field happened with commonly available deep transcriptome wide RNA sequencing, where no prior information was needed and sensitivity was further increased (Wang et al., 2009, Guttman et al., 2010). An additional benefit of RNA-seq was to gain further information about the transcript structure of lncRNAs. Whereas 98% of them were found to be spliced, only 25% had multiple isoforms (Derrien et al., 2012). Furthermore, it was shown that lncRNAs are shorter than protein coding genes on average (median, 592 vs. 2,453 bp) but exons (median, 149 vs. 132 bp) as well as introns (median, 2,280 vs. 1,602 bp) are longer (Derrien et al., 2012). Comparison of lncRNA expression in 11 species and 8 tissue types found 11,000 primate specific lncRNAs. However, only 2,500 were conserved at sequence and expression level, suggesting a more rapid evolution compared to protein coding genes (Necsulea et al., 2014). Interestingly, lncRNAs of 9 tissues from 6 mammals clustered by tissue and not by species. Furthermore, lncRNAs conserved between mammals had higher sequence conservation in promoters and exons, were more proximate to protein coding genes, had fewer repetitive elements, were more often single-exonic transcripts and enriched in tissue-specific functions than younger lncRNAs (Washietl et al., 2014).

1.6.3 Mechanisms of lncRNAs function

While the number of annotated human lncRNAs is steadily increasing, the majority remains functionally uncharacterized (Mercer et al., 2009, Dinger et al., 2009). Recent efforts to functionally catalogue lncRNAs in databases provided functional associations for 294 lncRNAs (183 human lncRNAs) in the LncRNAdb database and 363 human lncRNAs in LncRNAWiki (Amaral et al., 2011, Quek et al., 2014, Ma et al., 2015). Their functions vary widely from transcriptional regulation over splicing regulation and translational control to scaffolding in protein complexes (Ma et al., 2015).

By far, most lncRNAs with known function fall in the category of transcription regulation. One mode of action is to directly regulate one specific target gene. As an example, ncRNA-CCND1 mediates the repression of cyclin D1 by recruiting the RNA binding protein TLS, which inhibits the activity of the transcription factor CREBBP and the histone acetyltransferase p300 at the cyclin D1 locus (Wang et al., 2008). Another example is the lncRNA Evf-2 that functions as transcriptional coactivator of Dlx5 by recruiting the transcription factor Dlx2 to the Dlx5 locus (Feng et al., 2006). The other mode is to regulate groups of genes or entire chromosomes as in the case of Xist. During early embryonic stem cell differentiation Xist is expressed from the X chromosome to be silenced. Its expression is followed by the loss of active chromatin marks like H3K9ac and H3K4me and an induction of repressive marks as H3K27me3, H3K9me and H4K20me1. In consequence, the marked chromosome is entirely inactivated with an overall lack of transcription (Wutz and Gribnau, 2007).

In splicing, lncRNAs can interfere with the splicing machinery at a given exon or intron to change the splicing outcome. The lncRNA ZEB2-AS1, for example, is in antisense direction to ZEB2, overlapping the 5' splice site of an intron in the 5'UTR of ZEB2. Upon ZEB2-AS1 expression, the 5' splice site is not accessible and in turn the intron is retained (Beltran et al., 2008). In a similar way, the antisense lncRNA Rev-ErbAa2 affects splicing of ErbAa2 (Munroe and Lazar, 1991).

Indirectly, ZEB2-AS1 also affects translation as the related intron in the 5'UTR contains an internal ribosome entry site that is necessary for efficient translation. In a broader way, translation control was discovered in neurons, where mRNA needs to be translated to maintain synaptic plasticity. Here, the expression of the lncRNA BC1 was linked with translational repression to control dopamine D2 receptor-mediated transmission (Centonze et al., 2007).

Besides directly regulating biological processes, lncRNAs also fulfil structural functions. As scaffolds, they bring protein coding genes together to form complexes (Spitale et al., 2011, Rinn and Chang, 2012). One well-known example is the lncRNA TERC, which, as a RNA scaffold, assembles the telomerase complex (Zappulla and Cech, 2006). HOTAIR, another classic example of a lncRNA scaffold, can bind the polycomb group protein PRC2 in its 5' region and the histone demethylase LSD1 in its 3' region to silence the HOXD locus by H3K27 methylation and H3K4me2 demethylation (Rinn et al., 2007, Tsai et al., 2010). Further examples are ANRIL that brings PRC2 in proximity to PRC1 (Yap et al., 2010, Kotake et al., 2011) and Kcnq1ot1 that tethers PRC2 to G9a (Pandey et al., 2008)

1.6.4 LncRNAs in disease

As for functions of lncRNAs in normal development, only a small fraction of lncRNAs is investigated in their contribution to the onset and progression of disease. However, the amount of studies that link alteration in lncRNA expression to disease states is steadily increasing. Most reported associations were found in cancer, where dozens of lncRNAs have been reported to alter expression (Tsai et al., 2011, Rinn and Chang, 2012). Those lncRNAs were found to be regulated by tumor suppressors like p53, MYC, and NF-kB (Guttman et al., 2009, Huarte et al., 2010, Hung et al., 2011). Furthermore, lncRNAs with cyclic expression during normal cell cycle were found to have altered expression patterns in cancer cells (Hung et al., 2011). In human breast cancer about 25% of the patients have the lncRNA HOTAIR overexpressed, which is a predictor for metastatic risk and survival chances (Gupta et al., 2010). *In vitro* it could be demonstrated that overexpression of HOTAIR drives metastasis formation by rearranging PRC2 occupancy pattern to resemble embryonic fibroblasts. This helps cells to mimic morphological properties of the anatomic sites of invasion (Gupta et al., 2010). Similar alterations in HOTAIR expression were also found in colon and liver cancers (Kogo et al., 2011, Yang et al., 2011). Another study found extensive alterations in lncRNA expression in prostate cancer, where the lncRNA PCAT-1 was identified as a marker for prognosis outcome. Due to their secondary structure, lncRNAs are stable in human body fluids, which allows for non-invasive and cost efficient testing (Prensner et al., 2011). The aforementioned lncRNA ANRIL, that has scaffolding functions, has been associated with both cancer and cardiovascular diseases (Burd et al., 2010). Besides ANRIL, several other

lncRNAs were found to be involved in heart failure, cardiac autophagy and myocardial infarction (Lorenzen and Thum, 2016).
In kidney disease, Xist and Neat1 were associated with membranous nephropathy (Huang et al., 2014) whereas in diabetes mellitus MALAT1 upregulation in endothelial cells, subjected to high glucose treatment, was associated with upregulated expression of genes that regulate inflammation like SAA3, TNF and IL-6 (Puthanveetil et al., 2015).
As the brain is one of the tissues with the highest amount of lncRNAs expressed, there were various instances of neurological disorders and diseases linked to lncRNAs dysregulation (Briggs et al., 2015). In schizophrenia recent evidence suggests that defects in regulation of the lncRNA GOMAFU cause aberrant splicing of the mediator molecule DISC1 and the receptor tyrosine kinase ERRB4 (Barry et al., 2014). Interestingly, GOMAFU forms a ribonucleoprotein complex in the nucleus with the splicing factors SRSF1, SF-1, and QKI, which could be the dysregulated mechanism in schizophrenia. An association of GOMAFU with schizophrenia risk was shown by several studies independently, however, without mechanistic insights (Takahashi et al., 2003, Di Chiara et al., 2004, Albertson et al., 2006, Michelhaugh et al., 2011, Spadaro et al., 2015). In neurons, differentiated *in vitro*, it could be shown that downregulation of GOMAFU indeed causes splicing defects in DISC1 and ERRB4 that mimic those seen in schizophrenia patients (Barry et al., 2014). Besides schizophrenia, lncRNAs were also found in other neurologic diseases such as ADS, AD and neuropathic pain (Briggs et al., 2015).

1.7 CRISPR

1.7.1 CRISPR discovery

Probably the most impactful innovation in recent biology, next after high throughput sequencing methods, is the development of the CRISPR/Cas9 gene editing system. Originally, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) were found in bacteria and archaea (Barrangou, 2015), where they function as an adaptive immune system. In the spacer region between the repeats, short foreign DNA fragments from previous exposure to e.g. viruses and plasmids were found (Mojica et al., 2000, Pourcel et al., 2005, Mojica et al., 2005, Bolotin et al., 2005, Barrangou et al., 2007, Marraffini and Sontheimer, 2008, Marraffini and Sontheimer, 2010). CRISPR-associated system (Cas) proteins utilize this foreign genetic information to recognize and cut new encounters of foreign DNA or RNA, depending on the Cas protein (Mohanraju et al., 2016).

1.7.2 Development of CRISPR/Cas9

The CRISPR system of *Streptococcus pyogenes* containing a CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA) was further simplified to a single guide RNA (sgRNA) that works together with the endonuclease Cas9. With replacing the guide region, essentially any DNA stretch can be targeted to induce double strand breaks (Jinek et al., 2012). However, optimal performance is limited to DNA stretches with certain sequence properties, e.g. GC content. The system was rapidly utilized to modify human cells (Cong et al., 2013, Mali et al., 2013) and most common model organisms (Gratz et al., 2013, Friedland et al., 2013, Jiang et al., 2013, Wang et al., 2013).

1.7.3 CRISPR and previous methods

When researchers aimed to suppress gene expression before CRISPR was available, the method of choice was RNA interference (RNAi), where, instead of knocking out a gene from the genome, its transcripts were targeted. While it was easy and cheap to target several genes or screen whole genomes, the effect was limited to a couple of days and instead of depleting gene expression, RNA abundances were only reduced. Transcription activator-like effector nucleases (TALENs), which emerged after RNAi, affected DNA with full disruption of gene expression. However, for each target a specific protein needed to be produced, which made the procedure less cost efficient than RNAi and CRISPR.

1.7.4 Increased accessibility and alternative CRISPR applications

Advancement in CRISPR development made it easier for researchers to apply the technology and explore new applications. Predefined CRISPR libraries are now commercially available to target the entire coding genome in human and mouse (Shalem et al., 2014). Furthermore, there are the alternative applications CRISPRi that allow a transient knockdown, as with RNAi, and CRISPRa for activation. For those applications a catalytically dead version of Cas9 (dCas9) is used that does not induce DNA cleavage but instead allows to deliver proteins to defined genomic loci (Qi et al., 2013). Bringing transcription factors or chromatin remodelling factors to a site of interest allows to activate or repress genes in a non-invasive manner. These methods allow to find and study regulatory elements, like enhancers, that would suffer less from induced frameshifts resulting from double-strand-break repair used to disrupt protein coding gene function. In the first study performed that uses CRISPRi, 98,000 sgRNAs were designed to target a 1.29 Mb sized region around GATA1 and MYC (Fulco et al., 2016). Nine distal enhancers affecting proliferation of K562 cells through GATA1 and MYC expression regulation were identified. A further, smaller scale study targeted 15 super enhancers with 241 sgRNAs (Xie et al., 2017). CRISPRi and CRISPRa

were also used to study lncRNAs (Liu et al., 2017, Joung et al., 2017).
Another interesting approach to target lncRNAs is to entirely remove their promoter region. This method would help to avoid a collision of endogenous regulation with imposed one. A screen employing such a paired design identified 51 lncRNAs that affect cancer cell growth (Zhu et al., 2016). Recent advancements have been made to simplify the delivery of paired guide RNAs (pgRNAs) fully scalable from single locus studies to complex libraries targeting thousands of regions (Aparicio-Prat et al., 2015). The Double Excision CRISPR Knockout (DECKO) plasmid allows cloning of both sgRNAs in a single 165 bp oligonucleotide. But also the design of paired guides has been made simpler. With CRISPETa a pipeline for flexible and scalable pgRNA design, guide pairs for any number of target regions can be designed in parallel (Pulido-Quetglas et al., 2017). During the designing procedure CRISPETa compares them against a pre-computed off-target database. Furthermore, pre-designed libraries for different classes of protein coding and non-coding elements are available for human, mouse, zebrafish, *Drosophila melanogaster* and *Caenorhabditis elegans*.

1.8 Blood cell development

During hematopoiesis all blood cell types are generated from multipotent hematopoietic stem cells (HSC) found in the bone marrow (Morrison et al., 1995). In an asymmetric division, they renew themself but also produce precursor cells that lose the potential to self-renew. However, still being multipotent, they can give rise to different daughter cell types (Morrison and Kimble, 2006). The first major branching point appears with the separation into the common myeloid and common lymphoid progenitor cells (Figure 1.4). Towards terminally differentiated cell types like erythrocytes, macrophages or T-cells precursors lose the potential for different fates but gain specific morphological and functional properties.

Differentiated blood cells are classified into red blood cells, white blood cells and platelets. While lymphoid progenitors give only rise to a subgroup of white blood cells (B-cells, T-cells and natural killer cells), myeloid progenitors produce all subtypes. The above classification is based on the function of the cells. Erythrocytes (red blood cells) are the most abundant cell type in blood, responsible for its characteristic red color. As they are highly enriched with hemoglobin, their main purpose is the distribution of oxygen in the organism. Megakaryocyte-derived thrombocytes (platelets), lost the cell nucleus like erythrocytes. Their main function is to aggregate in cluts to stop bleeding after blood vessel injuries (Machlus et al., 2014). All remaining blood cell types are classified as white blood cells and fulfil functions in innate and adaptive immunity.

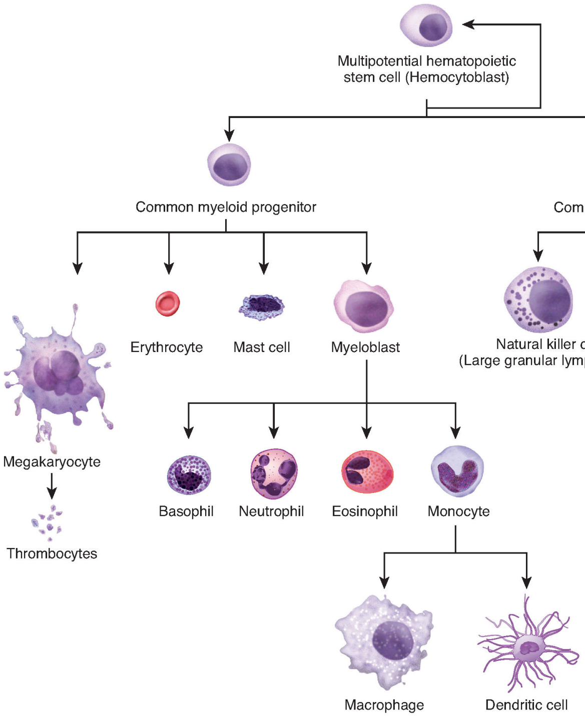


Figure 1.4: Hematopoietic cells differentiate from multipotent stem cells in the bone marrow to differentiated cell types released to blood (adapted from OpenStax Anatomy and Physiology Textbook Version 8.25)

1.8.1 The role of transcription factors in hematopoiesis

While growth factors like interleukins (IL) and colony-stimulating factors (CSF) deliver signals between cells to stimulate the differentiation and proliferation of a certain cell type, transcription factors are the executors, who take the signal and shape the transcriptome and thereby the morphology of the cell (Ketley and Newland, 1997, Nakajima, 2011). C/EBPa, from the CCAAT-enhancer-binding protein family, is one of the key transcription factors for hematopoiesis. Together with PU.1, it primes cells for myeloid differentiation and directs them further into the monocyte/granulocyte branch (Ohlsson et al., 2016, Pundhir et al., 2018). This is not a binary process where expressed or not expressed discriminates fates, the expression level matters as well for decisions. In HSCs, C/EBPa is expressed at low levels but upon depletion was shown to not just induce proliferation, but also trigger a loss of self-renewal (Ye et al., 2013, Hasemann et al., 2014). Furthermore, a high expression level of PU.1 is needed for the myeloid lineage, while low PU.1 levels lead to lymphoid cell development (Fiedler and

Brunner, 2012).

An indication that transcription factors not just direct cells into fates, but are actually required for their identity was given by the knockout of Pax5 in mature B-cells, which allowed them to dedifferentiate into uncommitted progenitors (Cobaleda et al., 2007). Enforced C/EBPa expression in B-cells boosted reprogramming into induced pluripotent stem cells along the same lines (Bueno et al., 2016).

The NF-kB protein complex is another crucial transcription regulator for hematopoiesis. Its activity is especially important in late hematopoietic development for almost all cell types, including erythrocytes, macrophages, dendritic cells, granulocytes, NK-cells, B and T-cells (Bottero et al., 2006). The function of NF-kB signalling is buffered by its different subunits to some extent. While mice deficient in p100 have reduced numbers of follicular B cells, a lack of cRel results in defects in germinal center B cells differentiation (Gerondakis et al., 1999). The lack of either p100, p105 or RelB results in an entire loss of marginal zone B-cells (Cariappa et al., 2000, Weih et al., 2001). A loss of multiple subunits has more severe consequences, e.g. the lack of p105 and p100 locks B-cells at an early stage of development in peripheral lymphoid organs, while the loss of p105 and p65 results in no lymphocytes in peripheral lymphoid organs at all (Horwitz et al., 1997, Gerondakis et al., 1999). Importantly, NF-kB signalling is not just crucial for differentiation but also for both B-cell activation and survival (Gerondakis and Siebenlist, 2010).

1.8.2 Intron retention in hematopoietic development

As briefly mentioned in the intron retention section, IR was observed in mammalian blood cell development. Actually, to our knowledge, blood together with neuronal tissue has the highest amount of publications reporting IR in mammals. So far it is mainly described to change during development in the terminal differentiation of myeloid cells. Specifically, it increases during terminal erythrocyte and granulocyte differentiation (Wong et al., 2013, Pimentel et al., 2016, Edwards et al., 2016). In granulocytes it was associated with a change in nuclear morphology induced by intron retention in the transcript of the lamina protein LMNB1 (Wong et al., 2013). For terminal erythropoiesis both publications found an association of genes affected by IR to splicing and iron homeostasis (Pimentel et al., 2016, Edwards et al., 2016). While iron homeostasis is essential for erythropoiesis, no functional mechanism indicating IR to be crucial for erythrocyte development could be shown, however. In contrast to increasing IR in terminal erythrocyte and granulocyte differentiation, progression from earlier stage megakaryocyte-erythrocyte progenitors to megakaryocytes and erythrocytes revealed a decrease in IR levels for most introns (Edwards et al., 2016). While so far there are no reports about IR changes in lymphoid differentiation, an increase in IR was observed after T-cell activation (Ni et al., 2016). In addition, a change in IR levels was associated with the RNA binding protein hnRNPLL (Cho et al., 2014).

1.8.3 LncRNAs in hematopoietic development

Although only a minority of the known lncRNAs are functionally characterized in mammals so far, some of them were found to be important for hematopoietic cell differentiation and function. In hematopoietic stem cells it has been shown that expression of the lncRNA H19 is important to suppress the insulin-like growth factor 1 (Igf1) and its receptor Igf1r to keep the HSCs in their quiescent state (Keniry et al., 2012, Venkatraman et al., 2013). In a manuscript identifying 323 novel lncRNAs in HSCs, two more lncRNAs (LncHSC-1 and LncHSC-2) were found to impact HSC self-renewal (Luo et al., 2015). Further down the differentiation tree into the myeloid lineage, lncRNAs, fulfilling regulatory functions, were found in many sub-branches. In macrophages the lncRNA lincRNA-Cox2 was found to modulate the expression of immune response genes during inflammatory response (Carpenter et al., 2013) while, lnc-DC, exclusively expressed in dendritic cells, was found to impair normal dendritic cell differentiation upon depletion (Wang et al., 2014). Furthermore, HOX antisense intergenic RNA myeloid 1 (HOTAIRM1) was found to activate the HOX genes HOXA1 and HOXA4 important for granulocyte differentiation (Zhang et al., 2009) while LincRNA-EPS was found prevent apoptosis during terminal differentiation of erythroblasts (Hu et al., 2011). Similar to intron retention, less lncRNAs are described for the lymphoid branch. While in type 2 helper T-cells (Th2) the lncRNA Linc-MAF-4 recruits LSD1 and EZH2 to the promoter of MAF and thereby represses it, which in turn skews T-cell differentiation toward the Th2 phenotype (Ranzani et al., 2015), less is known about lncRNA function in B-cells.

Chapter 2

INTRON RETENTION IN BLOOD CELLS

Intron retention (IR) is a potential mechanism between transcription and translation to alter the outcome of gene expression in a temporal and quantitative manner. So far, IR was observed in a wide set of tissues of multiple vertebrate species in a relatively low fraction of all transcribed genes (Braunschweig et al., 2014). However, most publications to date found higher levels of IR in developmental processes and differentiation, especially in neuronal tissues and blood (Wong et al., 2013, Pimentel et al., 2016, Edwards et al., 2016, Middleton et al., 2017). While differential IR was associated with blood cell differentiation for granulocytes, erythrocytes and megakaryocytes, a general overview, of how much the different cell types are affected compared to each other, is missing. Furthermore, it is unclear if IR preferentially affects the myeloid branch to which all of the aforementioned cells belong, or if it is also present in the lymphoid branch. In this part of the thesis we are going to address the following questions:

1. How do differentiated blood cell types compare in IR levels?
2. Are there branches of hematopoietic development with differential IR not characterized?
3. Which cell type specific processes and properties are affected by IR?

4. How do splicing modulators relate to IR events?

Intron retention is tightly associated with regulation of splicing factors and proliferative activity during B-cell development.

Ullrich S, Guigó R. Dynamic changes in intron retention are tightly associated with regulation of splicing factors and proliferative activity during B-cell development. Nucleic Acids Res. 2020 Feb 20;48(3):1327–40. DOI: 10.1093/nar/gkz1180

Chapter 3

LNCRNAS IN TRANSDIFFERENTIATION

Similarly to the role of AS, the role of the non coding genome during differentiation in the hematopoietic lineage is poorly characterized. In this context, deep transcriptome sequencing that greatly facilitated the exploration of alternative splicing events also advanced the exploration of the non-coding genome. In principle, every *de novo* annotation study in a cell type or process slightly off the well explored cell culture path finds a couple to hundreds of new non-coding transcripts. In recent annotations of the human genome about 15,000 long non-coding RNAs (lncRNAs) are reported, however, less than 5% of them are functionally characterized (Amaral et al., 2011, Quek et al., 2014, Ma et al., 2015, https://www.gencodegenes.org/stats/current.html). One challenge in genomics today, with steadily increasing numbers of new transcripts, is to scale approaches for functional characterization of those transcripts. This section of the thesis aims to develop a CRISPR/Cas9-based workflow to delete basically any stretch of genomic DNA (<5 kb) for many sites in parallel, and to apply it specifically to differentiation in the hematopoietic lineage. More specifically in this section of the thesis we are going to focus on the following technical and biological aspects:

1. Development of a method for CRISPR/Cas9 screening with paired guide RNAs to excise any set of regulatory regions in the genome

2. Development of a method for efficient quantification of the sequencing output of paired guide screens

3. Exploration of lncRNA and protein coding gene expression patterns during transdifferentiation of (B-cell like) BLaER1 cells to macrophages

4. Disruption of lncRNA and protein coding gene expression by CRISPR knockout and monitoring the effect on transdifferentiation

Screening for novel regulators that affect speed and efficiency of transdifferentiation from B-cell like BLaER1 cells to macrophages with CRISPR/Cas9.

Sebastian Ullrich, Carme Arnan, Alexandre Esteban, Ramil Nurtdinov, Sílvia Pérez-Lluch, Rory Johnson, Roderic Guigó.

Introduction

The hematopoietic system has been widely used to understand differentiation processes of pluripotent cells into differentiated cells due to its accessibility for experimental investigation. Initially, it was believed that transcription factors guide cells in a stepwise process into determined fates by altering their transcriptome (Morrison et al., 1995, Kondo et al., 1997, Akashi et al., 2000a, Akashi et al., 2000b, Chao et al., 2008, Doulatov et al., 2010, Notta et al., 2011). However, recently with the use of single cell data it was demonstrated that those transitions are more fluid than initially expected (Velten et al., 2017).

Furthermore, it was shown that those transitions can not just occur down the hematopoietic differentiation tree but also, by induction, from one branch to another. Either expression of raf/ras oncogenes or an activated form of the M-CSF receptor (M-CSFR) could transdifferentiate cells already committed to B-cell differentiation (lymphoid lineage) to closely related macrophages (myeloid lineage) with, however, low efficiency (Klinken et al., 1988, Borzillo et al., 1990).

In later work, it could be demonstrated that a single transcription factor C/EBPa induced by estradiol was sufficient to transdifferentiate murine B-cells into macrophages with 100% efficiency within 2-3 days (Xie et al., 2004, Bussmann et al., 2009). During the transdifferentiation process, it is crucial to shut down the B-cell related expression program and activate the macrophage related. Along these lines, it was found that PU.1, a downstream transcription factor activated by C/EBPa, however, needed for both B-cells and macrophages (Schebesta et al., 2002), is a major determinant for shaping the transcriptome in one direction or the other (DeKoter and Singh, 2000). More recently in a screening of human lymphoma and leukemia B-cell lines, the lymphoblastic leukemia B-cell line BLaER1 was found to transdifferentiate to macrophages upon induced activation of C/EBPa within 5-7 days (Rapino et al., 2013). While transdifferentiation happens efficiently in both mouse and human, it remains elusive what causes the differences in the speed of transdifferentiation and which downstream targets of C/EBPa and PU.1 are essential for transdifferentiation.

In the following, we identified genes with peaking or increasing expression profiles during the course of transdifferentiation in human. We created a library with guide pairs for CRISPR/Cas9 excision of 163 lncRNA transcription start sites and coding regions in the initial exons of 939 protein coding genes from the above selected genes. Out of the targeted genes, we identified 28 candidate genes potentially affecting transdifferentiation.

Results

Selection of lncRNAs and protein coding genes with activated expression pattern

In principle, lineage specific key transcription factors drive differentiation processes by regulating a wide set of downstream genes to change cellular properties. However, it requires experimental interference to understand which of those genes are essential for speed and efficiency of the differentiation. To study the downstream regulatory network in a well established differentiation setup, we chose the transdifferentiation model of B-cell like lymphoma cells to macrophages that leads to nearly 100% efficient conversion to macrophages. For induction of the process, a Burkitt Lymphoma derived cell line BLaER1 was created that stably expresses a fusion protein of CEBPa with an estrogen receptor hormone binding domain. When β-estradiol is added to the cell, it binds to the fusion protein and induces its translocation into the nucleus where it induces the transcriptional program leading to macrophage properties and morphology (Figure 1a). In addition, a supply of IL-3 and M-CSF in the culture medium is needed but not sufficient to stimulate transdifferentiation.

During the 7 day period needed in human cells for the transdifferentiation changes in the cell population to occur, they can be monitored by flow cytometry. CD19, a cell surface molecule found on B-cells, decreases its abundance, while MAC1, a cell surface marker for macrophage identification, increases it (Figure 1b).

To investigate the transcriptomic transformation during transdifferentiation, we performed RNA-seq of the cells at 12 time points throughout the process in two biological replicates. In principle, we assumed that there are two types of interesting expression profiles that we want to investigate further. On the one hand, genes that are peaking in their expression throughout the process with a potential role in transforming cells into a transition state before turning into macrophages. On the other hand, genes that are upregulated with potential functions for the new transcriptomic identity. We determined subsets of lncRNAs (rep1 n=642, rep2 n=536) and protein coding genes (rep1 n=4804, rep2 n=4552) with minimum peak expression values as well as expression changes outlined in the methods section. Especially, to select genes that peak early in the process, we clustered them with k-means clustering in sufficiently large subclusters (16 lncRNA clusters and 36 protein coding clusters) (Figure S1). From genes with an overlap in the temporal shape of expression between the replicates, we selected 163 lncRNAs and 939 protein coding genes for targeting with CRISPR/Cas9 (Figure 1c and 1e). Peaking lncRNAs had their highest median expression at 36 hours with

1.6 FPKM, while upregulated lncRNAs had a higher median expression of 2.79 FPKM at the terminal time point (Figure 1d). For protein coding genes where overall expression is higher, we also found that peaking genes had a lower median expression at peak than upregulated at 7 days (18.79 FPKM, 23.98 FPKM respectively) (Figure 1f).

Design of the library and screening

With the shortlisted lncRNAs and protein coding genes, we used CRISPETa (Pulido-Quetglas et al., 2017) to design paired guide RNAs (pgRNAs) to target the TSS region for lncRNAs and exonic regions for protein coding genes (Figure 2a). For each gene we designed 10 guide pairs. In addition to the targets, we designed guide pairs for (rat-CEBPa, human-CEBPa, SPI1, ITGAM, eGFP and mCherry), each of them targeted with 50 guide pairs. As negative controls, we designed 10 guide pairs each for 100 intergenic regions. Then we ordered a library of about 12,000 oligonucleotides of 165 bp in size containing both pgRNAs. The library was cloned into the plasmid pDECKO (Figure 2b).

For performing the experiment, we infected BLaER1 cells, stably expressing Cas9, with the pDECKO lentiviral library at a low lentiviral dose (Figure 2c). After 20 days of selection for cells containing the plasmid, we induced the transdifferentiation by adding β-estradiol, IL-3 and M-CSF. After 3 and 6 days (respectively), we FACS sorted cells into populations that were differentiated (high) or retarded (low), regarding their expression of the B-cell and macrophage specific surface markers (CD19 and Mac1 respectively). We performed the experiment in two biological replicates (Figure S2). Due to slightly different distributions of the cells in the two experiments, the gates from FACS sorting differ, especially for the differentiated sub-population.

After extraction of the genomic DNA from each fraction, we performed two PCR steps in order to amplify the integrated pDECKO plasmid. In the first PCR step, we added staggered oligos to avoid the same bases being read for the constant region during illumina sequencing to minimize technical issues during base calling (Figure 2d and 2e). In the second PCR step, we added Illumina barcoded oligos. We then pooled the libraries to have about 20 million reads per sorted subfraction and sequenced on the Illumina platform.

Paired guide quantification pipeline with adjustable matching precision

Unlike for sgRNA screens, quanitification of the sequenced pgRNAs is not that straightforward, as pairing needs to be kept, while single guide designs were used in multiple guide pairs for a given target gene. In particular, for developing a pipeline that efficiently quantifies the representation of each pair, we needed to address two major issues. The first issue was, due to repetitive structures in the plasmid, that primers for amplification needed to be placed about 100 bp away from pgRNA1, which resulted in poor quality of the sequences derived from the end of the 150 bp reads. Thus, the pipeline needed to be adjustable to mismatches resulting from the poor read quality, in order to get sufficient numbers of reads for having enough statistical power to detect targets with as weak effects as it would be expected from lncRNAs with low expression. The use of staggered oligos made the quantification less straightforward, as the position inside the sequenced reads was not fixed. However, it was needed to minimize technical issues during sequencing. To maximize the quantification outcome, we located a 4 bp constant sequence upstream of the pgRNAs, in the expected region in the reads and extracted the pgRNA lying after it. We then merged both pgRNAs, while pgRNA1 from the reverse strand was reverse complemented to have both sequences in the same orientation (Figure 3a). For mapping with STAR mapper (the most widely used mapper for RNA-seq samples), we created an artificial chromosome for each guide pair from the guide design provided by CRISPETa and mapped the merged reads against it. Due to the low memory footprint of the artificial genome, this quantification strategy can be applied even on mobile computers with moderate specifications (minimum requirements: single core CPU, 4GB RAM, 10GB disk space). From the resulting BAM files generated by STAR, we aggregated the mapped reads to count tables containing the representation of each guide pair in the analyzed sample.

Running the pipeline without allowing for any mismatches, we could only make use of about 25 to 30% of the reads. Hence, we investigated how many mismatches are tolerated without resulting into too many reads to multiple guide pairs (Figure S3a). Allowing for one more mismatch each resulted in a steep increase of mapped reads until a saturation point is reached between 10-15 mismatches, depending on the sample. For further analysis, we allowed for a maximum of 13 mismatches to stay below 1% of multi-mapped reads for all samples of both replicates (Figure S3c). Spearman correlation values of 0.95 - 1.00 between samples mapped with zero mismatches, compared with up to 13 mismatches, justified the usage of the quantification data with substantially more reads and therefore higher statistical power (Figure S3b). For the samples used in further analysis, we started with 20 to 35 million reads, of which we mapped on average 57% for replicate 1 and 62% for replicate 2 (Figure 3b and S3c). Clustering of Spearman correlation values for both replicates of all samples used for further analysis resulted in aggregation of the replicates from samples with retarded cells by time (D3_low and D6_low) (Figure 3c). Differences in the distributions of transdifferentiated cells and therefore different gate settings for FACS sorting resulted in clustering by replicate (Figure 3c and S2). Inspecting the representation of guide pairs at the time when transdifferentiation was induced resulted in a relatively low dynamic range of the distribution within the technical limitations that was similar to the initial representation of the plasmid library amplified (Figure 3d). In contrast, during the course of transdifferentiation (3 and 6 days), a small fraction of guide pairs gets enriched while most others get depleted (Figure 3d).

Identification of lncRNAs and protein coding genes that retard transdifferentiation

With the counts from both replicates (for 3 and 6 days each), we computed the differentiation retarding effect (DRE), which is the ratio of normalized counts from the sub population with retarded retention (low) divided by the counts from the transdifferentiated population (high).

DRE values > 1 indicate that disrupting the correct expression of the targeted gene reduces the efficiency in which cells transdifferentiate from BLaER1 cells to macrophages.

As a sanity check, we compared DRE values between both replicates. For rat-CEBPa, which we used as one of the positive controls, we observe DRE values > 1 for all tested guide pairs. Furthermore, DRE values correlate between the biological replicates (r=0.41, Spearman) (Figure 4a). Due to high sequence conservation between human and rat, rat-CEBPa targets the rat CEBPa estrogen-receptor fusion-protein DNA sequence as well as the intrinsic human-CEBPa. pgRNAs targeting mCherry (as a negative control) are not expected to affect transdifferentiation. In both replicates only few (3 out of 50) have DRE values >1 (Figure 4a). Along the same lines, the majority of a set of 1000 pgRNA, designed to target 100 intergenic regions, had very low DRE values (Figure 4a). For both, mCherry and intergenic regions, we did not observe positive correlations between the replicates (0.005, 0.008, Spearman), which makes it more likely that unexpectedly high DRE values come from random virus integration and other methodological limitations.

To visually inspect the change from 3 to 6 days and the separation between positive and negative controls, we plotted the average count distribution of both replicates for the retarded fraction against the differentiated fraction (Figure 4b). CEBPa, as a positive control (green), visually separates well from negative ones, targeting intergenic regions (red). The diagonal, which represents equal distribution between transdifferentiated cells and retarded ones, is covered by pgRNA guide pairs at 3 days after induction of the experiment. With further progress, after 6 days, most intergenic targets shifted towards higher cell proportions in transdifferentiated fractions. As a consequence, separation between positive and negative controls increases.

In order to find potential targets affecting the transdifferentiation, we selected all guide pairs with DRE values (mean of both replicates) in the highest decile (for 3 days DRE > 1.89, 6 days DRE > 0.44) (Figure 4c). For 3 and 6 days separately, we required potential targets to have at least 2 identical pgRNA pairs in that upper decile for both replicates. For the 3 day (6 day) time point this resulted in 18 (50) lncRNAs and 86 (135) protein coding genes. Comparing the distribution for all pgRNAs of all selected targets with positive and negative controls revealed significant differences between their distributions (both sided t-test) (Figure 4d).

Finally, we compared the potential targets with the output of MAGeCK, a tool to analyse CRISPR screening experiments, and overlapped the initially selected targets with ones that had at least a p-value < 0.05 from the MAGeCK output (both time points were treated separately) (Figure 4e).

Overall, we observed stronger effects for protein coding genes, but nevertheless, after careful inspection of all targets, we suspect that LINC00847 and RP11-84C10.2 are the lncRNA with the highest potential to affect transdifferentiation efficiency. For protein coding genes, we hypothesize that FURIN, CEACAM1 and NFE2 have potential on actively impacting transdifferentiation efficiency. FURIN is an amino acid cleaving enzyme that processes pre-proteins in major proteins and thereby activates them (Wise et al., 1990, Kiefer et al., 1991). The cell-cell adhesion molecule CEACAM1 was found to have regulatory functions in

T-cells and could potentially be important for the adhesion of macrophages at sites of infection (Nagaishi et al., 2006). The transcription factor NFE2 was found to be essential for regulating erythroid and megakaryocytic maturation and differentiation, but also impacts the renewal of hematopoietic stem cells (Shivdasani 2001, Gasiorek and Blank 2015, Di Tullio et al., 2017).

Discussion

Questioning which noncoding and coding genes impact the efficiency of transdifferentiation from B-cell like BLaER1 cells to macrophages, we selected about 1,100 potential target genes and screened for retardation of transdifferentiation at an intermediate time point (3 days) of the process and closer towards the end (6 days). While the initial idea was to use predominantly non-coding target genes, the lack of lncRNAs changing expression and being expressed at least at one time point above a considerable level (>1 RPKM) led us to target 85% percent protein coding genes to take full advantage of the oligonucleotide library size.

With no established strategy for quantification of the sequencing output available for a paired guide design, we developed a pipeline easy to run on moderate hardware that maps the reads against the screening library with adjustable tolerance for mismatching nucleotides. However, due to limitations of the plasmid design, especially pgRNA1 falls in a suboptimal end region of the read sequence resulting in up to 70% of read loss.

In general, the library complexity was high at the beginning of the experiment and decreased towards the end as expected. That effect could be caused by the impairment of transdifferentiation and the higher proliferative potential of the lymphoma derived BLaER1 cells, compared to quiescent macrophages. In addition, DRE values were highly affected by insertion effects of the virus, delivering the pgRNA pairs. This was not very obvious in CEBPa where, due to its central importance, all guide pairs had high DRE values and correlated quite well. However, for SPI1 and more obvious for IL3RA, the latter was not designed as a control initially, only a fraction of the guide pairs showed a retarding effect, while others had lower DRE values than pgRNA pairs targeting intergenic regions. Also, reproducibility between replicates was lower for both. Even more for the potential targets, where we in some cases observed very high DRE values for one guide design that could not be reproduced in the second replicate and vice versa for hits in the second replicate. In those cases, we observed binary effects of one design working very well, while all others did not show a retarding effect on transdifferentiation. We believe that these results stem from the disruptive effect of virus integration into the human genome. Therefore, for targets being selected, we required them to have at least two designs with DRE values among the top 10% for the given time point for both biological replicates to be selected. By requiring them, in addition, to come from the same pgRNA pair, we assumed this to be an indication of a working design. As a supplemental strategy for analyzing the data, we used MAGeCK implemented merely to quantify CRISPR screens performed with sgRNAs. While we could not derive positive hits form MAGeCK considering commonly used cutoffs of 0.01 or 0.05

for p-values corrected for multiple testing, as only controls surpassed that level, we saw some indications for potential targets inspecting the qq-plots for targets sheering out from the expected p-value distribution. Among them were FURIN, CEACAM1 and NFE2, which we consider the most promising targets, but unfortunately no lncRNAs. A drawback regarding MAGeCK seems to us that reproducibility of guide designs between the replicates was not given as much weight as we would have wanted. For that reason, we used the combined strategy, taking from the top of the DRE value distribution and intersecting with MAGeCK output.

As the refinement of the positive targets needs further validation, we are planning on retesting the selected targets individually to reach further conclusions on their transdifferentiation retarding potential.

Methods

Experimental methods

Library cloning

A ssDNA library composed by 12,000 oligos of 165 nt was purchased from Twist Biosciences. The library was amplified to obtain dsDNA using emulsion PCR as described in Schütze et al., 2011, and cloned into pDECKO_mCherry vector (Addgene 78534) following the 2 cloning steps described in Aparicio-Prat et al., 2015. ENDURA electrocompetent cells (Bionova Scientifica) were used to ensure high efficiency transformation and avoid recombination errors. Several transformations were done in parallel and for the 1st step of cloning (intermediate plasmid) about 486,450 bacterial colonies were collected and processed together in a maxiprep. To eliminate the background (empty plasmid), we took advantage of that insert-1 (in the intermediate plasmid) contains unique restriction sites (EcoRI and BamHI) which are not present in the original backbone. Digesting the intermediate plasmid resulted in a linear product that we could distinguish from the circular empty backbone and purify it in a gel. For the 2nd step of cloning, 50 ng of BsmbI-digested intermediate plasmid was mixed with 1 ul annealed Insert-2 (diluted 1:20) and 1 ul of T4 DNA ligase (Thermo Scientific) and incubated for 4h at 22ºC. Several transformations with ENDURA electrocompetent cells were done in parallel and for the 2nd step of cloning (final plasmid) more than 107,650 bacterial colonies were collected and processed together in a maxiprep. The final maxiprep library was deep sequenced to check for the quality and representation of the different constructs.

Cell culture and lentivirus production

Human BLaER1 cells (Rapino et al. 2013) were kindly provided by Thomas Graf (CRG, Barcelona) and grown in RPMI medium (Invitrogen) supplemented with 10% heat-inactivated foetal bovine serum (FBS), 2 mM L-glutamine, and 100 U/ml penicillin G sodium (Rapino et al. 2013). BLaER1 cells were first infected with a plasmid containing

Cas9 fused to BFP and blasticidin resistance (Addgene 78545), selected for more than 5 days with blasticidin (15 µg/ml) and sorted using a BD FACS Aria instrument. These cells, stably expressing Cas9 were then infected with the pDECKO library. For lentivirus production, we performed 80 co-transfections of HeK293T virus packaging cells with 3 ug of the pDECKO_mCherry plasmid library and 2.25 ug of the packaging plasmid pVsVg (Addgene 8484) and 750 ng of psPAX2 (Addgene 12260) using Lipofectamine 2000 (according to manufacturer's protocol). Transfection media was changed on the following day to RPMI. In total, 400 ml of viral supernatant was collected 48h post transfection and used for overnight infection of 90x10E6 BLaER1-Cas9 cells at a density of 250.000 cells/ml. The percentage of infection was checked as the percentage of mCherry+ cells with a Fortesa cell cytometer. Infection rate ranged between 2%-4%. Cells were selected with puromycine (2 ug/ml) and blasticidine (20 ug/ml) for 18-19 days. 15 million of the BLaER1-Cas9 library infected cells were induced for transdifferentiation into macrophages. After incubation for 3 days/6 days, as described previously (Rapino et al. 2013), they were collected for FACS sorting.

FACS sorting

30x10E6 cells were counted and resuspended in 300 ul PBS+3% FBS in the presence of FcR Block reagent. Cells were incubated for 10 minutes and 5 ul of the anti-CD19 antibody conjugated with BV510 (562947, Becton Dickinson) and 5 ul of anti-MAC1 antibody conjugated with PE-Cy7 (25-0118-41, Labclinics) were added. Cells were incubated for 30 minutes, washed with PBS and resuspended in 2 ml of PBS+3%FBS. Topro-3 was added as a viability marker. Cells were sorted in a BD FACS Aria instrument.

Sample processing for sequencing

Genomic DNA was extracted from the FACS sorted cells with the GeneJET Genomic DNA purification kit (Thermo Scientific). A first PCR step was done by Phusion polymerase (Thermo Fisher) using 500 ng of genomic DNA and staggered oligos (Table S1) with the presence of 6% DMSO, annealing temperature of 60ºC and a total of 20 cycles of amplification. Up to 6 PCR reactions were combined, the amplicons were gel-purified and 2 ng was used as a template for a second PCR. For the second PCR step, we used Illumina barcoded oligos (Table S2), an annealing temperature of 60ºC and a total of 8 cycles of amplification. Samples were purified with Agencourt Ampure beads (Beckman Coulter), quantified with a Qubit fluorometer (Thermo Scientific) and checked for quality in a Bioanalyzer (Agilent) prior to be sequenced in an Illumina Hiseq 2500 (150-paired end sequencing).

Computational methods

Target gene selection from transcriptomics data

The selection of target genes for the CRISPR screen was based on RNA-seq data sampled at 12 time points (0h, 3h, 6h, 9h, 12h, 18h, 24h, 36h, 48h, 72h, 120h, 168h) during

transdifferentiation of human BLaER1 cells to macrophages. Two biological replicates were analyzed separately. Resulting target genes were overlapped between the two replicates. For both protein coding genes and lncRNAs GENCODE annotation v22 was used as gene model.
The 19,814 protein coding genes from gencode v22 were filtered for a minimum average expression of at least 1 FPKM and at least 4x fold change between highest and lowest expression value along the temporal profile, resulting in 4,804 genes for replicate 1 and 4,552 for replicate 2 remaining. Those genes were clustered separately for each replicate into 36 expression profiles with k-means clustering in R. All profiles with peaking or increasing expression were pooled within each replicate and then intersected between the replicates, resulting in 939 genes used for screening.
For long non-coding RNAs the biotypes: processed transcript, 3prime overlapping ncrna, sense intronic, antisense, macro lncRNA, lincRNA, non-coding and sense overlapping were selected from gencode v22 resulting in 14,855 lncRNAs. They were filtered to be non overlapping and 5 kb from other genes on the same strand and 50 bp on the opposite strand in their TSS. Furthermore, they were required to have a minimum average expression of 0.1 FPKM, a minimum of 1 FPKM at any of the time points and at least a two fold change between minimum and maximum expression, resulting in 642 lncRNAs for replicate 1 and 536 for replicate 2. Those lncRNAs were clustered separately for each replicate into 16 expression profiles (due to lower gene count over protein coding genes) with k-means clustering in R. All profiles with peaking or increasing expression were pooled within each replicate and then intersected between the replicates, resulting in 163 lncRNAs used for screening.

Guide RNA design

CRISPETa (Pulido-Quetglas et al., 2017) was used to design gRNA pairs. For protein coding genes, the exonic region was targeted, while for lncRNAs the promoter/TSS region was targeted.

Creation of an artificial genome

Based on the initial pgRNA library with two guides per target, a concatenated 41 bp sequence of the two pgRNAs (pgRNA1 21 bp, pgRNA2 20bp) was created and converted into FASTA format. STAR mapper (2.4.2a) was used to index the genome with adjusting the standard settings by the following parameter for small genomes:

```
--genomeSAindexNbases 6
```

In the resulting genome, each pgRNA pair represented one out of 11,666 chromosomes with a length of 41 bp.

Extraction of pgRNA guide regions from Illumina reads

Dynamic trimming of Illumina reads was done in perl by pattern matching the insertion site of the pgRNAs in the plasmid sequence (“ACCG” for pgRNA1 in the window of 15-55 bp of read2, “AAAC” for pgRNA2 in the window of 100-150 bp of read1). The extracted 20 bp fastq sequences for the pgRNA2 were reverse complemented and concatenated to the 21 bp fastq sequences for the pgRNA1. Fusion reads with fewer than 20 bp sequence length were filtered out.

Mapping of index reads against the artificial genome

Mapping was performed with STAR version 2.4.2a with the following parameters:

```
STAR --runMode alignReads --runThreadN 8 --genomeDir /users/resources/genome
--readFilesCommand zcat --readFilesIn pgRNA1_pgRNA2.fastq.gz --alignIntronMax
1 --outSAMtype BAM SortedByCoordinate --outSAMunmapped Within
--limitBAMsortRAM 3000000000 --outFilterMultimapNmax 1
--outFilterMismatchNmax 11 --outFilterMatchNmin 30
--outFilterMatchNminOverLread 0.1 --outFilterMismatchNoverLmax 0.9
--outFilterScoreMinOverLread 0.1
```

Due to the various PCR amplifications and the limitations with the position within the 150 bp reads, especially for pgRNA2, we adjusted the mapping parameters to allow for a maximum of 13 mismatches within the 41 bp sequence. After a previous parameter scanning within a range of 0-25 mismatches, we chose a maximum of 13 mismatches to stay below 1% multimaps within the mapped libraries. The count for each guide pair within the mapped libraries was aggregated from the BAM files with samtools.

Analysis of the read counts

The count tables generated from the BAM files were filtered for guide pairs having at least 5 counts in the initial sample at t=0h, to ensure a minimum representation at the beginning of the experiment. For both biological replicates, the ratio of the FACS sorted retarded (low) over differentiated fraction (high) was computed for both 3d and 6d. From the distribution of ratios form each of the 12k guide pairs, all guide designs found above the 90th percentile were selected. Guide pairs for which both biological replicates of each time point had at least 2 guide designs for a given target above these 90th percentile were overlapped with a run of MAGeCK v0.5.7 (Li et al., 2014) for both timepoints separately. MAGeCK was run in its test mode, where significant targets are estimated from prequalified count tales. For the merge with the previous by ratio selected targets a FDR < 0.25 was required.

Table S1: Staggered Oligos

Stag0nt_F	TTCAGACGTGTGCTCTTCCGATCTGTGGAAAGGACGAAACACCg
Stag1nt_F	TTCAGACGTGTGCTCTTCCGATCT**H**GTGGAAAGGACGAAACACCg
Stag2nt_F	TTCAGACGTGTGCTCTTCCGATCT**HM**GTGGAAAGGACGAAACACCg
Stag3nt_F	TTCAGACGTGTGCTCTTCCGATCT**HMM**GTGGAAAGGACGAAACACCg
Stag4nt_F	TTCAGACGTGTGCTCTTCCGATCT**NNMM**GTGGAAAGGACGAAACACCg
Stag5nt_F	TTCAGACGTGTGCTCTTCCGATCT**NNMMC**GTGGAAAGGACGAAACACCg
Stag0nt_R	CCCTACACGACGCTCTTCCGATCTcaagatctagttacgccaagcttAAA
Stag1nt_R	CCCTACACGACGCTCTTCCGATCT**D**caagatctagttacgccaagcttAAA
Stag2nt_R	CCCTACACGACGCTCTTCCGATCT**DK**caagatctagttacgccaagcttAAA
Stag3nt_R	CCCTACACGACGCTCTTCCGATCT**DKK**caagatctagttacgccaagcttAAA
Stag4nt_R	CCCTACACGACGCTCTTCCGATCT**NNKT**caagatctagttacgccaagcttAAA
Stag5nt_R	CCCTACACGACGCTCTTCCGATCT**NNBBT**caagatctagttacgccaagcttAAA

Table S2: Illumina Oligos (barcode in bracket).

TS-HT-D5x-96-f	AATGATACGGCGACCACCGAGATCTACAC(ttggcaga) ACACTCTTTCCCTACACG ACGCTCTTC
TS-HT-D7x-r	CAAGCAGAAGACGGCATACGAGAT(xxxxxxxx) GTGACTGGAGTTCAGACG TGTGCTCTTC
Replicate 1 Barcodes	**Barcode sequence**
TS-HT-D7x-6-r	CTGAGGTT
TS-HT-D7x-12-r	CGGTTGGT
TS-HT-D7x-60-r	TGACCAGC
TS-HT-D7x-18-r	CCGGTTCT
TS-HT-D7x-24-r	AATGCAAT
TS-HT-D7x-17-r	GCTCCAGT
TS-HT-D7x-55-r	TTAGTTGC
TS-HT-D7x-67-r	TCAGATAC
TS-HT-D7x-62-r	TTCAAGCC
TS-HT-D7x-64-r	GACTAACC
TS-HT-D7x-79-r	CAGAGAGA
Replicate 2 Barcodes	**Barcode sequence**
TS-HT-D7x-1-r	CGTTGGTT
TS-HT-D7x-9-r	TGGTTCTT
TS-HT-D7x10-r	GTCTTCTT
TS-HT-D7x-22-r	GAACCGAT
TS-HT-D7x-32-r	ACTTACGG
TS-HT-D7x-43-r	CTTGATAG
TS-HT-D7x-61-r	GCAACGCC
TS-HT-D7x-78-r	TCGAACGA
TS-HT-D7x-80-r	GCCATTCA
TS-HT-D7x-82-r	CCTGCTCA
TS-HT-D7x-93-r	GTCGCGAA

Citations

Akashi K, Reya T, Dalma-Weiszhausz D, Weissman IL (2000a) Lymphoid precursors. Curr Opin Immunol. 2000 Apr;12(2):144-50. doi: 10.1016/S0952-7915(99)00064-3

Akashi K, Traver D, Miyamoto T, Weissman IL. (2000b) A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature. 2000;404:193–197. doi: 10.1038/35004599

Aparicio-Prat E, Arnan C, Sala I, Bosch N, Guigó R, Johnson R (2015) DECKO: Single-oligo, dual-CRISPR deletion of genomic elements including long non-coding RNAs. BMC Genomics. 2015 Oct 23;16:846. doi: 10.1186/s12864-015-2086-z.

Borzillo GV, Ashmun RA, Sherr CJ (1990) Macrophage lineage switching of murine early pre-B lymphoid cells expressing transduced fms genes. Mol Cell Biol. 1990 Jun;10(6):2703-14.

Bussmann LH, Schubert A, Vu Manh TP, De Andres L, Desbordes SC, Parra M, Zimmermann T, Rapino F, Rodriguez-Ubreva J, Ballestar E, Graf T (2009). A robust and highly efficient immune cell reprogramming system. Cell Stem Cell 5, 554–566. doi: 10.1016/j.stem.2009.10.004

Chao MP, Seita J, Weissman IL (2008) Establishment of a normal hematopoietic and leukemia stem cell hierarchy. Cold Spring Harb Symp Quant Biol. 2008;73:439–449. doi: 10.1101/sqb.2008.73.031.

DeKoter RP and Singh H (2000) Regulation of B lymphocyte and macrophage development by graded expression of PU.1. Science. 2000 May 26;288(5470):1439-41. doi: 10.1126/science.288.5470.1439

Di Tullio A, Passaro D, Rouault-Pierre K, Purewal S, Bonnet D (2017) Nuclear Factor Erythroid 2 Regulates Human HSC Self-Renewal and T Cell Differentiation by Preventing NOTCH1 Activation. Stem Cell Reports. 2017 Jul 11;9(1):5-11. doi: 10.1016/j.stemcr.2017.05.027.

Doulatov S, Notta F, Eppert K, Nguyen LT, Ohashi PS, Dick JE (2010) Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. Nat Immunol. 2010 Jul;11(7):585-93. doi: 10.1038/ni.1889.

Gasiorek JJ and Blank V (2015) Regulation and function of the NFE2 transcription factor in hematopoietic and non-hematopoietic cells. Cell Mol Life Sci. 2015 Jun;72(12):2323-35. doi: 10.1007/s00018-015-1866-6. Epub 2015 Feb 27.

Kiefer MC, Tucker JE, Joh R, Landsberg KE, Saltman D, Barr PJ (1991) Identification of a second human subtilisin-like protease gene in the fes/fps region of chromosome 15. DNA Cell Biol. 1991 Dec;10(10):757-69. DOI: 10.1089/dna.1991.10.757

Klinken SP, Alexander WS, Adams JM (1988) Hematopoietic lineage switch: v-raf oncogene converts Emu-myc transgenic B cells into macrophages. Cell. 1988 Jun 17;53(6):857-67. doi: 10.1016/S0092-8674(88)90309-1

Kondo M, Weissman IL, Akashi K (1997) Identification of clonogenic common lymphoid progenitors in mouse bone marrow. Cell. 1997;91:661–672. doi: 10.1016/S0092-8674(00)80453-5

Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, Irizarry RA, Liu JS, Brown M, Liu XS (2014) MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol. 2014;15(12):554.

Morrison S, Uchida N, Weissman I (1995) The biology of hematopoietic stem cells. Annu Rev Cell Dev Biol. 1995;11:35–71. doi: 10.1146/annurev.cb.11.110195.000343

Nagaishi T, Iijima H, Nakajima A, Chen D, Blumberg RS (2006) Role of CEACAM1 as a regulator of T cells. Ann N Y Acad Sci. 2006 Aug;1072:155-75.

Notta F, Doulatov S, Laurenti E, Poeppl A, Jurisica I, Dick JE (2011) Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. Science. 2011 Jul 8;333(6039):218-21. doi: 10.1126/science.1201219.

Pulido-Quetglas C, Aparicio-Prat E, Arnan C, Polidori T, Hermoso T, Palumbo E, Ponomarenko J, Guigo R, Johnson R (2017) Scalable Design of Paired CRISPR Guide RNAs for Genomic Deletion. PLoS Comput Biol. 2017 Mar 2;13(3):e1005341. doi: 10.1371/journal.pcbi.1005341.

Rapino F, Robles EF, Richter-Larrea JA, Kallin EM, Martinez-Climent JA, Graf T (2013) C/EBPα induces highly efficient macrophage transdifferentiation of B lymphoma and leukemia cell lines and impairs their tumorigenicity. Cell Rep. 2013 Apr 25;3(4):1153-63. doi: 10.1016/j.celrep.2013.03.003.

Schebesta M, Heavey B, Busslinger M (2002) Transcriptional control of B-cell development. Curr Opin Immunol. 2002 Apr;14(2):216-23. doi: 10.1016/S0952-7915(02)00324-2

Schütze T, Rubelt F, Repkow J, Greiner N, Erdmann VA, Lehrach H, Konthur Z, Glökler J (2011) A streamlined protocol for emulsion polymerase chain reaction and subsequent purification. Anal Biochem. 2011 Mar 1;410(1):155-7. doi: 10.1016/j.ab.2010.11.029.

Shivdasani RA (2001) Molecular and transcriptional regulation of megakaryocyte differentiation. Stem Cells. 2001;19(5):397-407.

Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, Hirche C, Lutz C, Buss EC, Nowak D, Boch T, Hofmann WK, Ho AD, Huber W, Trumpp A, Essers MA, Steinmetz LM. (2017) Human haematopoietic stem cell lineage commitment is a continuous process. Nat Cell Biol. 2017 Apr;19(4):271-281. doi: 10.1038/ncb3493. Epub 2017 Mar 20.

Wise RJ, Barr PJ, Wong PA, Kiefer MC, Brake AJ, Kaufman RJ (1990) Expression of a human proprotein processing enzyme: correct cleavage of the von Willebrand factor precursor at a paired basic amino acid site. Proc Natl Acad Sci U S A. 1990 Dec;87(23):9378-82.

Xie H, Ye M, Feng R, Graf T (2004) Stepwise reprogramming of B cells into macrophages. Cell. 2004 May 28;117(5):663-76. doi: 10.1016/S0092-8674(04)00419-2

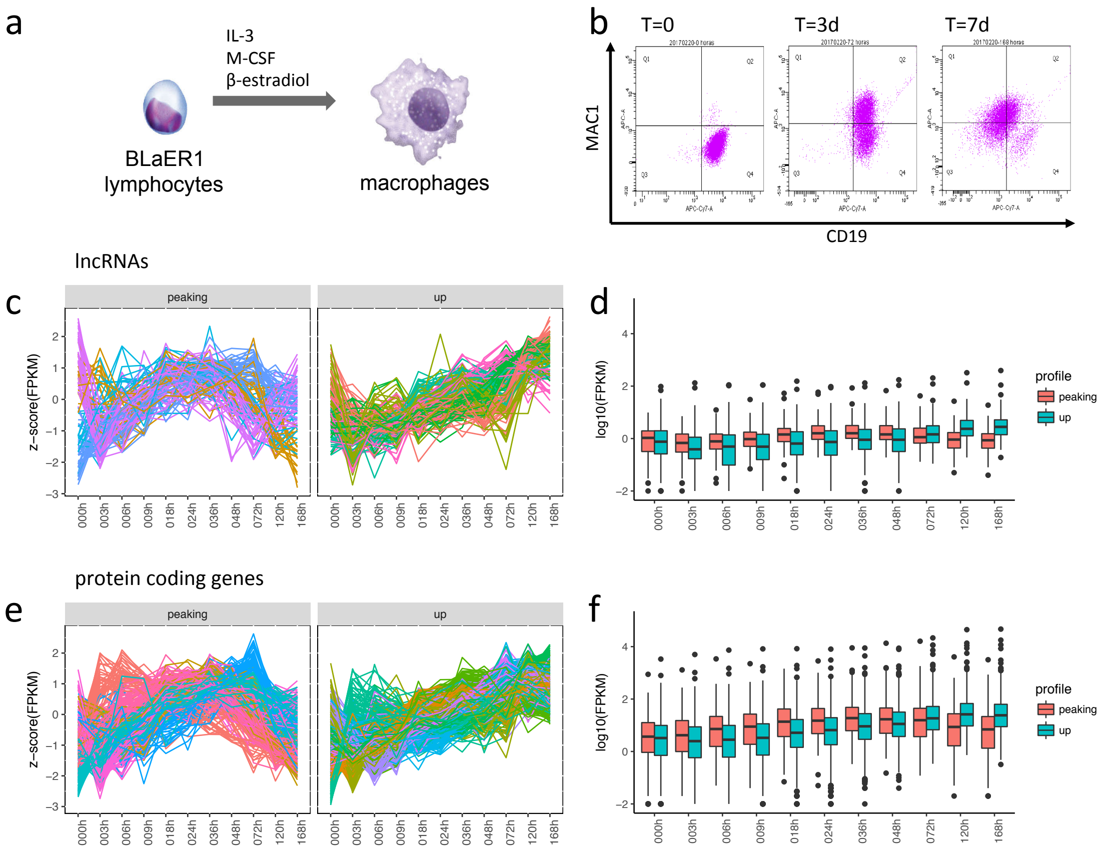


Figure 1. Transdifferentiation of B-cell like BLaER1 cells into macrophages is accompanied by a dynamic transcriptomic remodelling of the cells. (a) B-cell like BLaER1 lymphocytes transdifferentiate to macrophages in the presence of Interleukin 3 (IL-3) and Macrophage colony-stimulating factor (M-CSF) upon β-estradiol induced release of CEBPa to the nucleus. (b) Flow cytometric analysis of cell surface markers at 0, 3 and 7 days after induced transdifferentiation. CD19 was used as a marker for B-cell identity and MAC1 as a marker for macrophage identity. BLaER1 cells reside in the lower right corner with high CD19 and low MAC1 abundance and vice versa for macrophages. (c) Merged clusters (k-means, 16 initial clusters) of lncRNA (n=163) expression profiles with peaking and upregulated expression during transdifferentiation. FPKM values were log10 transformed before the normalisation by z-score. (d) Log10 transformed expression profiles of the 163 lncRNAs. (e) Merged clusters (k-means, 36 initial clusters) of protein coding gene (n=939) expression profiles with peaking and upregulated expression during transdifferentiation. FPKM values were log10 transformed before the normalisation by z-score. (f) Log10 transformed expression profiles of the 939 protein coding genes.

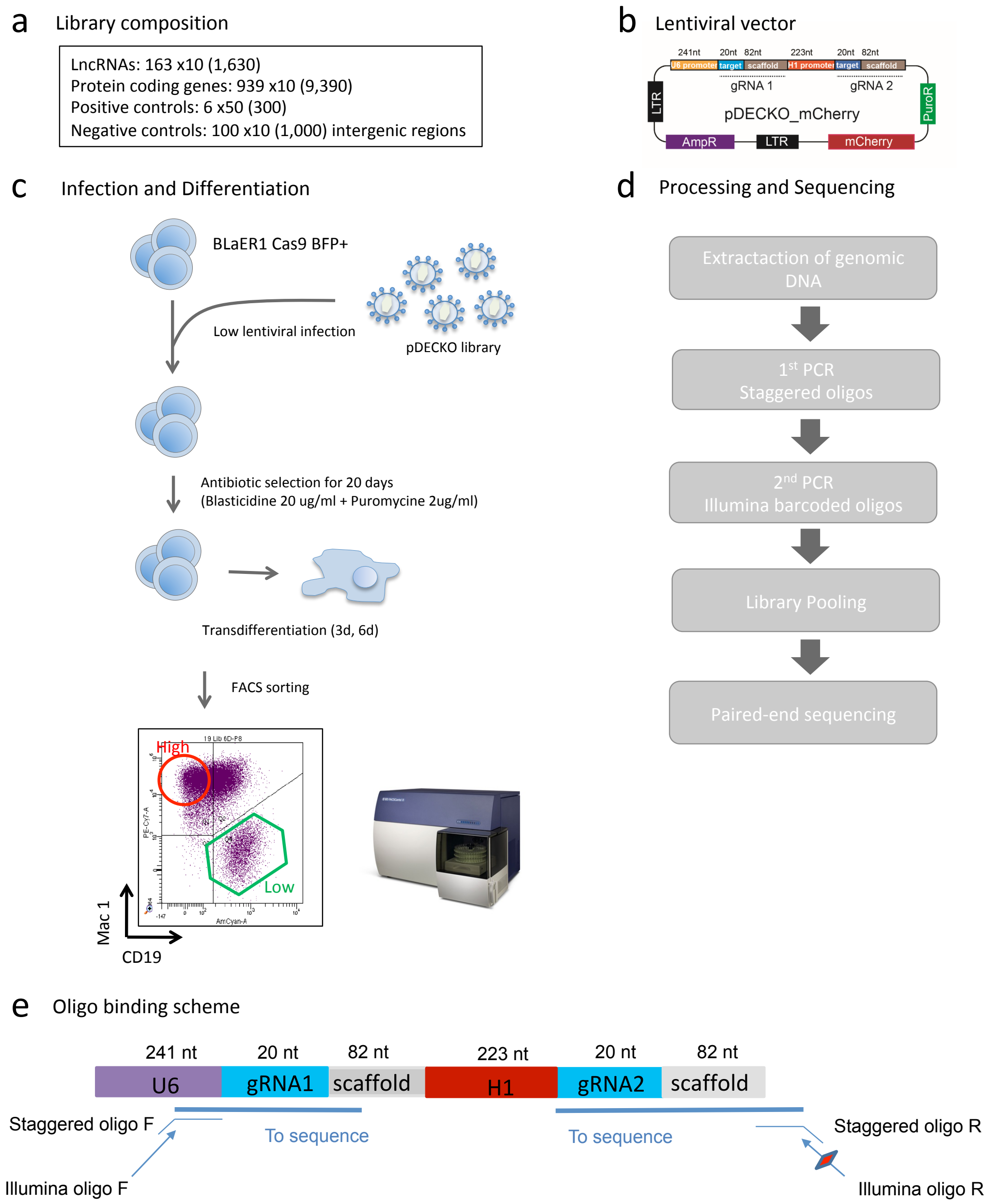


Figure 2. pDECKO CRISPR Library . (a) Library composition. (b) pDECKO vector scheme. (c) General workflow of the experiment. BLaER1 Cas9 BFP+ cells were low infected with lentiviral pDECKO library. Cells were double selected with antibiotics for 20 days and differentiated to macrophages for 3 days and 6 days. Cells were labelled with specific antibodies against surface markers CD19 (for lymphocytes) and Mac-1 (for macrophages). High, intermediate and low differentiated populations were monitored and harvested by FACS sorting. (d) Cell processing and deep sequencing. Genomic DNA of cells was extracted and PCR amplified in 2 steps. The 1st PCR step was done with specific staggered oligos and the 2nd PCR step was done with Illumina oligos that introduce a barcode for pooling several samples. Barcodes were sequenced with 150 bp paired-end Illumina sequencing. (e) Scheme of oligo binding sites in the pDECKO construct.

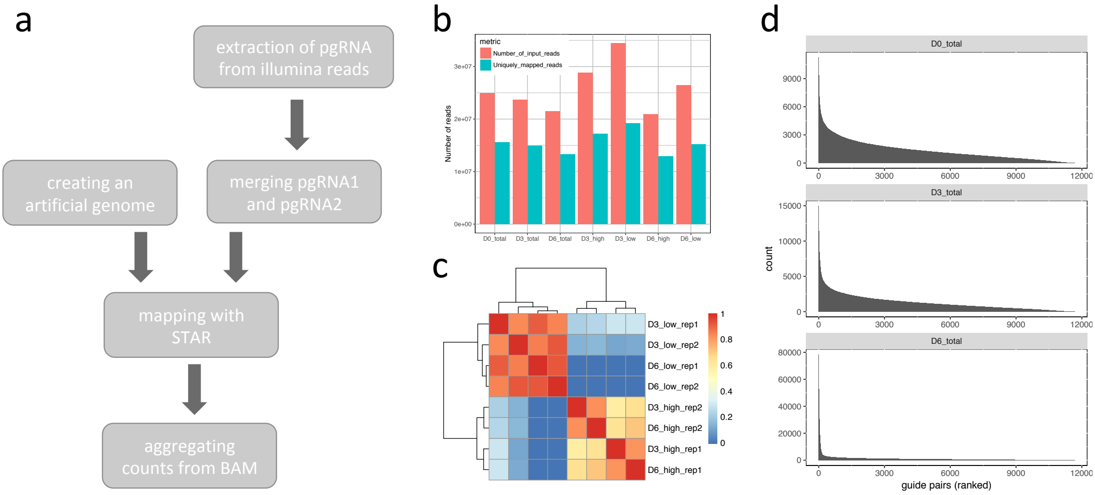


Figure 3. Adjusted quantification procedure for efficient quantification of paired pgRNA screen. (a) Schematic flow diagram displaying the steps from pgRNA FASTQ sequences to count tables per guide pair. In short, pgRNA sequences are extracted from FASTQ sequences by finding the proximal conserved plasmid sequence. pgRNA2 is reverse complemented (only needed for paired end sequencing) and merged with pgRNA1. Both are mapped as one sequence to the merged expected sequences converted into artificial chromosomes with STAR mapper. Count tables are generated from BAM files by aggregation. (b) Read counts from the samples used for accessing targets that affect transdifferentiation efficiency. Initial read counts per sample range between 20 to 35 million reads of which on average 55% were mapped against perfect library sequences with not more than 13 mismatches. Counts represent the average of both replicates. (c) Spearman correlation of both replicates for the samples used for accessing targets that affect transdifferentiation efficiency. (d) Ranked distribution of counts per pgRNA guide pair in the control samples at 0, 3 and 6 days that contain all cells independent of B-cell and macrophage marker abundance.

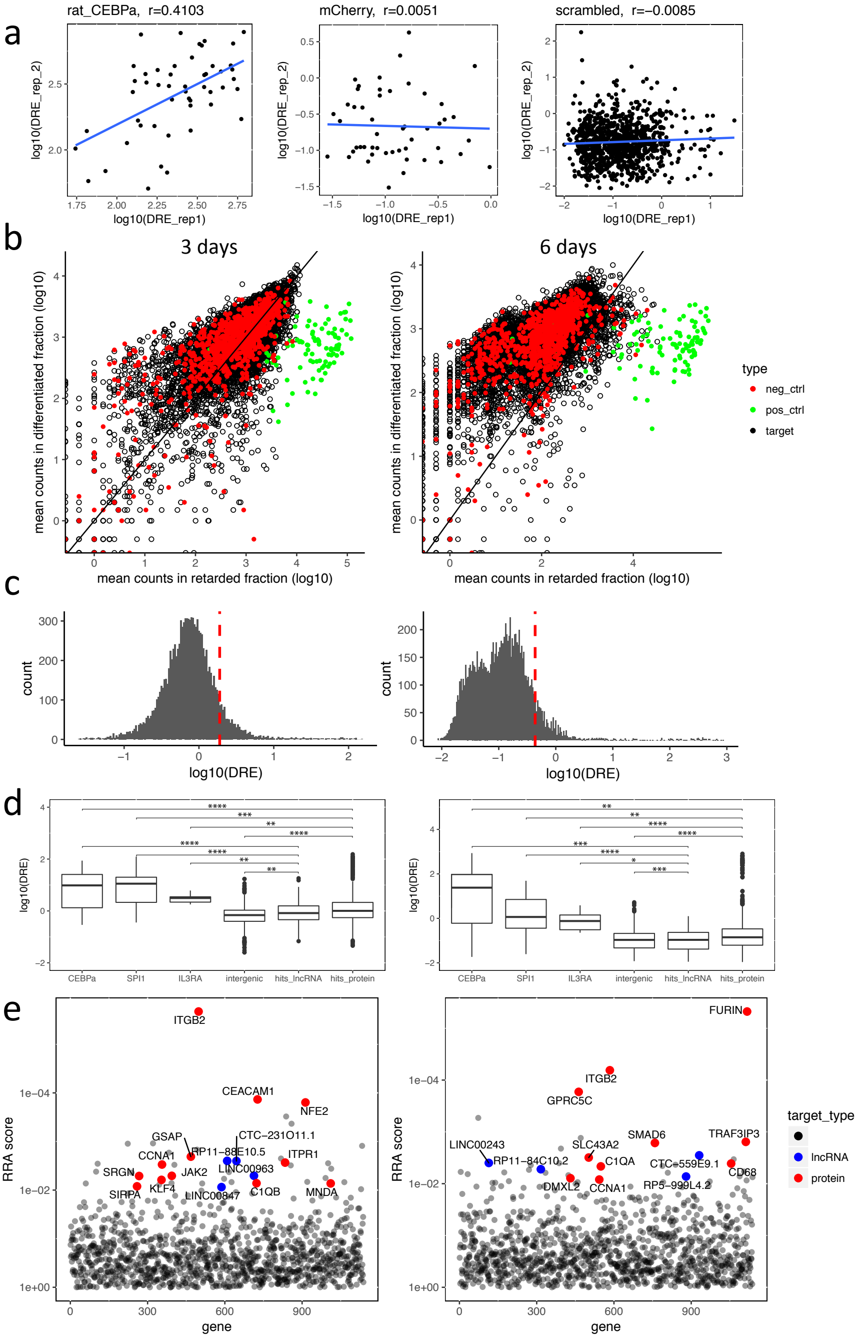
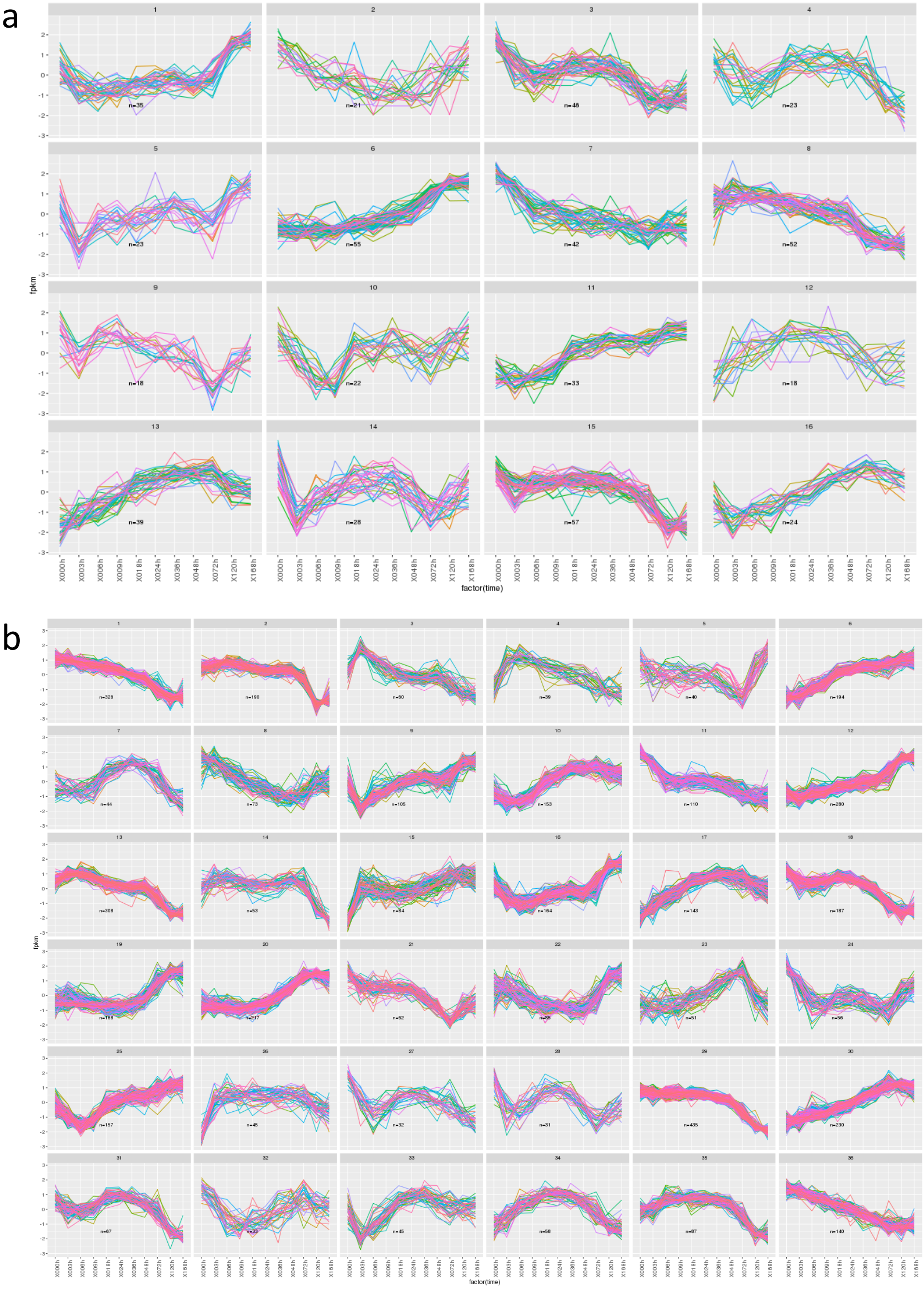
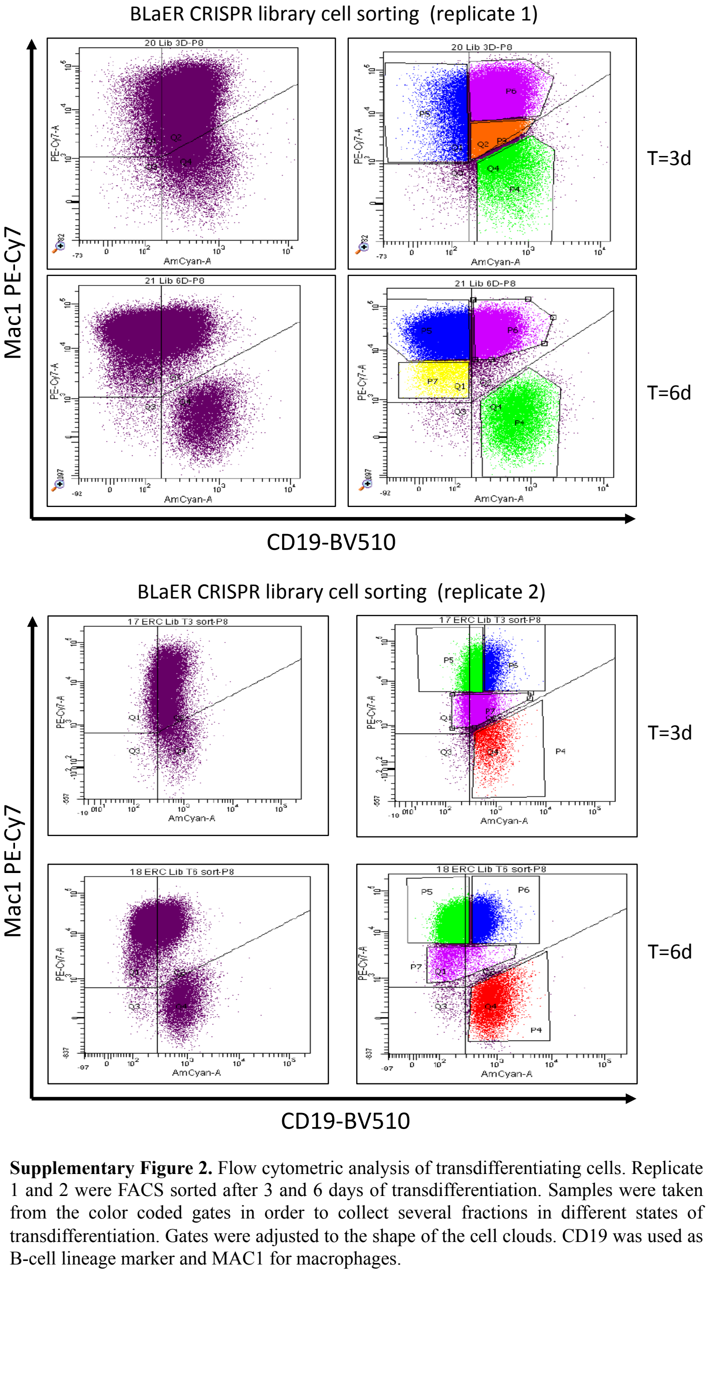


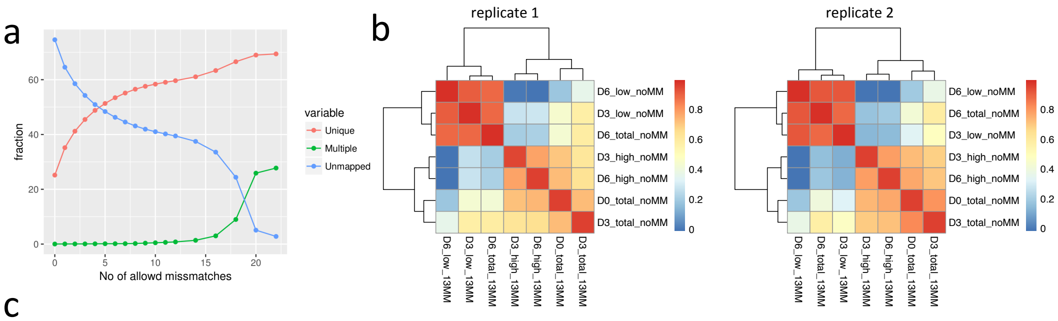
Figure 4. Identification of lncRNAs and protein coding genes that retard transdifferentiation (a) Correlations of differentiation retarding effect (DRE) between replicates for three controls after 6 days of transdifferentiation induction. Rat CEBPa serves as positive while mCherry and a set of targeted intergenic regions (scramble) serve as negative controls. DRE is the ratio of reads from retarded versus transdifferentiated fractions. Spearman correlation values are stated above. (b) Scatterplot of log10 transformed counts in retarded versus differentiated fractions. Guide pairs targeting CEBPa in green and intergenic in red. (c) Distribution of DRE values. Highest decile is marked by dashed red line. Values are means of both replicates. (d) Comparison of DRE from guide pairs targeting lncRNAs and protein coding genes of the highest decile with controls. Values are means of both replicates. (e) RRA scores (Measure of effect strength, aggregate over all guides for each target) for all targets computed by mageck. Targets with at least two identical guide pairs in highest decile for both biological replicates are highlighted.



Supplementary Figure 1. Clusters of lncRNAs and protein coding genes with similar expression characteristics during transdifferentiation of BLeAR1 cells to macrophages. (a) lncRNAs clustered into 16 expression profiles by k-means clustering. (b) Protein coding genes clustered into 36 expression profiles by k-means clustering. FPKM values were log10 transformed and normalized by z-transformation.



Supplementary Figure 2. Flow cytometric analysis of transdifferentiating cells. Replicate 1 and 2 were FACS sorted after 3 and 6 days of transdifferentiation. Samples were taken from the color coded gates in order to collect several fractions in different states of transdifferentiation. Gates were adjusted to the shape of the cell clouds. CD19 was used as B-cell lineage marker and MAC1 for macrophages.



c

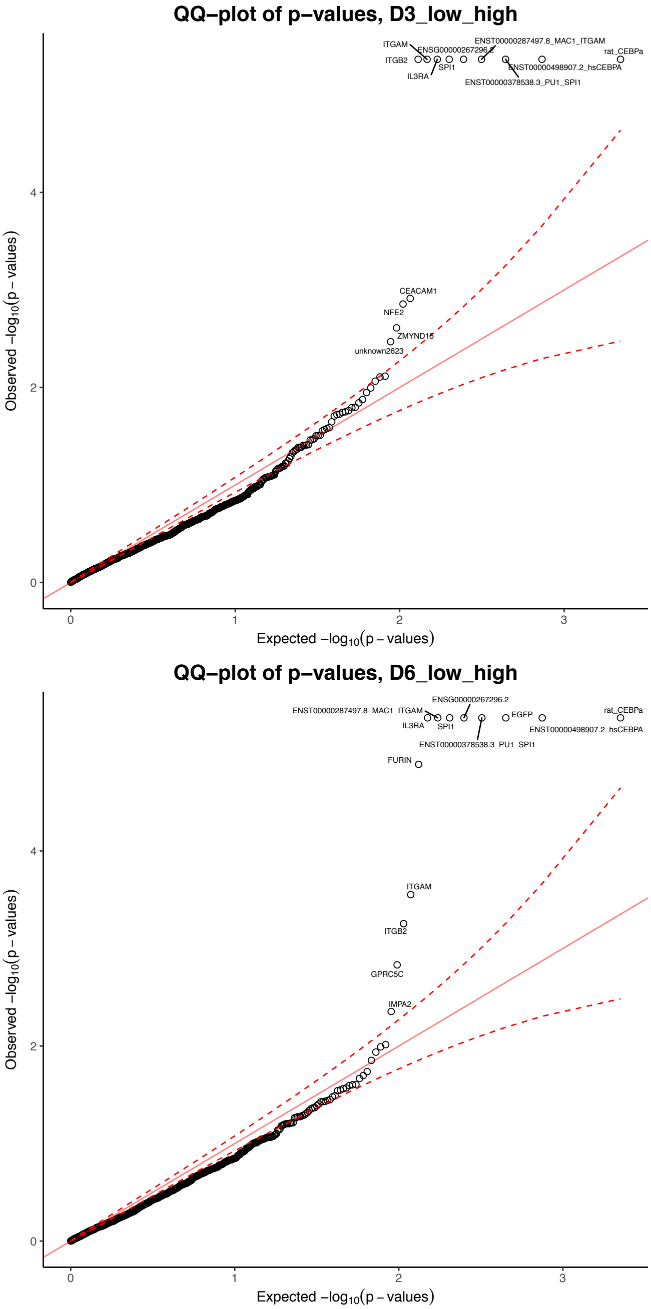
replicate 1

Number_of_input_reads	28360044	26764175	23174328	20461721	25660393	35163035	33261946	23457263	17622205	21620659	22693503
Average_input_read_length	40	40	40	40	40	40	40	40	40	40	40
Uniquely_mapped_reads_number	15755565	16007907	13648684	12123869	14590551	19804799	18582918	13083276	10320734	12840942	13366676
Uniquely_mapped_reads_%	55.56%	59.81%	58.90%	59.25%	56.86%	56.32%	55.87%	55.77%	58.57%	59.39%	58.90%
Average_mapped_length	40.58	40.66	40.59	40.61	40.63	40.6	40.61	40.59	40.6	40.62	40.62
Number_of_splices:_Total	0	0	0	0	0	0	0	0	0	0	0
Number_of_splices:_Annotated_(sjdb)	0	0	0	0	0	0	0	0	0	0	0
Number_of_splices:_GT/AG	0	0	0	0	0	0	0	0	0	0	0
Number_of_splices:_GC/AG	0	0	0	0	0	0	0	0	0	0	0
Number_of_splices:_AT/AC	0	0	0	0	0	0	0	0	0	0	0
Number_of_splices:_Non-canonical	0	0	0	0	0	0	0	0	0	0	0
Mismatch_rate_per_base,_%	2.95%	2.72%	2.97%	2.95%	2.83%	2.86%	2.96%	2.97%	2.95%	2.87%	2.88%
Deletion_rate_per_base	0.04%	0.05%	0.05%	0.06%	0.05%	0.05%	0.05%	0.05%	0.05%	0.05%	0.05%
Deletion_average_length	1	1	1	1	1	1	1	1	1	1	1
Insertion_rate_per_base	0.06%	0.08%	0.08%	0.07%	0.08%	0.08%	0.08%	0.08%	0.07%	0.08%	0.07%
Insertion_average_length	1.32	1.42	1.43	1.4	1.45	1.43	1.43	1.45	1.4	1.44	1.44
Number_of_reads_mapped_to_multiple_loci	0	0	0	0	0	0	0	0	0	0	0
%_of_reads_mapped_to_multiple_loci	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Number_of_reads_mapped_to_too_many_loci	235855	140376	120157	106017	131841	195573	165894	116831	102014	107748	108742
%_of_reads_mapped_to_too_many_loci	0.83%	0.52%	0.52%	0.52%	0.51%	0.56%	0.50%	0.50%	0.58%	0.50%	0.48%
%_of_reads_unmapped:_too_many_mismatches	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
%_of_reads_unmapped:_too_short	43.56%	39.55%	40.52%	40.15%	42.54%	43.06%	43.55%	43.65%	40.78%	40.03%	40.53%
%_of_reads_unmapped:_other	0.05%	0.12%	0.07%	0.08%	0.08%	0.06%	0.08%	0.07%	0.07%	0.08%	0.09%

replicate 2

Number_of_input_reads	26644511	26966759	25395692	24415795	24981137	26829569	24517712	20150115	29973705	33691107	30855643
Average_input_read_length	40	40	40	40	40	40	40	40	40	40	40
Uniquely_mapped_reads_number	17547956	17617769	16316787	15763071	15778659	17218229	14626243	12993696	19594707	18564557	13898156
Uniquely_mapped_reads_%	65.86%	65.33%	64.25%	64.56%	63.16%	64.18%	59.66%	64.48%	65.37%	55.10%	45.04%
Average_mapped_length	40.76	40.76	40.77	40.75	40.75	40.77	40.77	40.76	40.78	40.74	40.74
Number_of_splices:_Total	0	0	0	0	0	0	0	0	0	0	0
Number_of_splices:_Annotated_(sjdb)	0	0	0	0	0	0	0	0	0	0	0
Number_of_splices:_GT/AG	0	0	0	0	0	0	0	0	0	0	0
Number_of_splices:_GC/AG	0	0	0	0	0	0	0	0	0	0	0
Number_of_splices:_AT/AC	0	0	0	0	0	0	0	0	0	0	0
Number_of_splices:_Non-canonical	0	0	0	0	0	0	0	0	0	0	0
Mismatch_rate_per_base,_%	1.73%	1.77%	1.76%	1.83%	1.90%	1.69%	1.71%	1.82%	1.69%	1.78%	1.83%
Deletion_rate_per_base	0.05%	0.06%	0.05%	0.06%	0.06%	0.05%	0.03%	0.05%	0.06%	0.05%	0.05%
Deletion_average_length	1	1	1	1	1	1	1	1	1	1	1
Insertion_rate_per_base	0.08%	0.08%	0.07%	0.08%	0.08%	0.08%	0.05%	0.09%	0.08%	0.08%	0.08%
Insertion_average_length	1.43	1.43	1.42	1.4	1.45	1.45	1.25	1.47	1.42	1.44	1.38
Number_of_reads_mapped_to_multiple_loci	0	0	0	0	0	0	0	0	0	0	0
%_of_reads_mapped_to_multiple_loci	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Number_of_reads_mapped_to_too_many_loci	131929	138169	138888	116321	124514	124723	169568	99229	150453	216878	164455
%_of_reads_mapped_to_too_many_loci	0.50%	0.51%	0.55%	0.48%	0.50%	0.46%	0.69%	0.49%	0.50%	0.64%	0.53%
%_of_reads_unmapped:_too_many_mismatches	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
%_of_reads_unmapped:_too_short	33.55%	34.06%	35.10%	34.82%	36.18%	35.25%	39.55%	34.90%	34.02%	44.15%	54.34%
%_of_reads_unmapped:_other	0.09%	0.10%	0.10%	0.15%	0.16%	0.11%	0.10%	0.12%	0.10%	0.11%	0.09%

Supplementary Figure 3. Statistics on quantification of pgRNA representation in the screening. (a) Uniquely mapped, multi-mapped and unmapped reads as a function of allowed mismatches during quantification of the sequenced samples in a range of 0 to 26 mismatches. (b) Spearman correlations of guide pair quantification between runs allowing for up to 13 mismatches and only allowing perfect matches. Values for identical samples were ranging between 0.95 and 1. (c) Detailed mapping statistics for both replicates in the quantification runs allowing for up to 13 mismatches.



Supplementary Figure 4. QQ-plots of target genes depleted by CRISPR/Cas9. QQ-plots are based on p-values generated by MAGeCK v0.5.7 (Li et al., 2014). MAGeCK was run in test mode comparing read counts of pgRNA pairs from differentiated versus undifferentiated fractions. Targets with fewer than two guide pairs, indicated as working by MAGeCK were removed. Samples collected after 3 days (upper panel) or 6 days (lower panel) of transdifferentiation.

Chapter 4

IHEC CONTRIBUTION

During the work in the core processing team for RNA-seq data in the Blueprint Epigenome consortia, I got involved in the IHEC umbrella organisation that, among other goals, aims to unify approaches to epigenetic data production, processing and analysis (Stunnenberg et al., 2016). As transcriptomic data is always a relevant complementation of ChIP-seq and DNA methylation data, we aimed to also find a unified approach for RNA-seq. The idea was, if all international members process their data in a unified way, data is comparable between centers, which facilitates the exchange of data without the requirement of local reprocessing. As a consequence, it would not just save researchers time but also avoid redundant computational costs and storage space. Furthermore, we aimed to provide the guidelines and processing pipelines to the community to use as a general standard.

In the framework of the quality standards sub-group, I was responsible for the RNA-seq data. I aggregated a set of more than 60 measures to the five most predictive ones, to characterize a transcriptomic data set in regard to most common biases causing poor quality data. Together with Sitanshu Gakkhar, I provided guidelines about how to compute the metrics and a bash script for computing them on any given BAM file.

Together with Emilio Palumbo, we merged the metrics computation into modular containerized integrative pipeline that can be executed in various computational environments to produce identical results.

RNA sequencing quality control metrics definition

Sebastian Ullrich[1], Sitanshu Gakkhar[3], Martin Hirst[2,3], Roderic Guigo[1]

1.Computational RNA Biology Group, Center for Genomic Regulation (CRG), Carrer del Dr. Aiguader, 88, 08003 Barcelona, Spain.
2.Department of Microbiology and Immunology, Michael Smith Laboratories, Centre for High-Throughput Biology, University of British Columbia, 2125 East Mall, Vancouver BC V6T1Z4, Canada.
3. Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, 675 W. 10th Avenue, Vancouver, BC V5Z 1L3, Canada.

This document aims to define the RNA Sequencing Quality Control Core Metrics and how to compute them. In general we highly recommend using only chastity passed reads.

1. Input RNA quality

The quality of the input sample RNA material is crucial for obtaining meaningful estimates for gene expression. As a wieldy used measure the **RNA integrity number** (RIN) indicates the degradation status of a given RNA sample. GTEx (www.gtexportal.org/) suggests a threshold for primary samples of 6 (range: 0 – 10) For samples coming from cell culture higher values should be expected (above 8).

2. Genomic contamination

Besides degradation of the input material, RNA samples can have remaining DNA molecules due to problems in the purification procedure. Those contaminations can be evaluated by the amount of reads from intergenic regions. The suggested measure "**proportion of intergenic reads**" is calculated as the following:

$$\frac{\text{number of mapped reads not overlaping with any transcript coordinates (including introns)} +/- \text{L}}{\text{number of all mapped reads per sample}}$$

L : Overhang extending transcript region by +/-500bp

Limitations mostly depend on the used annotation and are therefore prone to bias by novel transcripts. As a common standard we recommend to use the provided BED file derived from gencode annotation version 22 http://www.gencodegenes.org/releases/22.html

3. Library enrichment (riboZero, polyA+)

Ribosomal RNAs are highly abundant in most cells and would take a high fraction of the sequenced reads if they were not removed, resulting in a low representation of less abundant transcripts. No matter if cells are enriched for polyadenylated tails or specifically depleted for ribosomal RNAs, levels of ribosomal RNAs should be very low in the resulting libraries. In order to access successful depletion we recommend to use the "**fraction of reads mapping to ribosomal genes**" as measure.

$$\frac{\text{number of reads mapped to ribosomal genes}}{\text{number of mapped reads to the entire genome}}$$

We recommend using the Fasta files (humRibosomal.fa, hum5SrDNA.fa) provided within Illumina iGenomes for the corresponding genome assembly.

http://support.illumina.com/sequencing/sequencing_software/igenome.html

4. Library amplification/diversity

Most common library preparation procedures include PCR amplification steps of the fragments to increase the amount of material for the sequencing process. Over-amplification can occur when the number of cycles is high and the input material had a low RNA concentration. Resulting libraries have a low diversity of fragments whereas duplicates with identical start and end positions are abundant. We suggest using the “**fraction of reads flagged as duplicates**” as measure for library diversity, which is computed as the following:

$$\frac{\text{number of mapped reads}^{*}\text{ with the same start and end position as any other read from the same sample}}{\text{number of all mapped reads per sample}}$$

* read pairs in the case of paired end data

5. Mappability

The amount of reads of a RNA sequencing sample that can be mapped to the genome of the given species can be biased by a variety of factors from contamination by RNA or DNA of other samples coming from different organisms to errors during base calling. As a measure of those biases we suggest to use the “**proportion of mapped reads**“ which is computed as the following:

$$\frac{\text{number of all mapped reads per sample}}{\text{number of all reads in the provided sample fastq library file}}$$

The amount of added spike-ins can account for significant amount of reads not being mapped if not added to the genome towards which mapping is done.

Chapter 5

DISCUSSION

This work comprises the investigation of two different aspects of alternative gene regulation during hematopoiesis - intron retention in mature mRNA and regulatory lncRNAs. As results and pitfalls of both parts were discussed before in the relevant sections each, this discussion aims to integrate the observations of our work into the current research landscape and to give an outlook on future prospects.

The most important aims of modern genetics/genomics are to understand how information is translated from the DNA blueprint to complex phenotypes on an organismic scale and to dissect all information carrying genetic elements on the small scale. Furthermore, how these gene regulatory processes unfold in development and differentiation and what are the causes if something goes wrong along the lines, leading to disease and aging. Those functional pieces do not just comprise genes that create morphological structures, when translated to proteins, but also non-coding genes that contribute to protein complex formation and gene regulation. Besides those transcribed elements there are many more structural elements like enhancers and promoters that are themselves only subtypes of regulatory sequences in the DNA. Furthermore, there are regions of the genome biochemically modified like CpG islands and regions that contain information for transcripts to be bound by RNA binding proteins, e.g. in splicing regulatory elements within introns.

In order to gain a fundamental understanding of such regulatory structures and processes, large consortia have been formed to collect and analyse data. And indeed, their work led to a significant gain of information about the layers in which information is encoded in the DNA sequence in the past two decades. The human genome project, together with similar projects for other organisms, provided the basic sequence layer of the DNA. Based on that, projects such as ENCODE and FANTOM cataloged functional elements (Kawai et al., 2001, Birney et al., 2007, Djebali et al., 2012, Dunham et al., 2012). Which finally led to steadily updated gene annotations by e.g. GENCODE (Harrow et al., 2006, Harrow et al., 2012). In addition, FANTOM aimed to get transcriptomics data of as many cell types as possible, complemented by the human body map (HBM) project. The genotype tissue expression (GTEx) project aimed to get tissue expression data like the HBM project, but from different individuals to take genomic variations into account that allowed to find expression and splicing quantitative trait loci (eQTLs, sQTLs) (Li et al., 2017, Tan et al., 2017). In addition to genotype and transcriptomic data, projects like Roadmap (Bernstein et al., 2010), Blueprint Epigenome project (Adams et al., 2012), DEEP (Perner et al., 2014, Wallner et al., 2016) and ENCODE gathered epigenomic data on a variety of primary and cell cultured cells.

What all the projects have in common is that they provide a valuable resource of information for the community; their limitation is, however, that they are mostly descriptive, taking a static snapshot from the cell type or tissue they sample.

This limitation has been adjusted for in newer phases of ENCODE and FANTOM as well as in the Blueprint Epigenome project, where differentiation and developmental processes were traced. The value of taking time into consideration in this data becomes obvious, when comparing to other work that continues to find new transcript isoforms and not yet annotated lncRNAs, when a stimulus is provided to cells or they follow a differentiation path.

Regarding the first project presented in the thesis, focusing on intron retention, a background level of IR has been observed that is largely conserved in many vertebrate species (Braunschweig et al., 2014, Middleton et al., 2017). However, those retention events might not contribute much to regulation, as they appear to be rather static in their processing. Again, in those cases findings are mostly descriptive and valuable as a resource, but do not provide clues about biological consequences of IR. On the contrary, several other papers found IR shaping the transcriptomes of developing cells in brain and blood cell differentiation (Wong et al., 2013, Pimentel et al., 2016, Edwards et al., 2016). In blood cell differentiation observations were limited to the myeloid lineage. In our analysis, we found IR to appear in some branches of hematopoietic development not previously explored, including lymphoid cells development. In B-cells we observed an increase of IR from the bone marrow residing precursor to marginal zone B-cells, found in secondary lymphoid organs. Cells that undergo affinity maturation in the germinal center reaction displayed a sharp decline of IR. Increasing proliferation of those cells that undergo affinity maturation is strongly anti-correlated to the change in IR. Similarly, we observed negative correlations between proliferation and IR for granulocytes. In agreement with our observations in B-cells, differentiating to plasma cells we observed a loss of IR for monocytes differentiating towards terminally differentiated macrophages and dendritic cells.

On the contrary, erythrocytes and granulocytes displayed an increase of IR towards terminally differentiated stages of development.

It is important to mention that blood cells are rather easy to take samples from, compared to other mammalian tissues, as they are not adherent and do not aggregate to complex tissues morphologies. For that reason, the occurrence of IR during development and differentiation in other tissues is probably even less understood and potentially hides further clues about the biological relevance of IR. In the past, approaches have been made to experimentally dissect complex tissues with laser micro dissection or to computationally deconvolute the acquired data with linear models, with limitations on both sides (Simone et al., 1998, Zhao and Simon, 2010). A promising new avenue, for developmental processes and tissues hard to sample form, is three dimensional cell culture, where cells grow to complex shapes in a matrix-like culture medium. With this method, organs have been regrown *in vitro* as less complex smaller models. Among them, models for liver, the nervous, and the gastrointestinal system have been established (Dedhia et al., 2016, Bijsmans et al., 2017, Di Lullo and Kriegstein, 2017).

For lncRNAs, the second project in the thesis, it is even more true that the functional characterization lags behind the discovery of new lncRNAs. Currently, the human genome annotations contain more than 15,000 lncRNAs (GENCODE v28), from which only a small fraction is functionally explored (<5%) and with almost every deep transcriptomic study, exploring a new biological process, more lncRNAs are added. One of the major limitations is the throughput, in which lncRNAs can be screened in developmental and differentiation processes that they potentially affect. Whole organism knockouts (Zan et al., 2003), previously used to characterize the function of genes, are time-consuming to establish for mammalian model organisms, like mice, and can not be scaled for parallel investigation of hundreds to thousands of novel transcripts. The other approach, culturing cells *in vitro*, scales well for bigger sets of transcripts to investigate, but has limitations for delivering the perturbation. Short hairpin RNAs (shRNAs) as well as short interfering RNAs (siRNAs), used to deplete transcripts of interest, have limitations for infection/transfection efficiency, as well as for knockdown strength and duration.

A game changer for such screening efforts was the reengineering of the bacterial pathogen defence system CRISPR to target and disrupt any genomic locus. CRISPR is now becoming the standard for knockout screening, with libraries targeting the entire coding genome of humans and the most used model organisms being commercially available. Targeting non-coding genes is, however, still in its early phase, as a frame shift inducing double strand break is not sufficient to disrupt their function, more often based on their secondary structure. The key for targeting such transcripts is to remove fragments of the genome entirely, either the whole transcript or the promoter/TSS region, which is more practical for transcripts spanning genomic regions larger than 5 kilobases. Along those lines, the work presented in the second project within the thesis manuscript provides a simple way to set up and perform such CRISPR screens. With the presented method, lncRNA or enhancers can be targeted, employing commonly available tools like CRISPETa (Pulido-Quetglas et al., 2017), for designing guide RNA pairs and the pDECKO plasmid (Aparicio-Prat et al., 2015), for inserting them into the cell to be screened.

As a proof of concept, we applied the method to screen for potential regulators involved in reshaping the transcriptome during transdifferentiation of B-cell like BLaER1 cells to macrophages. We identified a set of 34 genes, consisting of lncRNAs and protein coding genes that we plan to validate further. A method similar to ours has been provided by Zhu et al. 2016. While they commercially provide the pooled library for the 671 lncRNAs, we believe that our method is easier to adapt for any set of genetic elements in any organism of interest. Furthermore, we developed an efficient pipeline to quantify the sequencing output of such pooled screens on average office hardware.

A further interesting direction regarding screening functional genomic elements are the methods CRISPRi and CRISPRa (Liu et al., 2017, Joung et al., 2017). In both methods the catalytic subunit of the cleavage enzyme is mutated and instead genomic targeting is used to deliver activating or deactivating proteins like histone remodelers to increase or repress expression/accessibility (Qi et al., 2013). However, all the methods that bring guide RNAs into cells by virus infec-

tion have the limitation that integration into the genome happens at random sites, causing false positive hits, when functional genomic structures are disrupted. The workaround so far was to screen with multiple guide designs (also to adjust for off target effects) in multiple biological or technical replicates. An interesting direction for further development would be the site specific integration of the plasmids as outlined by Recchia et al. in 2004. And in addition, to implement automated methods for screening organoid arrays with CRISPR methods, to target biological structures less accessible in the past.
As it was already indicated by previous work, the distinct layers of gene regulation like epigenetic remodelling, transcription factor activation and recruitment, lncRNAs function and alternative splicing are all interconnected. In this sense, the overall goal for understanding how complex organisms are build form genetic information would be to integrate all that layers to a global model.

Chapter 6

CONCLUSIONS

By investigating intron processing in cells of the hematopoietic lineage and screening for novel regulators that affect timing and efficiency of B-cell like BLaER1 cells transdifferentiating into functional macrophages, we reached the following conclusions:

- Retained introns alter the transcriptomes of cell types in both branches of hematopoietic differentiation, myeloid and lymphoid, with neutrophiles, B-cells and monocytes being most affected. Terminally differentiated and precursor cells are affected alike.
- In B-cells we observed that intron retention increases from early precursors in the bone marrow towards maturated cells, found in the blood and lymphoid organs. Follicular and marginal zone B-cells extracted from spleen had the highest intron retention values.
- Germinal center B-cells, undergoing affinity maturation, display a sharp decline in intron retention that reduces even further in antibody producing plasma cells. Memory B-cells, also derived from germinal center cells, regain high retention patterns.
- Genes enriched for introns with high differential IR during B-cell differentiation were almost exclusively associated with GO categories related to pre-mRNA processing and splicing.

- Developmental stages of B-cells highly impacted by intron retention were the most proliferative inactive stages. We also observed these negative correlation of IR and proliferation in granulocyte differentiation.

- For the majority of SR and hnRNP genes we observed a reduced expression in developmental stages with high IR. Several of those splicing regulators had an significant enrichment of binding sites in the introns with highest differential retention.

- For the introns with the highest differential IR we observed significantly higher sequence conservation among placentalia compared to introns with lower differential IR levels.

- During human BLaER1 to macrophage transdifferentiation we identified a set of 163 lncRNAs and 939 protein coding genes that have either a peaking or increasing expression pattern.

- Targeting the TSS (lncRNAs) or the coding sequences (protein coding genes) of those regulated genes led to a set of 28 potential regulators that may affect the efficiency of the transdifferentiation process.

Bibliography

Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S, Bird A, Bock C, Boehm B, Campo E, Caricasole A, Dahl F, Dermitzakis ET, Enver T, Esteller M, Estivill X, Ferguson-Smith A, Fitzgibbon J, Flicek P, Giehl C, Graf T, Grosveld F, Guigo R, Gut I, Helin K, Jarvius J, Küppers R, Lehrach H, Lengauer T, Lernmark Å, Leslie D, Loeffler M, Macintyre E, Mai A, Martens JH, Minucci S, Ouwehand WH, Pelicci PG, Pendeville H, Porse B, Rakyan V, Reik W, Schrappe M, Schübeler D, Seifert M, Siebert R, Simmons D, Soranzo N, Spicuglia S, Stratton M, Stunnenberg HG, Tanay A, Torrents D, Valencia A, Vellenga E, Vingron M, Walter J, Willcocks S (2012) BLUEPRINT to decode the epigenetic signature written in blood. Nat Biotechnol. 2012 Mar 7;30(3):224-6. doi: 10.1038/nbt.2153.

Ajith S, Gazzara MR, Cole BS, Shankarling G, Martinez NM, Mallory MJ, Lynch KW (2016) Position-dependent activity of CELF2 in the regulation of splicing and implications for signal-responsive regulation in T cells. RNA Biol. 2016 Jun 2;13(6):569-81. doi: 10.1080/15476286.2016.1176663. Epub 2016 Apr 20.

Albertson DN, Schmidt CJ, Kapatos G, Bannon MJ (2006) Distinctive profiles of gene expression in the human nucleus accumbens associated with cocaine and heroin abuse. Neuropsychopharmacology 31, 2304–2312. doi:10.1038/sj.npp.1301089

Allfrey VG, Faulkner R, Mirsky AE. (1964) Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc Natl Acad Sci USA. 1964 May;51:786-94.

Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS (2011). lncRNAdb: a reference database for long noncoding RNAs. Nucleic Acids Res 39: D146-151.

Ameur A, Zaghlool A, Halvardson J, Wetterbom A, Gyllensten U, Cavelier L, Feuk L. (2011) Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nat Struct Mol Biol. 2011 Nov 6;18(12):1435-40. doi: 10.1038/nsmb.2143.

Amit M, Donyo M, Hollander D, Goren A, Kim E, Gelfman S, Lev-Maor G, Burstein D, Schwartz S, Postolsky B, Pupko T, Ast G (2012) Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. Cell Rep. 2012 May 31;1(5):543-56. doi: 10.1016/j.celrep.2012.03.013.

Aparicio-Prat E, Arnan C, Sala I, Bosch N, Guigó R, Johnson R (2015) DECKO: Single-oligo, dual-CRISPR deletion of genomic elements including long non-coding RNAs. BMC Genomics. 2015 Oct 23;16:846. doi: 10.1186/s12864-015-2086-z.

Avery OT, MacLeod, CM, McCarty M (1944) Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III. The Journal of Experimental Medicine. 79 (2): 137–58. doi:10.1084/jem.79.2.137.

Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA (2004) Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol. 2004 Jun;14(3): 283-91.

Baralle FE, Giudice J (2017) Alternative splicing as a regulator of development and tissue identity. Nat Rev Mol Cell Biol. 2017 Jul;18(7):437-451. doi: 10.1038/nrm.2017.27. Epub 2017 May 10.

Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ, Frey BJ (2010) Deciphering the splicing code. Nature. 2010 May 6;465(7294):53-9. doi: 10.1038/nature09000.

Barrangou R (2015) The roles of CRISPR-Cas systems in adaptive immunity and beyond. Current Opinion in Immunology. 32: 36–41. doi:10.1016/j.coi.2014.12.008

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science. 315 (5819): 1709–12. doi:10.1126/science.1138140.

Barry G, Briggs JA, Vanichkina DP, Poth EM, Beveridge NJ, Ratnu VS, Nayler SP, Nones K, Hu J, Bredy TW, Nakagawa S, Rigo F, Taft RJ, Cairns MJ, Blackshaw S, Wolvetang EJ, Mattick JS (2014) The long noncoding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. Mol. Psychiatry 19, 486–494. doi: 10.1038/mp.2013.45

Bashaw GJ and Baker BS (1995) The msl-2 dosage compensation gene of Drosophila encodes a putative DNA-binding protein whose expression is sex specifically regulated by Sex-lethal. Development. 1995 Oct;121(10):3245-58.

Bell TJ, Miyashiro KY, Sul JY, Buckley PT, Lee MT, McCullough R, Jochems J, Kim J, Cantor CR, Parsons TD, Eberwine JH (2010) Intron retention facilitates splice variant diversity in calcium-activated big potassium channel populations. Proc Natl Acad Sci U S A. 2010 Dec 7;107(49):21152-7. doi: 10.1073/pnas.1015264107.

Bell TJ, Miyashiro KY, Sul JY, McCullough R, Buckley PT, Jochems J, Meaney DF, Haydon P, Cantor C, Parsons TD, Eberwine JH (2008) Cytoplasmic BK(Ca) channel intron-containing mRNAs contribute to the intrinsic excitability of hippocampal neurons. Proc Natl Acad Sci U S A. 2008 Feb 12;105(6):1901-6. doi: 10.1073/pnas.0711796105.

Beltran M, Puig I, Peña C, García JM, Alvarez AB, Peña R, Bonilla F, de Herreros AG (2008) A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. Genes Dev. 22 (6): 756–69. doi:10.1101/gad.455708

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar

SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara E Catenazzi M, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurles ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008 Nov 6;456(7218):53-9. doi: 10.1038/nature07517.

Bentmann E, Haass C, Dormann D (2013) Stress granules in neurodegeneration - lessons learnt from TAR DNA binding protein of 43kDa and fused in sarcoma. FEBS J. 2013 Sep;280(18):4348-70. doi: 10.1111/febs.12287.

Beraldi R, Li X, Martinez Fernandez A, Reyes S, Secreto F, Terzic A, Olson TM, Nelson TJ (2014) Rbm20-deficient cardiogenesis reveals early disruption of RNA processing and sarcomere remodeling establishing a developmental etiology for dilated cardiomyopathy. Hum Mol Genet. 2014 Jul 15;23(14):3779-91. doi: 10.1093/hmg/ddu091. Epub 2014 Feb 28.

Bergeron D, Pal G, Beaulieu YB, Chabot B, Bachand F (2015) Regulated Intron Retention and Nuclear Pre-mRNA Decay Contribute to PABPN1 Autoregulation. Mol Cell Biol. 2015 Jul;35(14):2503-17. doi: 10.1128/MCB.00070-15.

Berget SM (1995) Exon recognition in vertebrate splicing. J Biol Chem 1995, 270:2411–2414. doi: 10.1074/jbc.270.6.2411

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, Farnham PJ, Hirst M, Lander ES, Mikkelsen TS, Thomson JA (2010) The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010 Oct;28(10):1045-8. doi: 10.1038/nbt1010-1045.

Beyer AL and Osheim YN (1988) Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev. 2, 754–765.

Bijsmans IT, Milona A, Ijssennagger N, Willemsen EC, Ramos Pittol JM, Jonker JW, Lange K, Hooiveld GJ, van Mil SW (2017) Characterization of stem cell-derived liver and intestinal organoids as a model system to study nuclear receptor biology. Biochim Biophys Acta. 2017 Mar;1863(3):687-700. doi: 10.1016/j.bbadis.2016.12.004.

Bill BR, Lowe JK, DyBuncio CT, Fogel BL (2013) Orchestration of neurodevelopmental programs by RBFOX1: implications for autism spectrum disorder. Int Rev Neurobiol. 2013;113:251-67. doi: 10.1016/B978-0-12-418700-9.00008-3.

Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermüller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaöz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Löytynoja A, Whelan S, Pardi F, Massingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Seringhaus M, Church D, Rosenbloom K, Kent WJ, Stone EA; NISC Comparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children's Hospital Oakland Research Institute, Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameur A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CW, Ng P, Shahab A, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S, Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Zhang X, Xu M, Haidar JN, Yu Y, Ruan Y, Iyer VR, Green

RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K, Thakkapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, de Bakker PI, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyras E, Hallgrímsdóttir IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B, de Jong PJ (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007 Jun 14;447(7146):799-816. doi: 10.1038/nature05874

Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. Annual Review of Biochemistry. 72 (1): 291–336. doi:10.1146/annurev.biochem.72.121801.161720.

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology. 151 (Pt 8): 2551–61. doi:10.1099/mic.0.28048-0

Boothby TC, Zipper RS, van der Weele CM, Wolniak SM (2013) Removal of retained introns regulates translation in the rapidly developing gametophyte of Marsilea vestita. Dev Cell. 2013 Mar 11;24(5):517-29. doi: 10.1016/j.devcel.2013.01.015.

Borsani G, Tonlorenzi R, Simmler MC, Dandolo L, Arnaud D, Capra V, Grompe M, Pizzuti A, Muzny D, Lawrence C, Willard HF, Avner P, Ballabio A (1991) Characterization of a murine gene expressed from the inactive X chromosome. Nature 351:325–329. doi:10.1038/351325a0

Bottero V, Withoff S, Verma IM (2006) NF-kappaB and the regulation of hematopoiesis. Cell Death Differ. 2006 May;13(5):785-97. doi: 10.1038/sj.cdd.4401888

Braunschweig U, Barbosa-Morais NL, Pan Q, Nachman EN, Alipanahi B, Gonatopoulos-Pournatzis T, Frey B, Irimia M, Blencowe BJ (2014) Widespread intron retention in mammals functionally tunes transcriptomes. Genome Res. 2014 Nov;24(11):1774-86. doi: 10.1101/gr.177790.114.

Briggs JA, Wolvetang EJ, Mattick JS, Rinn JL, Barry G (2015) Mechanisms of Long Non-coding RNAs in Mammalian Nervous System Development, Plasticity, Disease, and Evolution. Neuron. 2015 Dec 2;88(5):861-877. doi: 10.1016/j.neuron.2015.09.045.

Brivanlou AH, Darnell JE (2002) Signal transduction and the control of gene expression. Science. 2002 Feb 1;295(5556):813-8.

Brockdorff N, Ashworth A, Kay GF, Cooper P, Smith S, McCabe VM, Norris DP, Penny GD, Patel D, Rastan S (1991) Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature 351:329–331. doi:10.1038/351329a0

Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF (1991) A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature 349:38–44. doi:10.1038/349038a0

Buckley PT, Lee MT, Sul JY, Miyashiro KY, Bell TJ, Fisher SA, Kim J, Eberwine JH (2011) Cytoplasmic intron sequence-retaining transcripts can be dendritically targeted via ID element retrotransposons. Neuron. 2011 Mar 10;69(5):877-84. doi: 10.1016/j.neuron.2011.02.028.

Bueno C, Sardina JL, Di Stefano B, Romero-Moya D, Muñoz-López A, Ariza L, Chillón MC, Balanzategui A, Castaño J, Herreros A, Fraga MF, Fernández A, Granada I, Quintana-Bustamante O, Segovia JC, Nishimura K, Ohtaka M, Nakanishi M, Graf T, Menendez P (2016) Reprogramming human B cells into induced pluripotent stem cells and its enhancement by C/EBPα. Leukemia. 2016 Mar;30(3):674-82. doi: 10.1038/leu.2015.294.

Buljan M, Chalancon G, Eustermann S, Wagner GP, Fuxreiter M, Bateman A, Babu MM (2012) Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. Mol Cell. 2012 Jun 29;46(6):871-83. doi: 10.1016/j.molcel.2012.05.039.

Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z, Sharpless NE (2010) Expression of Linear and Novel Circular Forms of an INK4/ARF-Associated Non-Coding RNA Correlates with Atherosclerosis Risk. PLoS Genet. 2010 Dec; 6(12): e1001233. doi: 10.1371/journal.pgen.1001233

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes & Development. 25 (18): 1915–27. doi:10.1101/gad.17446611.

Cariappa A, Liou HC, Horwitz BH, Pillai S (2000) Nuclear factor kappa B is required for the development of marginal zone B lymphocytes. J Exp Med. 2000 Oct 16;192(8):1175-82.

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, Kodzius R, Shimokawa K, Bajic VB, Brenner SE, Batalov S, Forrest AR, Zavolan M, Davis MJ, Wilming LG, Aidinis V, Allen JE, Ambesi-Impiombato A, Apweiler R, Aturaliya RN, Bailey TL, Bansal M, Baxter L, Beisel KW, Bersano T, Bono H, Chalk AM, Chiu KP, Choudhary V, Christoffels A, Clutterbuck DR, Crowe ML, Dalla E, Dalrymple BP, de Bono B, Della Gatta G, di Bernardo D, Down T, Engstrom P, Fagiolini M, Faulkner G, Fletcher CF, Fukushima T, Furuno M, Futaki S, Gariboldi M, Georgii-Hemming P, Gingeras TR, Gojobori T, Green RE, Gustincich S, Harbers M, Hayashi Y, Hensch TK, Hirokawa N, Hill D, Huminiecki L, Iacono M, Ikeo K, Iwama A, Ishikawa T, Jakt M, Kanapin A, Katoh M, Kawasawa Y, Kelso J, Kitamura H, Kitano H, Kollias G, Krishnan SP, Kruger A, Kummerfeld SK, Kurochkin IV, Lareau LF, Lazarevic D, Lipovich L, Liu J, Liuni S, McWilliam S, Madan Babu M, Madera M, Marchionni L, Matsuda H, Matsuzawa S, Miki H, Mignone F, Miyake S, Morris K, Mottagui-Tabar S, Mulder N, Nakano N, Nakauchi H, Ng P, Nilsson R, Nishiguchi S, Nishikawa S, Nori F, Ohara O, Okazaki Y, Orlando V, Pang KC, Pavan WJ, Pavesi G, Pesole G, Petrovsky N, Piazza S, Reed J, Reid JF, Ring BZ, Ringwald M, Rost B, Ruan Y, Salzberg SL, Sandelin A, Schneider C, Schönbach C, Sekiguchi K, Semple CA, Seno S, Sessa L, Sheng Y, Shibata Y, Shimada H, Shimada K, Silva D, Sinclair B, Sperling S, Stupka E, Sugiura K, Sultana R, Takenaka Y, Taki K, Tammoja K, Tan SL, Tang S, Taylor MS, Tegner J, Teichmann SA, Ueda HR, van Nimwegen E, Verardo R, Wei CL, Yagi K, Yamanishi H, Zabarovsky E, Zhu S, Zimmer A, Hide W, Bult C, Grimmond SM, Teasdale RD, Liu ET, Brusic V, Quackenbush J, Wahlestedt C, Mattick JS, Hume DA, Kai C, Sasaki D, Tomaru Y, Fukuda S, Kanamori-Katayama M, Suzuki M, Aoki J, Arakawa T, Iida J, Imamura K, Itoh M, Kato T, Kawaji H, Kawagashira N, Kawashima T, Kojima M, Kondo S, Konno H, Nakano K, Ninomiya N, Nishio T, Okada M, Plessy C, Shibata K, Shiraki T, Suzuki S, Tagami M, Waki K, Watahiki A, Okamura-Oho Y, Suzuki H, Kawai J, Hayashizaki Y; FANTOM Consortium; RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group) (2005) The transcriptional landscape of the mammalian genome. Science. 309 (5740): 1559–63. doi:10.1126/science.1112014. PMID 16141072.

Carpenter S, Aiello D, Atianand MK, Ricci EP, Gandhi P, Hall LL, Byron M, Monks B, Henry-Bezy M, Lawrence JB, O'Neill LA, Moore MJ, Caffrey DR, Fitzgerald KA (2013) A long noncoding RNA mediates both activation and repression of immune response genes. Science. 2013 Aug 16;341(6147):789-92. doi: 10.1126/science.1240925.

Cartault F, Nava C, Malbrunot AC, Munier P, Hebert JC, N'guyen P, Djeridi N, Pariaud P, Pariaud J, Dupuy A, Austerlitz F, Sarasin A (2011) A new XPC gene splicing mutation has lead to the highest worldwide prevalence of xeroderma pigmentosum in black Mahori patients. DNA Repair (Amst). 2011 Jun 10;10(6):577-85. doi: 10.1016/j.dnarep.2011.03.005.

Cartegni L, Chew SL, Krainer AR (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet. 2002 Apr;3(4):285-98. doiI: 10.1038/nrg775

Cascino I, Fiucci G, Papoff G, Rubert, G (1995) Three functional soluble forms of the human apoptosis-inducing Fas molecule are produced by alternative splicing. J. Immunol. 1995;154:2706–2713.

Caspersson T, Schultz J (1939) Pentose nucleotides in the cytoplasm of growing tissues. Nature. 143 (3623): 602–3.

Cazzola M, Della Porta MG, Malcovati L (2013) The genetic basis of myelodysplasia and its clinical relevance. Blood 122, 4021–4034 2013. doi: 10.1182/blood-2013-09-381665

Centonze D, Rossi S, Napoli I, Mercaldo V, Lacoux C, Ferrari F, Ciotti MT, De Chiara V, Prosperetti C, Maccarrone M, Fezza F, Calabresi P, Bernardi G, Bagni C (2007) The brain cytoplasmic RNA BC1 regulates dopamine D2 receptor-mediated transmission in the striatum. The Journal of Neuroscience. 27 (33): 8885–92. doi:10.1523/JNEUROSCI.0548-07.2007

Charizanis K, Lee KY, Batra R, Goodwin M, Zhang C, Yuan Y, Shiue L, Cline M, Scotti MM, Xia G, Kumar A, Ashizawa T, Clark HB, Kimura T, Takahashi MP, Fujimura H, Jinnai K, Yoshikawa H, Gomes-Pereira M, Gourdon G, Sakai N, Nishino S, Foster TC, Ares M Jr, Darnell RB, Swanson MS (2012) Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. Neuron. 2012 Aug 9;75(3):437-50. doi: 10.1016/j.neuron.2012.05.029.

Chau A, Kalsotra A (2015) Developmental insights into the pathology of and therapeutic strategies for DM1: back to the basics. Dev Dyn. 2015 Mar;244(3):377-90. doi: 10.1002/dvdy.24240.

Chen L, Kostadima M, Martens JHA, Canu G, Garcia SP, Turro E, Downes K, Macaulay IC, Bielczyk-Maczynska E, Coe S, Farrow S, Poudel P, Burden F, Jansen SBG, Astle WJ, Attwood A, Bariana T, de Bono B, Breschi A, Chambers JC, Consortium B, Choudry FA, Clarke L, Coupland P, van der Ent M, Erber WN, Jansen JH, Favier R, Fenech ME, Foad N, Freson K, van Geet C, Gomez K, Guigo R, Hampshire D, Kelly AM, Kerstens HHD, Kooner JS, Laffan M, Lentaigne C, Labalette C, Martin T, Meacham S, Mumford A, Nürnberg S, Palumbo E, van der Reijden BA, Richardson D, Sammut SJ, Slodkowicz G, Tamuri AU, Vasquez L, Voss K, Watt S, Westbury S, Flicek P, Loos R, Goldman N, Bertone P, Read RJ, Richardson S, Cvejic A, Soranzo N, Ouwehand WH, Stunnenberg HG, Frontini M, Rendon A (2014) Transcriptional diversity during lineage commitment of human blood progenitors. Science. 2014 Sep 26;345(6204):1251033. doi: 10.1126/science.1251033.

Cho V, Mei Y, Sanny A, Chan S, Enders A, Bertram EM, Tan A, Goodnow CC, Andrews TD (2014) The RNA-binding protein hnRNPLL induces a T cell alternative splicing program delineated by differential intron retention in polyadenylated RNA. Genome Biol. 2014 Jan 29;15(1):R26. doi: 10.1186/gb-2014-15-1-r26.

Clark F and Thanaraj TA (2002) Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. Human Molecular Genetics, Volume 11, Issue 4, 15 February 2002, Pages 451–464, doi: 10.1093/hmg/11.4.451

Coady TH, Manley JL (2015) ALS mutations in TLS / FUS disrupt target gene expression. Genes Dev. 2015 Aug 15;29(16):1696-706. doi: 10.1101/gad.267286.115. Epub 2015 Aug 6.

Cobaleda C, Jochum W, Busslinger M (2007) Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. Nature. 2007 Sep 27;449(7161):473-7. doi: 10.1038/nature06159

Comstock CE, Augello MA, Benito RP, Karch J, Tran TH, Utama FE, Tindall EA, Wang Y, Burd CJ, Groh EM, Hoang HN, Giles GG, Severi G, Hayes VM, Henderson BE, Le Marchand L, Kolonel LN, Haiman CA, Baffa R, Gomella LG, Knudsen ES, Rui H, Henshall SM, Sutherland RL, Knudsen KE (2009) Cyclin D1 splice variants:

polymorphism, risk, and isoform-specific regulation in prostate cancer. Clin Cancer Res. 2009 Sep 1;15(17):5338-49. doi: 10.1158/1078-0432.CCR-08-2865.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. Science. 339 (6121): 819–23. doi:10.1126/science.1231143.

Crick FH (1958) On Protein Synthesis. Symposia of the Society for Experimental Biology, Number XII: The Biological Replication of Macromolecules. Cambridge University Press. pp. 138–163.

Cuenca-Bono B, García-Molinero V, Pascual-García P, Dopazo H, Llopis A, Vilardell J, Rodríguez-Navarro S (2011) SUS1 introns are required for efficient mRNA nuclear export in yeast. Nucleic Acids Res. 2011 Oct;39(19):8599-611. doi: 10.1093/nar/gkr496.

Dai Z, Dai X (2012) Nuclear colocalization of transcription factor target genes strengthens co regulation in yeast. Nucleic Acids Res. 2012 Jan;40(1):27-36. doi: 10.1093/nar/gkr689. Epub 2011 Aug 31.

De Conti L, Baralle M, Buratti E (2013) Exon and intron definition in pre-mRNA splicing. Wiley Interdiscip Rev RNA. 2013 Jan-Feb;4(1):49-60. doi: 10.1002/wrna.1140. Epub 2012 Oct 8.

de Wit E, de Laat W. (2012) A decade of 3C technologies: insights into nuclear organization. Genes Dev. 2012 Jan 1;26(1):11-24. doi: 10.1101/gad.179804.111.

Dedhia PH, Bertaux-Skeirik N, Zavros Y, Spence JR (2016) Organoid Models of Human Gastrointestinal Development and Disease. Gastroenterology. 2016 May;150(5):1098-1112. doi: 10.1053/j.gastro.2015.12.042.

Dekker J, Rippe K, Dekker M, Kleckner N. (2002) Capturing chromosome conformation. Science 2002 Feb 15;295(5558):1306-11.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhattar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigó R (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012 Sep;22(9):1775-89. doi: 10.1101/gr.132159.111.

Devaux Y, Zangrando J, Schroen B, Creemers EE, Pedrazzini T, Chang CP, Dorn GW, Thum T, Heymans S, Cardiolinc network (2015) Long noncoding RNAs in cardiac development and ageing. Nat Rev Cardiol. 2015 Jul;12(7):415-25. doi: 10.1038/nrcardio.2015.55.

Di Chiara G, Bassareo V, Fenu S, De Luca MA, Spina L, Cadoni C, Acquas E, Carboni E, Valentini V, Lecca D (2004). Dopamine and drug addiction: the nucleus accumbens shell connection. Neuropharmacology 47 (Suppl 1), 227–241. doi: 10.1016/j.neuropharm.2004.06.032

Di Lullo E and Kriegstein AR (2017) The use of brain organoids to investigate neural development and disease. Nat Rev Neurosci. 2017 Oct; 18(10): 573–584. doi: 10.1038/nrn.2017.107

Dillman AA, Hauser DN, Gibbs JR, Nalls MA, McCoy MK, Rudenko IN, Galter D, Cookson MR (2013) mRNA expression, splicing and editing in the embryonic and adult mouse cerebral cortex. Nat Neurosci. 2013 Apr;16(4):499-506. doi: 10.1038/nn.3332. Epub 2013 Feb 17.

Dinger ME, Amaral PP, Mercer TR, Mattick JS (2009) Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications. Briefings in Functional Genomics & Proteomics. 8 (6): 407–23. doi:10.1093/bfgp/elp038

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guigó R, Gingeras TR (2012) Landscape of transcription in human cells. Nature. 489 (7414): 101–8. doi:10.1038/nature11233

Dostie J, Bickmore WA (2012) Chromosome organization in the nucleus - charting new territory across the Hi-Cs. Curr Opin Genet Dev. 2012 Apr;22(2):125-31. doi: 10.1016/j.gde.2011.12.006. Epub 2012

Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee BK, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shoresh N, Simon JM, Song L, Trinklein ND, Altshuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong X, Dunham I, Ernst J, Furey TS, Gerstein M, Giardine B, Greven M, Hardison RC, Harris RS, Herrero J, Hoffman MM, Iyer S, Kellis M, Khatun J, Kheradpour P, Kundaje A, Lassmann T, Li Q, Lin X, Marinov GK, Merkel A, Mortazavi A, Parker SC, Reddy TE, Rozowsky J, Schlesinger F, Thurman RE, Wang J, Ward LD, Whitfield TW, Wilder SP, Wu W, Xi HS, Yip KY, Zhuang J, Pazin MJ, Lowdon RF, Dillon LA, Adams LB, Kelly CJ, Zhang J, Wexler JR, Green ED, Good PJ, Feingold EA, Bernstein BE, Birney E, Crawford GE, Dekker J, Elnitski L, Farnham PJ, Gerstein M, Giddings MC, Gingeras TR, Green ED, Guigó R, Hardison RC, Hubbard TJ, Kellis M, Kent W, Lieb JD, Margulies EH, Myers RM, Snyder M, Stamatoyannopoulos JA, Tenenbaum SA, Weng Z, White KP, Wold B, Khatun J, Yu Y, Wrobel J, Risk BA, Gunawardena HP, Kuiper HC, Maier CW, Xie L, Chen X, Giddings MC, Bernstein BE, Epstein CB, Shoresh N, Ernst J, Kheradpour P, Mikkelsen TS, Gillespie S, Goren A, Ram O, Zhang X, Wang L, Issner R, Coyne MJ, Durham T, Ku M, Truong T, Ward LD, Altshuler RC, Eaton ML, Kellis M, Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Batut P, Bell I, Bell K, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena HP, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Li G, Luo OJ, Park E, Preall JB, Presaud K, Ribeca P, Risk BA, Robyr D, Ruan X, Sammeth M, Sandhu KS, Schaeffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Hayashizaki Y, Harrow J, Gerstein M, Hubbard TJ, Reymond A, Antonarakis SE, Hannon GJ, Giddings MC, Ruan Y, Wold B, Carninci P, Guigó R, Gingeras TR, Rosenbloom KR, Sloan CA, Learned K, Malladi VS, Wong MC, Barber GP, Cline MS, Dreszer TR, Heitner SG, Karolchik D, Kent W, Kirkup VM, Meyer LR, Long JC, Maddren M, Raney BJ, Furey TS, Song L, Grasfeder LL, Giresi PG, Lee BK, Battenhouse A, Sheffield NC, Simon JM, Showers KA, Safi A, London D, Bhinge AA, Shestak C, Schaner MR, Kim SK, Zhang ZZ, Mieczkowski PA, Mieczkowska JO, Liu Z, McDaniell RM, Ni Y, Rashid NU, Kim MJ, Adar S, Zhang Z, Wang T, Winter D, Keefe D, Birney E, Iyer VR, Lieb JD, Crawford GE, Li G, Sandhu KS, Zheng M, Wang P, Luo OJ, Shahab A, Fullwood MJ, Ruan X, Ruan Y, Myers RM, Pauli F, Williams BA, Gertz J, Marinov GK, Reddy TE, Vielmetter J, Partridge E, Trout D, Varley KE, Gasper C, Bansal A, Pepke S, Jain P, Amrhein H, Bowling KM, Anaya M, Cross MK, King B, Muratet MA, Antoshechkin I, Newberry KM, McCue K, Nesmith AS, Fisher-Aylor KI, Pusey B, DeSalvo G, Parker SL, Balasubramanian S, Davis NS, Meadows SK, Eggleston T, Gunter C, Newberry J, Levy SE, Absher DM, Mortazavi A, Wong WH, Wold B, Blow MJ, Visel A, Pennachio LA, Elnitski L, Margulies EH, Parker SC, Petrykowska HM, Abyzov A, Aken B, Barrell D, Barson G, Berry A, Bignell A, Boychenko V, Bussotti G, Chrast J, Davidson C, Derrien T, Despacio-Reyes G, Diekhans M, Ezkurdia I, Frankish A, Gilbert J, Gonzalez JM, Griffiths E, Harte R, Hendrix DA, Howald C, Hunt T, Jungreis I, Kay M, Khurana E, Kokocinski F, Leng J, Lin MF, Loveland J, Lu Z, Manthravadi D, Mariotti M, Mudge J, Mukherjee G, Notredame C, Pei B, Rodriguez JM, Saunders G, Sboner A, Searle S, Sisu C, Snow C, Steward C, Tanzer A, Tapanari E, Tress ML, van Baren MJ, Walters N, Washietl S, Wilming L, Zadissa A, Zhang Z, Brent M, Haussler D, Kellis M, Valencia A, Gerstein M, Reymond A, Guigó R, Harrow J, Hubbard TJ, Landt SG, Frietze S, Abyzov A, Addleman N, Alexander RP, Auerbach RK, Balasubramanian S, Bettinger K, Bhardwaj N, Boyle AP, Cao AR, Cayting P, Charos A, Cheng Y, Cheng C, Eastman C, Euskirchen G, Fleming JD, Grubert F, Habegger L, Hariharan M, Harmanci A, Iyengar S, Jin VX, Karczewski KJ, Kasowski M, Lacroute P, Lam H, Lamarre-Vincent N, Leng J, Lian J, Lindahl-Allen M, Min R, Miotto B, Monahan H, Moqtaderi Z, Mu XJ, O'Geen H, Ouyang Z, Patacsil D, Pei B, Raha D, Ramirez L, Reed B, Rozowsky J, Sboner A, Shi M, Sisu C, Slifer T, Witt H, Wu L, Xu X, Yan KK, Yang X, Yip KY, Zhang Z, Struhl K, Weissman SM, Gerstein M, Farnham PJ, Snyder M, Tenenbaum SA, Penalva LO, Doyle F, Karmakar S, Landt SG, Bhanvadia RR, Choudhury A, Domanus M, Ma L, Moran J, Patacsil D, Slifer T, Victorsen A, Yang X, Snyder M, Auer T, Centanin L, Eichenlaub M, Gruhl F, Heermann S, Hoeckendorf B, Inoue D, Kellner T, Kirchmaier S, Mueller C, Reinhardt R, Schertel L, Schneider S, Sinn R, Wittbrodt B, Wittbrodt J, Weng Z, Whitfield TW, Wang J, Collins PJ, Aldred SF, Trinklein ND, Partridge EC, Myers RM, Dekker J, Jain G, Lajoie BR, Sanyal A, Balasundaram G, Bates DL, Byron R, Canfield TK, Diegel MJ, Dunn D, Ebersol AK, Frum T, Garg K, Gist E, Hansen R, Boatman L, Haugen E, Humbert R, Jain G, Johnson AK, Johnson EM, Kutyavin TV, Lajoie BR, Lee K, Lotakis D, Maurano MT, Neph SJ, Neri FV, Nguyen ED, Qu H, Reynolds AP, Roach V, Rynes E, Sabo P, Sanchez ME, Sandstrom RS, Sanyal A, Shafer AO, Stergachis AB, Thomas S, Thurman RE, Vernot B, Vierstra J, Vong S, Wang H, Weaver MA, Yan Y, Zhang M, Akey JM, Bender M, Dorschner MO, Groudine M, MacCoss MJ, Navas P, Stamatoyannopoulos G, Kaul R, Dekker J, Stamatoyannopoulos JA, Dunham I, Beal K, Brazma A, Flicek P, Herrero J, Johnson N, Keefe D, Lukk M, Luscombe NM, Sobral D, Vaquerizas JM, Wilder SP, Batzoglou S, Sidow A, Hussami N, Kyriazopoulou-Panagiotopoulou S, Libbrecht MW, Schaub MA, Kundaje A, Hardison RC, Miller W, Giardine B, Harris RS, Wu W, Bickel PJ, Banfai B, Boley NP, Brown JB, Huang H, Li Q, Li JJ, Noble WS, Bilmes JA, Buske OJ, Hoffman MM, Sahu AD, Kharchenko PV, Park PJ, Baker D, Taylor J, Weng Z, Iyer S, Dong X, Greven M, Lin X, Wang J, Xi HS, Zhuang J, Gerstein M, Alexander RP, Balasubramanian S, Cheng C, Harmanci A, Lochovsky L, Min R, Mu XJ, Rozowsky J, Yan KK, Yip KY, Birney E (2012) An integrated encyclopedia of DNA elements in the human genome. Nature. 489 (7414): 57–74. doi:10.1038/nature11247

Durairaj G, Garg P, Bhaumik SR (2009) Nuclear export of mRNA and its regulation by ubiquitylation, RNA Biology, 6:5, 531-535, DOI: 10.4161/rna.6.5.10078

Dvinge H and Bradley RK (2015) Widespread intron retention diversifies most cancer transcriptomes. Genome Med. 2015 May 15;7(1):45. doi: 10.1186/s13073-015-0168-9.

Dye MJ, Gromak N, Proudfoot NJ. (2006) Exon tethering in transcription by RNA polymerase II. Mol Cell. 2006 Mar 17;21(6):849-59. DOI: 10.1016/j.molcel.2006.01.032

Echeverria GV, Cooper TA (2012) RNA-binding proteins in microsatellite expansion disorders: mediators of RNA toxicity. Brain Res. 1462, 100–111 2012. Epub 2012 Feb 22. doi: 10.1016/j.brainres.2012.02.030

Edwards CR, Ritchie W, Wong JJ, Schmitz U, Middleton R, An X, Mohandas N, Rasko JE, Blobel GA (2016) A dynamic intron retention program in the mammalian megakaryocyte and erythrocyte lineages. Blood. 2016 Apr 28;127(17):e24-e34. doi: 10.1182/blood-2016-01-692764.

Ellis JD, Barrios-Rodiles M, Colak R, Irimia M, Kim T, Calarco JA, Wang X, Pan Q, O'Hanlon D, Kim PM, Wrana JL, Blencowe BJ (2012) Tissue-specific alternative splicing remodels protein-protein interaction networks. Mol Cell. 2012 Jun 29;46(6):884-92. doi: 10.1016/j.molcel.2012.05.037.

ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. Science. 2004 Oct 22;306(5696):636-40. doi: 10.1126/science.1105136

Eswaran J, Horvath A, Godbole S, Reddy SD, Mudvari P, Ohshiro K, Cyanam D, Nair S, Fuqua SA, Polyak K, Florea LD, Kumar R (2013) RNA sequencing of cancer reveals novel splicing alterations. Sci Rep. 2013;3:1689. doi: 10.1038/srep01689.

Ethier SD, Miura H, Dostie J (2012) Discovering genome regulation with 3C and 3C-related technologies. Biochim Biophys Acta. 2012 May;1819(5):401-10. doi: 10.1016/j.bbagrm. 2011.12.004. Epub 2011 Dec 20.

Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD (2006) The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. Genes & Development. 20 (11): 1470–84. doi:10.1101/gad.1416106

Fiedler K and Brunner C (2012) The role of transcription factors in the guidance of granulopoiesis. Am J Blood Res. 2012;2(1):57-65.

Fields C, Adams MD, White O, Venter JC (1994) How many genes in the human genome? Nat Genet. 1994 Jul;7(3):345-6. doi: 10.1038/ng0794-345

Filichkin SA and Mockler TC (2012) Unproductive alternative splicing and nonsense mRNAs: a widespread phenomenon among plant circadian clock genes. Biol Direct. 2012 Jul 2;7:20. doi: 10.1186/1745-6150-7-20.

Flomen R and Makoff A (2011) Increased RNA editing in EAAT2 pre-mRNA from amyotrophic lateral sclerosis patients: involvement of a cryptic polyadenylation site. Neurosci Lett. 2011 Jun 22;497(2):139-43. doi: 10.1016/j.neulet.2011.04.047.

Fox-Walsh KL, Dou Y, Lam BJ, Hung SP, Baldi PF, Hertel KJ (2005) The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc Natl Acad Sci U S A 2005, 102:16176–16181. doi: 10.1073/pnas.0508489102

Friedland AE, Tzur YB, Esvelt KM, Colaiácovo MP, Church GM, Calarco JA (2013) Heritable genome editing in C. elegans via a CRISPR-Cas9 system. Nature Methods. 10 (8): 741–3. doi:10.1038/nmeth.2532.

Fulco CP, Munschauer M, Anyoha R, Munson G, Grossman SR, Perez EM, Kane M, Cleary B, Lander ES, Engreitz JM (2016) Systematic mapping of functional enhancer-promoter connections with CRISPR interference. Science. 2016 Nov 11;354(6313):769-773. doi: 10.1126/science.aag2445

Galante PAF, Sakabe NJ, Kirschbaum-Slager N, De Souza SJ (2004) Detection and evaluation of intron retention events in the human transcriptome. RNA. 2004 May; 10(5): 757–765. doi: 10.1261/rna.5123504

Gascard P, Bilenky M, Sigaroudinia M, Zhao J, Li L, Carles A, Delaney A, Tam A, Kamoh B, Cho S, Griffith M, Chu A, Robertson G, Cheung D, Li I, Heravi-Moussavi A, Moksa M, Mingay M, Hussainkhel A, Davis B, Nagarajan RP, Hong C, Echipare L, O'Geen H, Hangauer MJ, Cheng JB, Neel D, Hu D, McManus MT, Moore R, Mungall A, Ma Y, Plettner P, Ziv E, Wang T, Farnham PJ, Jones SJ, Marra MA, Tlsty TD, Costello JF, Hirst M (2015) Epigenetic and transcriptional determinants of the human breast. Nat Commun. 2015 Feb 18;6:6351. doi: 10.1038/ncomms7351.

Gaudreau MC, Heyd F, Bastien R, Wilhelm B, Möröy T (2012) Alternative splicing controlled by heterogeneous nuclear ribonucleoprotein L regulates development, proliferation, and migration of thymic pre-T cells. J Immunol. 2012 Jun 1;188(11):5377-88. doi: 10.4049/jimmunol.1103142. Epub 2012 Apr 20.

Gerondakis S and Siebenlist U (2010) Roles of the NF-kappaB pathway in lymphocyte development and function. Cold Spring Harb Perspect Biol. 2010 May;2(5):a000182. doi: 10.1101/cshperspect.a000182.

Gerondakis S, Grossmann M, Nakamura Y, Pohl T, Grumont R (1999) Genetic approaches in mice to understand Rel/NF-kappaB and IkappaB function: transgenics and knockouts. Oncogene. 1999 Nov 22;18(49):6888-95. doi: 10.1038/sj.onc.1203236

Giudice J and Cooper TA (2014) RNA-binding proteins in heart development. Adv Exp Med Biol. 2014;825:389-429. doi: 10.1007/978-1-4939-1221-6_11.

Giudice J, Xia Z, Wang ET, Scavuzzo MA, Ward AJ, Kalsotra A, Wang W, Wehrens XH, Burge CB, Li W, Cooper TA (2014) Alternative splicing regulates vesicular trafficking genes in cardiomyocytes during postnatal heart development. Nat Commun. 2014 Apr 22;5:3603. doi: 10.1038/ncomms4603.

Gontijo AM, Miguela V, Whiting MF, Woodruff RC, Dominguez M (2011) Intron retention in the Drosophila melanogaster Rieske Iron Sulphur Protein gene generated a new protein. Nat Commun. 2011;2:323. doi: 10.1038/ncomms1328.

Goodwin M, Mohan A, Batra R, Lee KY, Charizanis K, Fernández Gómez FJ, Eddarkaoui S, Sergeant N, Buée L, Kimura T, Clark HB, Dalton J, Takamura K, Weyn-Vanhentenryck SM, Zhang C, Reid T, Ranum LP, Day JW, Swanson MS (2015) MBNL Sequestration by Toxic RNAs and RNA Misprocessing in the Myotonic Dystrophy Brain. Cell Rep. 2015 Aug 18;12(7):1159-68. doi: 10.1016/j.celrep.2015.07.029.

Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, Wildonger J, O'Connor-Giles KM (2013) Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics. 194 (4): 1029–35. doi:10.1534/genetics.113.152710.

Gudipati RK, Xu Z, Lebreton A, Séraphin B, Steinmetz LM, Jacquier A, Libri D (2012) Extensive degradation of RNA precursors by the exosome in wild-type cells. Mol Cell. 2012 Nov 9;48(3):409-21. doi: 10.1016/j.molcel.2012.08.018.

Guo R, Zheng L, Park JW, Lv R, Chen H, Jiao F, Xu W, Mu S, Wen H, Qiu J, Wang Z, Yang P, Wu F, Hui J, Fu X, Shi X, Shi YG, Xing Y, Lan F, Shi Y (2014) BS69/ZMYND11 reads and connects histone H3.3 lysine 36 trimethylation-decorated chromatin to regulated pre-mRNA processing. Mol Cell. 2014 Oct 23;56(2):298-310. doi: 10.1016/j.molcel.2014.08.022.

Guo W, Schafer S, Greaser ML, Radke MH, Liss M, Govindarajan T, Maatz H, Schulz H, Li S, Parrish AM, Dauksaite V, Vakeel P, Klaassen S, Gerull B, Thierfelder L, Regitz-Zagrosek V, Hacker TA, Saupe KW, Dec GW, Ellinor PT, MacRae CA, Spallek B, Fischer R, Perrot A, Özcelik C, Saar K, Hubner N, Gotthardt M (2012) RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. Nat Med. 2012 May;18(5):766-73. doi: 10.1038/nm.2693.

Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010 Apr 15;464(7291):1071-6. doi: 10.1038/nature08975.

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009 Mar 12;458(7235):223-7. doi: 10.1038/nature07672.

Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C, Rinn JL, Lander ES, Regev A (2010) Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol. 2010 May;28(5):503-10. doi: 10.1038/nbt.1633.

Hang J, Wan R, Yan C, Shi Y (2015) Structural basis of pre-mRNA splicing. Science. 2015 Sep 11;349(6253):1191-8. doi: 10.1126/science.aac8159.

Harrow J, Denoeud F, Frankish A, Reymond A, Chen CK, Chrast J, Lagarde J, Gilbert JG, Storey R, Swarbreck D, Rossier C, Ucla C, Hubbard T, Antonarakis SE, Guigo R (2006) GENCODE: producing a reference annotation for ENCODE. Genome Biol. 2006;7 Suppl 1:S4.1-9. Epub 2006 Aug 7.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M, Rodriguez JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigó R, Hubbard TJ (2012) GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 2012 Sep;22(9):1760-74. doi: 10.1101/gr.135350.111.

Hasemann MS, Lauridsen FK, Waage J, Jakobsen JS, Frank AK, Schuster MB, Rapin N, Bagger FO, Hoppe PS, Schroeder T, Porse BT (2014) C/EBPα is required for long-term self-renewal and lineage priming of hematopoietic stem cells and for the maintenance of epigenetic configurations in multipotent progenitors. PLoS Genet. 2014 Jan;10(1):e1004079. doi: 10.1371/journal.pgen.1004079.

Häsler R, Kerick M, Mah N, Hultschig C, Richter G, Bretz F, Sina C, Lehrach H, Nietfeld W, Schreiber S, Rosenstiel P (2011) Alterations of pre-mRNA splicing in human inflammatory bowel disease. Eur J Cell Biol. 2011 Jun-Jul;90(6-7):603-11. doi: 10.1016/j.ejcb.2010.11.010.

Hegele A, Kamburov A, Grossmann A, Sourlis C, Wowro S, Weimann M, Will CL, Pena V, Luhrmann R, Stelzl U (2012) Dynamic protein-protein interaction wiring of the human spliceosome. Mol Cell 2012, 45:567–580. DOI: 10.1016/j.molcel.2011.12.034

Hershey AD, Chase M (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. The Journal of General Physiology. 36 (1): 39–56. doi:10.1085/jgp.36.1.39.

Hon CC, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJ, Gough J, Denisenko E, Schmeier S, Poulsen TM, Severin J, Lizio M, Kawaji H, Kasukawa T, Itoh M, Burroughs AM, Noma S, Djebali S, Alam T, Medvedeva YA, Testa AC, Lipovich L, Yip CW, Abugessaisa I, Mendez M, Hasegawa A, Tang D, Lassmann T, Heutink P, Babina M, Wells CA, Kojima S, Nakamura Y, Suzuki H, Daub CO, de Hoon MJ, Arner E, Hayashizaki Y, Carninci P, Forrest AR (2017) An atlas of human long non-coding RNAs with accurate 5' ends. Nature. 543 (7644): 199–204. doi:10.1038/nature21374

Horwitz BH, Scott ML, Cherry SR, Bronson RT, Baltimore D (1997) Failure of lymphopoiesis after adoptive transfer of NF-kappaB-deficient fetal liver cells. Immunity. 1997 Jun;6(6):765-72. doi: 10.1016/S1074-7613(00)80451-3

Hossain MA, Rodriguez CM, Johnson TL (2011) Key features of the two-intron Saccharomyces cerevisiae gene SUS1 contribute to its alternative splicing. Nucleic Acids Res. 2011 Oct;39(19):8612-27. doi: 10.1093/nar/gkr497.

Hu W, Yuan B, Flygare J, Lodish HF (2011) Long noncoding RNA-mediated anti-apoptotic activity in murine erythroid terminal differentiation. Genes Dev. 2011; 25(24): 2573–2578. doi: 10.1101/gad.178780.111

Huang YS, Hsieh HY, Shih HM, Sytwu HK, Wu CC (2014) Urinary Xist is a potential biomarker for membranous nephropathy. Biochem Biophys Res Commun. 2014 Sep 26;452(3):415-21. doi: 10.1016/j.bbrc.2014.08.077.

Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, Attardi LD, Regev A, Lander ES, Jacks T, Rinn JL (2010) A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell. 2010 Aug 6;142(3):409-19. doi: 10.1016/j.cell.2010.06.040.

Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbricht C, Wang P, Wang Y, Kong B, Langerød A, Børresen-Dale AL, Kim SK, van de Vijver M, Sukumar S, Whitfield ML, Kellis M, Xiong Y, Wong DJ, Chang HY (2011) Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet. 2011 Jun 5;43(7):621-9. doi: 10.1038/ng.848.

Hussein SM, Puri MC, Tonge PD, Benevento M, Corso AJ, Clancy JL, Mosbergen R, Li M, Lee DS, Cloonan N, Wood DL, Munoz J, Middleton R, Korn O, Patel HR, White CA, Shin JY, Gauthier ME, Lê Cao KA, Kim JI, Mar JC, Shakiba N, Ritchie W, Rasko JE, Grimmond SM, Zandstra PW, Wells CA, Preiss T, Seo JS, Heck AJ, Rogers IM, Nagy A

(2014) Genome-wide characterization of the routes to pluripotency. Nature. 2014 Dec 11;516(7530):198-206. doi: 10.1038/nature14046.

International Human Genome Sequencing Consortium (IHGSC) (2004) Finishing the euchromatic sequence of the human genome. Nature 431 (7011): 931–945. doi:10.1038/nature03001

Irimia M, Weatheritt RJ, Ellis JD, Parikshak NN, Gonatopoulos-Pournatzis T, Babor M, Quesnel-Vallières M, Tapial J, Raj B, O'Hanlon D, Barrios-Rodiles M, Sternberg MJ, Cordes SP, Roth FP, Wrana JL, Geschwind DH, Blencowe BJ (2014) A highly conserved program of neuronal microexons is misregulated in autistic brains. Cell. 2014 Dec 18;159(7):1511-23. doi: 10.1016/j.cell.2014.11.035.

Jaillon O, Bouhouche K, Gout JF, Aury JM, Noel B, Saudemont B, Nowacki M, Serrano V, Porcel BM, Ségurens B, Le Mouël A, Lepère G, Schächter V, Bétermier M, Cohen J, Wincker P, Sperling L, Duret L, Meyer E (2008) Translational control of intron splicing in eukaryotes. Nature. 2008 Jan 17;451(7176):359-62. doi: 10.1038/nature06495.

Janssens J, van Broeckhoven C (2013) Pathological mechanisms underlying TDP-43 driven neurodegeneration in FTLD-ALS spectrum disorders. Hum Mol Genet. 2013 Oct 15;22(R1):R77-87. doi: 10.1093/hmg/ddt349. Epub 2013 Jul 29.

Jelen, N, Ule J, Zivin M, Darnell, RB (2007) Evolution of Nova-dependent splicing regulation in the brain. PLoS Genet. 3, 1838–1847 2007. doi: 10.1371/journal.pgen.0030173

Jenuwein T and Allis D (2001) Translating the Histone Code. Science 10 Aug 2001: Vol. 293, Issue 5532, pp. 1074-1080 DOI: 10.1126/science.1063127

Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Research. 41 (20): e188. doi:10.1093/nar/gkt780.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 337 (6096): 816–21. doi:10.1126/science.1225829

Johnson DS, Mortazavi A, Myers RM, Wold B. (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science. 2007 Jun 8;316(5830):1497-502.

Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science. 2003;302:2141–2144.

Joung J, Engreitz JM, Konermann S, Abudayyeh OO, Verdine VK, Aguet F, Gootenberg JS, Sanjana NE, Wright JB, Fulco CP, Tseng YY, Yoon CH, Boehm JS, Lander ES, Zhang F (2017) Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. Nature. 2017 Aug 17;548(7667):343-346. doi: 10.1038/nature23451.

Jung H, Lee D, Lee J, Park D, Kim YJ, Park WY, Hong D, Park PJ, Lee E (2015) Intron retention is a widespread mechanism of tumor-suppressor inactivation. Nat Genet. 2015 Nov;47(11):1242-8. doi: 10.1038/ng.3414.

Jurica MS, Moore MJ (2003) Pre-mRNA splicing: awash in a sea of proteins. Mol Cell 2003, 12:5–14. doi: 10.1016/S1097-2765(03)00270-3

Kafasla P, Mickleburgh I, Llorian M, Coelho M, Gooding C, Cherny D, Joshi A, Kotik-Kogan O, Curry S, Eperon IC, Jackson RJ, Smith CW. (2012) Defining the roles and interactions of PTB. Biochem Soc Trans. 2012 Aug;40(4):815-20. doi: 10.1042/BST20120044.

Kalsotra A, Xiao X, Ward AJ, Castle JC, Johnson JM, Burge CB, Cooper TA (2008) A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. Proc Natl Acad Sci USA. 2008 Dec 23;105(51):20333-8. doi: 10.1073/pnas.0809045105. Epub 2008 Dec 15.

Kalyna M, Simpson CG, Syed NH, Lewandowska D, Marquez Y, Kusenda B, Marshall J, Fuller J, Cardle L, McNicol J, Dinh HQ, Barta A, Brown JW (2012) Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. Nucleic Acids Res. 2012 Mar;40(6):2454-69. doi: 10.1093/nar/gkr932.

Kan Z, States D, Gish W (2002) Selecting for functional alternative splices in ESTs. Genome Res. 2002 Dec; 12(12): 1837–1845. doi: 10.1101/gr.764102

Kawai J, Shinagawa A, Shibata K, Yoshino M, Itoh M, Ishii Y, Arakawa T, Hara A, Fukunishi Y, Konno H, Adachi J, Fukuda S, Aizawa K, Izawa M, Nishi K, Kiyosawa H, Kondo S, Yamanaka I, Saito T, Okazaki Y, Gojobori T, Bono H, Kasukawa T, Saito R, Kadota K, Matsuda H, Ashburner M, Batalov S, Casavant T, Fleischmann W, Gaasterland T, Gissi C, King B, Kochiwa H, Kuehl P, Lewis S, Matsuo Y, Nikaido I, Pesole G, Quackenbush J, Schriml LM, Staubli F, Suzuki R, Tomita M, Wagner L, Washio T, Sakai K, Okido T, Furuno M, Aono H, Baldarelli R, Barsh G, Blake J, Boffelli D, Bojunga N, Carninci P, de Bonaldo MF, Brownstein MJ, Bult C, Fletcher C, Fujita M, Gariboldi M, Gustincich S, Hill D, Hofmann M, Hume DA, Kamiya M, Lee NH, Lyons P, Marchionni L, Mashima J, Mazzarelli J, Mombaerts P, Nordone P, Ring B, Ringwald M, Rodriguez I, Sakamoto N, Sasaki H, Sato K, Schönbach C, Seya T, Shibata Y, Storch KF, Suzuki H, Toyo-oka K, Wang KH, Weitz C, Whittaker C, Wilming L, Wynshaw-Boris A, Yoshida K, Hasegawa Y, Kawaji H, Kohtsuki S, Hayashizaki Y; RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium (2001) Functional annotation of a full-length mouse cDNA collection. Nature. 409 (6821): 685–690. doi:10.1038/35055500

Keniry A, Oxley D, Monnier P, Kyba M, Dandolo L, Smits G, Reik W (2012) The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. Nat Cell Biol. 2012 Jun 10;14(7):659-65. doi: 10.1038/ncb2521.

Ketley NJ and Newland AC (1997) Haemopoietic growth factors. Postgrad Med J. 1997 Apr; 73(858): 215–221. doi:10.1136/pgmj.73.858.215

Khaladkar M, Buckley PT, Lee MT, Francis C, Eghbal MM, Chuong T, Suresh S, Kuhn B, Eberwine J, Kim J (2013) Subcellular RNA sequencing reveals broad presence of cytoplasmic intron-sequence retaining transcripts in mouse and rat neurons. PLoS One. 2013 Oct 3;8(10):e76194. doi: 10.1371/journal.pone.0076194.

Khodor YL, Menet JS, Tolan M, Rosbash M (2012) Cotranscriptional splicing efficiency differs dramatically between Drosophila and mouse. RNA. 2012 Dec;18(12):2174-86. doi: 10.1261/rna.034090.112. Epub 2012 Oct 24.

Khodor YL, Rodriguez J, Abruzzi KC, Tang CH, Marr MT, Rosbash M (2011) Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes Dev. 2011 Dec 1;25(23):2502-12. doi: 10.1101/gad.178962.111.

Knoll M, Lodish HF, Sun L (2015) Long non-coding RNAs as regulators of the endocrine system. Nat Rev Endocrinol. 2015 Mar;11(3):151-60. doi: 10.1038/nrendo.2014.229.

Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T, Tanaka F, Shibata K, Suzuki A, Komune S, Miyano S, Mori M (2011) Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. Cancer Res. 2011 Oct 15;71(20):6320-6. doi: 10.1158/0008-5472.CAN-11-1021.

Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat Rev Mol Cell Biol. 2013 Mar;14(3):153-65. doi: 10.1038/nrm3525. Epub 2013 Feb 6.

Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, Xiong Y (2011) Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15INK4B tumor suppressor gene Oncogene. Oncogene. 2011 Apr 21; 30(16): 1956–1962. doi: 10.1038/onc.2010.568

Kotovic KM, Lockshon D, Boric L, Neugebauer KM (2003) Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. Mol Cell Biol. 2003 Aug;23(16):5768-79.

Lacadie SA and Rosbash M (2005) Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. Mol. Cell 19, 65–75.

Lacroix M, Lacaze-Buzy L, Furio L, Tron E, Valari M, Van der Wier G, Bodemer C, Bygum A, Bursztejn AC, Gaitanis G, Paradisi M, Stratigos A, Weibel L, Deraison C, Hovnanian A (2012) Clinical expression and new SPINK5 splicing defects in Netherton syndrome: unmasking a frequent founder synonymous mutation and unconventional intronic mutations. J Invest Dermatol. 2012 Mar;132(3 Pt 1):575-82. doi: 10.1038/jid.2011.366.

Lai WKM, Pugh BF (2017) Understanding nucleosome dynamics and their links to gene expression and DNA replication. Nat Rev Mol Cell Biol. 2017 Sep;18(9):548-562. doi: 10.1038/nrm.2017.47. Epub 2017 May 24.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowki J; International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature. 2001 Feb 15;409(6822):860-921. doi: 10.1038/35057062

Lang KM, Spritz RA. (1983) RNA splice site selection: evidence for a 5 leads to 3 scanning model. Science 1983, 220:1351–1355. DOI: 10.1126/science.6304877

Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature. 2007 Apr 19;446(7138):926-9. DOI:10.1038/nature05676

Lee JA, Tang ZZ, Black DL (2009) An inducible change in Fox‑1/A2BP1 splicing modulates the alternative splicing of downstream neuronal target exons. Genes Dev. 23, 2284–2293 2009. doi: 10.1101/gad.1837009. Epub 2009 Sep 17.

Li S, Guo W, Dewey CN, Greaser ML (2013) Rbm20 regulates titin alternative splicing as a splicing repressor. Nucleic Acids Res. 2013 Feb 1;41(4):2659-72. doi: 10.1093/nar/gks1362. Epub 2013 Jan 9.

Li X, Kim Y, Tsang EK, Davis JR, Damani FN, Chiang C, Hess GT, Zappala Z, Strober BJ, Scott AJ, Li A, Ganna A, Bassik MC, Merker JD; GTEx Consortium; Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VARI; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration & Visualization—EBI; Genome Browser Data Integration & Visualization—UCSC Genomics Institute, University of California Santa Cruz, Hall IM, Battle A, Montgomery SB (2017) The impact of rare variation on gene expression across tissues. Nature. 2017 Oct 11;550(7675):239-243. doi: 10.1038/nature24267.

Liang F, Holt I, Pertea G, Karamycheva S, Salzberg SL, Quackenbush J (2000) Gene index analysis of the human genome estimates approximately 120,000 genes. Nat Genet. 2000 Jun;25(2):239-40. doi: 10.1038/76126

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J (2009) Comprehensive mapping of long range interactions reveals folding principles of the human genome. Science. 2009 Oct 9; 326(5950): 289–293. doi: 10.1126/science.1181369

Lim LP and Burge CB (2001) A computational analysis of sequence features involved in recognition of short introns. PNAS September 25, 2001 98 (20) 11193-11198; doi: 10.1073/pnas.201407298

Ling JP, Pletnikova O, Troncoso JC, Wong PC (2015) TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. Science. 2015 Aug 7;349(6248):650-5. doi: 10.1126/science.aab0983.

Listerman I, Sapra AK, Neugebauer KM (2006) Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. Nat Struct Mol Biol. 2006 Sep;13(9):815-22. Epub 2006 Aug 20.

Liu SJ, Horlbeck MA, Cho SW, Birk HS, Malatesta M, He D, Attenello FJ, Villalta JE, Cho MY, Chen Y, Mandegar MA, Olvera MP, Gilbert LA, Conklin BR, Chang HY, Weissman JS, Lim DA (2017) CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. Science. 2017 Jan 6;355(6320). pii: aah7111. doi: 10.1126/science.aah7111.

Lorenzen JM, Thum T (2016) Long noncoding RNAs in kidney and cardiovascular diseases. Nat Rev Nephrol. 2016 Jun;12(6):360-73. doi: 10.1038/nrneph.2016.51.

Lu F, Gladden AB, Diehl JA (2003) An alternatively spliced Cyclin D1 isoform, Cyclin D1b, is a nuclear oncogene. Cancer Res. 2003 Nov 1;63(21):7056-61.

Lunghi M, Galizi R, Magini A, Carruthers VB, Di Cristina M (2015) Expression of the glycolytic enzymes enolase and lactate dehydrogenase during the early phase of Toxoplasma differentiation is regulated by an intron retention mechanism. Mol Microbiol. 2015 Jun;96(6):1159-75. doi: 10.1111/mmi.12999.

Luo M, Jeong M, Sun D, Park HJ, Rodriguez BA, Xia Z, Yang L, Zhang X, Sheng K, Darlington GJ, Li W, Goodell MA (2015) Long non-coding RNAs control hematopoietic stem cell function. Cell Stem Cell. 2015 Apr 2;16(4):426-38. doi: 10.1016/j.stem.2015.02.002.

Ma L, Bajic VB, Zhang Z (2013) On the classification of long non-coding RNAs. RNA Biol. 2013 Jun;10(6):925-33. doi: 10.4161/rna.24604.

Ma L, Li A, Zou D, Xu X, Xia L, Yu J, Bajic VB, Zhang Z (2015) LncRNAWiki: harnessing community knowledge in collaborative curation of human long non-coding RNAs. Nucleic Acids Research. 43: D187–92. doi:10.1093/nar/gku1167

Maatz H, Jens M, Liss M, Schafer S, Heinig M, Kirchner M, Adami E, Rintisch C, Dauksaite V, Radke MH, Selbach M, Barton PJ, Cook SA, Rajewsky N, Gotthardt M, Landthaler M, Hubner N (2014) RNA-binding protein RBM20 represses splicing to orchestrate cardiac pre-mRNA processing. J Clin Invest. 2014 Aug;124(8):3419-30. doi: 10.1172/JCI74523. Epub 2014 Jun 24.

Machlus KR, Thon JN, Italiano JE (2014) Interpreting the developmental dance of the megakaryocyte: a review of the cellular and molecular processes mediating platelet formation. British Journal of Haematology. 165 (2): 227–36. doi:10.1111/bjh.12758

Madhani HD, Guthrie C (1994) Dynamic RNA-RNA interactions in the spliceosome. Annu Rev Genet 1994, 28:1–26. DOI: 10.1146/annurev.ge.28.120194.000245

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. Science. 339 (6121): 823–6. doi:10.1126/science.1232033

Mallory MJ, Allon SJ, Qiu J, Gazzara MR, Tapescu I, Martinez NM, Fu XD, Lynch KW (2015) Induced transcription and stability of CELF2 mRNA drives widespread alternative splicing during T-cell signaling. Proc Natl Acad Sci U S A. 2015 Apr 28;112(17):E2139-48. doi: 10.1073/pnas.1423695112. Epub 2015 Apr 13.

Maquat LE (2004) Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol. 2004 Feb;5(2):89-99.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005 Sep 15;437(7057):376-80.

Marraffini LA and Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science. 322 5909: 1843–5. doi: 10.1126/science.1165771

Marraffini LA and Sontheimer EJ (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nature Reviews Genetics. 11 (3): 181–90. doi:10.1038/nrg2749.

Martin JA and Wang Z (2011) Next-generation transcriptome assembly. Nat Rev Genet. 2011 Sep 7;12(10):671-82. doi: 10.1038/nrg3068.

Martinez NM, Agosto L, Qiu J, Mallory MJ, Gazzara MR, Barash Y, Fu XD, Lynch KW (2015) Widespread JNK-dependent alternative splicing induces a positive feedback loop through CELF2-mediated regulation of MKK7 during T-cell activation. Genes Dev. 2015 Oct 1;29(19):2054-66. doi: 10.1101/gad.267245.115.

Maselli RA, Arredondo J, Nguyen J, Lara M, Ng F, Ngo M, Pham JM, Yi Q, Stajich JM, McDonald K, Hauser MA, Wollmann RL (2014) Exome sequencing detection of two untranslated GFPT1 mutations in a family with limb-girdle myasthenia. Clin Genet. 2014 Feb;85(2):166-71. doi: 10.1111/cge.12118.

Masood N, Malik FA, Kayani MA (2012) Unusual intronic variant in GSTP1 in head and neck cancer in Pakistan. Asian Pac J Cancer Prev. 2012;13(4):1683-6.

Matlin AJ, Clark F, Smith CWJ (2005) Understanding alternative splicing: towards a cellular code. Nature Reviews. 6 (5): 386–398. doi:10.1038/nrm1645.

Mendel G (1866) Versuche über Pflanzen-Hybriden. Verh. Naturforsch. Ver. Brünn 4: 3–47 (in English in 1901, J. R. Hortic. Soc. 26: 1–32).

Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: insights into functions. Nature Reviews. Genetics. 10 (3): 155–9. doi:10.1038/nrg2521.

Michelhaugh SK, Lipovich L, Blythe J, Jia H, Kapatos G, Bannon MJ (2011) Mining Affymetrix microarray data for long non-coding RNAs: altered expression in the nucleus accumbens of heroin abusers. J. Neurochem. 116, 459–466 doi: 10.1111/j.1471-4159.2010.07126.x.

Middleton R, Gao D, Thomas A, Singh B, Au A, Wong JJ, Bomane A, Cosson B, Eyras E, Rasko JE, Ritchie W (2017) IRFinder: assessing the impact of intron retention on mammalian gene expression. Genome Biol. 2017 Mar 15;18(1):51. doi: 10.1186/s13059-017-1184-4.

Modrek B and Lee C (2002) A genomic view of alternative splicing. Nat Genet. 2002 Jan;30(1):13-9. DOI: 10.1038/ng0102-13

Mohanraju P, Makarova KS, Zetsche B, Zhang F, Koonin EV, van der Oost J (2016) Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science. 353 (6299): aad5147. doi:10.1126/science.aad5147

Mojica FJ, Díez-Villaseñor C, García-Martínez J, Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. Journal of Molecular Evolution. 60 (2): 174–82. doi:10.1007/s00239-004-0046-3.

Mojica FJ, Díez-Villaseñor C, Soria E, Juez G (2000) Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Molecular Microbiology. 36 (1): 244–6. doi:10.1046/j.1365-2958.2000.01838.x

Moran Y, Weinberger H, Reitzel AM, Sullivan JC, Kahn R, Gordon D, Finnerty JR, Gurevitz M (2008) Intron retention as a posttranscriptional regulatory mechanism of neurotoxin expression at early life stages of the starlet anemone Nematostella vectensis. J Mol Biol. 2008 Jul 11;380(3):437-43. doi: 10.1016/j.jmb.2008.05.011.

Morris KV, Mattick JS (2014) The rise of regulatory RNA. Nat Rev Genet. 2014 Jun;15(6):423-37. doi: 10.1038/nrg3722.

Morrison J and Kimble J (2006) Asymmetric and symmetric stem-cell divisions in development and cancer. Nature. 441 (7097): 1068–74. doi:10.1038/nature04956

Morrison SJ, Uchida N, Weissman IL (1995) The biology of hematopoietic stem cells. Annu Rev Cell Dev Biol. 1995;11:35-71. doi: 10.1146/annurev.cb.11.110195.000343

Mullis Kary B. et al. (1986) Process for amplifying, detecting, and/or-cloning nucleic acid sequences U.S. Patent 4,683,195

Munroe SH, Lazar MA (1991) Inhibition of c-erbA mRNA splicing by a naturally occurring antisense RNA. The Journal of Biological Chemistry. 266 (33): 22083–6.

Nakajima H (2011) Role of transcription factors in differentiation and reprogramming of hematopoietic cells. Keio J Med. 2011;60(2):47-55.

Narlikar GJ, Fan HY, Kingston RE (2002) Cooperation between complexes that regulate chromatin structure and transcription. Cell. 2002 Feb 22;108(4):475-87.

Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grützner F, Kaessmann H (2014) The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature. 2014 Jan 30;505(7485):635-40. doi: 10.1038/nature12943.

Ner-Gaon H, Halachmi R, Savaldi-Goldstein S, Rubin E, Ophir R, Fluhr R (2004) Intron retention is a major phenomenon in alternative splicing in Arabidopsis. The Plant Journal (2004) 39, 877–885 doi: 10.1111/j.1365-313X.2004.02172.x

Ni T, Yang W, Han M, Zhang Y, Shen T, Nie H, Zhou Z, Dai Y, Yang Y, Liu P, Cui K, Zeng Z, Tian Y, Zhou B, Wei G, Zhao K, Peng W, Zhu J (2016) Global intron retention mediated gene regulation during CD4+ T cell activation. Nucleic Acids Res. 2016 Aug 19;44(14):6817-29. doi: 10.1093/nar/gkw591.

Nilsen TW (2003) The spliceosome: the most complex macromolecular machine in the cell? Bioessays 2003, 25:1147–1149. DOI: 10.1002/bies.10394

Ohlsson E, Schuster MB, Hasemann M, Porse BT (2016) The multifaceted functions of C/EBPα in normal and malignant haematopoiesis. Leukemia. 2016 Apr;30(4):767-75. doi: 10.1038/leu.2015.324.

Pachnis V, Belayew A, Tilghman SM (1984) Locus unlinked to alpha-fetoprotein under the control of the murine raf and Rif genes. Proc Natl Acad Sci U S A. 1984 Sep; 81(17): 5523–5527.

Palazzo AF, Mahadevan K, Tarnawsky SP (2013) ALREX-elements and introns: two identity elements that promote mRNA nuclear export. Wiley Interdiscip Rev RNA. 2013 Sep-Oct;4(5):523-33. doi: 10.1002/wrna.1176.

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nature Genetics volume 40, pages 1413–1415 (2008) doi:10.1038/ng.259

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell. 2008;32:232–46. doi: 10.1016/j.molcel.2008.08.022.

Pandya-Jones A and Black DL (2009) Co‐transcriptional splicing of constitutive and alternative exons. RNA. 2009 Oct;15(10):1896-908. doi: 10.1261/rna.1714509. Epub 2009 Aug 5.

Perner J, Lasserre J, Kinkley S, Vingron M, Chung HR (2014) Inference of interactions between chromatin modifiers and histone modifications: from ChIP-Seq data to chromatin-signaling. Nucleic Acids Res. 2014 Dec 16;42(22):13689-95. doi: 10.1093/nar/gku1234.

Pimentel H, Parra M, Gee SL, Mohandas N, Pachter L, Conboy JG (2016) A dynamic intron retention program enriched in RNA processing genes regulates gene expression during terminal erythropoiesis. Nucleic Acids Res. 2016 Jan 29;44(2):838-51. doi: 10.1093/nar/gkv1168.

Pourcel C, Salvignol G, Vergnaud G (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology. 151 (Pt 3): 653–63. doi:10.1099/mic.0.27437-0.

Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, Laxman B, Asangani IA, Grasso CS, Kominsky HD, Cao X, Jing X, Wang X, Siddiqui J, Wei JT, Robinson D, Iyer HK, Palanisamy N, Maher CA, Chinnaiyan AM (2011) Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. Nat Biotechnol. 2011 Jul 31;29(8):742-9. doi: 10.1038/nbt.1914.

Pulido-Quetglas C, Aparicio-Prat E, Arnan C, Polidori T, Hermoso T, Palumbo E, Ponomarenko J, Guigo R, Johnson R (2017) Scalable Design of Paired CRISPR Guide RNAs for Genomic Deletion. PLoS Comput Biol. 2017 Mar 2;13(3):e1005341. doi: 10.1371/journal.pcbi.1005341.

Pundhir S, Bratt Lauridsen FK, Schuster MB, Jakobsen JS, Ge Y, Schoof EM, Rapin N, Waage J, Hasemann MS, Porse BT (2018) Enhancer and Transcription Factor Dynamics during Myeloid Differentiation Reveal an Early Differentiation Block in Cebpa null Progenitors. Cell Rep. 2018 May 29;23(9):2744-2757. doi: 10.1016/j.celrep.2018.05.012.

Puthanveetil P, Chen S, Feng B, Gautam A, Chakrabarti S (2015) Long non-coding RNA MALAT1 regulates hyperglycaemia induced inflammatory process in the endothelial cells. J Cell Mol Med. 2015 Jun;19(6):1418-25. doi: 10.1111/jcmm.12576.

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 152 (5): 1173–83. doi:10.1016/j.cell.2013.02.022

Quek XC, Thomson DW, Maag JL, Bartonicek N, Signal B, Clark MB, Gloss BS, Dinger ME (2014) lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. Nucleic Acid Res 43: D168-D173

Ranzani V, Rossetti G, Panzeri I, Arrigoni A, Bonnal RJ, Curti S, Gruarin P, Provasi E, Sugliano E, Marconi M, De Francesco R, Geginat J, Bodega B, Abrignani S, Pagani M (2015) The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. Nat Immunol. 2015 Mar;16(3):318-325. doi: 10.1038/ni.3093.

Ravasi T, Suzuki H, Pang KC, Katayama S, Furuno M, Okunishi R, Fukuda S, Ru K, Frith MC, Gongora MM, Grimmond SM, Hume DA, Hayashizaki Y, Mattick JS (2006) Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. Genome Research. 16 (1): 11–9. doi:10.1101/gr.4200206.

Razin SV, Gavrilov AA, Pichugin A, Lipinski M, Iarovaia OV, Vassetzky YS (2011) Transcription factories in the context of the nuclear and genome organization. Nucleic Acids Res. 2011 Nov; 39(21): 9085–9092. doi: 10.1093/nar/gkr683

Recchia A, Perani L, Sartori D, Olgiati C, Mavilio F (2004) Site-specific integration of functional transgenes into the human genome by adeno/AAV hybrid vectors. Mol Ther. 2004 Oct;10(4):660-70. doi: 10.1016/j.ymthe.2004.07.003

Reddy AS, Rogers MF, Richardson DN, Hamilton M, Ben-Hur A (2012) Deciphering the plant splicing code: experimental and computational approaches for predicting alternative splicing and splicing regulatory elements. Front Plant Sci. 2012 Feb 7;3:18. doi: 10.3389/fpls.2012.00018.

Remy E, Cabrito TR, Batista RA, Hussein MA, Teixeira MC, Athanasiadis A, Sá-Correia I, Duque P (2014) Intron retention in the 5'UTR of the novel ZIF2 transporter enhances translation to promote zinc tolerance in arabidopsis. PLoS Genet. 2014 May 15;10(5):e1004375. doi: 10.1371/journal.pgen.1004375.

Ren S, Peng Z, Mao JH, Yu Y, Yin C, Gao X, Cui Z, Zhang J, Yi K, Xu W, Chen C, Wang F, Guo X, Lu J, Yang J, Wei M, Tian Z, Guan Y, Tang L, Xu C, Wang L, Gao X, Tian W, Wang J, Yang H, Wang J, Sun Y (2012) RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicings. Cell Res. 2012 May;22(5):806-21. doi: 10.1038/cr.2012.30.

Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. Annu Rev Biochem. 2012;81:145-66. doi: 10.1146/annurev-biochem-051410-092902.

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 129 (7): 1311–23. doi:10.1016/j.cell.2007.05.022

Robberson BL, Cote GJ, Berget SM (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol Cell Biol 1990, 10:84–94. doi: 10.1128/MCB.10.1.84

Rocchi V, Janni M, Bellincampi D, Giardina T, D'Ovidio R (2012) Intron retention regulates the expression of pectin methyl esterase inhibitor (Pmei) genes during wheat growth and development. Plant Biol (Stuttg). 2012 Mar;14(2):365-73. doi: 10.1111/j.1438-8677.2011.00508.x.

Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, Cortese MS, Sickmeier M, LeGall T, Obradovic Z, Dunker AK. (2006) Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. Proc Natl Acad Sci U S A. 2006 May 30;103(22):8390-5. Epub 2006 May 22.

Roy SW, Irimia M (2008) Intron mis-splicing: no alternative? Genome Biol. 2008;9(2):208. doi: 10.1186/gb-2008-9-2-208.

Sakabe N and de Souza S (2007) Sequence features responsible for intron retention in human. BMC Genomics. 2007 Feb 26;8:59. doi.org/10.1186/1471-2164-8-59

Sakharkar MK, Perumal BS, Sakharkar KR, Kangueane P (2005) An analysis on gene architecture in human and mouse genomes. In Silico Biol 2005, 5:347–365.

Salomonis N, Nelson B, Vranizan K, Pico AR, Hanspers K, Kuchinsky A, Ta L, Mercola M, Conklin BR (2009) Alternative splicing in the differentiation of human embryonic stem cells into cardiac precursors. PLoS Comput Biol. 2009 Nov;5(11):e1000553. doi: 10.1371/journal.pcbi.1000553. Epub 2009 Nov 6.

Sammeth M, Foissac S, Guigó R (2008) A general definition and nomenclature for alternative splicing events. PLoS Comput. Biol. 4 (8): e1000147. doi: 10.1371/journal.pcbi.1000147.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America. 74 (12): 5463–7. doi:10.1073/pnas.74.12.5463

Saredi S, Ardissone A, Ruggieri A, Mottarelli E, Farina L, Rinaldi R, Silvestri E, Gandioli C, D'Arrigo S, Salerno F, Morandi L, Grammatico P, Pantaleoni C, Moroni I, Mora M (2012) Novel POMGNT1 point mutations and intragenic rearrangements associated with muscle-eye-brain disease. J Neurol Sci. 2012 Jul 15;318(1-2):45-50. doi: 10.1016/j.jns.2012.04.008.

Schuster SC (2008) Next-generation sequencing transforms today's biology. Nat Methods. 2008 Jan;5(1):16-8. doi: 10.1038/nmeth1156.

Scotter EL, Chen HJ, Shaw CE (2015) TDP-43 proteinopathy and ALS: insights into disease mechanisms and therapeutic targets. Neurotherapeutics. 2015 Apr;12(2):352-63. doi: 10.1007/s13311-015-0338-x.

Sebé-Pedrós A, Irimia M, Del Campo J, Parra-Acero H, Russ C, Nusbaum C, Blencowe BJ, Ruiz-Trillo I (2013) Regulated aggregative multicellularity in a close unicellular relative of metazoa. Elife. 2013 Dec 24;2:e01287. doi: 10.7554/eLife.01287.

Seo PJ, Kim MJ, Ryu JY, Jeong EY, Park CM (2011) Two splice variants of the IDD14 transcription factor competitively form nonfunctional heterodimers which may regulate starch metabolism. Nat Commun. 2011;2:303. doi: 10.1038/ncomms1303.

Seo PJ, Park MJ, Lim MH, Kim SG, Lee M, Baldwin IT, Park CM (2012) A self-regulatory circuit of CIRCADIAN CLOCK-ASSOCIATED1 underlies the circadian clock regulation of temperature responses in Arabidopsis. Plant Cell. 2012 Jun;24(6):2427-42. doi: 10.1105/tpc.112.098723.

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelson T, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F (2014) Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells Science. 2014 Jan 3; 343(6166): 84–87. doi: 10.1126/science.1247005

Shankarling G, Cole BS, Mallory MJ, Lynch KW (2014) Transcriptome-wide RNA interaction profiling reveals physical and functional targets of hnRNP L in human T cells. Mol Cell Biol. 2014 Jan;34(1):71-83. doi: 10.1128/MCB.00740-13. Epub 2013 Oct 28.

Shapiro IM, Cheng AW, Flytzanis NC, Balsamo M, Condeelis JS, Oktay MH, Burge CB, Gertler FB (2011) An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. PLoS Genet. 2011 Aug;7(8):e1002218. doi: 10.1371/journal.pgen.1002218.

Simon JM, Hacker KE, Singh D, Brannon AR, Parker JS, Weiser M, Ho TH, Kuan PF, Jonasch E, Furey TS, Prins JF, Lieb JD, Rathmell WK, Davis IJ (2014) Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects. Genome Res. 2014 Feb;24(2):241-50. doi: 10.1101/gr.158253.113.

Simone NL, Bonner RF, Gillespie JW, Emmert-Buck MR, Liotta LA (1998) Laser-capture microdissection: opening the microscopic frontier to molecular analysis. Trends Genet. 1998 Jul;14(7):272-6. doi: 10.1016/S0168-9525(98)01489-9

Singh RK, Cooper TA (2012) Pre-mRNA splicing in disease and therapeutics. Trends Mol. Med. 18, 472–482 .2012 doi: 10.1016/j.molmed.2012.06.006

Solomon DA, Wang Y, Fox SR, Lambeck TC, Giesting S, Lan Z, Senderowicz AM, Conti CJ, Knudsen ES (2003) Cyclin D1 splice variants. Differential effects on localization, RB phosphorylation, and cellular transformation. J Biol Chem. 2003 Aug 8;278(32):30339-47. DOI: 10.1074/jbc.M303969200

Spadaro PA, Flavell CR, Widagdo J, Ratnu VS, Troup M, Ragan C, Mattick JS, Bredy TW (2015) Long noncoding RNA-directed epigenetic regulation of gene expression is associated with anxiety-like behavior in mice. Biol. Psychiatry, S0006-3223(15)00095-5. doi: 10.1016/j.biopsych.2015.02.004.

Spitale RC, Tsai MC, Chang HY (2011) RNA templating the epigenome: long noncoding RNAs as molecular scaffolds. Epigenetics. 2011 May;6(5):539-43. doi:10.4161/epi.6.5.15221

Stunnenberg HG, International Human Epigenome Consortium, Hirst M (2016) The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. Cell. 2016 Nov 17;167(5):1145-1149. doi: 10.1016/j.cell.2016.11.007.

Sun S, Ling SC, Qiu J, Albuquerque CP, Zhou Y, Tokunaga S, Li H, Qiu H, Bui A, Yeo GW, Huang EJ, Eggan K, Zhou H, Fu XD, Lagier-Tourenne, Cleveland DW (2015) ALS-causative mutations in FUS/TLS confer gain and loss of function by altered association with SMN and U1-snRNP. Nat Commun. 2015 Jan 27;6:6171. doi: 10.1038/ncomms7171.

Takahashi S, Ohtsuki T, Yu SY, Tanabe E, Yara K, Kamioka M, Matsushima E, Matsuura M, Ishikawa K, Minowa Y, Noguchi E, Nakayama J, Yamakawa-Kobayashi K, Arinami T, Kojima T (2003) Significant linkage to chromosome 22q for exploratory eye movement dysfunction in schizophrenia. Am. J. Med. Genet. B Neuropsychiatr. Genet. 123B, 27–32. doi: 10.1002/ajmg.b.10046

Tan MH, Li Q, Shanmugam R, Piskol R, Kohler J, Young AN, Liu KI, Zhang R, Ramaswami G, Ariyoshi K, Gupte A, Keegan LP, George CX, Ramu A, Huang N, Pollina EA, Leeman DS, Rustighi A, Goh YPS; GTEx Consortium; Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VARI; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration & Visualization—EBI; Genome Browser Data Integration & Visualization—UCSC Genomics Institute, University of California Santa Cruz, Chawla A, Del Sal G, Peltz G, Brunet A, Conrad DF, Samuel CE, O'Connell MA, Walkley CR, Nishikura K, Li JB (2017) Dynamic landscape and regulation of RNA editing in mammals. Nature. 2017 Oct 11;550(7675):249-254. doi: 10.1038/nature24041.

Taub FE, DeLeo JM, Thompson EB (1983) Sequential comparative hybridizations analyzed by computerized image processing can identify and quantitate regulated RNAs. DNA. 1983;2(4):309-27.

The ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermüller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaöz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Löytynoja A, Whelan S, Pardi F, Massingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Seringhaus M, Church D, Rosenbloom K, Kent WJ, Stone EA, NISC Comparative Sequencing Program, Baylor

College of Medicine Human Genome Sequencing Center, Washington University Genome Sequencing Center, Broad Institute, Children's Hospital Oakland Research Institute, Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameur A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CW, Ng P, Shahab A, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S, Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Zhang X, Xu M, Haidar JN, Yu Y, Ruan Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K, Thakkapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, de Bakker PI, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyras E, Hallgrímsdóttir IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B and de Jong PJ (2012). An integrated encyclopedia of DNA elements in the human genome. Nature. 489 (7414): 57–74. doi:10.1038/nature11247

Tilgner H, Knowles DG, Johnson R, Davis CA, Chakrabortty S, Djebali S, Curado J, Snyder M, Gingeras TR, Guigó R (2012) Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co‑transcriptional in the human genome but inefficient for lncRNAs. Genome Res. 2012 Sep;22(9):1616-25. doi: 10.1101/gr.134445.111.

Tollervey JR, Curk T, Rogelj B, Briese M, Cereda M, Kayikci M, König J, Hortobágyi T, Nishimura AL, Zupunski V, Patani R, Chandran S, Rot G, Zupan B, Shaw CE, Ule J (2011) Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. Nat Neurosci. 2011 Apr;14(4):452-8. doi: 10.1038/nn.2778. Epub 2011 Feb 27.

Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY (2010) Long noncoding RNA as modular scaffold of histone modification complexes. Science. 329 (5992): 689–93. doi:10.1126/science.1192002

Tsai MC, Spitale RC, Chang HY (2011) Long intergenic noncoding RNAs: new links in cancer progression. Cancer Res. 2011;71:3–7. doi: 10.1158/0008-5472.CAN-10-2483

Ule J, Stefani G, Mele A, Ruggiu M, Wang X, Taneri B, Gaasterland T, Blencowe BJ, Darnell RB (2006) An RNA map predicting Nova-dependent splicing regulation. Nature 444, 580–586 2006 DOI: 10.1038/nature05304

Vaquerizas JM, Akhtar A, Luscombe NM (2011) Large-scale nuclear architecture and transcriptional control. Subcell Biochem. 2011;52:279-95. doi: 10.1007/978-90-481-9069- 0_13.

Vargas DY, Shah K, Batish M, Levandoski M, Sinha S, Marras SA, Schedl P, Tyagi S (2011) Single-molecule imaging of transcriptionally coupled and uncoupled splicing. Cell. 2011 Nov 23;147(5):1054-65. doi: 10.1016/j.cell.2011.10.024.

Venkatraman A, He XC, Thorvaldsen JL, Sugimura R, Perry JM, Tao F, Zhao M, Christenson MK, Sanchez R, Yu JY, Peng L, Haug JS, Paulson A, Li H, Zhong XB, Clemens TL, Bartolomei MS, Li L (2013) Maternal imprinting at the H19-Igf2 locus maintains adult haematopoietic stem cell quiescence. Nature. 2013 Aug 15;500(7462):345-9. doi: 10.1038/nature12303.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X (2001) The sequence of the human genome. Science. 291 (5507): 1304–1351. doi:10.1126/science.1058040

Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, Geschwind DH (2011) Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature. 2011 May 25;474(7351):380-4. doi: 10.1038/nature10110.

Wahl MC, Will CL, Luhrmann R (2009) The spliceosome: design principles of a dynamic RNP machine. Cell 2009, 136:701–718. DOI: 10.1016/j.cell.2009.02.009

Wallner S, Schröder C, Leitão E, Berulava T, Haak C, Beißer D, Rahmann S, Richter AS, Manke T, Bönisch U, Arrigoni L, Fröhler S, Klironomos F, Chen W, Rajewsky N, Müller F, Ebert P, Lengauer T, Barann M, Rosenstiel P, Gasparoni G, Nordström K, Walter J, Brors B, Zipprich G, Felder B, Klein-Hitpass L, Attenberger C, Schmitz G, Horsthemke B (2016) Epigenetic dynamics of monocyte-to-macrophage differentiation. Epigenetics Chromatin. 2016 Jul 29;9:33. doi: 10.1186/s13072-016-0079-z.

Wan Y, Wu CJ (2013) SF3B1 mutations in chronic lymphocytic leukemia. Blood 121, 4627–4634 2013. doi: 10.1182/ blood-2013-02-427641

Wang BB and Brendel V (2006) Genomewide comparative analysis of alternative splicing in plants. Proc Natl Acad Sci USA. 2006 May 2;103(18):7175-80. DOI: 10.1073/pnas.0602039103

Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell. 153 (4): 910–8. doi:10.1016/j.cell.2013.04.025

Wang P, Xue Y, Han Y, Lin L, Wu C, Xu S, Jiang Z, Xu J, Liu Q, Cao X (2014) The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. Science. 2014 Apr 18;344(6181):310-3. doi: 10.1126/science.1251456.

Wang P, Yan B, Guo JT, Hicks C, Xu Y (2005) Structural genomics analysis of alternative splicing and application to isoform structure modeling. Proc Natl Acad Sci U S A. 2005 Dec 27;102(52):18920-5. Epub 2005 Dec 14.

Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R (2008) Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature. 454 (7200): 126–30. doi:10.1038/nature06992

Wang Z, Burge CB (2008) Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. RNA 2008 May; 14(5): 802–813. doi: 10.1261/rna.876308

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009 Jan;10(1):57-63. doi: 10.1038/nrg2484.

Washietl S, Kellis M, Garber M (2014) Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. Genome Res. 2014 Apr;24(4):616-28. doi: 10.1101/gr.165035.113.

Watson JD, Crick FH (1953) Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature. 171 (4356): 737–8. doi:10.1038/171737a0

Weigel NL, Moore NL. (2007) Steroid receptor phosphorylation: a key modulator of multiple receptor functions. Mol Endocrinol. 2007 Oct;21(10):2311-9. Epub 2007 May 29.

Weih DS, Yilmaz ZB, Weih F (2001) Essential role of RelB in germinal center and marginal zone formation and proper expression of homing chemokines. J Immunol. 2001 Aug 15;167(4):1909-19.

Wells QS, Becker JR, Su YR, Mosley JD, Weeke P, D'Aoust L, Ausborn NL, Ramirez AH, Pfotenhauer JP, Naftilan AJ, Markham L, Exil V, Roden DM, Hong CC (2013) Whole exome sequencing identifies a causal RBM20 mutation in a large pedigree with familial dilated cardiomyopathy. Circ Cardiovasc Genet. 2013 Aug;6(4):317-26. doi: 10.1161/CIRCGENETICS.113.000011. Epub 2013 Jul 16.

Weyn-Vanhentenryck SM, Mele A, Yan Q, Sun S, Farny N, Zhang Z, Xue C, Herre M, Silver PA, Zhang MQ, Krainer AR, Darnell RB, Zhang C (2014) HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. Cell Rep. 2014 Mar 27;6(6):1139-1152. doi: 10.1016/j.celrep.2014.02.005. Epub 2014 Mar 6.

Whiteside ST, Goodbourn S. (1993) Signal transduction and nuclear targeting: regulation of transcription factor activity by subcellular localisation. J Cell Sci. 1993 Apr;104 (Pt 4):949-55.

Wong JJ, Gao D, Nguyen TV, Kwok CT, van Geldermalsen M, Middleton R, Pinello N, Thoeng A, Nagarajah R, Holst J, Ritchie W, Rasko JEJ (2017) Intron retention is regulated by altered MeCP2-mediated splicing factor recruitment. Nat Commun. 2017 May 8;8:15134. doi: 10.1038/ncomms15134.

Wong JJ, Ritchie W, Ebner OA, Selbach M, Wong JW, Huang Y, Gao D, Pinello N, Gonzalez M, Baidya K, Thoeng A, Khoo TL, Bailey CG, Holst J, Rasko JE (2013) Orchestrated intron retention regulates normal granulocyte differentiation. Cell. 2013 Aug 1;154(3):583-95. doi: 10.1016/j.cell.2013.06.052.

Wutz A, Gribnau J (2007) X inactivation Xplained. Current Opinion in Genetics & Development. 17 (5): 387–93. doi:10.1016/j.gde.2007.08.001

Xiangyue Wu and Gary Brewer (2012) The Regulation of mRNA Stability in Mammalian Cells: 2.0. Gene 2012 May 25; 500(1): 10–21. doi: 10.1016/j.gene.2012.03.021

Xie S, Duan J, Li B, Zhou P, Hon GC (2017) Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. Mol Cell. 2017 Apr 20;66(2):285-299.e5. doi: 10.1016/j.molcel.2017.03.007.

Xiong J, Lu X, Zhou Z, Chang Y, Yuan D, Tian M, Zhou Z, Wang L, Fu C, Orias E, Miao W (2012) Transcriptome analysis of the model protozoan, Tetrahymena thermophila, using

Deep RNA sequencing. PLoS One. 2012;7(2):e30630. doi: 10.1371/journal.pone.0030630.

Yamamoto ML, Clark TA, Gee SL, Kang JA, Schweitzer AC, Wickrema A, Conboy JG (2009) Alternative pre-mRNA splicing switches modulate gene expression in late erythropoiesis. Blood. 2009 Apr 2; 113(14): 3363–3370. Prepublished online 2009 Feb 4. doi: 10.1182/blood-2008-05-160325

Yan L, Yang M, Guo H, Yang L, Wu J, Li R, Liu P, Lian Y, Zheng X, Yan J, Huang J, Li M, Wu X, Wen L, Lao K, Li R, Qiao J, Tang F (2013) Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nature Structural & Molecular Biology. 20 (9): 1131–9. doi:10.1038/nsmb.2660

Yang Z, Zhou L, Wu LM, Lai MC, Xie HY, Zhang F, Zheng SS (2011) Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. Ann Surg Oncol. 2011 May;18(5):1243-50. doi: 10.1245/s10434-011-1581-y.

Yap K, Lim ZQ, Khandelia P, Friedman B, Makeyev EV (2012) Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. Genes Dev. 2012 Jun 1;26(11):1209-23. doi: 10.1101/gad.188037.112.

Yap KL, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, Zhou MM (2010) Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol Cell. 2010 Jun 11;38(5):662-74. doi: 10.1016/j.molcel.2010.03.021.

Yarosh CA, Tapescu I, Thompson MG, Qiu J, Mallory MJ, Fu XD, Lynch KW (2015) TRAP150 interacts with the RNA-binding domain of PSF and antagonizes splicing of numerous PSF-target genes in T cells. Nucleic Acids Res. 2015 Oct 15;43(18):9006-16. doi: 10.1093/nar/gkv816. Epub 2015 Aug 10.

Ye M, Zhang H, Amabile G, Yang H, Staber PB, Zhang P, Levantini E, Alberich-Jordà M, Zhang J, Kawasaki A, Tenen DG (2013) C/EBPa controls acquisition and maintenance of adult haematopoietic stem cell quiescence. Nat Cell Biol. 2013 Apr;15(4):385-94. doi: 10.1038/ncb2698.

Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M, Chalkidis G, Suzuki Y, Shiosaka M, Kawahata R, Yamaguchi T, Otsu M, Obara N, Sakata-Yanagimoto M, Ishiyama K, Mori H, Nolte F, Hofmann WK, Miyawaki S, Sugano S, Haferlach C, Koeffler HP, Shih LY, Haferlach T, Chiba S, Nakauchi H, Miyano S, Ogawa S (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 478, 64–69 2011. doi: 10.1038/nature10496

Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ (2005) Genome-scale identification of nucleosome positions in S. cerevisiae. Science. 2005 Jul 22;309(5734):626-30. Epub 2005 Jun 16.

Zan Y, Haag JD, Chen KS, Shepel LA, Wigington D, Wang YR, Hu R, Lopez-Guajardo CC, Brose HL, Porter KI, Leonard RA, Hitt AA, Schommer SL, Elegbede AF, Gould MN (2003) Production of knockout rats using ENU mutagenesis and a yeast-based screening assay. Nature Biotechnology. 21 (6): 645–51. doi:10.1038/nbt830

Zappulla DC, Cech TR (2006) RNA as a flexible scaffold for proteins: yeast telomerase and beyond. Cold Spring Harb Symp Quant Biol. 2006;71:217-24. doi:10.1101/sqb.2006.71.011

Zhang J, Liu H, Liu Z, Liao Y, Guo L, Wang H, He L, Zhang X, Xing Q (2013) A Functional Alternative Splicing Mutation in AIRE Gene Causes Autoimmune Polyendocrine Syndrome Type 1. PLoS One. 2013; 8(1): e53981. doi: 10.1371/journal.pone.0053981

Zhang MQ (1998) Statistical features of human exons and their flanking regions. Hum Mol Genet 1998, 7:919–932.

Zhang Q, Li H, Jin H, Tan H, Zhang J, Sheng S (2014) The global landscape of intron retentions in lung adenocarcinoma. BMC Med Genomics. 2014 Mar 20;7:15. doi: 10.1186/1755-8794-7-15.

Zhang X, Lian Z, Padden C, Gerstein MB, Rozowsky J, Snyder M, Gingeras TR, Kapranov P, Weissman SM, Newburger PE (2009) A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. Blood. 2009 Mar 12;113(11):2526-34. doi: 10.1182/blood-2008-06-162164.

Zhao Y and Simon R (2010) Gene expression deconvolution in clinical samples. Genome Med. 2010; 2(12): 93. doi: 10.1186/gm214

Zhou S, Yang Y, Scott MJ, Pannuti A, Fehr KC, Eisen A, Koonin EV, Fouts DL, Wrightsman R, Manning JE, Lucchesi JC (1995) Male-specific lethal 2, a dosage compensation gene of Drosophila, undergoes sex-specific regulation and encodes a protein with a RING finger and a metallothionein-like cysteine cluster. EMBO J. 1995 Jun 15;14(12):2884-95.

Zhu S, Li W, Liu J, Chen CH, Liao Q, Xu P, Xu H, Xiao T, Cao Z, Peng J, Yuan P, Brown M, Liu XS, Wei W (2016) Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. Nat Biotechnol. 2016 Dec;34(12):1279-1286. doi: 10.1038/nbt.3715.