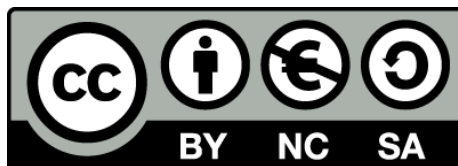




Variació genòmica de les Sintases d'Òxid Nítric i de regions associades amb Risc Cardiovascular en Poblacions Humanes

Robert Carreras Torres



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- NoComercial – Compartir Igual 3.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - NoComercial – Compartir Igual 3.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-NonCommercial-ShareAlike 3.0. Spain License.**

**Variació genòmica de les Sintases d'Òxid Nítric
i de regions associades amb Risc Cardiovascular
en Poblacions Humanes**

Tesi Doctoral presentada per
Robert Carreras Torres

Per optar al grau de Doctor per la Universitat de Barcelona
Programa de Doctorat en Biodiversitat

Dirigida pel Dr. Pedro Moral Castrillo, professor de la Unitat
d'Antropologia, Departament de Biologia Animal, Universitat de Barcelona,
i pel Dr. Marc Via García, professor del Departament de Psiquiatria i
Psicobiologia Clínica, Universitat de Barcelona.

Pedro Moral Castrillo
Director

Marc Via García
Director

Robert Carreras Torres
Estudiant de doctorat

Als meus pares, Àngela i Joaquim,
i a la meva germana, Marina.

*"Gilgamesh, per què vagues d'un costat a un altre?
No aconseguiràs la vida que persegueixes.
Quan els déus van crear la humanitat,
van decidir que el seu destí fos morir
i van reservar la Vida per a si mateixos.
Tu, Gilgamesh, omple't el ventre, gaudeix de dia i de nit.
Celebra cada dia una alegre festa, balla i juga dia i nit.
Posa't vestits flamants, renta't el cap i banya't.
Atén el nen que t'agafa de la mà i alegra't,
alegra't abraçant la teva dona.
Ja que aquest és el destí de l'home."*

Tauleta X; columna III
La Epopeia de Gilgamesh

*"Gilgamesh, where are you hurrying to?
You will never find that life for which you are looking.
When the gods created man they allotted to him death,
but life they retained in their own keeping.
As for you, Gilgamesh, fill your belly with good things;
day and night, night and day, dance and be merry, feast and rejoice.
Let your clothes be fresh, bathe yourself in water,
cherish the little child that holds your hand,
and make your wife happy in your embrace;
for this too is the lot of man."*

Table X, column III
The Epic of Gilgamesh

Agraïments

Girona, 8 d'Abril de 2014.

A l'hora d'escriure aquests agraïments repasso aquests últims anys en què vaig iniciar-me en el món de la recerca i puc dir que n'he vist de tots colors, he viscut en considerables llocs i he gaudit d'uns dels períodes més importants de la meva vida. Gràcies a totes aquestes experiències he tingut la oportunitat de gaudir i conèixer un grup de persones que m'ha ajudat d'una forma difícil de compensar a evolucionar i millorar com a persona i científic. És gràcies a persones com elles que un se n'adona que sol no s'arriba enlloc. Primer de tot, donar les gràcies al director de la tesi, el Dr. Pedro Moral. Moltes gràcies Pedro per confiar en mi, per animar-me en moments difícils, per escoltar els meus dubtes i propostes, i per ajudar-me a definir i concretar la meva ment dispersa. Gràcies per convertir-me en un investigador.

També vull donar les gràcies al co-director de la tesi, el Dr. Marc Via. Gràcies Marc per tots els consells, propostes i ajudes que he rebut. Veritablement m'has ajudat a superar molts obstacles, inclús quan estaves a EEUU, encara no ens coneixíem i jo feia el projecte del màster.

Gràcies a totes les persones del grup d'en Pedro que en formen part o n'han participat puntualment. A la Dra. Esther Esteban per la teva elegància en resoldre obstacles. A la Daniela per les hores compartides dins i fora la facultat. A Yorgos, Magda G, Emili, Ares, Josep, Magda R, Irene, Albert, Lucía, Reyhane, Miguel, Cristina, Toni, Neus, Natalia, Meri, Laura, Valeria, Wifak, Roudha, Abir... per les hores de treball, dinars, congressos i vivències divertides, intenses i algunes de surrealistes.

Gràcies a tots els professors, doctorands i estudiants del departament d'Antropologia pels dinars, cafès, cerveses, shows i partits de futbol compartits.

I would also like to thank Prof. Cecile Janssens and Prof. Laurent Excoffier for supervising my work stays at the Erasmus Medical Centre of Rotterdam, the Netherlands, and at the Institute of Ecology and Evolution of Bern, Switzerland, respectively. I want to thank all the people I met there for sharing meals, beers, laughs, hangovers, holidays, and even scientific discussions.

I will never forget them all.

En el terreny personal, vull donar enormement les gràcies a la meva mare Àngela i al meu pare Joaquim, que ens va abandonar a mitjans de la tesi. A la Marina, la millor germana que es pot tenir. També a la resta de la família, especialment a la iaia Francisca, que també ens va deixar fa pocs mesos. No puc evitar les llàgrimes quan escric aquestes últimes línies.

Gràcies Tània per compartir les meves il·lusions. A la Vanessa per ser-hi sempre per compartir cigarret i conversa. A tots els amics i amigues de Salt i Girona per les innombrables vivències i activitats des de ja fa bastants anys. Us estimo a tots.

CONTINGUTS

TAULA DE CONTINGUTS

1. INTRODUCCIÓ	19
1.1 Malalties cardiovasculars com a fenotip complex	21
1.1.1 Bases fisiològiques de les malalties cardiovasculars.....	21
1.1.2 Factors implicats en la malaltia coronària complexa: Gens i Ambient	26
1.2 Elements genètics estudiats: gen <i>F13A1</i> i gens <i>NOS</i>	29
1.3 Bases de la diversitat genètica	31
1.3.1 Organització del genoma humà.....	31
1.3.2 Generació de variació genètica: mutació i recombinació.....	31
1.3.3 Tipus de variació genètica o polimorfismes.....	33
1.4 Variació genètica i malaltia: Epidemiologia genètica	38
1.4.1 Identificació dels factors de risc genètics.....	38
1.4.2 Estimació de risc individual a la malaltia cardiovascular.....	42
1.4.3 Distribució poblacional de la malaltia cardiovascular i factors de risc	43
1.5 Genètica de Poblacions Humanes.....	47
1.5.1 Distribució de la diversitat genètica en les poblacions humanes.....	47
1.5.1.1 Processos de diferenciació genètica: deriva genètica, selecció i migració	50
1.5.1.2 Diversitat genètica en humans.....	52
1.5.1.3 Mesures de distribució i estructuració genètica.....	54
1.5.1.4 Distribució i estructuració genètica global i del continent europeu.....	56
1.5.2 Moviments poblacionals al llarg de la Història i les seves empremtes al genoma	60
1.5.2.1 Moviments poblacionals a Europa i al Mediterrani.....	64
1.6 Poblacions estudiades: Europa continental i àrea mediterrània	68
2. OBJECTIUS	73
2.1 Objectius de la tesi.....	75
3. RESULTATS.....	79
3.1 Informe del director de tesi del factor d'impacte dels articles publicats	81
3.2 Resultats I: Carreras-Torres et al., 2010	83

3.2.1 Resum en Català	83
3.2.2 Informe dels directors de tesi de la participació del doctorant en l'article i de la singularitat de l'article en la elaboració de la tesi.....	87
3.2.3 Article I.....	89
3.3 Resultats II: Carreras-Torres et al., 2014.....	95
3.3.1 Resum en Català	95
3.3.2 Informe dels directors de tesi de la participació del doctorant en l'article i de la singularitat de l'article en la elaboració de la tesi.....	99
3.3.3 Article II.....	101
3.4 Resultats III: Carreras-Torres et al., (article III).....	135
3.4.1 Resum en Català	135
3.4.2 Informe dels directors de tesi de la participació del doctorant en l'article i de la singularitat de l'article en la elaboració de la tesi.....	139
3.4.3. Article III	141
4. DISCUSSIÓ.....	177
4.1 Discussió general	179
4.2 Identificació de polimorfismes del gen <i>F13A1</i> de risc per l'infart de miocardi en població espanyola i possible interacció genètica de risc dins del gen.	179
4.2.1 Marcadors de risc per infart de miocardi al gen <i>F13A1</i> en població espanyola	179
4.2.2 Interacció de la variació genètica de risc dins del gen <i>F13A1</i>	180
4.2.3 Efecte poblacional en la identificació de marcadors de risc amb rol interactiu	182
4.3 Identificació de factors de risc en els gens <i>NOS</i> i distribució poblacional de risc a Europa i la Mediterrània en relació a la incidència coronària.	184
4.3.1 Identificació de marcadors de risc genètic dels gens <i>NOS</i> per infart de miocardi en població europea	184
4.3.2 Habilitat predictiva de la puntuació de risc genètic dels gens <i>NOS</i>	185
4.3.3 Distribució poblacional del GRS a través d'Europa i la Mediterrània.....	186
4.3.4 Correlació del GRS i de la variació genètica individual amb la incidència coronària descrita a Europa	187
4.3.5 Incidència coronària poblacional., predisposició genètica i hàbits i estil de vida	189
4.3.6 Utilitat dels GRS per futurs estudis eco-epidemiològics.....	191

4.4 Diferenciació poblacional reflectida pels gens <i>NOS</i> a Europa i la Mediterrània i identificació dels processos demogràfics i/o adaptatius en relació a la incidència coronària.	193
4.4.1 Estimació del grau de diferenciació genètica entre les poblacions analitzades, especialment d'Europa, Nord d'Àfrica i Orient Mitjà.	193
4.4.2 Processos d'adaptació genètica al medi reflectits en els gens <i>NOS</i>	194
4.4.3 Estructuració genètica observada a Europa i la Mediterrània per les regions genòmiques dels gens <i>NOS</i>	195
4.4.4 Components “ancestrals” estimats del patrimoni genètic actual de les poblacions d'Europa i la Mediterrània.	196
4.4.5 Interpretació dels processos poblacionals que originaren l'estructuració genètica en gradients ortogonals i la distribució de components ancestrals observats	199
4.4.6 Estimació de l'estructuració poblacional de l'arquitectura genètica de les malalties coronàries i processos poblacionals que determinarien la diversa incidència coronària a les poblacions europees	201
5. CONCLUSIONS.....	205
5.1 Conclusions d'aquesta tesi.....	207
5.2 Conclusions of this thesis	209
6. REFERÈNCIES	211
7. APÈNDIX.....	221
7.1 Apèndix 1. Material suplementari de Carreras-Torres et al., 2014	223
7.2 Apèndix 2. Material suplementari de Carreras-Torres et al., (article 3).....	247

Prefaci

Les poblacions humanes han estat objecte d'estudi des que les inicials civilitzacions s'esforçaren per justificar la seva presència i destí. La diversitat observada en els diferents grups humans ha estat àmpliament documentada per antropòlegs segons els coneixements i tècniques disponibles en cada moment. A mesura que les tècniques moleculars han fet accessible la informació continguda en el nostre genoma, vells models de descripció fenotípica i història poblacional han estat substituïts o complementats per models més sofisticats.

Actualment, un dels aspectes en els quals s'està invertint molt d'esforç és en la modelització dels fenotips anomenats complexos. Entre aquests fenotips, fruit de la interacció d'un elevat nombre de components genètics i ambientals, hi trobem moltes malalties que tant poden aparèixer en un estadi inicial de la vida com que es manifestin a edats tardanes. Alguns d'aquests fenotips presenten grans diferències entre les poblacions humanes, i per tant, també els seus factors causants. Per tal d'entendre la distribució observada d'aquests fenotips i els seus factors determinants, és necessari fer una ullada a la història demogràfica de les poblacions. Per entendre el present i obtenir-ne un benefici futur, hem de conciliar-nos amb el passat.

En aquest context, el treball presentat en aquesta tesi és un estudi antropogenètic de diferents gens, el gen del factor de coagulació XIII (*F13A1*) i els gens de les sintases d'òxid nítric (*NOS*), implicats en un fenotip complex amb variació poblacional com són les malalties cardiovasculars. Les nostres poblacions d'estudi es situen al voltant de la Mediterrània i al continent europeu, on, tot i tractar-se d'una àrea geogràfica no molt extensa, la variació en la presència de les malalties cardiovasculars és considerable. En la primera secció, la Introducció, situarem el treball

present dins d'un context general. La Introducció té sis parts; en la primera part s'introdueix les malalties cardiovasculars com a fenotip complex i se'n descriu els seus causants principals per tal de generar una visió de la problemàtica cardiovascular. La segona i la tercera part es centren en la informació genètica estudiada, els components genètics *F13A1* i *NOS* i els tipus de polimorfismes, alhora es dóna una visió de com es genera variació en la nostra estructura genètica. En la quarta, es dóna una pinzellada d'una àmplia àrea de coneixement i recerca com és l'epidemiologia genètica, amb les seves estratègies per tal de generar un model explicatiu dels fenotips complexos utilitzant l'exemple de les malalties cardiovasculars. En la cinquena part, es realitza una aproximació a una altra gran àrea de coneixement i recerca com és la genètica de poblacions humanes, amb les seves eines d'estudi de la distribució de la variació present en les poblacions humanes i amb les inferències poblacionals que s'obtenen, posant com exemple les poblacions d'Europa i la Mediterrània. Finalment, la sisena part descriu les mostres poblacionals utilitzades en la majoria de treballs presentats en aquesta tesi.

La Introducció és seguida per les fites a assolir en aquest treball en la secció d'Objectius. La secció de Resultats conté quatre apartats. El primer és un informe dels supervisors d'aquesta tesi sobre la qualitat dels articles publicats. I en els apartats restants, cadascun conté un dels articles amb el seu resum en català. El primer article descriu la identificació de variació de risc en el gen *F13A1* per malalties cardiovasculars en població espanyola. El segon article detalla la identificació de factors de risc en el gens *NOS* per malalties cardiovasculars en població europea, i com la seva distribució es correlaciona amb la incidència cardiovascular observada en el continent europeu. El tercer article estima les relacions entre les poblacions d'Europa i

la Mediterrània obtingudes per un conjunt de variació genètica de risc i no de risc de les regions genòmiques *NOS*, i la seva relació amb la incidència cardiovascular. Finalment, les segueixen les seccions de Discussió i Conclusions.

1.INTRODUCCIÓ

1.1 Malalties cardiovasculars com a fenotip complex

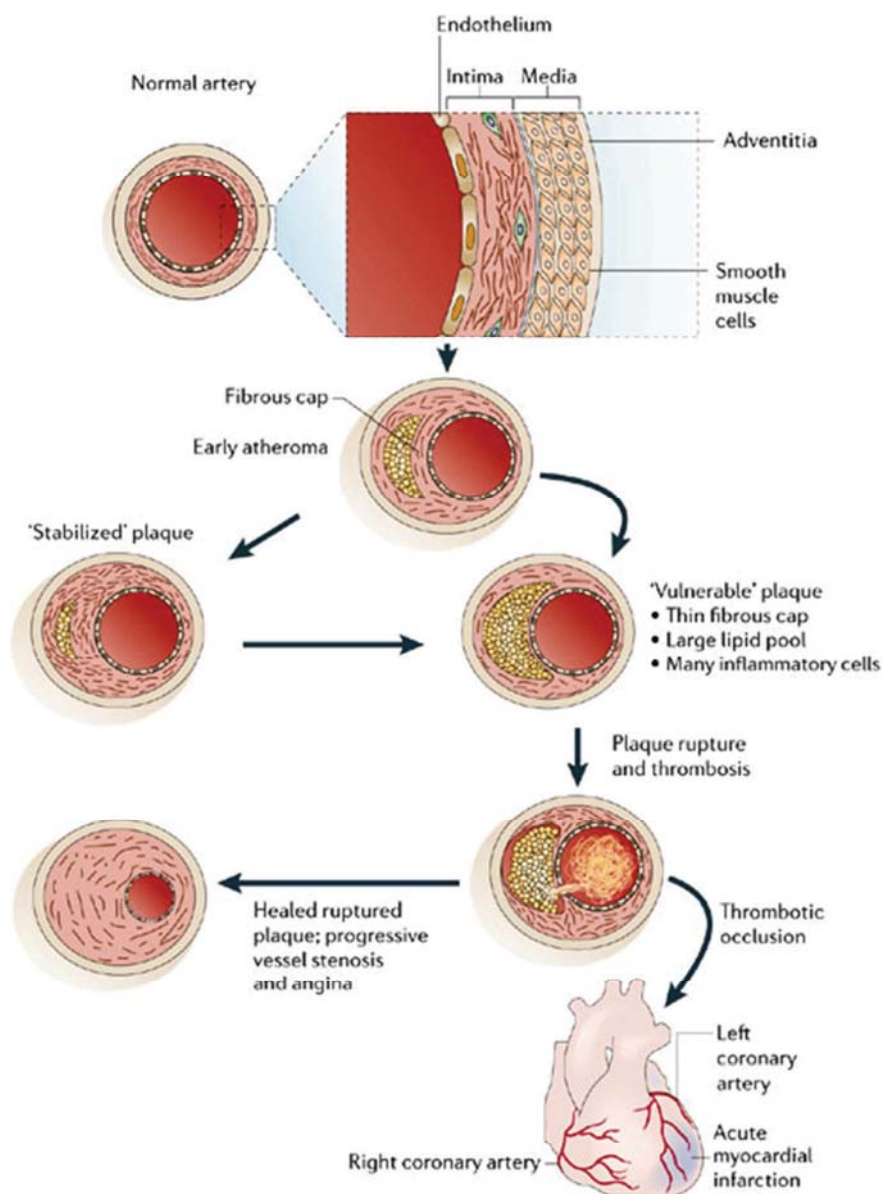
Aquesta tesi doctoral gira al voltant de les malalties cardiovasculars com a fenomen biològic que es presenta amb considerable freqüència a les poblacions humanes. Per tal d'entendre'n la actual presència i distribució a través de la geografia, aquest treball n'intenta identificar algunes de les causes genètiques i descriure la seva distribució poblacional des del prisma de la història demogràfica recent de les poblacions humanes del continent europeu i de la Mediterrània.

Es coneix com a fenotip l'expressió biològica observable d'un caràcter o conjunt de caràcters. Quan aquesta expressió és el resultat de la interacció de diversos factors es parla de fenotip complex. Les malalties cardiovasculars són expressions biològiques que acaben manifestant-se quan es donen una sèrie de condicions més o menys encadenades o, de forma paral·lela, després de l'acumulació de diversos factors normalment difícils de controlar. Per aquest motiu, a les malalties cardiovasculars, junt amb d'altres, són anomenades malalties complexes.

1.1.1 Bases fisiològiques de les malalties cardiovasculars

Les bases fisiològiques que intervenen en el desenvolupament d'una malaltia cardiovascular són diverses, però un procés clau és l'anomenat aterosclerosi. Aquest procés patològic consisteix en la formació de plaques ateroscleròtiques a les parets dels vasos sanguinis, que en redueixen el flux degut a l'estrenyiment i enduriment dels vasos (Figura 1). La progressiva expansió d'aquestes lesions en estadis avançats pot produir una ruptura de la paret vascular, i una oclusió del vas degut a un efecte trombòtic. Finalment, la interrupció del flux sanguini pot produir necrosis en el teixit afectat. Quan aquest procés es desenvolupa en una arteria coronària s'anomena CAD (de

l'anglès *Coronary Artery Disease*). La trombosis i la necrosis del teixit muscular cardíac (miocardi) acaben produint un infart de miocardi (MI de l'anglès *Myocardial Infarction*), més comunament conegut com a atac de cor [1] (Figura 1).



Copyright © 2006 Nature Publishing Group
Nature Reviews | Genetics

Figura 1- Estadis de l'aterosclerosi [1].

La formació de les plaques ateroscleròtiques és deguda inicialment a una elevada absorció de lípids (triglicèrids i colesterol) per part de l'endoteli vascular que s'acumulen a la matriu endotelial del vas sanguini. Aquesta elevada absorció pot ser estimulada per una disfunció de l'endoteli (afectat pel tabaquisme o la diabetis) i/o per nivells elevats de lípids en sang (producte d'una deficiència en el transport de les lipoproteïnes i síntesi dels lípids, o una ingesta elevada de greixos) [2] (Figura 2). Aquests lípids dipositats a la matriu del vas són absorbits i oxidats per macròfags (glòbuls blancs), que esdevenen les anomenades cèl·lules escumoses, i per cèl·lules musculars, que proliferen i entren en apoptòsis (mort cel·lular). Aquest procés d'inflamació, proliferació i apoptòsis acaba generant la matriu fibrosa, endurida i farcida de lípids coneguda com placa ateroscleròtica.

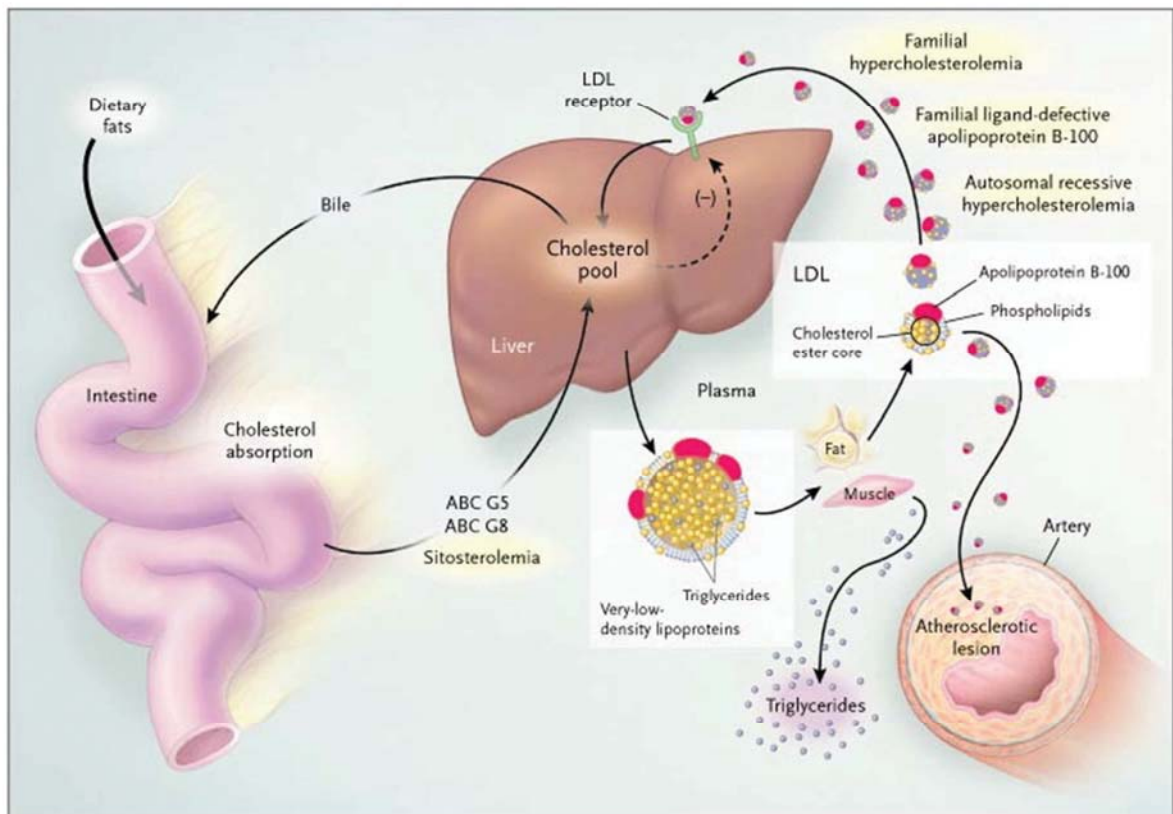


Figura 2 – Absorció, síntesi i excreció del colesterol [2].

Aquestes plaques ateroscleròtiques són altament susceptibles de patir ruptures i d'alliberar el seu contingut al torrent sanguini. L'alliberament dels diferents components de les plaques al torrent sanguini desencadena una sèrie de reaccions conegudes com a cascada de coagulació en les que intervenen les cèl·lules plaquetàries i múltiples factors de coagulació, que poden acabar formant un trombe o coàgul de fibrina [3] (Figura 3). La regulació de la cascada de coagulació, junt amb la fibrinòlisi (reabsorció del coàgul), està dissenyada per tal de generar una resposta adequada i evitar hemorràgies o trombosis no desitjades.

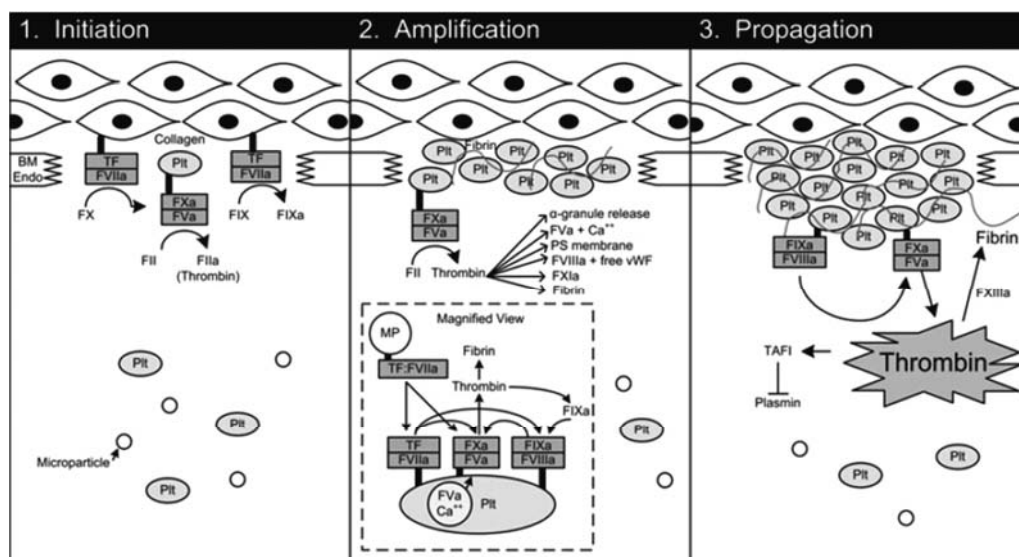


Figura 3- Estadis de la coagulació [3].

Aquest procés global de regulació per tal de mantenir la integritat d'un sistema circulatori tancat i amb una pressió vascular adequada s'anomena hemostàsia. En l'hemostàsia de la pressió vascular són múltiples les vies de regulació. Aquestes inclouen tant la regulació a nivell endocrí, com seria el mecanisme de renina-angiotensina, com la regulació mitjançant la contracció o relaxació de la paret muscular del vas sanguini, com seria la via de l'òxid nítric [4] (Figura 4).

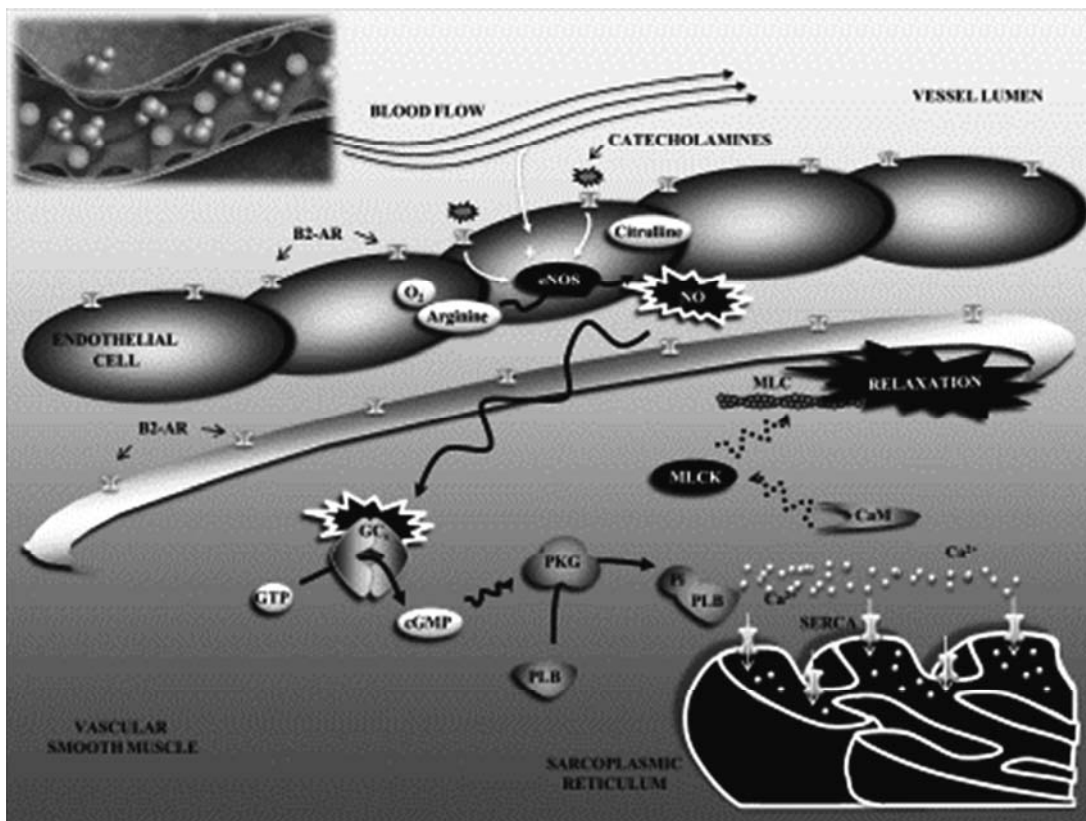


Figura 4 – Generació i alliberament d'òxid nítric i inducció de la relaxació muscular [4].

En concret, la disponibilitat d'òxid nítric en el torrent sanguini és considerat un fenomen clau del trencaclosques cardiovascular, ja que no només es troba involucrat en la flexibilitat de la paret muscular per tal de respondre a variacions de pressió, sinó que aquest procés evita possibles ruptures de la paret. A més, l'òxid nítric es troba involucrat dins de processos anteriorment anomenats com en la inducció de la resposta inflamatòria generada pels macròfags i en la inhibició de l'agregació plaquetària en la formació dels coàguls sanguinis. Pel fet que l'òxid nítric es troba involucrat en bastants processos relacionats amb la malaltia cardiovascular, els gens responsables de la disponibilitat d'òxid nítric són un dels elements estudiats en aquesta tesi doctoral.

1.1.2 Factors implicats en la malaltia coronària complexa: Gens i Ambient

Tal i com la Organització Mundial de la Salut (OMS o WHO, de l'anglès *World Health Organization*) ha constatat, els factors de risc més importants identificats per les malalties cardiovasculars són deguts als hàbits i estils de vida. Dietes no saludables, inactivitat física, tabaquisme o un consum no moderat d'alcohol, es troben presents en el 80% dels casos de malaltia coronària [5]. Aquests hàbits generen desequilibris fisiològics que acaben expressant-se com a sobrepès, hipertensió, hiperglucèmia i colesterolèmia. Aquestes expressions biològiques també s'anomenen fenotips, o endofenotips en el cas que generalment precedeixin fenotips més complexos com seria l'aparició d'una malaltia cardiovascular.

Tot i així, els antecedents o història familiar respecte la malaltia també és un factor de risc important. L'agregació familiar apareix quan existeix major freqüència de malaltia en parents pròxims a individus amb la malaltia. Aquest fet indica que existeix un component hereditari. La molècula responsable de l'herència, DNA (de l'anglès *Deoxyribonucleic Acid*), conté la informació estructural i funcional dels diferents elements biològics que conformen els éssers vius en forma de gens. Per tant, aquest component hereditari és atribuïble als gens.

La quantificació del component hereditari es pot realitzar mitjançant estudis de bessons monozigòtics i dizigòtics que creixen conjuntament o per separat. Així, s'ha estimat que l'heretabilitat dels casos d'infart es troba en un 60% en homes i un 40% en dones [6–8]. De la mateixa manera, els desequilibris fisiològics (endofenotips) generats pels mals hàbits i estils de vida també presenten un component hereditari. A la taula 1 es pot observar una llista de factors de risc ambientals i endofenotips de risc amb els seus

components hereditaris. Per tant, això ens indica que la combinació d'una predisposició genètica amb una exposició a l'ambient (mals hàbits en el cas de CAD) són els factors causants de les malalties complexes com és el cas de la CAD.

Factors de risc amb component genètic	Heretabilitat
Infart de Miocardi	25-60%
Colesterol Total	40-60%
Colesterol HDL	45-75%
Triglicèrids Totals	40-80%
Índex de massa corporal	25-60%
Pressió sanguínia sistòlica	50-70%
Pressió sanguínia diastòlica	50-65%
Lp(a)	90%
Homocisteïna	45%
Diabetis Tipus II	40-80%
Fibrinogen	20-50%

Taula 1 – Heretabilitat de les malalties cardiovasculars i els seus factors de risc [9].

Tot i que la contribució genètica a la malaltia es pot quantificar, la identificació del nombre de factors genètics implicats i l'efecte de la seva contribució són aspectes actualment encara per resoldre. Recentment, s'han incrementat les evidències que la susceptibilitat genètica a les malalties complexes involucren molts gens, inclosos en la patofisiologia molecular, la majoria dels quals tenen un efecte o contribució modesta. Un model teòric

que explica aquest tipus de contribució genètica s'anomena “malaltia comuna – variant comuna” (CDCV de l'anglès *common disease-common variant*). Aquesta hipòtesi prediu que fenotips complexos habitualment presents en la població són causats per múltiples variants genètiques també de freqüència comuna o elevada en la població, cadascuna contribuint en una proporció molt petita al fenotip final [10]. Estudis de simulació computacional han intentat estimar el nombre total de variants genètiques que conformarien l'arquitectura genètica de les CAD. Aquestes simulacions prediuen que el nombre necessari rondaria des dels pocs centenars de variants genètiques de baixa freqüència poblacional i elevada contribució al fenotip, fins a diversos centenars o pocs milers de variants d'elevada presència en la població i modesta contribució a la malaltia [11,12].

Finalment, degut a l'elevada complexitat que presenta la malaltia cardiovascular, la presència clínica de la malaltia pot ser deguda o desencadenada a partir de múltiples processos. Cada pacient pot haver-hi arribat amb una combinació quasi única de factors i cada pacient pot presentar diferents proporcions de factors ambientals i factors genètics, amb un fenotip final majoritàriament degut a factors ambientals, a factors genètics, o a una contribució similar d'ambdós.

1.2 Elements genètics estudiats: gen *F13A1* i gens *NOS*

El treball desenvolupat en aquesta tesi doctoral s'ha centrat en l'estudi de variació genètica present en diferents regions genòmiques. Aquestes regions comprenen el gen del factor de coagulació XIII (*F13A1*) i els tres gens de les sintases d'òxid nítric (*NOS* de l'anglès *Nitric Oxide Synthases*).

- **Gen *F13A1***

El gen *F13A1* codifica per la subunitat A del factor de coagulació XIII (FXIII). Aquest factor de coagulació és una pro-transglutaminasa essencial per l'equilibri hemostàtic. El FXIII és activat per la trombina junt amb la presència de calci (Ca^{2+}) i juga un paper molt important en les etapes finals de la coagulació sanguínia i en la regulació de la fibrinòlisi. El FXIII activat (FXIIIa) uneix covalentment monòmers de fibrina adjacents i estabilitza l'estructura del coàgul sanguini [13] (Figura 3). El FXIII està format per dues subunitats A (unitats funcionals) i dues subunitats B (molècules portadores). El gen *F13A1* es troba localitzat en la regió cromosòmica 6p25-p24, mentre que el gen *F13B*, que codifica per la subunitat B, es troba localitzat en la regió genòmica 1q31-q32.1. La variació genètica en el gen *F13A1* ha estat prèviament associada amb malalties aterotrombòtiques i variacions en els nivells i activitat de FXIII en sang [14,15].

- **Gens *NOS***

Els gens de la sintases de òxid nítric (*NOS*) es troben relacionats amb un fenomen clau del trencaclosques cardiovascular: la disponibilitat d'òxid nítric. L'òxid nítric (NO de l'anglès *Nitric Oxide*) es troba implicat en la regulació de l'hemostàsia de la pressió vascular, la perfusió tissular i

l'agregació plaquetària [16,17] (Figura 4). Gens candidats per un efecte patològic del NO són la família dels gens *NOS*. Existeixen tres gens de les sintases que codifiquen per tres isoformes de l'enzim: *NOS endotelial* (*eNOS* o *NOS3*), *NOS neuronal* (*nNOS* o *NOS1*) i *NOS induïble* (*iNOS* o *NOS2A*). *NOS3* i *NOS1* s'expressen constitutivament en les cèl·lules musculars de l'endoteli dels vasos sanguinis i al teixit neuronal, respectivament [16]. Ambdós són regulades a través de la unió reversible del conjunt calci-calmodulina. Per contra, *NOS2A* és activada en situacions crítiques a través de senyals inflamatoris, principalment en l'endoteli dels vasos sanguinis i en els macròfags. Però, a la vegada, alts nivells de l'enzim induïts per l'expressió de *NOS2A* poden generar citotoxicitat i interferir en les activitats beneficioses de l'enzim expressat constitutivament [16]. A més dels gens *NOS*, existeix el gen *NOS3AS* o *ATG9B* que codifica per un mRNA antisentit anomenat sONE que regula post-transcripcionalment l'expressió de *NOS3* [17]. El gen *ATG9B* es troba localitzat a la cadena de DNA complementària a la que trobem el gen *NOS3* en la regió cromosòmica 7q36, de manera que el mRNA antisentit sONE i el mRNA de *NOS3* es solapen i són complementaris en 662 nucleòtids. Els gens *NOS1* i *NOS2A* es localitzen a les regions cromosòmiques 12q24 i 17q11 respectivament.

D'acord amb el buscador Human Genome Epidemiology (HuGE) Navigator (www.hugenavigator.net), el *NOS3* es el segon gen amb més publicacions relacionant-lo amb la CAD, amb 139 articles, i el quart per l'infart de miocardi amb 75 articles (Abril, 2014). El gen *NOS3* també ha estat estudiat en casos afectats per hipertensió i diabetis. Tot i així, els resultats obtinguts no han acabat de ser concloents degut a la diversitat de resultats [18–23]. Els gens *NOS1* i *NOS2*, a més a més, s'han vist associats a malalties com asma, esquizofrènia, Parkinson i Alzheimer [24–27].

1.3 Bases de la diversitat genètica

L'estructura de la molècula de DNA va ser proposada pels doctors Watson i Crick l'any 1953. La molècula està formada per dues seqüències de nucleòtids definides per la base nitrogenada que porta incorporat cada nucleòtid. Les dues seqüències són complementàries de manera que les bases adenina (A) i timina (T) es complementen entre elles tal i com ho fan citosina (C) i guanina (G).

1.3.1 Organització del genoma humà

El 98% de les molècules de DNA es troben aparellades, i organitzades en 22 parelles de cromosomes autosòmics i el cromosoma X, aquest últim present en una còpia en homes i dues en dones. Els parells de cromosomes provenen un de cada progenitor i, per tant, cada individu conté la mateixa informació per duplicat. El 2% restant comprèn els peculiars segments de DNA només heretats d'un dels dos progenitors (uniparentals): un es el cromosoma Y, que és transmès de pares a fills masculins ja que és el determinant del sexe, mentre que l'altre és el DNA mitocondrial (mtDNA), una molècula circular present tant en homes com en dones, però transmès només de mares a fills (masculins i femenins).

1.3.2 Generació de variació genètica: mutació i recombinació

A mesura que el DNA és transmès a través de les generacions, es produeixen canvis en la seqüència degut a dos fenòmens bàsics: mutació i recombinació.

Els canvis de seqüència on es generen noves variants s'anomenen mutacions i poden ser produïts per factors mutagènics (radiació, virus, químics) o per errors de còpia durant la replicació del DNA. A nivell

evolutiu només es tenen en compte aquelles mutacions que es generen en les cèl·lules reproductores (via germinal) i, per tant, són transmeses a la descendència i passen a formar part de la població. Aquest mecanisme és l'únic que genera nous polimorfismes, o sia noves variacions genètiques alternatives (al·lels) en una posició de la seqüència, i fa que les seqüències entre individus es vagin diferenciant.

Un altre factor que també genera variació genètica no és la generació de noves variants sinó combinacions diferents entre les ja existents. La combinació de variacions genètiques en una seqüència ve donada per la molècula rebuda d'un dels progenitors. Però existeix un mecanisme anomenat recombinació que entrecreu molècules de DNA aparellades que contenen la mateixa informació però amb diferents variants genètiques (informació duplicada on cada còpia prové d'un dels dos progenitors) i genera noves combinacions de variants genètiques que són transmeses en bloc a la següent generació. Per tant, per cada cromosoma, la molècula de DNA que es transmet a la següent generació és una combinació de les dues molècules rebudes de la generació anterior. Aquest procés de recombinació i la combinació en què generalment es troben les diferents variants genètiques és matemàticament mesurable i s'anomena lligament.

Aquest fenomen de recombinació es dona en els cromosomes autosòmics i en el cromosoma X (només en dones). Per tant, l'estudi del mtDNA i del cromosoma Y només ens dona informació dels llinatges uniparentals matern i patern respectivament. En canvi, la diversitat autosòmica és biparental i ens dona informació sobre la ascendència genètica dels dos progenitors a la vegada.

1.3.3 Tipus de variació genètica o polimorfismes

Grans avenços s'han realitzat en els últims 25 anys pel que fa l'estudi de la variació genètica humana. Diferents tècniques de genotipació han permès descobrir moltes i diferents variacions genètiques. Segurament, la tècnica més popular, i emprada per obtenir els resultats d'aquesta tesi doctoral, ha estat l'anomenada PCR (de l'anglès *Polymerase Chain Reaction*), que permet obtenir moltes còpies d'una regió genòmica per la seva posterior determinació genètica. Tot i així, han estat les tècniques de seqüenciació les que han permès tenir la millor visió general, fins al moment, del nombre total de polimorfismes genètics i de la seva combinació en el genoma.

A principis del anys 90 del segle passat, es va engegar el projecte internacional Genoma Humà (*Human Genome Project*) per tal d'obtenir les primeres seqüències completes del genoma humà [28,29]. Es va trigar més de 10 anys en obtenir el primer esborrany del genoma complert d'un individu (Figura 5), i cap a 15 per obtenir la versió definitiva. Ara sabem que el genoma humà haploide conté 3.000 milions (3×10^9) de parells de bases, i que els aproximadament 25.000 gens identificats n'ocupen tan sols el 2%. A mesura que el projecte Genoma Humà avançava, les tècniques de seqüenciació progressaven radicalment. De manera que a principis del segle XXI, un cop obtingut un genoma sencer de referència i amb les noves tècniques de seqüenciació, es va proposar el projecte 1000 Genomes.

El projecte 1000 Genomes (Figura 6) consisteix en seqüenciar milers d'individus d'arreu del món per tal d'obtenir una referència de la quantitat i tipus de variació genètica present en la espècie humana i el patró geogràfic global que presenta. En l'actualitat, i en el marc d'aquest projecte, s'han descrit 1092 seqüències complertes d'individus d'arreu del món i en els

pròxims anys se n'esperen unes 1500 més [30].



Figura 5 – Portades de les revistes Science i Nature publicant els resultats de Projecte Genoma Humà

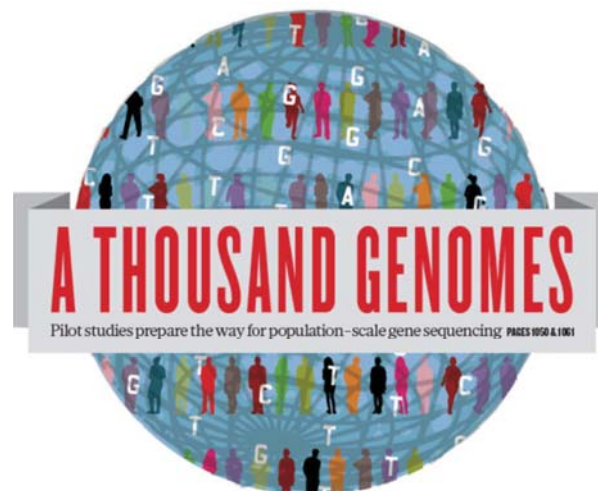


Figura 6 – Portada de Nature publicant els resultats del Projecte 1000 Genomes

- *Polimorfismes d'una sola base o SNPs* (de l'anglès *Single Nucleotide Polymorphism*).

Representen la variació genètica més comú, regularment distribuïda al llarg de tot el genoma, i es caracteritzen per la substitució d'una base de la

seqüència per una altra. Aquest tipus de variació genètica ha estat utilitzada en l'actual tesi doctoral.

Degut a les seves característiques d'estructura química, les bases T i C són anomenades pirimidines, mentre que les A i G s'anomenen purines. Quan una base és substituïda per una de la seva mateixa condició, aquest canvi s'anomena transició, mentre que quan és substituïda per una altre, la substitució s'anomena transversió. Independentment de la seva classificació aquests polimorfismes són bàsicament bial·lèlics, presentant una de les dues formes possibles (Figura 7).

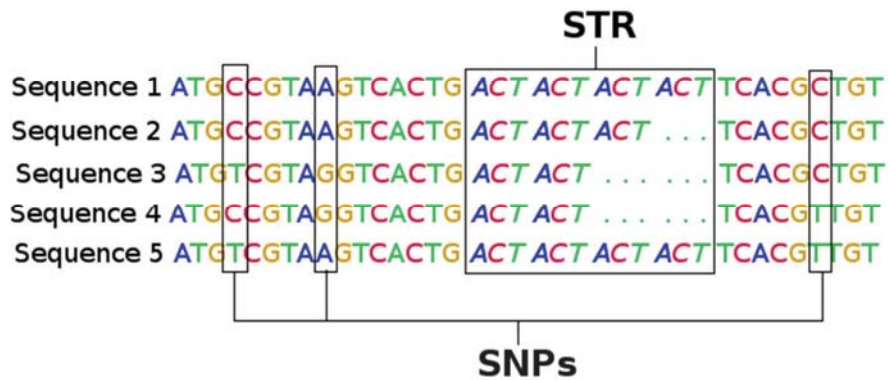


Figura 7 – Exemples de variació en la seqüència dels tipus de polimorfismes estudiats en aquesta tesi doctoral, SNPs i STRs.

Fins a l'actualitat, s'han descrit un total de 38 milions de polimorfismes SNP, amb una mitjana de 4 milions d'SNPs per individu [30]. La majoria es troben en regions no codificants (sense gens), però els que es troben en les seqüències codificants poden afectar la seva expressió i/o alterar les característiques de l'element estructural o funcional (proteïna) pel qual codifiquen. Polimorfismes que no alteren el producte del gen s'anomenen *sinònims*, mentre que els que modifiquen aminoàcids de la

proteïna resultant s'anomenen *no-sinònims*. Els SNPs *sense sentit* produeixen una interrupció en la formació de la proteïna, i els que es troben en regions reguladores del gen poden alterar-ne l'expressió. S'estima que de mitjana cada individu és portador d'entre 50 i 100 polimorfismes SNPs que causarien malalties hereditàries si es trobessin en les dues còpies del gen.

- ***Repeticions en tàndem de nombre variable o VNTR*** (de l'anglès *Variable Number Tandem Repeat*)

Aquests polimorfismes genètics es caracteritzen pel nombre de repeticions en què s'observa la seqüència o motiu de repetició. Segons si el motiu de repetició és de 2 a 10 parells de bases s'anomenen *microsatèl·lits* o STRs (de l'anglès *Short Tandem Repeats*). Si la seqüència repetida és de 10 a 100 parells de bases, llavors s'anomenen *minisatèl·lits*. Aquests marcadors genètics es troben àmpliament distribuïts pel genoma i són deguts a errors en la replicació del DNA. La característica més especial que presenten els polimorfismes VNTR és que generalment són multial·lèlics, ja que el motiu de la seqüència es pot trobar repetit des de 2 fins a diverses desenes de vegades (Figura 7). Cada nombre de repeticions observat designa un al·lel diferent pel mateix polimorfisme. Degut al seu elevat nombre d'al·lels, els microsatèl·lits o STRs són el tipus de polimorfismes utilitzats en identificació de mostres forenses, ja que el perfil genètic resultant és únic per cada individu, mentre que dos individus presentaran diferents graus de similitud segons el tipus de parentiu entre els dos.

En aquesta tesi doctoral s'han estudiat polimorfismes STR que presenten com a motiu de repetició seqüències de tres, quatre i cinc bases. Així, són considerats tri-, tetra- i penta-nucleòtids, respectivament. La majoria d'aquest elements es consideren neutrals ja que no es troben en

zones estrictament codificants, però poden trobar-se implicats en la regulació de l'expressió gènica i causar variacions fenotípiques.

- ***Insercions/Delecions***

Aquestes variacions genètiques comporten la inserció o deleció (INDEL) de seqüència i es classifiquen segons la mida de la seqüència inserida o suprimida. També són bial·lèlics ja que es poden presentar o descriure com a presència o absència de la seqüència determinada.

Petits INDELS poden variar des de 1 base fins a pocs centenars de bases. Aquests són majoritàriament deguts a errors de replicació i poden comportar alteracions en l'expressió, estructura o funció dels productes gènics. Hi ha descrits fins a 1.4 milions de petits INDELS, amb una mitjana de 340 mil per individu.

Els elements inserits més grans en mida, de 300 a 10.000 parelles de bases, s'anomenen *elements mòbils*. Tenen la peculiaritat que poden “saltar” de la regió on es troben i ubicar-se inclús en un altre cromosoma. Aquests elements són molt diversos i es classifiquen segons la mida i la manera de “transposar-se”. S'estima que el 45% de tota la seqüència genòmica prové d'algun d'aquests elements mòbils tot i que només 140 mil elements mòbils presenten variació entre individus, amb una mitjana de 700 insercions/delecions per individu.

Tot i els diferents tipus i nombre de diferències genètiques en l'espècie humana, és important destacar que les seqüències de DNA de totes les poblacions humanes són un 99.8% idèntiques entre dos individus qualsevol [30]. Però cal recordar que, si bé les diferències genòmiques entre individus són petites, aquestes són estadísticament significatives.

1.4 Variació genètica i malaltia: Epidemiologia genètica

Segons la *International Epidemiology Association*; “*Epidemiology: The study of the occurrence and distribution of health-related states or events in specified populations, including the study of the determinants influencing such states, and the application of this knowledge to control the health problems.*” [31]. Per tant, l'epidemiologia es dedica a l'estudi de la distribució de les malalties, les seves freqüències (incidència i prevalença), les causes i el control dels factors relacionats. A més a més, l'epidemiologia es considera la ciència bàsica per a la medicina preventiva i una font d'informació per a la formulació de polítiques de salut pública.

Conseqüentment, l'epidemiologia genètica enfoca el seu estudi en la identificació dels factors genètics relacionats amb les malalties i n'estudia la seva distribució poblacional i la seva interacció amb la resta de factors implicats.

1.4.1 Identificació dels factors de risc genètics

Un cop obtingudes evidències d'un probable component genètic com a causant d'una malaltia complexa, el següent pas és la identificació i localització dels gens i variacions genètiques causants.

L'estratègia bàsica utilitzada, i emprada en aquesta tesi doctoral, és l'estudi de gens candidats per implicació en l'etiologia de la malaltia. Recentment, s'ha aplicat la estratègia d'analitzar variants genètiques distribuïdes al llarg de tot el genoma, independentment de si es troben en regions amb gens rellevants o no. Els estudis on s'aplica aquesta estratègia s'anomenen GWAS (de l'anglès *Genome-Wide Association Study*).

Independentment de l'estratègia, la metodologia de detecció d'una variant genètica relacionada amb la malaltia en una població és anomenada

estudi d'associació. Aquests estudis poblacionals d'associació indiquen que aquella variant genètica és la que contribueix directament a modificar el fenotip o, més habitualment, es troba en lligament amb la variant realment causant de la contribució a la malaltia.

Existeixen diferents tipus de dissenys i tests d'associació genètica, però ens centrarem en els més comuns i alhora utilitzats en aquesta tesi doctoral:

- **Estudi Cas - Control**

Aquest disseny basat en la població general consisteix en la recollida d'una mostra d'individus afectats (casos) i no afectats (controls) per la malaltia. Els estudis cas-control són els més utilitzats en epidemiologia genètica per la relativa facilitat de recollir mostres de pacients afectats i població general de control. Per tal d'equiparar al màxim el grup d'afectats i el grup control i obtenir individus amb característiques similars únicament diferenciats en l'estat de la malaltia, les mostres solen recollir-se aparellades. De manera que les mostres aparellades reflecteixen proporcions similars d'edat, sexe i altres característiques que es tinguin en compte. Aquest tipus de disseny estima els efectes de les variants genètiques sobre el fenotip malaltia a través de la diferent proporció en què les variants genètiques són observades en el grup de casos i el grup de controls. Aquesta estimació de l'efecte genètic sobre el fenotip ve determinat pel paràmetre *Odds Ratio* (OR) i quantifica la probabilitat d'ocurrència del fenotip o malaltia en els portadors del factor genètic respecte la població general.

	Casos	Controls	
Presència de al·lel de risc	A	B	% presència de al·lel de risc general
Absència de al·lel de risc	C	D	% absència de al·lel de risc general
	% total de casos	% total de controls	1

On *Odds Ratio*; $OR = (A/B) / (C/D)$

El principal problema que presenta aquest disseny és que, si casos i controls no han estat adequadament aparellats, les diferents freqüències genètiques entre els dos grups poden ser degudes a altres factors no tinguts en compte durant l'anàlisi i no a la malaltia. Aquest factors que generen confusió s'anomenen d'estructuració poblacional ja que structuren la població en subgrups amb possibles diferències genètiques que detectaríem en l'anàlisi, però sense estar relacionades amb la malaltia. El factor de confusió més habitual sol ser l'estructuració genètica de la població (el concepte estructuració genètica serà detallat dins el següent apartat 1.5 d'aquesta secció d'introducció).

- **Test de desequilibri de la transmissió (TDT)**

Dins dels dissenys basats en famílies es troba el test de desequilibri de la transmissió al·lèlica o TDT. Els tests basats en famílies tenen la gran avantatge que no es veuen afectats per l'estructuració poblacional, ja que la unitat d'anàlisi no és la mostra general sinó cada mostra familiar individualment. En aquest cas, les mostres són trios d'individus afectats de la malaltia i els seus progenitors. Aquest test s'anomena de desequilibri de

transmissió, ja que davant d'una transmissió genètica a l'atzar d'un pare/mare heterozigot (amb dos variants o al·lels diferents en un polimorfisme), s'espera que cada al·lel sigui transmès a la següent generació amb una probabilitat del 50%. Quan s'observin proporcions de transmissions d'al·lels significativament diferents al 50% en la mostra seran indicatius de transmissió no aleatòria entre progenitors i fills afectats de la malaltia. En la figura 8 pot veure-se'n un exemple, on en una mostra de 5 trios pels quals s'estudia un marcador genètic amb 3 al·lels diferents (A, B i C) l'al·lel A és transmès als fills afectats en 4 ocasions de les 5 possibles, sent major que les 2 o 3 vegades que s'esperaria que s'observés transmissió per atzar.

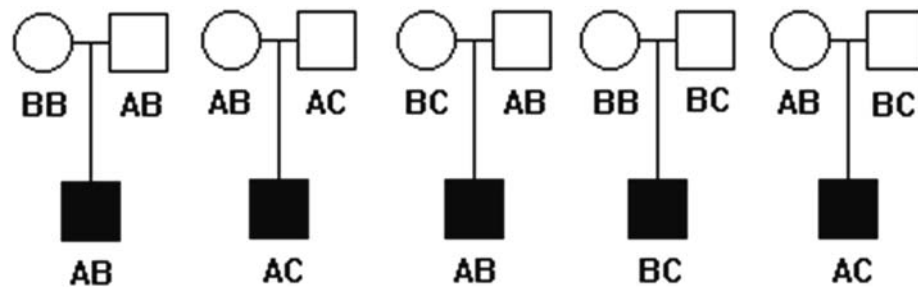


Figura 8 - Exemple de TDT o test de desequilibri de transmissió (<http://www.smd.qmul.ac.uk/statgen/dcurtis/lectures/assoc5.html>).

A data d'avui, centenars de variants genètiques han estat identificades com a associades a la malaltia cardiovascular. Aquestes variacions genètiques s'han trobat tant en gens implicats en mecanismes fisiològics relacionats, com en regions genètiques aparentment no funcionals. Tot i així, el principal problema que presenten aquests estudis d'associació és la falta de replicació de resultats en diferents poblacions, cosa que dificulta la interpretació dels resultats.

1.4.2 Estimació de risc individual a la malaltia cardiovascular

La identificació dels factors de risc tradicionals (TRF de l'anglès *Traditional Risk Factors*) ha millorat el grau de detecció d'individus amb elevat risc cardiovascular. Entre aquests factors tradicionals es troben els no modificables (com el sexe i l'edat) i els modificables. Els factors de risc modificables són els que acaben expressant-se quan es tenen mals hàbits de vida (sobrepès, hipertensió, hiperglucèmia, colesterolèmia, etc.). Ja que una correcció en els hàbits (dieta, exercici, consum reduït d'alcohol, pràctica de la paciència, etc..) genera una disminució dels nivells de risc. Diferents consorcis internacionals de recerca han contribuït a desenvolupar taules d'estimació del risc basats en els TRF, tal com el *Framingham Risk Score* [32], el *Reynolds Risk Score* [33], el *Prospective Cardiovascular Munster Heart Study* (PROCAM) [34] i el *Systematic Coronary Risk Evaluation* (SCORE) [35]. Aquestes taules d'estimació del risc tenen una habilitat de predicció entre moderada i bona [6]. Tot i així, al voltant del 15-20% dels afectats cardiovasculars són mal classificats com a individus de baix risc per aquestes taules basades en els TRF [7].

La presència del component genètic en aquesta malaltia ens dóna una bona hipòtesi de base de perquè individus sense símptomes previs acaben mostrant la malaltia, i perquè individus amb els mateixos hàbits de vida presenten diferents nivells dels anomenats TRF. A mesura que s'han anat identificant factors genètics de risc, la voluntat de millorar les taules de predicció de risc ha anat en augment. Però, diferents causes no han permès encara la millora d'aquestes taules de risc. Els principals factors que ho dificulten són la petita contribució individual a la malaltia que mostren les variants genètiques de risc (<1%) i la falta de replicació de resultats en les diferents poblacions [36–38]. Inclús les puntuacions de risc genètiques (GRS

de l'anglès *Genetic Risk Score*) calculades a partir de desenes o un centenar de variants genètiques de risc no han estat del tot satisfactòries ja que milloraven la classificació entre malalts i no malalts en algunes poblacions però no en d'altres [39–48].

La porció del fenotip que roman inexplicada és anomenada “*missing heritability*” o heretabilitat perduda o amagada, ja que molt probablement és deguda a la porció del component genètic encara no identificat [36,37]. Aquesta dificultat en la identificació de tots els factors genètics implicats és deguda a la petita contribució de cada un d'ells i/o al fet que les causes genètiques siguin de freqüència molt baixa en la població general i diferents entre els malalts. Ambdós casos només són detectables estadísticament amb mostres epidemiològiques molt grans i donen una visió d'un fenotip molt complex de determinar i predir [8,49].

1.4.3 Distribució poblacional de la malaltia cardiovascular i factors de risc

Les malalties cardiovasculars presenten una prevalença (presència) en la població bastant considerable del 6-10% segons la regió geogràfica continental. Ja des dels anys 50 del segle passat, estudis epidemiològics a Europa han destacat diferències geogràfiques en la incidència (aparició) de les malalties coronàries. Aquestes diferències es basen en un gradient on els valors més baixos es troben al sud-oest del continent (Espanya, França) i incrementen fins a valors quatre vegades superiors al nord/nord-est (Regne Unit, Finlàndia) [50–52] (Figura 9).

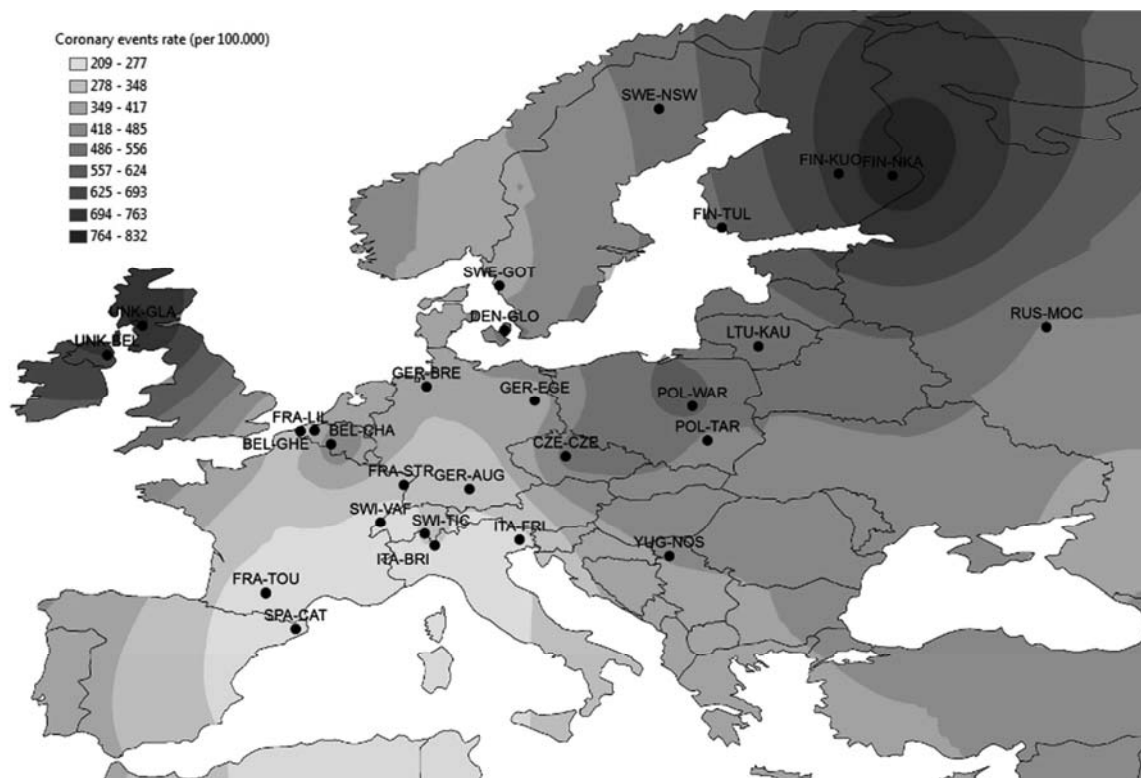


Figura 9 – Distribució de la incidència cardiovascular estimada pel projecte MONICA de la OMS. La incidència cardiovascular és determinada pel nombre d'accidents coronaris per cada 100000 habitants [51].

Alhora, aquests estudis epidemiològics han estimat la distribució poblacional dels factors de risc tradicionals per tal d'observar si expliquen el gradient de mortalitat coronària. Probablement, la major contribució en aquest camp d'estudi l'ha aportat el projecte internacional MONICA de la OMS [51], on es va estudiar l'evolució de la malaltia cardiovascular i de factors de risc associats al llarg d'uns 20 anys en un total de 38 poblacions en 21 països del continent europeu. En aquest projecte, es van registrar els casos mortals i no mortals d'infart de miocardi, i els nivells mitjans de colesterol total, pressió sanguínia sistòlica, índex de massa corporal i tabaquisme en les poblacions estudiades. Aquest i altres projectes han valorat la contribució de

la distribució poblacional dels factors de risc clàssics en un 30-40% del total de variació observada en malaltia coronària. Al mateix temps, es va observar una major incidència cardiovascular al Regne Unit i Finlàndia en comparació amb França que no es corresponia amb una diferent proporció de factors de risc, ja que aquestes poblacions presentaven nivells similars, especialment de consum de greixos d'origen animal [53–55]. Aquesta contradicció ecològica es va anomenar paradoxa francesa, sud europea o mediterrània. Però més recentment, s'ha proposat com a factor que podria explicar aquesta paradoxa geogràfica el consum de diferents tipus d'alcohol, especialment el vi, que afavoriria les poblacions mediterrànies de patir malalties cardiovasculars [56].

El fet que les distribucions poblacionals dels factors de risc tradicionals no expliquin la totalitat de les diferències geogràfiques en malaltia coronària suggereix que les distribucions poblacionals de factors de risc genètic podrien ajudar a explicar-les. Poques aproximacions s'han realitzat en aquest àmbit. A data d'avui, un parell d'estudis han intentat correlacionar les freqüències poblacionals d'algunes variants genètiques identificades com a risc amb la incidència cardiovascular al continent europeu. Únicament s'ha trobat correlacionada positivament la isoforma E4 del gen *APOE*. Per contra, altres variants de risc es van trobar correlacionades negativament. Aquests resultats van estendre la paradoxa mediterrània al component genètic, ja que poblacions amb freqüències més baixes de variants genètiques de risc presentaven incidències coronàries més elevades [57,58].

Probablement, una alternativa més realista per tal d'estimar la contribució de la distribució genètica de risc a la incidència cardiovascular poblacional seria utilitzar *scores* de risc genètic (GRS). En aquesta tesi

doctoral, es proposa l'ús dels GRS per tal de poder aglutinar en un sol paràmetre tota la informació genètica que es vulgui analitzar, de la mateixa manera que s'utilitza en la valoració del risc a nivell individual. D'aquesta manera en cada població estudiada es podria calcular el GRS de cada individu i fer-ne una mitja poblacional, que podria ser comparada entre poblacions. Aquests GRS poblacionals no dependrien de l'efecte d'una sola variant genètica i la seva distribució poblacional podria ser comparada amb la distribució de la malaltia.

Pel que sabem a dia d'avui, cap estudi epidemiològic previ ha analitzat els GRS com a avaluadors del risc genètic d'una població a patir malalties cardiovascular.

1.5 Genètica de Poblacions Humanes

La genètica de poblacions humanes és l'estudi de les diferències genètiques entre individus humans d'una mateixa població i/o diferents poblacions. Generalment, el patró de variació entre individus sol ser similar dins d'una mateixa població i sol presentar diferències similars entre individus de dos poblacions diferents. Per tant, la similitud en la variació entre individus d'una mateixa població es pot considerar com el patró de variació representatiu d'aquella població. D'aquesta manera, la genètica de poblacions pot mesurar el patró de variació entre poblacions característic per cada regió genòmica o gen, i la seva comparació amb d'altres regions o gens ens pot informar de l'efecte de les forces evolutives, com la potencial selecció, sobre aquests gens. Tanmateix, l'anàlisi comparatiu del patró global o mitjà de variació genòmica permet obtenir inferències demogràfiques, es a dir, la reconstrucció de la història genètica de les poblacions humanes (migracions, expansions, colonitzacions) [59,60].

Un dels objectius de la recerca realitzada en aquesta tesi és reivindicar la contribució que la genètica de poblacions pot realitzar en el camp de l'epidemiologia genètica a l'hora de descriure la distribució poblacional dels factors genètics de risc per malalties complexes, determinar la seva contribució a la incidència poblacional i estimar els processos evolutius i/o moviments poblacionals que els han modelat.

1.5.1 Distribució de la diversitat genètica en les poblacions humanes

Cada polimorfisme genètic pot ser descrit en una població segons la diversitat en què és observat, o sia el nombre d'al·lels que presenta i la freqüència en que s'hi troben. Per exemple, un SNP (polimorfisme bial·lèlic) que presenta els al·lels A i G, es pot trobar en unes determinades freqüències

de A i G (p.e. A 30% i G 70%), que representen la probabilitat d'observar els al·lels en una població, i són generalment simbolitzades algebraicament com a p i q :

$$p(A) = p = 0.3$$

$$p(G) = q = 1 - p = 0.7$$

Però els al·lels no són observats directament de forma individual en les poblacions, sinó que són observats en genotips, combinacions de dos al·lels ja que la majoria de cromosomes es troben aparellats i existeixen dues còpies de la seqüència (una provinent de cada progenitor). Els genotips poden ser homozigots (si presenten la mateixa variant genètica en les dues còpies) o heterozigots (si presenten diferents variants genètiques). Per tant, les freqüències al·lèliques venen determinades per les freqüències genotípiques, o sia, el nombre de vegades que un al·lel és observat (dues en cada homozigot i una en cada heterozigot) comparat amb el total d'al·lels a la població (dues vegades el nombre de genotips totals) :

Si en una població: $N_{AA} = 9$, $N_{AG} = 42$ i $N_{GG} = 49$, llavors $N_{total} = 100$

$$p(A) = p = (2N_{AA} + N_{AG}) / 2N_{total} = p(AA) + p(AG)/2$$

$$p(A) = p = (2*9 + 42) / 2*100 = 0.09 + 0.42/2 = 0.09 + 0.21 = 0.3$$

$$p(G) = q = (2N_{GG} + N_{AG}) / 2N_{total} = p(GG) + p(AG)/2$$

$$p(G) = q = (2*49 + 42) / 2*100 = 0.49 + 0.42/2 = 0.49 + 0.21 = 0.7$$

Aquesta propietat que s'observa de determinar les freqüències al·lèliques a partir de les genotípiques, també es dona en sentit contrari (determinar les freqüències genotípiques a partir de les al·lèliques) si es compleixen una sèrie de requisits en la població sota l'anomenat equilibri de

Hardy-Weinberg (HW). La llei de Hardy-Weinberg prediu que en una població, amb una mida suficientment gran on l'aparellament entre individus és aleatori i en absència de selecció, migració i mutació, les freqüències genotípiques es poden derivar de les freqüències al·lèliques i es mantenen constants de generació en generació. A més a més, en una població que no es trobi en equilibri HW, en una generació de panmíxia es restauren les condicions d'equilibri. Una manera intuïtiva de visualitzar la situació d'equilibri HW seria la amb la següent taula de contingència on es descriuria la combinació aleatòria de gàmetes o al·lells independentment dels genotips:

	p(A)	p(G)
p(A)	p(A)p(A)	p(A)p(G)
p(G)	p(G)p(A)	p(G)p(G)

On les freqüències genotípiques d'una població en equilibri de HW són descrites de la següent manera:

$$p(AA) = p(A)p(A) = p(A)^2$$

$$p(AA) = 0.3 \cdot 0.3 = 0.3^2 = 0.09$$

$$p(AG) = p(A)p(G) + p(G)p(A) = 2p(A)p(G)$$

$$p(AG) = 0.3 \cdot 0.7 + 0.7 \cdot 0.3 = 2 \cdot 0.3 \cdot 0.7 = 0.42$$

$$p(GG) = p(G)p(G) = p(G)^2$$

$$p(GG) = 0.7 \cdot 0.7 = 0.7^2 = 0.49$$

Una mesura global de la diversitat d'una població és l'anomenada heterozigozitat, que representa la proporció mitjana d'heterozigots per tots els polimorfismes analitzats. Aquesta mesura d'heterozigozitat informa de la diversitat genètica i és comparable entre poblacions.

1.5.1.1 Processos de diferenciació genètica: deriva genètica, selecció i migració

Els canvis en les freqüències al·lèliques dels polimorfismes ja existents en les poblacions biològiques són deguts a forces o processos evolutius lligats als moviments demogràfics com la selecció, la deriva genètica o la influència migratòria.

- **Deriva genètica**

La deriva genètica és la diferenciació poblacional no direccional per efecte de mostratge, és a dir, a la generació següent només trobem una porció de la variació genètica de la generació anterior. L'efecte de la deriva és particularment evident en poblacions un nombre reduït d'individus. Una situació que afavoreix la deriva és el cas de l'efecte fundador, on a partir d'un subgrup poblacional molt reduït, que freqüentment no és representatiu de la població parental, es genera una nova població. La probabilitat de fixació genètica (pèrdua d'un dels dos al·lèls, en el cas de polimorfismes bial·lèlics, i per tan de diversitat) és molt elevada. La conseqüència a nivell poblacional és la reducció de la diversitat genètica intra-poblacional i l'increment de la diversitat inter-poblacional.

- **Selecció**

La selecció és la diferenciació poblacional dirigida com a conseqüència d'una adaptació al medi. La selecció en les poblacions ve donada pel fet que en una població els diferents genotips presenten unes taxes de supervivència i/o reproducció diferencials. En els processos selectius, un grup d'individus és capaç de generar més descendència, reproduint-se amb més freqüència o generant descendència més fèrtil o amb

menys mortalitat associada al medi, que un altre grup d'individus. La selecció actua directament sobre els fenotips (i només indirectament sobre els genotips associats a aquests fenotips) i en un context ambiental determinat. El resultat és el canvi en la composició genètica de les successives generacions. La probabilitat de fixació genètica depèn de la magnitud de l'efecte selectiu.

Un aspecte particular i important en l'espècie humana d'adaptació al medi serien els processos de colonització de nous territoris i els canvis produïts en l'economia de subsistència degut a les adquisicions culturals. Un cas relativament recent seria l'adquisició de l'agricultura durant el Neolític. L'augment de producció de recursos alimentaris va augmentar la possibilitat de supervivència i va permetre la presència de poblacions amb un nombre més elevat d'individus, i també comportà un contacte constant amb els animals. Aquests factors van tenir importants conseqüències en la incidència de malalties infeccioses ja que van facilitar-ne la transferència de paràsits d'animals a humans. Entre les implicacions genètiques d'aquesta adaptació hi trobem la persistència de la lactasa (capacitat de digerir llet i làctics en edat adulta) i segurament d'altres adquisicions immunològiques.

- **Migració**

Els moviments migratoris en totes les seves modalitats (continus o puntuals, flux genètic o moviments en massa) han contribuït a modular la composició genètica de les poblacions provocant-ne la diferenciació. Les seves conseqüències són l'augment de la diversitat genètica intra-poblacional i la disminució de la diversitat inter-poblacional. En els humans aquest ha estat un dels fenòmens més importants alhora de descriure el paisatge genètic que observem avui dia ja que, un cop colonitzats tots els territoris,

les diferents revolucions culturals (agricultura, era dels metalls, expansions de civilitzacions, etc..) han generat uns fluxos genètics entre poblacions completament allunyades geogràficament incomparables amb cap altre espècie biològica.

1.5.1.2 Diversitat genètica en humans

A finals dels segle passat, per tal d'explorar la magnitud de la diversitat genètica en humans i la seva distribució, es va originar el projecte Diversitat del Genoma Humà (HGDP de l'anglès *Human Genome Diversity Project*) on es van col·leccionar mostres de DNA de 1050 individus procedents de 52 grups ètnics diferents distribuïts pels cinc continents [61], mostres sobre els que s'han anat analitzant diferent nombre i tipus de polimorfismes.

Un dels primers resultats a destacar que van sorgir d'aquest projecte és la relació inversament proporcional entre la diversitat genètica intrapoblacional (heterozigositat) i la distància de cada població a l'est d'Àfrica [62] (Figura 10). Aquesta relació observada va ser interpretada com una confirmació més de la hipòtesi "Origen Africà" (*Out of Africa* en anglès), on l'espècie humana anatómicament moderna es va originar a l'est d'Àfrica central fa 150.000 anys i, a través de successives migracions, acabà poblant tota la superfície planetària. Seguint la hipòtesi prevalent, coneguda com a efecte fundador en sèrie, en el curs de l'expansió humana sobre el planeta, subgrups de població abandonaven les poblacions parentals africanes per tal d'establir nous assentaments en territoris encara no colonitzats. Aquests subgrups de poblacions eren portadors només d'una part de la diversitat genètica total de la colònia parental, produint una reducció progressiva de la diversitat genètica poblacional a mesura que incrementava

la distància geogràfica a l'origen de la diversitat. En la figura 11 es pot observar la localització geogràfica de les 52 mostres poblacionals dels *Human Genome Diversity Project*, alhora que s'observen rutes migratòries estimades en la colonització dels continents.

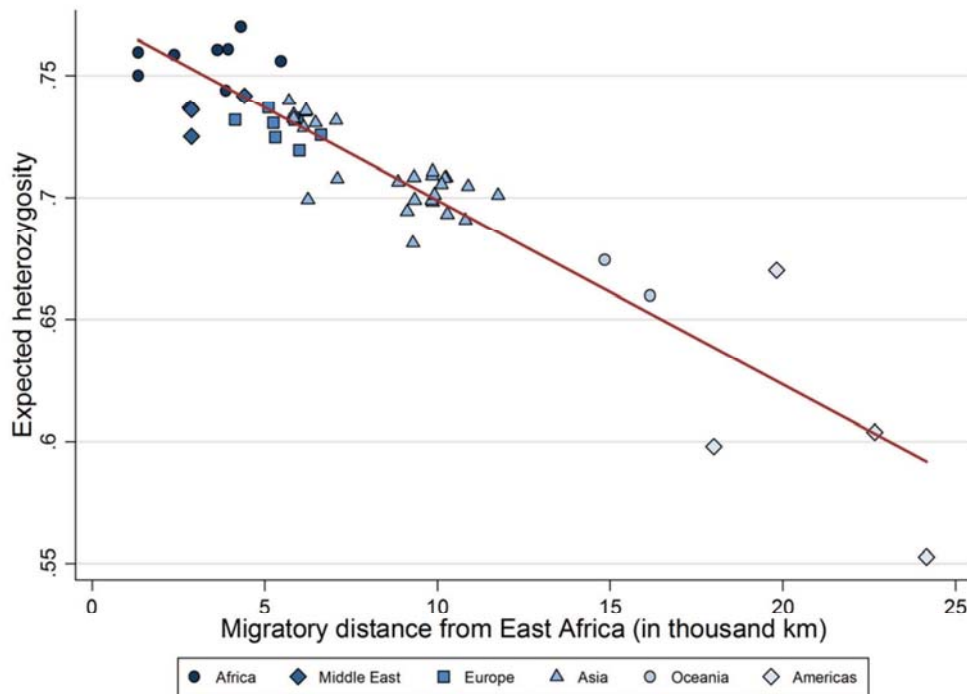


Figura 10- Relació entre heterozigositat (diversitat genètica) esperada i distància geogràfica a l'Est d'Àfrica central de les 52 poblacions del projecte HGD [63].

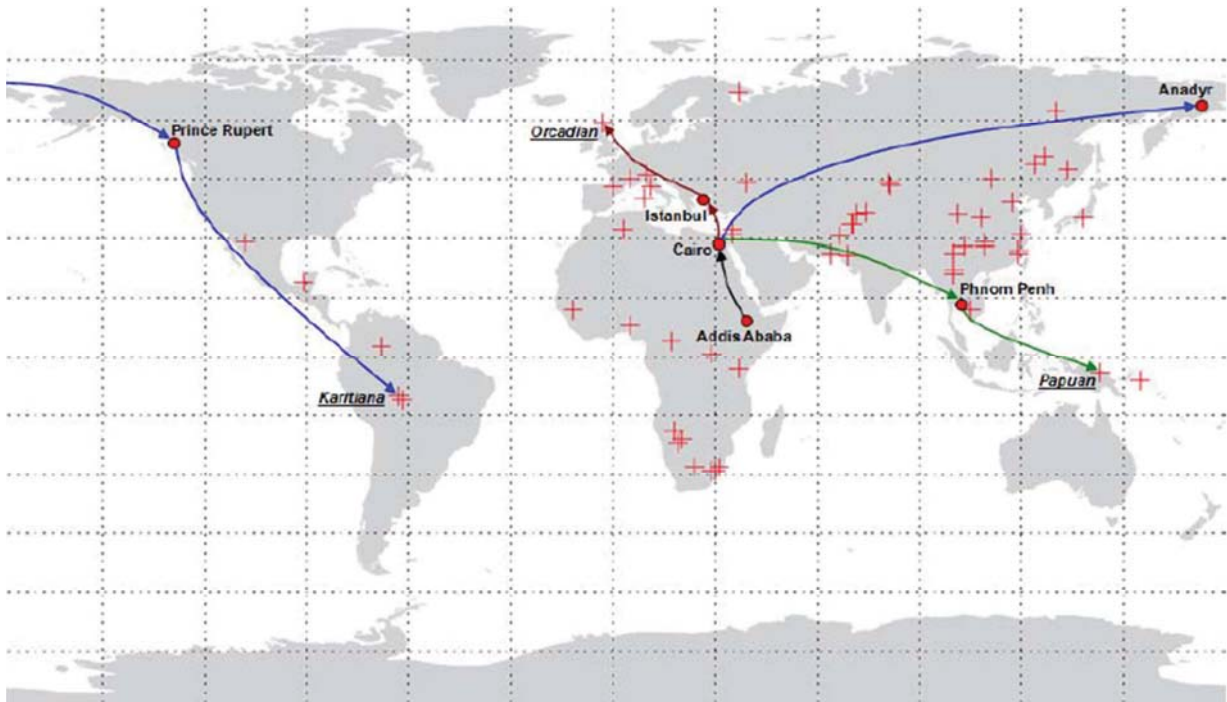


Figura 11 - Distribució geogràfica dels 52 grups ètnics que inclosos al projecte HGD i rutes migratòries estimades en la colonització dels continents [63].

1.5.1.3 Mesures de distribució i estructuració genètica

Comparant genèticament diferents poblacions, independentment de si són més o menys diverses genèticament, les freqüències dels polimorfismes estudiats poden ser bastant similars o bastant diferents. En comparació amb l'exemple mencionat anteriorment, en una altra població les freqüències dels dos al·lels exemple poden ser diferents (A 60% i G 40%). Aquesta diferència entre poblacions en la freqüència dels polimorfismes acaba dibuixant una distribució de les freqüències genètiques a través de la geografia poblacional que pot ser mesurada i representada de diverses formes, donant una idea de les relacions genètiques poblacionals i de la possible estructuració genètica.

Una de les mesures més comuns de diferenciació o estructuració genètica entre poblacions és l'anomenat índex de fixació o F_{ST} . Aquest índex

és una mesura de la variància de les freqüències al·lèliques entre poblacions o subgrups de poblacions, és a dir, és una estimació de la diferenciació genètica deguda a les diferències entre poblacions.

$$F_{ST} = \text{Variància entre poblacions } (S) / \text{variància total } (T)$$

Aquest índex varia entre 0 i 1, on 0 significaria absència de diferenciació genètica entre les poblacions analitzades. La resta de proporció de la variància total és l'atribuïda a diferències genètiques entre individus dins de les poblacions. A la vegada, la variància entre les poblacions (F_{ST}) pot ser descomposta o jerarquitzada en variància entre diferents grups de poblacions (ex. entre continents) (F_{SC}) i/o dins d'un continent (F_{CT})

$$F_{ST} = F_{SC} + F_{CT}$$

La magnitud d'aquestes variàncies pot ser testada estadísticament, per tal de comprovar si la diferenciació genètica observada és significativament diferent a una diferenciació genètica nul·la (H_0) entre poblacions, dins dels continents i entre continents.

Per tal d'observar gràficament les diferències poblacionals, es pot representar la matriu de distàncies genètiques F_{ST} entre cada un dels individus o poblacions. Aquesta matriu de distàncies genètiques és projectada en pocs eixos de variació o dimensions (normalment dos) mitjançant un anàlisi d'escalat multidimensional (MDS de l'anglès *Multidimensional Scaling*). L'anàlisi MDS assigna a cada població o individu unes coordenades dels eixos o dimensions estimades que poden ser representades gràficament en uns eixos de coordenades. Un altre mètode pel

qual s'obté una representació similar de semblança genètica entre poblacions o individus és l'anàlisi de components principals (PCA, de l'anglès *Principal Component Analysis*), on la variació genètica observada es reduïda a poques components que poden ser representades també en un eix de coordenades [64].

Finalment, una altra anàlisi molt emprada per tal d'estimar l'estructuració genètica poblacional és l'estimació del nombre de grups o "clusters" genètics a partir dels quals es podria haver generat tota la diversitat trobada i la seva proporció en els individus estudiats ignorant inicialment la població de procedència. Aquest mètode estima un nombre de grups, clústers o "pools" genètics que poden ser interpretats com els originals o ancestrals dels quals els individus actuals en són originaris. També n'estima la proporció de cadascun d'ells en els individus proporcionant una idea del component genètic dels individus en funció de les porcions relatives dels components ancestrals [65]. Els individus són posteriorment agrupats en poblacions i s'observa com els individus de cada població comparteixen proporcions similars de components ancestrals. D'aquesta manera es pot identificar poblacions representatives de "pools" genètics ancestrals o originaris i poblacions fruit de la barreja poblacional amb proporcions de components genètics característics.

1.5.1.4 Distribució i estructuració genètica global i del continent Europeu

Gràcies a projectes on s'han genotipat diferent nombre i tipus de polimorfismes en poblacions d'arreu del món, entre ells el projecte *Human Genome Diversity*, s'ha estimat que la diferenciació genètica entre individus de diferents poblacions a nivell mundial és considerablement baixa. Com es

pot observar a la taula 2, la magnitud de diferenciació genètica entre poblacions de tots els continents es troba al voltant del 10%, el que significa que el 90% de tota la variabilitat genètica entre individus és observada dins de les poblacions. Dins d'aquest 10% de variació entre poblacions, les diferències genètiques més grans es troben entre grups de poblacions continentals (~8%), mentre que la variació entre poblacions d'un mateix continent és considerablement baixa (~2%). Tot i que aquestes magnituds de diferenciació genètica entre poblacions presenten valors baixos, són estadísticament significatives.

Poblacions	Polimorfismes	Dins poblacions	Entre poblacions dins continents	Entre continents
Tots els continents	377 STRs	94.1	2.4	3.6
	100 ALUs	88.6	1.9	9.6
	650K SNPs	88.9	2.1	9
Europa	377 STRs	99.3	0.7	
	100 ALUs	99	1	
Àfrica	377 STRs	96.9	3.1	
	100 ALUs	95.8	4.2	
Àsia	377 STRs	98.7	1.3	
	100 ALUs	97.7	2.3	
Oceania	377 STRs	93.6	6.4	
Amèrica	377 STRs	88.4	11.6	

Taula 2 - Diferenciació genètica poblacional (%) a partir de l'anàlisi molecular de la variància (AMOVA) [66–68].

Com es pot observar en els valors de diferenciació genètica entre poblacions d'un mateix continent, Europa i Àsia són els continents on la diferenciació genètica entre poblacions és més baixa (1-2%). Els valors més

elevats que presenta el continent africà (3-4%) s'han relacionat amb la possible major antiguitat d'aquestes poblacions. En l'altre extrem, els valors més alts de diferenciació entre poblacions d'un mateix continent es presenten en Oceania i Amèrica (6-11%), degut a una diferenciació poblacional per reducció de la diversitat durant el procés colonitzador amb un posterior aïllament genètic.

A Europa, tot i la poca diferenciació genètica poblacional, aquesta diferenciació és significativa i pot arribar a agrupar els individus segons la regió geogràfica reflectint d'aquesta manera la geografia europea (Figura 12). La representació observada en la figura 12 correspon a una anàlisi de components principals (PCA) de la variació genètica observada en 500,000 SNPs de 3192 individus europeus del projecte *Population Reference Sample* (POPRES) [69]. La variació genètica dels individus és reduïda a dos dimensions i els individus queden agrupats segons la seva semblança genètica, que ahora es correspon geogràficament.

La gran part de variació genètica observada a nivell de tot el genoma presenta un patró poblacional selectivament neutre [70]. Per tant, podem afirmar que les causes principals de diferenciació genètica entre poblacions humanes han estat la mutació i la deriva genètica. La selecció ha contribuït en processos i regions genètiques puntuals, però forma part de l'excepció dins del genoma. Aquests processos de diferenciació selectivament neutres són els principals causants de les diferenciacions continentals inicials. En canvi, els posteriors fluxos migratoris, especialment dins dels continents, han modulats la baixa diferenciació poblacional (F_{CT}) i el contínuum genètic observat.

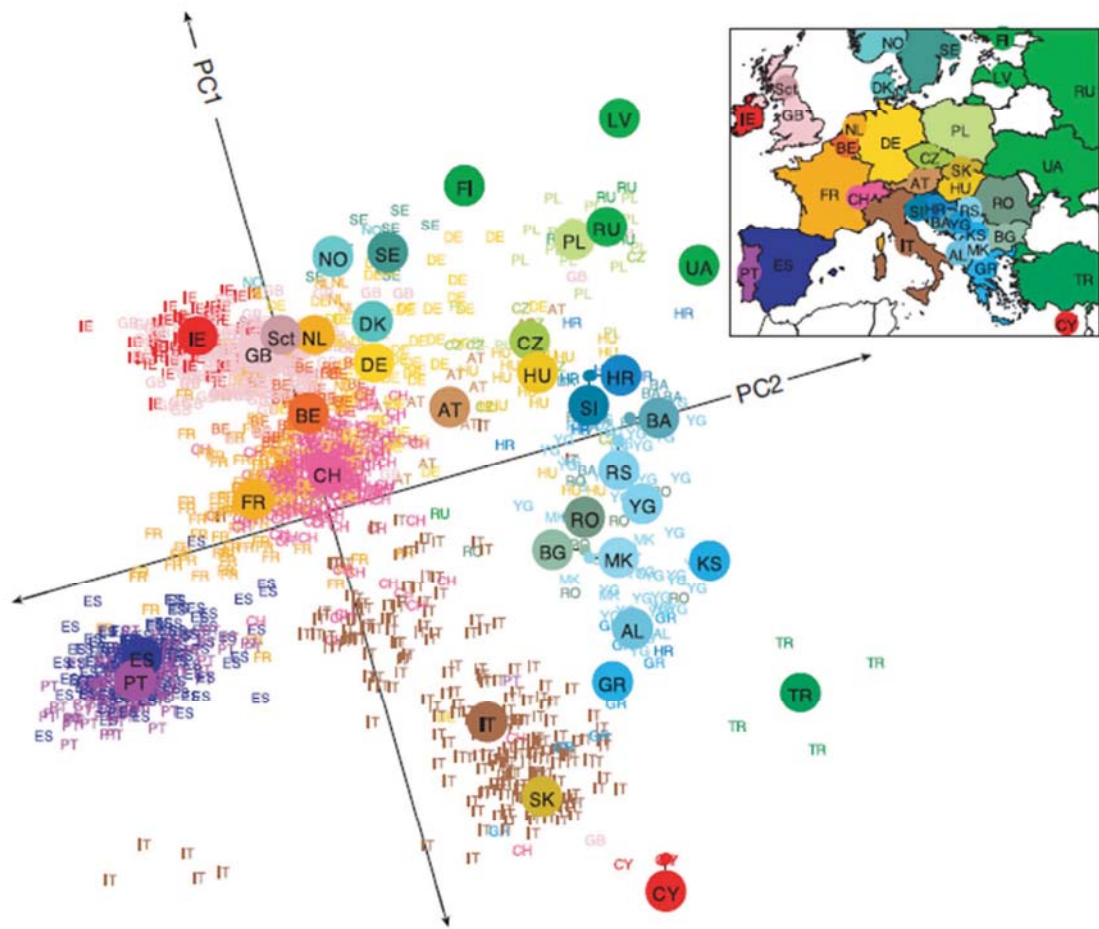


Figura 12 – Representació de l'estructuració genètica, mitjançant un anàlisi de components principals (PCA), observada en 500,000 SNPs de 3192 individus europeus del projecte *Population Reference Sample* (POPRES) [69].

Aquest procés es pot visualitzar fàcilment mitjançant l'estimació dels anomenats components genètics (conceptualment ancestrals) en 938 individus del projecte *Human Genome Diversity* genotipats per 650,000 SNPs assumint la presència de 7 components genètics (K) [67] (Figura 13). En aquesta figura es pot observar com cada individu és representat per una línia vertical fraccionada en segments de colors la longitud dels quals corresponen a la proporció inferida de cada component ancestral. La

identificació poblacional dels individus s'ha realitzat amb posterioritat a l'estimació ancestral i només s'ha utilitzat per tal d'ordenar els individus en el gràfic. S'observa com cada regió geogràfica té el seu component genètic “ancestral” propi. També s'observa com els individus d'una mateixa població presenten proporcions similars dels diferents components ancestrals, convertint aquestes proporcions en característiques de la població. Les poblacions del Pròxim Orient, tot i presentar el seu component genètic propi, són les que presenten major influència d'altres components genètics, degut a la seva situació geogràfica tradicionalment de trànsit per moltes expansions culturals.

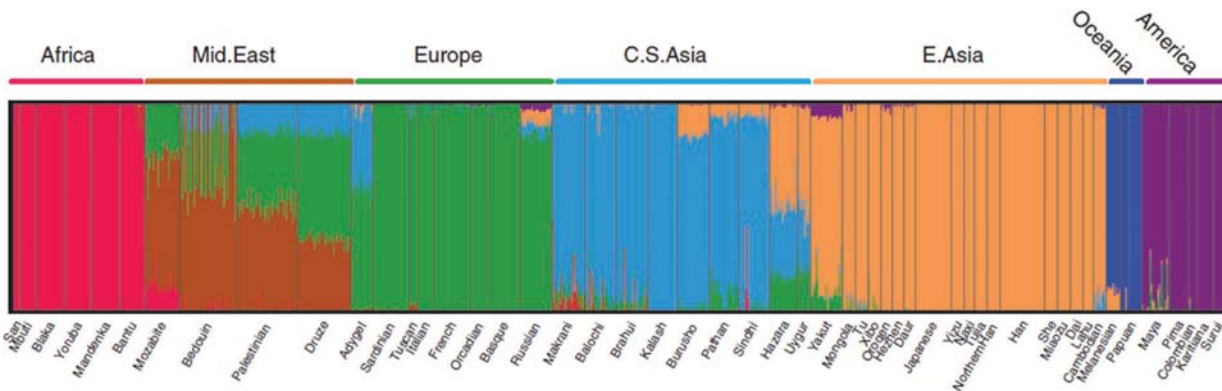


Figura 13 – Estimació de components ancestrals en 938 individus del projecte HGD assumint K (components ancestrals) igual a 7 [67].

1.5.2 Moviments poblacionals al llarg de la Història i les seves empremtes al genoma

Les estructures genètiques observades en les poblacions utilitzant aquests tipus d'anàlisi no aporten informació temporal, per tant, no es pot determinar quan es van generar. Una dificultat afegida rau en què la distribució i diversitat actuals no són fruit d'un únic fenomen, sinó de múltiples fenòmens que es sobreposen parcialment els uns als altres, des dels orígens fins els últims processos de migració i/o diferenciació, cadascun

deixant el seu eco en el patró de diversitat i distribució actuals.

Pel que fa a la distribució geogràfica dels marcadors uniparentals (mtDNA i cromosoma Y), degut a la absència de recombinació en la seqüència es pot estimar l'edat de la seqüència mitjançant la freqüència de mutació de les variants genètiques. Observant la distribució geogràfica de les seqüències es poden fer inferències dels fenòmens passats que millor expliquin la distribució observada. Però només es tracta de la informació d'un sol locus, en cada cas, que s'hereten via materna (mtDNA) o paterna (Chr Y).

En el cas de la variació autosòmica la possibilitat de disposar de informació genètica independent augmenta bastants ordres de magnitud, si bé la recombinació complica bastant la datació dels patrons de variació. Tot i així, es poden fer inferències sobre l'edat de divergència entre components genètics ancestrals a partir de distàncies F_{ST} entre components, mides efectives de població i temps de generació.

A mesura que s'han anat inferint moviments poblacionals descrits per la genètica, aquests corroboraven o complementaven el registre de troballes arqueològiques que fins el moment donaven un model als poblaments i migracions humanes. Tal i com s'ha vist, l'origen africà de l'espècie humana moderna fa 150.000 anys rau en l'elevada diversitat genètica present en el continent africà [71], i en una sèrie de migracions que sortiren d'Àfrica pel nord-est i poblaren Àsia central (fa 60.000 anys), Europa (fa 40.000 anys), l'est d'Àsia (40.000 anys) i més tard Amèrica (fa 15.000 anys) (Figura 14).

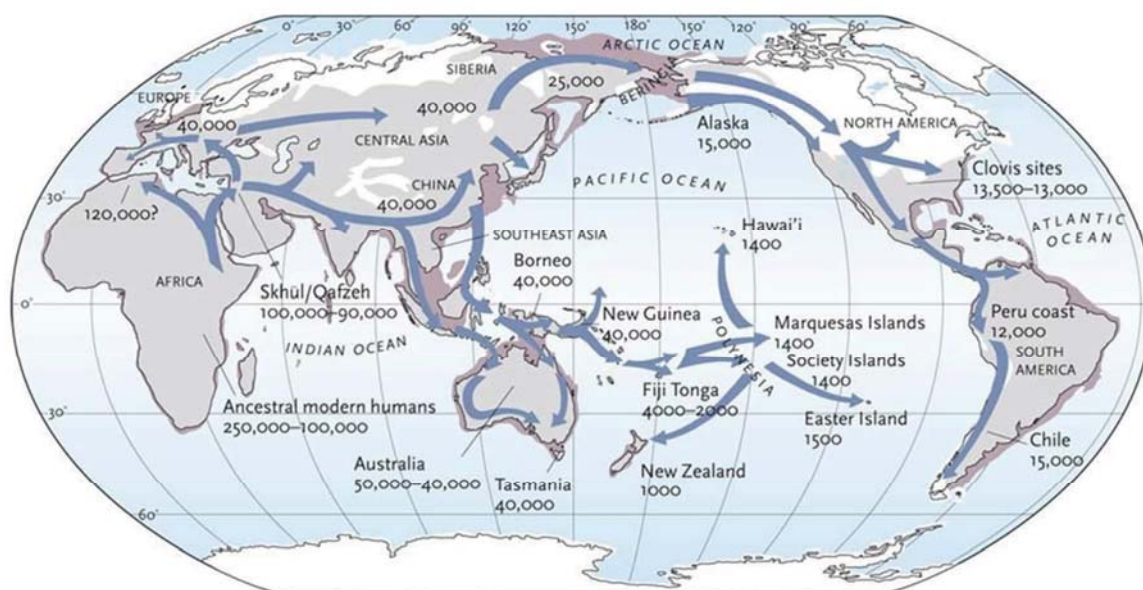


Figura 14 - Rutes migratòries de l'espècie humana moderna que poblaren el planeta. (<http://www.studyblue.com/notes/note/n/anthr-100-study-guide-2012-13-mariellasquire/deck/9715563>)

S'estima que les poblacions humanes inicials que sortiren de l'Àfrica pel nord-est i colonitzaren tot el territori no superaven els 1000 individus en nombre. Aquest processos migratoris es caracteritzen pel que s'anomenen efectes fundadors on el grup migrant no era genèticament representatiu ni tan divers com l'original, generant patrons de diversitat genètica inferiors a mesura que les poblacions s'escindien i s'allunyaven d'Àfrica [62]. Però en instal·lar-se a les noves àrees colonitzades, aquestes poblacions experimentaven un creixement demogràfic considerable. S'estima que el nombre d'humans colonitzant tots els continents fa 10.000 anys deuria ser ja d'entre 1 i 15 milions d'individus [72].

El fet diferencial d'espècie que va permetre colonitzar tots els racons del planeta van ser les innovacions culturals i socials per tal d'adaptar-se als nous ambients i territoris. L'aprofitament dels recursos característics de cada regió es va optimitzar generant eines més sofisticades, inicialment de pedra,

noves eines de pesca, l'arc i la fletxa, i elaborant diferents peces de vestir per protegir-se del clima. Però, sobretot, la millora en la navegació va permetre la colonització de les illes de Nova Guinea i, posteriorment, de la Polinèsia.

Tot aquest procés inicial de colonització es va dur a terme amb una economia caçadora-recol·lectora, on els assentaments de poblacions eren nòmades i es movien regularment per tal d'obtenir l'aliment.

Posteriorment, la domesticació dels primers conreus de cereals (blat, blat de moro i arròs) i les primeres cries d'animals aparegueren dins d'un mateix interval de temps (fa 12–10.000 anys) i d'una forma independent en diferents regions del planeta. Aquesta inicial producció d'aliments, que després esdevindria l'expansió de l'agricultura i la ramaderia, i per tant del Neolític, apareix al Pròxim Orient i Turquia, a la Xina (nord i sud independentment), al Sàhara, a Nova Guinea, a Mèxic i a l'oest d'Àfrica central. Aquesta adquisició tècnica i cultural comportà un canvi important en la dieta de les poblacions i un increment del sedentarisme. Les poblacions patiren un creixement demogràfic que augmentà la densitat poblacional i originaren noves migracions que expandiren la tècnica neolítica a les regions veïnes [73]. S'estima que aquesta cultura neolítica de producció de l'aliment es va expandir a un ritme de 1-2 km per any, fet que s'ha interpretat tant per l'expansió demogràfica dels agricultors com per l'adquisició de la tècnica per part de les poblacions caçadores-recol·lectores veïnes.

Aquest increment de la densitat poblacional també comportà l'especialització del treball, l'estratificació social i les primeres societats complexes amb diferents formes de govern. Posteriors revolucions tècniques permeteren noves expansions poblacionals, com l'adquisició de nous metalls (dins de l'edat del bronze fa 5000 anys i l'edat del ferro fa 3500 anys) i la domesticació del cavall i del camell. Els augments de població,

l'estratificació social i les revolucions tecnològiques van propiciar l'aparició de grans guerres entre governs veïns.

Per tant, els moviments poblacionals al llarg de la història de les poblacions humanes han dibuixat fortament la distribució genètica actual de la nostra espècie sobre el planeta.

1.5.2.1 Moviments poblacionals a Europa i al Mediterrani

Observant la distribució genètica al continent europeu i al nord d'Àfrica, destaca que els eixos de variació genètica segueixen els eixos geogràfics nord-sud i est-oest [69,74–77]. De fet, les primeres descripcions geogràfiques dels polimorfismes proteïcs clàssics ja mostraven un gradient continental des del Pròxim Orient i Turquia cap al nord-oest continental europeu (Illes Britàniques) i cap al Sàhara nord-africà tot seguint la costa africana. Aquesta distribució genètica ha estat tradicionalment interpretada com el reflex de les rutes migratòries adoptades per les poblacions humanes tant des de l'inicial poblament continental per part de les societats caçadores-recol·lectores fa 40.000 anys, com durant el desenvolupament de l'agricultura fa aproximadament 11.000 anys, ja que ambdós partiren de la regió oriental de la Mediterrània. Pel que fa als patrons de variació amb més caràcter nord-sud, aquests han estat tradicionalment interpretats com el reflex del repoblament del continent europeu des dels refugis glacials del sud d'Europa, després de l'últim període glacial fa 15.000 quan les regions més septentrionals d'Europa es van fer habitables per l'espècie humana [72,78].

Hi ha diverses qüestions referents a com es van desenvolupar aquest moviments poblacionals o culturals que han estat ampli tema de debat i investigació.

Segurament, la principal qüestió és si durant l'expansió de l'agricultura les poblacions neolítiques substituïren completament les poblacions paleolítiques ja presents en el territori (difusió dèmica), o bé, si van ser les poblacions paleolítiques les que adquiriren la nova tècnica neolítica de les poblacions veïnes (difusió cultural). Un model o l'altre suposaria una composició genètica de les poblacions europees actuals ben diferent. El model de difusió dèmica comportaria un gran substrat genètic original de les poblacions neolítiques del Pròxim Orient, mentre que el model de difusió cultural comportaria un gran substrat genètic procedent de les poblacions paleolítiques ja presents en el continent europeu abans de l'arribada de l'agricultura. Alguns estudis han intentat resoldre aquesta qüestió [79–81], i han descrit una aportació neolítica a la composició genètica europea actual també amb un patró clinal, amb una forta contribució genètica a la regió dels Balcans (70%) que progressivament disminuiria fins arribar a la península Ibèrica (30%), sent la població del País Basc la que mostraria menys contribució neolítica de tot Europa. Actualment aquesta qüestió no es troba encara resolta. De fet, recentment han incrementat les evidències arqueològiques, i també algunes paleogenètiques, defensant la dispersió del Neolític en el continent europeu mitjançant una sèrie de moviments marítims i terrestres al llarg de dos rutes principals, una resseguint les costes mediterrànies i una altre avançant a través de les valls fluvials d'Europa continental [82] (Figura 15).



Figura 15 – Mapa d’Europa que il·lustra el temps d’arribada aproximat (en anys abans del present) de les primeres cultures neolítiques a cada regió [82].

Al centre d’Europa, la cultura neolítica aparegué fa 7.500 anys caracteritzada per la ceràmica d’estil *Linearbandkeramik* (LBK). Aquesta cultura neolítica probablement procedia de la regió nord dels Balcans i avançava seguint la ruta que marcava el riu Danubi. Per contra, l’emergència del Neolític a l’oest del Mediterrani es produí lleugerament abans (8-7.500 anys) caracteritzada per la ceràmica d’estil Cardial. Aquesta dispersió neolítica al Mediterrani arribà a les dues costes més occidentals del Mediterrani (Península Ibèrica i Magrib) en un període molt curt de temps i s’atribueix als colonitzadors anomenats *maritime pioneers*, capaços de recórrer grans distàncies marítimes [83].

Un altre tema de debat és la possible presència d'alguna barrera genètica al llarg del Mediterrani. La presència del Mediterrani (barrera geogràfica) entre les dues costes hauria originat un aïllament genètic entre les poblacions del nord i del sud diferenciant-les genèticament. Per contra, la història cultural compartida per les poblacions mediterrànies (moviments poblacionals i les civilitzacions) entre les dues costes haurien fet del Mediterrani un mar permeable al flux genètic entre les poblacions del nord i les del sud. Els estudis més recent que han abordat aquest tema apunten cap a una continuïtat genètica entre les poblacions del nord i del sud del Mediterrani, especialment en la regió més occidental (Península Ibèrica i Magrib) on la distància entre les dues costes és mínima [76,84,85].

1.6 Poblacions estudiades: Europa continental i àrea mediterrània

En aquesta tesi doctoral s'han caracteritzat genèticament les regions dels gens *NOS* d'un total de 1451 individus provinents de 35 poblacions humanes. Aquestes poblacions es troben distribuïdes a Europa (22 poblacions), Orient Mitjà (2), nord d'Àfrica (7), Àfrica sud sahariana (2) i Sud-Amèrica (2). A l'hora d'analitzar les dades, es va incloure la informació genètica de 4 poblacions europees del projecte 1000 Genomes junt amb les nostres dades. L'estratègia de mostrejat va consistir en tenir àmpliament representades les poblacions humanes de la regió mediterrània, alhora que també apareguessin representades poblacions nord-europees i algunes poblacions externes al marc europeu i mediterrani, com les poblacions sud-saharianes i ameríndies. Detalls de les mostres poblacionals poden observar-se a la Taula 3 i una representació de la localització geogràfica de les mostres europees i mediterrànies s'observa a la Figura 16.

Aquesta representació de poblacions del Mediterrani i Europa va ser recollida per tal d'abordar diferents qüestions poblacionals dins de l'àmbit de les malalties cardiovasculars com a fenotip complex de les poblacions.

D'una banda, el primer objectiu és descriure la distribució poblacional dels factors genètics identificats com de risc en els estudis epidemiològics. Tal i com s'ha explicat anteriorment, no només es pretén descriure la variació poblacional de les freqüències d'aquests factors de risc, sinó que es proposa descriure la variació poblacional dels *scores* de risc genètic (GRS) obtinguts a partir dels factors de risc individuals per tal de descriure la covariació geogràfica del GRS amb la incidència coronària poblacional descrita pel projecte MONICA de la OMS. Aquesta estratègia no ha estat utilitzada prèviament en cap estudi eco-epidemiològic fins a l'actualitat.

Origen geogràfic	País	Població	Codi	N	Lat	Lon	
Europa	Regne Unit	Òrcades	ORK	48	59.00	-3.00	
	Polònia	Varsòvia	POL	44	52.25	21.00	
	França	França nord - Lille	NFR	48	50.60	3.05	
		França sud - Tolosa	SFR	48	43.65	1.50	
		Còrsega nord	NCO	33	42.60	9.50	
		Còrsega sud	CCO	45	41.90	8.60	
	Espanya	País Basc -Guipúscoa	BQC	41	42.80	-2.60	
		Catalunya - Girona	CAT	43	41.95	2.75	
		Castella-Lleó est - Burgos/Valladolid	NSP	39	42.10	-3.73	
		Andalusia oest - Sevilla/Cadis	SSP	46	37.38	-5.97	
		Illes balears - Menorca	MEN	49	40.00	3.80	
	Itàlia	Itàlia central - Lazio	CIT	37	43.15	13.40	
		Itàlia sud – Nàpols	SIT	39	40.60	15.78	
		Sardenya nord	NSR	43	40.33	9.40	
		Sardenya sud	SSR	41	39.20	9.12	
		Sicília	SIC	32	38.20	12.50	
	Bòsnia-Hercegovina	Bòsnia-Hercegovina nord – Krajina	NBH	40	44.80	17.20	
		Bòsnia-Hercegovina sud – Costa adriàtica	SBH	37	43.40	17.80	
	Grècia	Grècia nord - Tesaloniki	NGR	48	40.60	22.98	
		Grècia sud – Atenes	SGR	43	37.50	22.45	
		Creta	CRT	44	35.25	25.00	
	Turquia	Població general	TRK	48	39.95	32.90	
	Orient Mitjà	Jordània	Població general	GJD	48	31.65	36.50
			Beduins de Jordània	BJD	42	30.20	35.75
	Nord d'Àfrica	Marroc	Alt Atlas - Asni	ASM	48	31.15	-8.00
			Alt Atlas - Khenifra	KHM	43	32.90	-5.60
			Baix Atlas - Bouhria	BOM	35	34.60	-3.69
Algèria		M'zab	MZB	50	28.00	3.00	
Tunísia		Tunísia centre-nord	NTS	47	36.80	10.16	
		Tunísia sud	STS	49	33.88	10.03	

	Egipte	Siwa	SIW	23	29.15	25.50
Sud Sàhara	Costa d'Ivori	Ahizi	AHZ	26	6.12	-3.82
	Nigèria	Bamileké	BAM	25	4.25	9.25
Sud Amèrica	Bolívia	Aimara	AYM	43	-16.63	-68.47
		Quítxua	QCH	36	-19.45	-66.20
Total mostres DNA				1451		
1000 Genomes	Regne Unit	GBR	GBR	89	52.22	-0.97
	Europe central	CEU	CEU	85	52.38	8.14
	Finlàndia	FIN	FIN	93	62.02	24.07
	Itàlia	TSI	TSI	98	43.20	11.67
Total 1000G				365		
TOTAL				1816		

Taula 3 - Detalls de les mostres poblacionals



Figura 16 - Distribució geogràfica de les mostres poblacionals europees i mediterrànies. La codificació es pot observar a la taula 3.

D'altra banda, el segon objectiu és identificar quins són els patrons de variació genètica que expliquen la diferent incidència cardiovascular en les poblacions europees. En apartats anteriors d'aquesta tesi, s'ha observat com la diferenciació genètica entre poblacions europees és petita però significativa, alhora que els patrons de distribució genètica al continent segueixen patrons geogràfics. Per tant, aquest fenomen genètic àmpliament conegut al continent europeu podria estar influenciant la variació en la incidència cardiovascular, ja que els factors de risc genètic es trobarien distribuïts seguint els mateixos patrons que la resta de variació genètica neutra no associada a cap fenotip. Per tal de determinar la validesa d'aquesta hipòtesi, es vol comprovar si els factors genètics identificats de risc tenen patrons poblacionals modulats per algun efecte selectiu o, per contra, presenten un patró de variació neutra similar al de la resta de variació genètica estudiada. I per una altra banda, també s'ha avaluat si paràmetres d'estructuració genètica obtinguts amb les nostres dades (com els descrits anteriorment: distàncies genètiques, components principals i components ancestrals) són capaços d'explicar la variació geogràfica de la incidència coronària aportada pel projecte MONICA de la OMS.

Estudiar la relació entre les distribucions poblacionals de la variació genètica i les malalties complexes des del prisma de la història demogràfica recent de les poblacions humanes del continent europeu i de la Mediterrània pot ser de gran ajuda per la comprensió de la història natural d'aquestes malalties a la Mediterrània i també per a la prevenció i tractament d'aquestes malalties.

2. OBJECTIUS

2.1 Objectius de la tesi

Aquesta tesi es centra en la exploració de la diversitat genètica del gen *F13A1* i de regions genòmiques corresponents als tres gens *NOS* i adjacents, mitjançant polimorfismes SNPs i STRs, en mostres epidemiològiques cardiovasculars, i en un conjunt de mostres poblacionals representatives del continent europeu i de la regió mediterrània, i la seva interpretació des del punt de vista de la incidència cardiovascular i la història demogràfica d'aquestes poblacions.

Amb aquests propòsits, l'estudi ha estat dissenyat per tal d'assolir els següents objectius:

- Identificar variació genètica de risc present al gen *F13A1* per infart de miocardi i el rol de les seves possibles interaccions en una mostra de 101 famílies amb fills afectats. Aquestes anàlisis epidemiològiques van incloure:
 - a) Identificar marcadors genètics de risc per infart de miocardi al gen *F13A1* en població espanyola.
 - b) Identificar interaccions entre els factors genètics dins del gen *F13A1* i la seva contribució al risc.
 - c) Comprendre el possible efecte poblacional a l'hora d'identificar marcadors de risc.
- Identificar variació genètica de risc present als gens *NOS* per infart de miocardi en diferents mostres cas-control europees, i descriure i analitzar la seva distribució poblacional i les seves conseqüències epidemiològiques a Europa i al Mediterrani, dins del marc de la diferent

incidència cardiovascular observada a Europa. El desenvolupament d'aquest objectiu s'ha portat a terme mitjançant els següents objectius específics:

- a) Determinar la variació d'un conjunt de SNPs de les regions genòmiques dels gens *NOS* i adjacents en mostres cas-control de la Península Ibèrica i comparar-la amb mostres cas-control europees, per tal d'identificar factors de risc per l'infart de miocardi.
 - b) Valorar la capacitat dels factors genètics de risc identificats per discernir entre afectats i no afectats mitjançant el càlcul d'una puntuació de risc.
 - c) Descriure la variació poblacional a través d'Europa i el Mediterrani dels marcadors de risc i també de la puntuació genètica de risc mitjana de cada població.
 - d) Correlacionar la puntuació genètica de risc dels gens *NOS* amb la incidència cardiovascular descrita al continent europeu i determinar la proporció de variació que n'explica.
 - e) Valorar les aportacions que l'ús dels GRS poden oferir a futurs estudis eco-epidemiològics.
- Descriure i analitzar la variació poblacional a Europa i al Mediterrani dels marcadors genètics (tant de risc com no de risc) dels gens *NOS* per tal d'identificar quins processos demogràfics i/o adaptatius poden haver modulats la variació observada i estudiar si aquests processos demogràfics poblacionals permeten explicar la diferent incidència cardiovascular al continent europeu. Aquestes anàlisis poblacionals s'han assolit amb els següents objectius específics:

- f) Determinar la variació d'un conjunt de 114 SNPs i 17 STRs en 1451 mostres procedents de 35 poblacions localitzades principalment a Europa, nord d'Àfrica i Orient Mitjà, amb algunes referències externes localitzades a l'Àfrica sud-sahariana i Sud Amèrica.
- g) Valorar l'efecte de l'adaptació al medi sobre els marcadors genètics estudiats (relacionats o no amb les malalties coronàries) mitjançant l'estudi de possibles senyals de selecció.
- h) Caracteritzar les relacions genètiques de les mostres europees i mediterrànies mitjançant distàncies genètiques i estimacions de distribució i estructuració genètica.
- i) Estimar i interpretar el nombre i proporció de contribucions genètiques "ancestrals" al patrimoni genètic actual de les poblacions estudiades.
- j) Estimar la contribució de l'estructuració genètica poblacional a la incidència coronària observada a Europa i valorar quins processos poblacionals ens explicarien les relacions observades.

3. RESULTATS

3.1 Informe dels directors de tesi del factor d'impacte dels articles publicats

La tesi doctoral “**Variació genòmica de les Sintases d’Òxid Nítric i de regions associades amb Risc Cardiovascular en Poblacions Humanes**” es basa en els resultats obtinguts per Robert Carreras Torres i presentats en tres articles, dos d’ells ja publicats en revistes científiques internacionals mentre que el tercer es troba llest per presentar a consideració editorial.

La importància dels resultats obtinguts es demostra per la qualitat de les revistes científiques:

Thrombosis Research és la revista científica oficial de la *European Thrombosis Research Organization* (ETRO). Aquesta revista es troba indexada a SCI i SSCI amb un factor d’impacte de 3.133 (2012) i classificada en l’àrea *Hematology* 25/67, Q2; i a l’àrea *Peripheral Vascular Disease* 18/68, Q2.

PLoS ONE és una revista científica multidisciplinària d’accés obert, revisada i arbitrada per experts de la organització *Public Library of Science* (PLOS). Aquesta revista es troba indexada a SCI i SSCI amb un factor d’impacte de 3.730 (2012) i classificada en el primer quartil (Q1) de l’àrea *Multidisciplinary Sciences*, posició 7/56.

Signat per Dr. Pedro Moral Castrillo i Dr. Marc Via García
Barcelona, 22 Abril 2014.

3.2 Resultats I: Carreras-Torres et al., 2010

**Allele-allele interaction within the *F13A1* gene:
A risk factor for Ischaemic Heart Disease in Spanish population**

Robert Carreras-Torres, Georgios Athanasiadis, Marc Via, Joan Trenchs,
Magda Gayà-Vidal, Josep Santamaria, Esther Esteban, Pedro Moral.

Thrombosis Research 126 (2010) e241-e245.

3.2.1 Resum en Català

Interacció al·lel-al·lel dins del gen *F13A1*:

Un factor de risc per l'infart de miocardi en la població espanyola

El factor plasmàtic de coagulació XIII (FXIII) és una pro-transglutaminasa essencial per l'equilibri hemostàtic amb un rol molt important en la coagulació sanguínia. La variació genètica present al gen del factor de coagulació (gen *F13A1*) pot trobar-se entre les causes dels processos moleculars que ocasionen una malaltia cardiovascular. Existeixen dos polimorfismes genètics SNPs (rs5985 i rs5982) en el gen *F13A1* que provoquen un canvi aminoacídic en el factor FXIII, modulant lleugerament la seva funció. El canvi al·lèlic Leu34 per Val34 del polimorfisme rs5985 provoca el canvi aminoacídic de leucina per valina, mentre que la variació al·lèlica Leu564 per Pro564 de rs5982 genera el canvi de leucina per prolina. Aquestes modificacions aminoacídiques en el FXIII s'han trobat prèviament relacionades amb un caràcter protector davant fenòmens trombòtics o de risc davant fenòmens hemorràgics, sobretot en poblacions nord-centre europees però no en poblacions del sud d'Europa, com és el cas d'Espanya.

En aquest article s'ha estudiat en una mostra de 101 famílies de l'àrea

de Barcelona la transmissió entre pares/mares i fills afectats d'infart de miocardi d'aquest dos polimorfismes bial·lèlics SNP (rs5985 i rs5982) juntament amb dos polimorfismes multial·lèlics de repeticions en tàndem STR (*F13A1*-STR i D6S477), que es troben flanquejant a banda i banda el gen *F13A1*. El test de desequilibri de la transmissió (TDT) es va aplicar tan a la mostra global com als subgrups de mostra obtinguts en agrupar els pacients segons els seus genotips dels polimorfismes SNPs.

Com a resultats es va observar que la variant al·lèlica Pro564 de rs5982 es trobava transmesa en 34 casos d'un total de 42, sense ser significativament diferent del 50% esperat. Mentre que en el subgrup de mostra on els fills afectats presentaven el genotip Val34-Val34 per rs5985, la transmissió de l'al·lel Pro564 de rs5982 fou significativament diferent a l'esperada a l'atzar amb 21 casos d'un total de 23. Aquest desequilibri de la transmissió de Pro564 observat sobretot en el grup de pacients Val34-Val34 suggereix una interacció entre aquestes dues variants genètiques (Pro564 i Val34) que incrementaria el risc de patir un infart de miocardi. Aquests resultats, conjuntament amb d'altres de la bibliografia, indicarien que una de les causes del paper pro-trombòtic del FXIII vindria donat per la combinació de les dues variants Val34 i Pro564 en una mateixa seqüència del gen *F13A1*. De la mateixa manera, un paper pro-hemorràgic de FXIII seria afavorit per la combinació Leu34 i Leu564, mentre les combinacions Val34 i Leu564, i Leu34 i Pro564 equilibrarien el FXIII. Per tant, efectivament aquests polimorfismes també es troben associats a d'infart de miocardi en la població espanyola però l'associació només es manifesta quan es consideren els dos polimorfismes conjuntament degut al seu rol interactiu dins de FXIII. Pel que fa als polimorfismes multial·lèlics STR, es va observar que l'al·lel més curt (3.2 repeticions) de *F13A1*-STR es trobava infratransmès, mentre

que l'al·lel més llarg, d'entre els de més freqüència poblacional (7 repeticions), s'observava sobretransmès tan individualment com combinats amb l'al·lel Pro564. Aquest STR es troba localitzat a la regió promotora del gen i els resultats, juntament amb d'altres de la bibliografia, podrien apuntar cap a un paper regulador de *F13A1*-STR en l'expressió del gen *F13A1* i, per tant, en la disponibilitat en sang del FXIII.

3.2.2 Informe dels directors de tesi de la participació del doctorant en l'article i de la singularitat de l'article en la elaboració de la tesi.

Dr. Pedro Moral Castrillo, professor del Departament de Biologia Animal de la Universitat de Barcelona, i Dr. Marc Via García, professor del Departament de Psiquiatria i Psicobiologia Clínica de la Universitat de Barcelona, supervisors de la tesi doctoral “**Variació genòmica de les Sintases d’Òxid Nítric i de regions associades amb Risc Cardiovascular en Poblacions Humanes**”, manifesten que l’estudiant de doctorat Robert Carreras Torres ha participat en el treball titulat “*Allele-allele interaction within the F13A1 gene: A risk factor for Ischaemic Heart Disease in Spanish population*”, publicat a la revista *Thrombosis Research*, en les següents tasques:

- Disseny del pla de recerca (90%).
- Genotipació dels STRs en el laboratori (100%).
- Creació de las bases de dades genètiques i selecció de resultats disponibles a la bibliografia per tal de realitzar comparacions (100%).
- Anàlisis estadístics (100%).
- Redacció de l'article (95%).

Cap dels coautors de l'article ha utilitzat els resultats d'aquest estudi de manera implícita o explícita per elaborar una altra tesi doctoral. Per tant, aquest article forma part de manera exclusiva de la tesi doctoral de Robert Carreras Torres.

Signat per Dr. Pedro Moral Castrillo i Dr. Marc Via García
Barcelona, 22 Abril 2014.



Contents lists available at ScienceDirect

Thrombosis Research

journal homepage: www.elsevier.com/locate/thromres

Letter to the Editors-in-Chief

Allele-allele interaction within the F13A1 gene: A risk factor for Ischaemic Heart Disease in Spanish population

To the Editor,

Atherothrombotic diseases are a group of diseases that cause the majority of population morbidity in industrialized countries. Genetic variation in causative or susceptible genes constitutes the basis of molecular mechanisms, that together with environmental factors, lead to disease development [1].

Plasma coagulation factor XIII (FXIII) is a pro-transglutaminase essential for haemostasis which, after being activated by thrombin and Ca^{2+} , plays an important role in the final stages of blood coagulation and in the regulation of fibrinolysis. Plasma FXIII consists of two A subunits and two B subunits (carrier molecule), encoded by genes located on chromosomes 6p25–p24 and 1q31–q32.1, respectively. Activated FXIII cross-links covalently adjacent fibrin monomers and stabilizes the structure of blood clot [2]. Genetic variation in the sequence of FXIII A subunit (F13A1) has been associated with atherothrombotic disease and functional changes in FXIII levels and activity.

The Val34Leu polymorphism (rs5985) is a non-synonymous single-nucleotide polymorphism (SNP) located in exon 2 of F13A1, three residues away from the thrombin cleavage site [3]. Studies of fibrin structure and clot permeability showed how Leu34 allele modulates the association of loss of permeability with high fibrinogen concentrations [4]. Individuals carrying Leu34Leu34 genotype produced fibrin with thicker fibres at elevated fibrinogen concentrations, preventing physiological conditions associated with thrombotic risk and pointing out a gene-environment interaction. Controversial results on its potential role in coronary artery disease (CAD) susceptibility have been reported. Whereas some epidemiological studies indicated a main protective role of Leu34 against thrombosis in different populations in the UK, Finland, Northern France, Northern Italy, and Turkey [5–10], it has not been confirmed in other populations, such as Spain, Southern France, Netherlands, Italy, Sardinia, and Hungary [11–17], as reviewed by Vokó et al [18].

Another important non-synonymous SNP, Pro564Leu (rs5982), located in exon 11 of F13A1 on the barrel 1 domain of the FXIII A subunit [3], is implicated in the interactions between the A and B subunits of FXIII [19], affecting the supply of activated FXIII. Although this polymorphism has been less studied, evidence of association of the Leu564 allele with hemorrhagic stroke has been reported [20].

As well as SNPs, two additional tetranucleotide microsatellite polymorphisms (STR) flanking the F13A1 coding sequence have also been described: the F13A1-STR [21] and the D6S477. The former is located 704 bp upstream from the transcription initiation site and the latter 3692 bp downstream the end of transcription. To date, there are no data from association studies for these microsatellite polymorphisms.

The lack of replication in genetic epidemiological studies could be attributed to gene-gene interactions. Moore [22] suggested that

epistasis is a ubiquitous component of the genetic architecture of common human disease and complex interactions are more important than independent gene effects. As interactions among polymorphisms in different gene regions, interactions among polymorphisms within the same gene could be important to test the contribution of a gene to the development of complex human diseases. Previous studies tested interaction of the Val34Leu and Pro564Leu polymorphisms of F13A1 gene with hemorrhagic stroke [20], ischaemic stroke and myocardial infarction [23] in young European-American women without finding any association. Although these two polymorphisms were jointly analysed in a European sample from the UK [24], they had not been tested for interaction.

In this context, this paper presents the first approach to assess the importance of allele interactions within the F13A1 gene in the susceptibility to CAD in a European population. This has been carried out through a family-based association study of the above reported makers in the F13A1 gene in a sample of ischemic heart disease (IHD) patients from Spain.

A total sample of 101 nuclear families (N=302 individuals) were recruited from the area of Barcelona, Spain, including classic cardiovascular risk factors related information. One individual diagnosed for IHD and both parents formed each family trio. When sampling from both parents was not possible (56 families), brothers and sisters were also studied to reconstruct parental genotypes and haplotypes. Patients were characterized by being younger than 55 and affected by MI or angina pectoris. MI was diagnosed by clinical, enzymatic (CPK levels >200 U/l), and electrocardiography changes typical in myocardial ischaemic injury and necrosis, and AP by coronariography with an obstruction of more than 50% in a main coronary artery. A comprehensive description of DNA extraction methods and lipid level measurements has been reported by Via et al [25].

Both SNPs (Val34Leu and Pro564Leu) were determined by PCR amplification followed by RFLP analysis and allele identification as described in Reiner et al [20].

F13A-STR and D6S477 microsatellite polymorphisms were typed by DNA fragment analysis [PCR conditions in a total volume of 20 µl: 25 ng of template DNA; 10 pmol of each primer; 0.25 mM of each dNTP; 2.5 mM MgCl₂; 1 unit of *Taq* DNA polymerase and 2 µl buffer 10X (Biotools B&M Labs S.A., Madrid, Spain)]. Used primers for the F13A1-STR (unlabelled forward primer 5'GAGGTTGCACTCCAGCCTTT3' and the labelled reverse primer 5'GCCCAAGGAAGATGAGTAA3' with fluorescent FAM) and for the D6S477 (the labelled forward primer 5'GATTTGCCATGATAGATGGC3' with fluorescent HEX and the unlabelled reverse primer 5'GGGGATATCTCAACAACC3'). Amplification program (initial 5 min denaturation step at 94 °C, 30 cycles of 1 min at 95 °C, 1 min at 62 °C for the F13A1 polymorphism and 58 °C for the D6S477 polymorphism, and 1 min at 72 °C, followed by a final step of 5 min at 72 °C). The size standard used in the capillary electrophoresis was CST ROX 70–500 (BioVentures, Inc., Murfreesboro, TN). DNA from homozygote individuals was directly sequenced to establish the exact number of repeats and control allele designations. Analyses were performed on an ABI PRISM 3700 analyser and results were visualized using the GeneMapper 3.0 software (Applied Biosystems,

Foster City, CA). Allele designation for D6S477 was according to Barral et al [26].

Allele frequencies were calculated through SPSS for Windows (SPSS Inc., Chicago, USA). Goodness-of-fit to Hardy-Weinberg law and linkage disequilibrium were calculated using Genepop v. 4.0 [27]. Genetix v. 4.04 [28] was used to estimate gene diversities and other standard population genetics statistics. Cardiovascular risk factors frequencies and lipid mean values were also calculated and analyzed by SPSS for Windows.

The transmission disequilibrium test (TDT) was performed by the PBAT/FBAT software [29]. Bonferroni correction was applied for multiple testing. Statistical power of the used sample was assessed in two different ways: i) an *a priori* value through QUANTO vs 1.2.3 software assuming a disease prevalence of 0.07 [30], and ii) an *a posteriori* estimation based on allele frequencies under an additive model through PBAT/FBAT software [29]. Potential allele interactions were tested stratifying the 101 family trios according to the patients' genotype. Allele relative risk (RR) and confidence interval (CI) were calculated with Epi InfoTM v. 3.4.3 [31].

Allele frequencies of the two SNPs and the two STRs in IHD patients, their parents, and the total sample are shown in Table 1. Non-biased heterozygosity ranged from 0.277 to 0.410 for SNPs and from 0.718 to 0.771 for STRs. In all cases, genotype distributions were in H-W equilibrium. No significant linkage disequilibrium (LD) between loci or significant differences between patient and parent subgroups were detected (data not shown). Allele combinations between the two SNP loci pointed out the Val34-Pro564 as the most frequent.

The distribution of several risk factors CAD is shown in Table 2. Heavy smoking (>15 cigarettes per day), alcohol consumption, fat-rich diet, and elevated triglyceride levels, as well as low HDL cholesterol levels, were characteristics of our IHD patients. Interest-

ingly, patients with Val34Val34 genotypes showed significant low levels of triglycerides.

As far as TDT analysis, "*a priori*" power estimations indicated that our sample had a power >80% to detect risks of around 2.8 using minor allele frequencies (MAF) \geq 5%, and 2.05 using SNP MAF frequencies \geq 15%.

TDT results for allele and allele combinations are shown in Table 1. Transmission disequilibrium was observed for F13A1-STR*3.2 and D6S477*18 alleles and Val34-Pro564 haplotype (Table 1). Also deviations were present for Pro564-F13A1-STR*3.2 (risk ratio (RR) = 0.2; 95%CI: 0.06-0.73) and Pro564-F13A-STR*7 (RR = 3.2; 95%CI: 1.67-6.47) allele combinations. Allele interaction results between the two SNP loci indicated a significant transmission deviation for Pro564 allele in the patient subgroup of Val34Val34 genotype (Table 3). Also transmission deviation was detected for the Pro564-F13A-STR*7 allele combination in patients with Val34Val34 genotype (RR = 6.0; 95%CI: 1.63-22.03). However, all STR-involved deviations were insignificant when the stringent Bonferroni correction was applied. Interaction between genetic factors and classic cardiovascular risk factors of Table 2 were not observed (data not shown).

In general, our results in IHD Spanish patients agree with published data reporting no remarkable individual effect of F13A1 polymorphisms in Mediterranean populations. However, positive significant association of Pro564 allele in Val34Val34 homozygote patients was found in this survey. Recently, it has increased awareness that genetic and environmental interactions are involved in susceptibility to atherothrombotic diseases. Some previous reported studies highlighted the importance of interactions between Val34Leu polymorphism and acquired/environmental factors, such as features of insulin resistance [32] and fibrinogen concentrations [17]. In this way, our association results suggest an allele-allele interaction within F13A1 that could help to explain previous controversial results on the genetic

Table 1
Allele and allele combination frequencies and TDT analyses of Val34Leu, Pro564Leu, F13A1-STR and D6S477 polymorphisms.

Locus	Allele or allele combination	Gene Frequencies			TDT analyses	
		Patients	Parents	Total sample	T/NT§	RR (95%CI) †
Val34Leu		(n=98)	(n=133)	(n=280)		
	Val34	0.714	0.744	0.734	20/16	1.25 (0.78-2.00)
Pro564Leu		(n=101)	(n=139)	(n=294)		
	Pro564	0.835	0.791	0.808	34/8	4.25 (2.24-8.06)
Haplotypes Val34Leu/ Pro564Leu	Val34-Pro564	0.620	0.580	0.607	36/9	4 (2.19-7.31)*
	Val34-Leu564	0.096	0.168	0.147	4/25	0.16 (0.06-0.40)
	Leu34-Pro564	0.210	0.214	0.192	10/10	1 (0.54-1.86)
	Leu34-Leu564	0.074	0.038	0.054	2/7	0.29 (0.08-1.02)
F13A1-STR		(n=93)	(n=199)	(n=255)		
	3.2	0.032	0.071	0.063	2/15	0.13 (0.04-0.50)**
	4	0.010	0.017	0.012	1/1	1 (0.48-2.10)
	5	0.247	0.235	0.239	15/18	0.83 (0.51-1.36)
	6	0.290	0.348	0.319	24/18	1.33 (0.86-2.06)
	7	0.371	0.289	0.322	21/11	1.91 (1.11-3.28)
	8	0.006	0.013	0.009	1/3	0.33 (0.06-1.99)
	16	0.032	0.013	0.024	3/1	3 (0.50-17.95)
	17	0.006	0.013	0.008	0/2	
	18	0.006	0	0.004		
D6S477		(n=65)	(n=96)	(n=200)		
	13	0.008	0.016	0.013	0/1	
	15	0.015	0.026	0.018	1/2	0.5 (0.08-2.99)
	16	0.023	0.031	0.033	2/2	1 (0.25-4.00)
	17	0.215	0.188	0.208	8/8	1 (0.50-2.00)
	18	0.208	0.188	0.190	14/3	4.67 (1.63-13.34)*
	19	0.346	0.375	0.368	8/17	0.47 (0.25-0.88)
	20	0.154	0.151	0.153	6/7	0.86 (0.40-1.86)
	21	0.015	0.016	0.010	1/1	1 (0.48-2.10)
	22	0.015	0.010	0.010	2/0	

§ Number of transmitted (T) and untransmitted (NT) alleles. † Relative Risk with 95% Confidence Interval. * p<0.05 before Bonferroni correction. ** p<0.01 before Bonferroni correction.

Table 2
Sample description: classic cardiovascular risk factor distribution in sample subgroups.

	Patient's subgroups according to genotype (n)								Whole patients (101)	Parents (144)	Siblings (57)	p value
	Val34 Val34 (50)	Val34 Leu34 (40)	Leu34 Leu34 (8)	p value	Pro564 Pro564 (68)	Pro564 Leu564 (32)	Leu564 Leu564 (1)	p value				
Gender (%)												
Male	91.7	92.3	87.5	ns	89.7	96.8	100.0	ns	92.1	36.8	42.1	<0.01
Age (mean years)												
	45.3 ± 7.0	43.0 ± 5.9	41.5 ± 8.2	ns	44.4 ± 6.9	43.3 ± 6.1	52.0	ns	44.2 ± 6.6	72.6 ± 7.2	43.8 ± 9.2	<0.01
BMI (%)												
<25	29.8	18.9	12.5		32.8	10.3			25.5	27.3	41.2	
25-30	46.8	75.7	75.0		53.7	69.0	100.0		59.2	50.8	35.3	
>30	23.4	5.4	12.5	ns	13.4	20.7		ns	15.3	21.2	23.5	ns
Cigarettes/Day (%)												
0	4.5	17.1			12.7	3.7			9.8	67.8	31.7	
<15	9.1	11.4			9.5	3.7	100.0		8.7	9.9	9.8	
15-25	29.5	28.6	71.4		31.7	29.6			30.4	9.1	24.4	
>25	56.8	42.9	28.6	ns	46.0	63.0		<0.05	51.1	13.2	34.1	<0.01
Alcohol consumption (%)												
No	39.1	40.0	28.6		37.5	39.3	100.0		39.4	61.3	39.0	
<30 mg/day	43.5	54.3	71.4		54.7	42.9			50.0	32.8	58.5	
>30 mg/day	17.4	5.7		ns	7.8	17.9		ns	10.6	5.9	2.4	<0.05
Diet (%)												
Normal	38.1	40.0	66.7		42.4	36.4			41.5	71.3	66.7	
Fat-rich	59.5	56.7	33.3		55.9	59.1			56.1	25.2	31	
Diabetic		3.3				4.5			1.2	2.6	2.4	
Others				ns	1.7			ns	1.2	0.9		<0.05
Hypertension (%)												
Yes	38.3	29.7	0.0	ns	28.4	34.5	0.0	ns	30.6	46.6	18.6	<0.05
Lipid mean levels (mean mg/dl)												
Total cholesterol	193.7 ± 41.3	194.2 ± 46.2	194.1 ± 59.6	ns	193.5 ± 41.7	198.8 ± 50.7	173.2	ns	194.3 ± 44.6	207.5 ± 42.7	208.1 ± 44.7	ns
HDL	35.3 ± 8.6	36.2 ± 8.2	34.3 ± 6.0	ns	35.5 ± 8.0	36.7 ± 9.2	37.0	ns	36.0 ± 8.4	46.8 ± 13.1	47.5 ± 14.9	<0.01
ApoB	94.4 ± 22.5	99.5 ± 29.5	99.4 ± 33.3	ns	96.5 ± 23.3	99.6 ± 32.9	103.3	ns	97.3 ± 26.5	99.1 ± 25.3	98.8 ± 28.8	ns
Triglyceride	111.8 ± 61.4	126.9 ± 53.0	158.2 ± 65.6	<0.05	119.6 ± 75.4	139.0 ± 64.3	151.0	ns	125.6 ± 71.7	110.0 ± 62.6	98.7 ± 57.8	<0.05

susceptibility of F13A1 gene. Physiological results have suggested that these alleles can be associated to thrombotic risk, through more effective binding of subunits A and B of the FXIII molecule by Pro564 allele [19] and disability at modulating clot permeability in risen fibrinogen conditions [4]. In other words, the combination of both alleles in F13A1 would lead to increase the risk of thrombosis, not always detected with simple genetic association analysis.

Similarly to studies linked to CAD disease, some population-based studies involving venous thromboembolism (VTE) also reported a protective effect of Leu34 allele in populations of Northern UK, Austria and France [33–36]. The latter studies failed to report associations in

the populations of Spain, Hungary and Southern Italy [11,37,38], as reviewed by Wells et al [39]. Besides, an increased risk of the Leu34 allele associated with intracranial hemorrhage was observed in populations of Northern UK and Northern Italy [9,40], but was not reported in populations of Spain, Austria and Poland [41–43]. Accordingly, the Leu34 allele has been reported both to be protective against thrombosis and involved in the pathogenesis of bleeding. Thus, the final physiological role, thrombotic or hemorrhagic, of the plasmatic FXIII could be due to combined effects of different polymorphic sites of F13A1. This might be the case of the interaction between Val34Leu and Pro564Leu polymorphisms found in this study that could help to explain previous controversial results on the genetic variation of the F13A1 gene with vascular disease. A balanced combination of prothrombotic (such as Pro564 plus Val34) and prohemorrhagic (such as Leu564 and Leu34) variants in different FXIII sites can lead to negative association results when only single FXIII gene variants are examined. This situation is reflected with the decrease of RR values detected for the Pro564 allele in the different Val34Leu genotypes (RR=10.5 in Val34Val34; RR=2.0 in Val34-Leu34). As far as the reported differences between Northern & Central Europe and Mediterranean populations, Moore suggested [22] that, under an interacting model, independent effects can only be detected depending on the allele frequencies in the population. So, the protective or risk effect of Leu34 allele described in the former can be masked in Mediterranean populations by other interacting polymorphisms such as Pro564 allele. Likewise, another factor that can also increase the probability of detecting independent

Table 3
Interaction analyses among Val34Leu and Pro564Leu polymorphisms.

Allele	Family sample according to patients' genotype (N° families)	TDT analyses	
		T/NT§	RR (95%CI) †
Pro564	Val34Val34 (50)	21/2	10.5 (2.78-39.71)#
	Val34Leu34 (40)	10/5	2 (0.90-4.45)
	Leu34Leu34 (8)	1/1	1 (0.14-7.10)
Val34	Pro564Pro564 (68)	13/9	1.4 (0.78-2.66)
	Pro564Leu564 (32)	7/7	1 (0.48-2.10)
	Leu564Leu564 (1)		

§ Number of transmitted (T) and untransmitted (NT) alleles. † Relative Risk with 95% Confidence Interval. # p<0.05.

associations is linkage disequilibrium. In this point, both low heterozygosity and high linkage disequilibrium are characteristic of Northern Europe as compared with Mediterranean region [44]; a fact that could explain positive associations in the former but not in the later.

Concerning STRs markers, we have detected a protective effect of the F13A1-STR*3.2 allele, the shortest in size, and a risk role of F13A1-STR*7 allele, the largest in size among common alleles. These findings are consistent with experimental data pointing out that the region around this STR is an effective enhancer of the F13A1 promoter [45]. F13A1-STR is located between DNA sequences for binding of myeloid-enriched factors (GATA-1 and Ets-1) responsible for the enhancer activity. F13A1-STR extension enlarges the distance between GATA-1 and Ets-1 binding site sequences possibly affecting the stability of binding factors and possibly varying the levels of FXIII expression.

The biological plausibility of our results makes them enough interesting to suggest additional studies testing the connection between the F13A1-STR variation, F13A1 levels and CAD risk. Likewise, the interactive effect between Pro564 and Val34 alleles detected in our IHD sample underlines the importance of considering allele-allele interaction models in future studies of genetic susceptibility to vascular diseases.

Conflict of interest statement

All authors declare that they have no competing interests.

Acknowledgments

This research has been supported by the Ministerio de Ciencia y Tecnología CGL2005-3391 and CGL2008-03955 projects; and the Ministerio de Ciencia y Tecnología Predoctoral Grant (BES-2009-016334) to RCT. All volunteers are gratefully acknowledged for sample donation.

References

- Keavney B. Genetic epidemiological studies of coronary heart disease. *Int J Epidemiol* 2002;31:730–6.
- Muszbec L, Yee VC, Hevessy Z. Blood coagulation factor XIII: structure and function. *Thromb Res* 1999;94:271–305.
- Yee VC, Pedersen LC, Le Trong I, Bishop PD, Stenkamp RE, Teller DC. Three-dimensional structure of a transglutaminase: Human blood coagulation factor XIII. *Proc Natl Acad Sci USA* 1994;91:7296.
- Lim BC, Ariens RA, Carter AM, et al. Genetic regulation of fibrin structure and function: complex gene-environment interactions may modulate vascular risk. *Lancet* 2003;361:1424–31.
- Kohler HP, Stickland MH, Ossei-Gerning N, Carter A, Mikkola H, Grant PJ. Association of a common polymorphism in the Factor XIII gene with myocardial infarction. *Thromb Haemost* 1998;79:8–13.
- Wartiovaara U, Perola M, Mikkola H, Tötterman K, Savolainen V, Penttilä A, et al. Association of FXIII Val34Leu with decreased risk of myocardial infarction in Finnish males. *Atherosclerosis* 1999;142:295–300.
- Kakko S, Elo T, Tapanainen JM, Huikuri HV, Savolainen MJ. Polymorphisms of genes affecting thrombosis and risk of myocardial infarction. *Eur J Clin Invest* 2002;32:643–8.
- Elbaz A, Poirier O, Canaple S, Chédru F, Cambien F, Amarencu P. The association between the Val34Leu polymorphism in the factor XIII gene and brain infarction. *Blood* 2000;95:586–91.
- Gemmati D, Serino ML, Ongaro A, Tognazzo S, Moratelli S, Resca R, et al. A common mutation in the gene for coagulation factor XIII-A (Val34Leu): a risk factor for primary intracerebral hemorrhage is protective against atherothrombotic diseases. *Am J Haematol* 2001;61:183–8.
- Hancer VS, Diz-Kucukkaya R, Bilge AK, Ozben B, Oncul A, Ergen G, et al. The association between factor XIII Val34Leu polymorphism and early myocardial infarction. *Circ J* 2006;70:239–42.
- Corral J, Gonzalez-Conejero R, Iñiesta JA, Rivera J, Martinez C, Vicente V. The FXIII Val34Leu polymorphism in venous and arterial thromboembolism. *Haematologica* 2000;85:293–7.
- Roldan V, Corral J, Marin F, Rivera J, Pineda J, González-Conejero R, et al. Role of factor XIII Val34Leu polymorphism in patients <45 years of age with acute myocardial infarction. *Am J Cardiol* 2003;91:1242–5.
- Canavy I, Henry M, Morange PE, Tiret L, Poirier O, Ebagosti A, et al. Genetic polymorphisms and coronary artery disease in the South of France. *Thromb Haemost* 2000;83:212–6.
- Doggen CJ, Reiner AP, Vos HL, Rosendaal FR. Two factor XIII gene polymorphisms associated with a structural and functional defect and the risk of myocardial infarction in men. *J Thromb Haemost* 2003;1:2056–8.
- Atherosclerosis, Thrombosis, and Vascular Biology Italian Study Group. No evidence of association between prothrombotic gene polymorphisms and the development of acute myocardial infarction at a young age. *Circulation* 2003;107:1117–22.
- Rubattu S, Di Angelantonio E, Nitsch D, Gigante B, Zanda B, Stanzione R, et al. Polymorphism in prothrombotic genes and their impact on ischemic stroke in a Sardinian population. *Thromb Haemost* 2005;93:1095–100.
- Bereczky Z, Balogh E, Katona E, Pocsai Z, Czuriga I, Széles G, et al. Modulation of the risk of coronary sclerosis/myocardial infarction by the interaction between factor XIII subunit A Val34Leu polymorphism and fibrinogen concentration in the high risk Hungarian population. *Thromb Res* 2007;120:567–73.
- Vokó Z, Bereczky Z, Katona E, Ádány R, Muszbek L. Factor XIII Val34Leu variant protects against coronary artery disease. *Thromb Haemost* 2007;97:458–63.
- Gallivan L, Markham AF, Anwar R. The Leu564 factor XIII variant results in significantly lower plasma factor XIII levels than the Pro564 variant. *Thromb Haemost* 1999;82:1368–70.
- Reiner AP, Schwartz SM, Frank MB, Longstreth Jr WT, Hindorf LA, Teramura G, et al. Polymorphisms of coagulation factor XIII subunit A and risk of nonfatal hemorrhagic stroke in young white women. *Stroke* 2001;32:2580–7.
- Polymeropoulos MH, Rath DS, Xiao H, Merril CR. Tetranucleotide repeat polymorphism at the human coagulation factor XIII A subunit gene (F13A1). *Nucleic Acid Res* 1991;19(15):4306.
- Moore JH. The Ubiquitous Nature of Epistasis in Determining Susceptibility to Common Human Diseases. *Hum Hered* 2003;56:73–82.
- Reiner AP, Frank MB, Schwartz SM, Linenberger ML, Longstreth Jr WT, Teramura G, et al. Coagulation factor XIII polymorphisms and the risk of myocardial infarction and ischaemic stroke in young women. *Br J Haematol* 2002;116:376–82.
- Kohler HP, Futers TS, Grant PJ. Prevalence of three common polymorphisms in the A-subunit gene of factor XIII in patients with coronary artery disease. *Thromb Haemost* 1999;81(4):511–5.
- Via M, López-Alomar A, Valveny N, González-Pérez E, Bao M, Esteban E, et al. Lack of association between eNOS gene polymorphisms and ischemic heart disease in the Spanish population. *Am J Med Genet A* 2003;116:243–8.
- Barral S, Lareu MV, Salas A, Carracedo A. Sequence variation of two hypervariable short tandem repeats at the D22S683 and D6S477 loci. *Int J Leg Med* 2000;113:146–9.
- Rousset F. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol Ecol* 2008;8:103–6.
- K. Belkhir, P. Borsa, L. Chikhi (1996–2004) GENETIX 4.04, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).
- Lange C, DeMeo D, Silverman EK, Weiss ST, Laird NM. PBAT: Tools for Family-Based Association Studies. *Am J Hum Genet* 2004;74(2):367–9.
- W.J. Guaderman, J.M. Morrison. QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies. <http://hydra.usc.edu/gxe>, 2006.
- Dean AG, Dean JA, Burton AH, Dicker RC. Epi InfoTM: a general purpose microcomputer program for health information systems. *Am J Prev Med* 1991;7:178–82.
- Warner D, Mansfield MW, Grant PJ. Coagulation Factor XIII and Cardiovascular Disease in UK Asian Patients Undergoing Coronary Angiography. *Thromb Haemost* 2001;85:408–11.
- Catto AJ, Kohler HP, Coore J, Mansfield MW, Stickland MH, Grant PJ. Association of a common polymorphism in the factor XIII gene with venous thrombosis. *Blood* 1999;93(3):906–8.
- Renner W, Köppel H, Hoffmann C, Schallmoser K, Stanger O, Toplak H, et al. Prothrombin G20210A, factor V Leiden, and factor XIII Val34Leu: common mutations of blood coagulation factors and deep vein thrombosis in Austria. *Thromb Res* 2000;99:35–9.
- Alhenc-Gelas M, Reny JL, Aubry ML, et al. The FXIII Val 34 Leu mutation and the risk of venous thrombosis. (Letter). *Thromb Haemost* 2000;84:1117–8.
- Le Gal G, Delahousse B, Lacut K, Malaviole V, Regina S, Blouch MT, et al. Fibrinogen Aalfa.Thr312Ala and factor XIII-A Val34Leu polymorphisms in idiopathic venous thromboembolism. *Thromb Res* 2007. doi:10.1016/j.thromres.2007.05.003.
- Balogh I, Szöke G, Kárpáti L, Wartiovaara U, Katona E, Komáromi I, et al. Val34Leu polymorphism of plasma factor XIII: biochemistry and epidemiology in familial thrombophilia. *Blood* 2000;96:2479–86.
- Margaglione M, Bossone A, Brancaccio V, Ciampa A, Di Minno G. Factor XIII Val34Leu polymorphism and risk of deep vein thrombosis. *Thromb Haemost* 2000;84:1118–9.
- Wells PS, Anderson JL, Scarvelis DK, Doucette SP, Gagnon F. Factor XIII Val34Leu variant is protective against venous thromboembolism: a HuGe review and meta-analysis. *Am J Epidemiol* 2006;164:101–9.
- Catto AJ, Kohler HP, Bannan S, Stickland M, Carter A, Grant PJ. Factor XIII Val 34 Leu: a novel association with primary intracerebral hemorrhage. *Stroke* 1998;29:813–6.
- Corral J, Iñiesta JA, González-Conejero R, Villalón M, Vicente V. Polymorphisms of clotting factor modify the risk for primary intracranial hemorrhage. *Blood* 2001;97:2979–82.
- Endler G, Funk M, Haering D, Lalouchek W, Lang W, Mirafzal M, et al. Is the factor XIII 34Val/Leu polymorphism a protective factor for cerebrovascular disease? *Br J Haematol* 2003;120(2):310–4.

- [43] Slowik A, Dziedzic T, Pera J, Figlewicz DA, Szczudlik A. Coagulation factor XIII Val34Leu polymorphism in patients with small vessel disease or primary intracerebral hemorrhage. *Cerebrovasc Dis* 2005;19(3):165–70.
- [44] Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, Caliebe A, et al. Correlation between Genetic and Geographic Structure in Europe. *Curr Biol* 2008;18:1241–8.
- [45] Kida M, Sourj M, Yamamoto M, Saito H, Ichinose A. Transcriptional regulation of cell type-specific expression of the TATA-less A subunit gene for human coagulation factor XIII. *J Bio Chem* 1999;274(10):6138–47.

Robert Carreras-Torres
Georgios Athanasiadis
*Unitat d'Antropologia, Facultat de Biologia,
Universitat de Barcelona, Barcelona, Spain*

Marc Via
*Institute for Human Genetics, University of California,
San Francisco, CA, USA*

Joan Trenchs
Magdalena Gayà-Vidal
Josep Santamaria
Esther Esteban
Pedro Moral*
*Unitat d'Antropologia, Facultat de Biologia,
Universitat de Barcelona, Barcelona, Spain*
*Corresponding author. Av. Diagonal 645. Dpt Biologia
Animal-Antropologia, 08028. Barcelona Spain. Tel.: +34 934021461.
E-mail address: pmoral@ub.edu (P. Moral).

9 February 2010

3.3 Resultats II: Carreras-Torres et al., 2014

Genetic risk score of *NOS* gene variants associated with myocardial infarction correlates with coronary incidence across Europe

Robert Carreras-Torres, Suman Kundu, Daniela Zanetti,

Esther Esteban, Marc Via, Pedro Moral.

PLoS ONE (2014). doi:10.1371/journal.pone.0096504

3.3.1 Resum en Català

La puntuació (*score*) de risc genètic de les variants dels gens *NOS* associades a l'infart de miocardi es correlaciona amb la incidència coronària a Europa

La incidència i mortalitat de les malalties coronàries al continent europeu presenta un gradient poblacional des del sud (Espanya i França), on els valors són més baixos, cap al nord (Finlàndia i Regne Unit), on són quatre vegades superiors. Aquesta gradació poblacional ha estat descrita, entre d'altres, pel projecte MONICA de l'OMS que va recollir dades de factors de risc anomenats tradicionals lligats a la dieta i a l'estil de vida. La distribució poblacional d'aquests factors de risc tradicionals només explica aproximadament un 30% de la incidència observada, i dóna peu a una fal·làcia ecològica, anomenada paradoxa Mediterrània, ja que poblacions com França i Finlàndia amb nivells similars de factors de risc presenten incidències ben diferents. Estudis anteriors han intentat explicar aquesta variació en la incidència a partir de les freqüències poblacionals de marcadors genètics de risc, però la majoria de marcadors observats no presenten correlacions poblacionals.

En aquest estudi es proposa l'ús de les puntuacions de risc genètic (GRS, de l'anglès *Genetic Risk Scores*) per tal d'aglutinar la informació genètica de risc en un sol paràmetre quantitatiu i observar la seva correlació poblacional amb la incidència cardiovascular. Amb aquest objectiu, els gens candidats escollits han estat els gens de les sintases d'òxid nítric (*NOS*, de l'anglès *Nitric Oxide Synthases*), ja que són els responsables d'un paper clau en el trencaclosques cardiovascular; la disponibilitat de l'òxid nítric. L'òxid nítric és responsable de l'hemostàsia vascular regulant la vasopressió, i també intervé en l'agregació plaquetària. Un conjunt de 61 SNPs presents als tres gens *NOS* (*NOS3*, *NOS1* i *NOS2A*) van ser analitzats per associació amb l'infart de miocardi en 2165 casos i 2153 controls. D'aquests, 4 SNPs genèticament independents ($LD < 0.2$) van presentar indicis d'associació amb la malaltia ($p < 0.1$). La puntuació de risc genètica a partir d'aquests 4 SNPs va explicar una contribució a l'infart de miocardi inferior a l'1%, i una capacitat de discernir entre casos i controls nul·la o molt limitada ($AUC < 0.53$). Per contra, en un conjunt de 34 mostres poblacionals d'Europa i la Mediterrània, la mitjana poblacional d'aquesta puntuació de risc genètica presentava una distribució concèntrica amb valors més baixos a la regió nord-oest del Mediterrani (est Pirineus, illes de Còrsega i Sardenya i sud d'Itàlia) i valors més elevats al allunyar-nos d'aquesta regió en funció de la distància geogràfica ($p = 0.024$). Aquest GRS poblacional es correlaciona positivament amb la incidència coronària registrada en 10 poblacions del projecte MONICA ($p < 0.01$) i n'explica fins a un 67% en dones i un 86% en homes ($p < 0.001$).

Aquesta gran contribució del GRS a la incidència coronària al continent europeu deu ser un reflex de la colinearitat d'aquesta puntuació genètica de risc amb molts altres factors de risc amb la mateixa distribució

poblacional. Tenint en compte que els factors de risc tradicionals només expliquen el 30% de la incidència, els factors de risc amb la mateixa distribució poblacional que el GRS i que n'ajuden a explicar fins al 65-85% deuen de ser també genètics. Aquest fet ens indica que els majors causants genètics de la malaltia cardiovascular es troben dins l'espectre de freqüència comuna ($>5\%$) i que l'estudi de la seva distribució, mitjançant els GRS, ens pot ser molt útil per elaborar estratègies de prevenció en salut pública.

3.3.2 Informe dels directors de tesi de la participació del doctorant en l'article i de la singularitat de l'article en la elaboració de la tesi

Dr. Pedro Moral Castrillo, professor del Departament de Biologia Animal de la Universitat de Barcelona, i Dr. Marc Via García, professor del Departament de Psiquiatria i Psicobiologia Clínica de la Universitat de Barcelona, supervisors de la tesi doctoral “**Variació genòmica de les Sintases d’Òxid Nítric i de regions associades amb Risc Cardiovascular en Poblacions Humanes**”, manifesten que l’estudiant de doctorat Robert Carreras Torres ha participat en el treball titulat “*Genetic risk score of NOS gene variants associated with myocardial infarction correlates with coronary incidence across Europe*”, publicat a la revista *PLoS ONE*, en les següents tasques:

- Disseny i selecció de les mostres i marcadors analitzats (90%).
- Recol·lecció i preparació de les mostres poblacionals (90%).
- Creació de las bases de dades genètiques i selecció de resultats disponibles a la bibliografia per tal de realitzar comparacions (100%).
- Anàlisis estadístics (100%).
- Redacció de l'article (95%).

Cap dels coautors de l'article ha utilitzat els resultats d'aquest estudi de manera implícita o explícita per elaborar una altra tesi doctoral. Per tant, aquest article forma part de manera exclusiva de la tesi doctoral de Robert Carreras Torres.

Signat per Dr. Pedro Moral Castrillo i Dr. Marc Via García
Barcelona, 22 Abril 2014.

3.3.3 Article II

Genetic risk score of *NOS* gene variants associated with myocardial infarction correlates with coronary incidence across Europe

Robert Carreras-Torres^a, Suman Kundu^b, Daniela Zanetti^a, Esther Esteban^{ac}, Marc Via^{d*}, Pedro Moral^{ac*}.

^aDepartament Biologia Animal - Antropologia, Universitat de Barcelona, Barcelona, Spain;

^bDepartment of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands;

^cInstitut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain.

^dDepartament Psiquiatria i Psicobiologia Clínica and Institute for Brain, Cognition and Behaviour (IR3C), Universitat de Barcelona, Barcelona, Spain;

* These authors contributed equally to this work.

Abstract

Coronary artery disease (CAD) mortality and morbidity is present in the European continent in a four-fold gradient across populations, from the South (Spain and France) with the lowest CAD mortality, towards the North (Finland and UK). This observed gradient has not been fully explained by classical or single genetic risk factors, resulting in some cases in the so called Southern European or Mediterranean paradox. Here we approached population genetic risk estimates using genetic risk scores (GRS) constructed with single nucleotide polymorphisms (SNP) from nitric oxide synthases (*NOS*) genes. These SNPs appeared to be associated with myocardial infarction (MI) in 2165 cases and 2153 controls. The GRSs were computed in 34 general European populations. Although the contribution of these GRS was lower than 1% between cases and controls, the mean GRS per

population was positively correlated with coronary incidence explaining 65-85% of the variation among populations (67% in women and 86% in men). This large contribution to CAD incidence variation among populations might be a result of colinearity with several other common genetic and environmental factors. These results are not consistent with the cardiovascular Mediterranean paradox for genetics and support a CAD genetic architecture mainly based on combinations of common genetic polymorphisms. Population genetic risk scores is a promising approach in public health interventions to develop lifestyle programs and prevent intermediate risk factors in certain subpopulations with specific genetic predisposition.

Keywords

European CAD incidence, genetic risk scores, population distribution, genetic Mediterranean paradox,

Introduction

The development of coronary artery diseases (CAD) is the result of complex interactions between numerous environmental factors and genetic variants at many loci. Consequently, understanding CAD needs a multidisciplinary research effort.

Initially, epidemiologic research was largely based on cohort studies and clinical trials identifying and quantifying the relative importance of risk factors. As the World Health Organization (WHO) stated, the main identified risks for heart disease are behavioral factors: unhealthy diet, physical inactivity, tobacco use or harmful use of alcohol are present in about 80% of coronary events [1]. Different consortia had contributed to the development

of estimation risk charts based on traditional risk factors (TRF), such as the Framingham Risk Score [2], the Reynolds Risk Score [3], the Prospective Cardiovascular Munster Heart Study (PROCAM) [4] and the Systematic Coronary Risk Evaluation (SCORE) system [5]. The prediction ability of these risk estimates is moderate-good [6]. Nevertheless, it has been estimated that nearly 15-20% of CAD patients are misclassified as “low risk” by TRF-based charts [7].

Genetics provided a plausible explanation for disease outcome in people without previous symptoms, and to the observed symptomatic variability in people exposed to similar behavioral risk factors. At the time that genetic disease architecture was partially unveiled, the idea of improving cardiovascular risk prediction was targeted. However, lack of replication, modest genetic risks and the small proportion of heritability explained by genome-wide association (GWA) studies have prevented the improvement of genetic CAD prediction [8,9]. Even polygenetic risk scores, proposed as a way of improving already existing estimation risk charts, have not been completely satisfactory in different epidemiologic samples [10–16]. All the approaches mentioned above used individuals as the units of analysis.

From another perspective, in which general populations were the units of analysis, ecological/epidemiological (from now on: eco-epidemiologic) studies have assessed the disease population burden through the geographic distribution of CAD incidence and risk parameters. In this field, an important contribution has been made by the international WHO MONICA Project [17] which surveyed 38 populations from 21 countries. As far as CAD mortality was concerned, early cross-sectional studies reported an existing four-fold gradient across populations in the European continent, from the South (Spain and France) with the lowest CAD mortality towards the North (Finland and

UK) [18–20]. Several studies have attempted to correlate this observed CAD incidence variation with the distribution of both traditional and genetic risk factors. It was assessed that classical risk factors contribute to 30-40% of CAD population incidence. Furthermore, an ecological fallacy was described when populations with remarkable differences in coronary mortality had similar classical risk factor levels, especially animal fat intake [21–23], leading to the idea of a French, southern European or Mediterranean paradox. However, a more recent study pointed towards wine consumption as an alternative explanation for this phenomenon [24]. The lack of strong correlations between CAD incidence and traditional risk factors suggests that genetic variation could be behind the interpopulation gradient of coronary mortality.

Some researchers have analyzed the geographic distribution of genetic risk variants to predict variation in both TRF and CAD mortality. So far, these studies have demonstrated that only the *APOE**E4 risk allele is clearly correlated with CAD incidence among MONICA populations in the European continent. The lack of correlation for the vast majority of tested genetic markers led to extending the Mediterranean paradox to genetics [25,26].

This study proposes an alternative approach to estimating the population genetic CAD burden using genetic risk scores (GRS). GRS appear to be a more realistic tool because they summarize the potential multiple risk genetic influences into a single quantitative parameter and do not depend on single genetic variants. As far as we know, no previous epidemiology studies have considered GRS as ecological risk predictors of CAD incidence. In order to describe geographic patterns of genetic risk variation, this study maps the population mean GRS using the geostatistical method known as

kriging. Kriging is a geostatistic method for interpolating the spatial distribution of a variable by means of linear regression. Contour maps depicting interpolated spatial distribution patterns have previously been used to represent biologic anthropological data [27].

In order to explore these population approaches, we focused on a key piece of the CAD jigsaw: the role of nitric oxide (NO) in regulation of vascular tone homeostasis, tissue perfusion, and platelet aggregation [28,29]. Three nitric oxide synthases (*NOS*) are responsible for NO availability: *endothelial NOS* (*eNOS* or *NOS3*), *neuronal NOS* (*nNOS* or *NOS1*) and *inducible NOS* (*iNOS* or *NOS2A*). *NOS3* and *NOS1* are constitutively expressed mainly in vessel endothelium and neuronal tissue, respectively [28]. Both are acutely regulated through reversible calcium-calmodulin binding. Conversely, *NOS2A* is activated through inflammatory signals in critical situations, mainly in the vessel endothelium and macrophages. However, induction of high-output *NOS2A* may lead to direct oxide cell toxicity or interfere with the beneficial activities of constitutive *NOS* isoforms [28]. Besides *NOS*, sONE is an antisense mRNA derived from a *NOS3AS* or *ATG9B* transcript unit on the complementary DNA strand from which the *NOS3* mRNA is transcribed. *ATG9B* and *NOS3* genes are oriented in a tail-to-tail configuration, and the mRNAs encoding sONE and *NOS3* overlap for 662 nucleotides. There is evidence supporting a role for *ATG9B* in the post-transcriptional regulation of *NOS3* expression [29]. According to the Human Genome Epidemiology (HuGE) Navigator browser (www.hugenavigator.net), the *NOS3* gene is the second most reported gene for CAD, with 134 related papers, and the fourth most reported for myocardial infarction, with 74 reports (February, 2014). Variation in *NOS3* has also been tested for hypertension and diabetes. However, large meta-

analysis on *NOS3* gene polymorphisms reported inconsistent results for CAD [30-33] and hypertension [34,35] showing an excess of positive results associated with small sized studies and Asian populations. *NOS1* and *NOS2A* genes have been associated with CAD, hypertension, inflammation and diabetes [36-39], but also with a broader spectrum of diseases. All the above mentioned association studies only considered a few polymorphisms per gene region, and no one surveyed these chromosomal regions with a dense genetic coverage.

In this context, the present work had three main objectives. The first objective was to assess the prediction ability of GRS computed from *NOS* risk variants detected by association analyses among CAD patients and control samples. The second objective was to estimate, for the first time, the population *NOS* CAD burden computing GRS in general population samples, and to describe geographic patterns of GRS across Europe and the Mediterranean area. The third objective was to assess whether the population GRS are able to predict ecological risk. With this aim, population GRS were correlated with population distribution of CAD incidence and other traditional risk factors reported by the MONICA Project.

Materials and methods

Ethics Statement

The study has been specifically approved by the Ethical Committee of the University of Barcelona (Institutional Review Board: IRB00003099) and all the participants provided a written informed consent.

Association and prediction analyses sample description

DNA samples of 324 myocardial infarction (MI) patients and 366 controls

from the general Spanish population, obtained from the Spanish National DNA Bank (NDB) (www.bancoadn.org), were genotyped in this study. This sample will be referred to from now on as NDB cardiovascular (NDBC) sample. Additionally, genotype data from four European matched case-control samples from the Myocardial Infarction Generation (MIGen) Consortium [FINRISK (Finland), MDCS (Sweden), ATVB (Italy) and Regicor (Spain)] were obtained through the database of Genotypes and Phenotypes (dbGAP; www.ncbi.nlm.nih.gov/gap) [40]. In summary, a total of 2575 MI cases and 2617 controls were used in this stage. Extensive details on the clinical characteristics of these samples have been previously described (www.bancoadn.org/en/introNCa.htm) [40]. Briefly, fatal and nonfatal MI were reported or diagnosed by general practitioners based on autopsy reports, electrocardiographic data, cardiac biomarkers, and additional clinical information.

Eco-epidemiologic analyses sample description

A total of 34 populations (n=1663 individuals) from Europe, North Africa, and the Middle East were analyzed (see Figure S1). Thirty populations (n=1298) corresponded to healthy unrelated individuals of both sexes that were genotyped in the present study and whose four grandparents had been born in the same geographical region. Additionally, genetic data from four other European samples from the 1000 Genomes Project [41] were included in the analyses.

Polymorphisms and genotyping

The NDBC sample (324 cases and 366 controls) and 1298 individuals from

the 30 general populations were genotyped for 78 single nucleotide polymorphisms (SNP) using a GoldenGate Genotyping Assay (Illumina Inc., San Diego, CA). This SNP set was selected as being representative of the common variation in the three genomic regions of the *NOS* genes, with an average coverage of 1 SNP every 5kb with a minor allele frequency higher than 0.05 (MAF>0.05) in the CEU population as reported in the HapMap project (www.hapmap.org). Out of the 78 determined SNPs, 13 were located in chromosome 7 spanning 41.4 kb in the *NOS3* and *ATG9B* genes region; 43 SNPs in chromosome 12 that include the *NOS1* gene along 177.4 kb, and 22 SNPs in chromosome 17 covering 92.2 kb in the *NOS2A* gene region. SNP details are shown in Supporting Table (ST) Tab 1.

Genotype data for the MIGen samples were generated in the corresponding original project using the Affymetrix 6.0 GeneChip [40].

Quality control and imputation

Genotyping rate, allele frequencies, and deviations from the Hardy-Weinberg equilibrium (HWE) were calculated using PLINK software [42]. SNPs with a genotyping rate lower than 0.75 or not polymorphic in any sample were removed from the analysis. Individuals with a genotyping rate lower than 0.75 or not genetically homogeneous compared with individuals of the same population group were also removed for the analysis. Missing genotypes were inferred using MACH 1.0 software [43] taking as reference the rest of the genotypes ascertained in the same population. Linkage disequilibrium was calculated and visualized using Haploview software [44].

Datasets from the MIGen study already included 27 out of the 78 SNPs in the *NOS* regions. In order to have the same genetic information, SNPs not directly genotyped in the MIGen samples were imputed using two different

imputation softwares, MACH 1.0 [43] and IMPUTE2 [45]. In both imputations the computational effort was controlled performing 200 algorithm iterations when phasing and imputing data sets, and considering 300 haplotypes to use as templates when phasing observed genotypes. This imputation effort is four times higher than the standard effort recommended by software developers. Phased chromosomes from the most similar 1000 Genomes Project samples were used as reference panels: the FIN sample for the FINRISK case-controls, the TSI sample for the ATVB and Regicor case-controls, and the CEU sample for the MDCS case-controls.

As a control approach to validating the genotyping strategy of this study (SNPs selected as representative of *NOS* regions common variation), in our population sample from Central Italy (CIT) we imputed all the variation described in TSI sample from the 1000 Genomes Project in the studied three chromosomal regions. And then we checked the imputation quality indices regarding allele frequency thresholds.

Association and prediction analysis

A two-step analysis of association and prediction was performed with the *PredictABEL* R package [46]. These analyses were performed in duplicate, in the MACH imputed data set and in the IMPUTE2 imputed data set. In the first step, associations were tested by logistic regression analysis in the three case-control samples with the largest sample size: FINRISK from northern Europe, and ATVB and Regicor from southern Europe. The other two case-control samples (MDCS and NDBC) were kept as cross-validating samples for the posterior prediction step. In the association analyses, only SNPs with a LD measure (r^2) lower than 0.8 between pairs and imputation quality indexes (r^2 for MACH 1.0 and i for IMPUTE2) higher than 0.6 in all three

used case-control samples and the two imputation methods were included. Estimates of beta coefficients for each SNP were obtained using multivariate logistic regression analyses and adjusted for age, gender and the remaining genetic variables. In order to get a single robust estimate of the level of association for each genetic marker, a meta-analysis of the three previous association analyses (n=4318) was conducted with the METAL software [47].

In the second step, NOS genetic risk scores for MI were computed in all five case-control samples. Risk scores were constructed using allele dosages of low P value ($p < 0.1$) risk alleles identified in both meta-analysis from MACH and IMPUTE2 data sets. Thus, homozygotes for the reference allele were coded as 0 and homozygotes for the risk allele as 2. The risk SNPs were pruned by LD (r^2) lower than 0.2 in order to obtain a set of unequivocally independent SNPs to calculate the risk scores. This LD pruning was performed by Tagger [48], implemented in Haploview [44], preferentially picking the SNPs with the lowest P value. As an approach to checking for the epidemiological relevance of the estimated risk scores, predictive models were constructed based only on these NOS risk scores in all five case-control samples. These models were performed to assess the fraction of interindividual variance of the MI affection status explained by NOS risk score through Nagelkerke's R^2 . Moreover, discrimination accuracy of the NOS risk score between patients and healthy controls was estimated as the area under the receiver operating characteristic (ROC) analysis curve (AUC) index.

Eco-epidemiologic analysis

NOS genetic risk scores for MI were computed in the general population

samples as previously described. *NOS* risk scores were tested for normality in each population sample using the *nortest* R package. Spatial distribution of mean risk score across populations was mapped using the geostatistical method known as kriging from the ArcGIS software (ESRI, Redlands, CA, USA). Since anisotropy was not detected in the semivariogram, we used the ordinary spherical interpolation kriging method [27]. The covariation of the observed spatial distribution with geography was assessed by Moran's I and Geary's C randomization tests for spatial autocorrelation [49] using *ade4* R package. Also, the spatial structure of mean risk scores was assessed using correlograms, which estimate autocorrelation coefficients for different spatial relationships, with the PASSaGE software [50]. Population pair relationships were classified in different classes representing increasingly larger distances. Autocorrelation coefficients were then calculated for each distance class and plotted against distance [51]. Data related to coronary event rates and prevalence of traditional risk factors in middle-aged individuals were compiled from the MONICA Project [52] for the 11 European populations genetically tested here. Among the genotyped populations in this work, four of them (POL, NFR, SFR, and CAT) had a MONICA counterpart and seven additional populations (ORK, GBR, CEU, FIN, TSI, CIT and NBH) had a MONICA population within a 200-km radius or from the same country. The CEU sample was considered counterpart of the MONICA German-Bremen population according to Lao et al. [53]. Average annual coronary event rates over 5 years and average levels of systolic blood pressure (SBP), total cholesterol (TCH), body-mass index (BMI), and daily smoking rate (SMK) by gender were obtained from Kuulasmaa et al. [52].

Spearman's correlation and univariate linear regression analyses were

performed to estimate the contribution of traditional risk factors to population variation in coronary event rates using the *stats* R package. Since *NOS* genes are involved in blood pressure homeostasis, *NOS* genetic parameters (population mean risk scores and allele frequencies) were also correlated and regressed with population variation in coronary event rate and systolic blood pressure levels. Moreover, since geography could underlie the distribution pattern of several environmental and genetic risk factors, latitude and longitude were also tested for correlation and regression with coronary event rates. Finally, multivariate regression analyses were performed with factors that were significant in univariate analyses in order to estimate the contribution of genetic risk factors beyond geography.

Results

Genotyping, quality control, and imputation.

Genotyping rates and status for the 78 SNPs initially tested in our samples are shown in ST Tab 1. Genotyping rates ranged from 81.01 to 91.21%. Ten SNPs were not successfully genotyped, and three SNPs were not polymorphic in the tested populations. These 13 SNPs were removed from the study. As for the data coming from the international project, genotyping status and imputation quality indexes are presented in ST Tab 1. Four SNPs had an imputation quality lower than 0.6 in at least one case-control sample used in the association analyses. Hence, 61 SNPs were considered consistent for analytical epidemiologic analyses.

After quality control, a total of 5096 samples for the epidemiologic survey and 1298 for the population analysis were included. The largest case-control sample was ATVB with a total of 3352 individuals, and the smallest samples were FINRISK and MDCS with 339 and 184 individuals (ST Tab 2). Among

the general populations, sample sizes ranged from 32 to 50 individuals except the populations from the 1000 Genomes project (n=85-98) as can be seen in ST Tab 3.

Minor allele frequencies (MAF) can be found in ST Tab 4 and 5 for case-control samples imputed with MACH and IMPUTE2 respectively, and in ST Tab 6 for population samples. None of the SNP showed significant departures from Hardy-Weinberg equilibrium expectations in any case-control or population sample (data not shown). Concerning linkage disequilibrium patterns, our data indicate that the three *NOS* genes are not regions of high LD. For instance, the LD pattern of the three *NOS* regions in the CEU sample can be visualized in Figure S2 for *NOS3* gene, Figure S3 for *NOS1* gene and Figure S4 for *NOS2A* gene. LD values were similar in the different case-control samples used in this study. After applying the LD pruning criteria for association analysis, 38 SNPs with low LD ($r^2 < 0.8$) were considered.

Assessing the validity of our genotyping strategy, 71% of the common variants (MAF>0.1) present in the 1000 Genomes TSI sample were imputed with high accuracy (MACH $r^2 > 0.75$) in our population samples from Central Italy, CIT (ST Tab 7). Hence, this result indicates that our genotyping strategy (1SNP each 5kb) efficiently captures more than 70% of the common variation reported by the 1000 Genomes Project in the three genomic regions.

Meta-analysis and interindividual prediction analysis.

Five SNPs had low P value ($p < 0.1$) in both MI meta-analysis from MACH and IMPUTE2 data sets, four in the *NOS3* and one in the *NOS1* gene regions (highlighted in ST Tab 4 and 5). After LD pruning criteria ($r^2 < 0.2$) to select

completely independent SNPs, four SNPs remained dropping one SNP from *NOS3*. Genetic effects of these four SNPs with their standard error were similar for meta-analyses from MACH and IMPUTE2 data sets (Table 1). Association parameters of all tested markers for both multivariate logistic regression analyses and for both meta-analysis are shown in ST Tab 4 and 5.

ID reference	Gene	Alleles	Risk Allele	MACH Effect	SE	P value	IMPUTE2 Effect	SE	P value
rs3793342	NOS3	A/G	G	0.18	0.10	0.071	0.20	0.10	0.046
rs1799983	NOS3	T/G	G	0.44	0.16	0.006	0.53	0.18	0.004
rs3918188	NOS3	A/C	C	0.29	0.09	0.001	0.30	0.09	0.001
rs3782219	NOS1	T/C	C	0.15	0.09	0.088	0.18	0.09	0.055

Table 1 – Meta-analyses association parameters of risk variants with P value<0.1 in both analysis (from MACH and IMPUTE2 datasets) after being pruned by LD lower than 0.2. SE: Standard Error.

The assessment of interindividual predictive ability was almost identical using MACH and IMPUTE2 imputations and showed limited power in differentiating between cases and controls (Table 2 for MACH data set, and ST Tab 8 for IMPUTE2 data set). The distribution of *NOS* risk scores was similar in cases and controls, showing almost overlapping distributions. Besides, *NOS* risk score explained less than 1% of interindividual variance in affection status according to the Nagelkerke’s R². *NOS* risk score was only a significant discriminating factor in the ATVB case-control sample. This was also reflected by the discrimination accuracy assessed through AUC index slightly higher than 0.5 in this sample.

Case-control sample	Mean risk score \pm SD [min - max]		Nagelk R ²	AUC [95%CI]
	Cases	Controls		
FINRISK (Finland)	6.11 \pm 0.80 [4.00 - 7.97]	6.05 \pm 0.85 [3.96 - 7.98]	<0.01	0.517 [0.455-0.578]
ATVB (Italy)	5.88 \pm 0.68 [3.17 - 7.81]	5.81 \pm 0.70 [3.60 - 7.83]	<0.01	0.527 [0.508-0.547]
Regicor (Spain)	5.79 \pm 0.70 [3.88 - 7.46]	5.88 \pm 0.71 [3.92 - 7.87]	<0.01	0.470 [0.425-0.515]
MDCS (Sweden)	5.84 \pm 0.72 [4.01 - 7.67]	5.96 \pm 0.64 [4.47 - 7.75]	0.01	0.448 [0.364-0.532]
NDBC (Spain)	5.79 \pm 0.84 [4.00 - 8.00]	5.85 \pm 0.88 [3.00- 8.00]	<0.01	0.469 [0.426-0.513]

Table 2 – *NOS* genetic risk score (GRS) distribution for cases and controls and discrimination accuracy for MACH imputed dataset. SD: Standard Deviation; Nagelkerke’s R²: explained interindividual variance of MI by *NOS* risk score predictive model; Nagelk R²: Nagelkerke’s R²; AUC: Area Under the receiver operating characteristic (ROC) curve; CI: Confidence Interval.

Population distribution of NOS risk scores

The geographical distribution of mean *NOS* risk scores across European and Mediterranean populations is represented in a smoothed spherical contour map in Figure 1. Population values are shown in ST Tab 9. The lowest score values corresponded to southwestern European populations, specifically to the islands of Corsica and Sardinia (<5.5 risk alleles), North-East Spain (5.65) and South Italy (5.69). Variation in the map fitted a global pattern of concentric distribution departing from a center of low risk score values in the North-West of the Mediterranean Basin and gradually increasing according to geographical distances. This concentric pattern significantly covaried with geography as reflected by the spatial autocorrelation analysis (p=0.024). At European continental level, the pattern is consistent with a gradual increase towards North and North-East, with highest values in Great Britain (6.04), Poland (6.07) and Finland (6.14). A similar cline is also observed in the northern shore of the Mediterranean, with increasing values from Spain to Turkey (6.14) and Middle-Eastern populations, GJD (6.02) and BJD

(6.22). This clinal pattern in the European continent and the Middle East was statistically assessed plotting Moran's I and Geary's C autocorrelation coefficients by distance between population pairs (Figure S5). Six distance classes of population pairs with an average of 30 observations per class were obtained. Autocorrelation coefficients for population pairs at short distances denoted significant positive autocorrelations while for population pairs at long distances Moran's I and Geary's C coefficients detected negative autocorrelations. Concerning the distribution between the northern and southern Mediterranean shores, a gradual variation can be observed in the westernmost part (Spain and Morocco), but the pattern is sharper in the central part of the region (i.e. Tunisia and Italy).

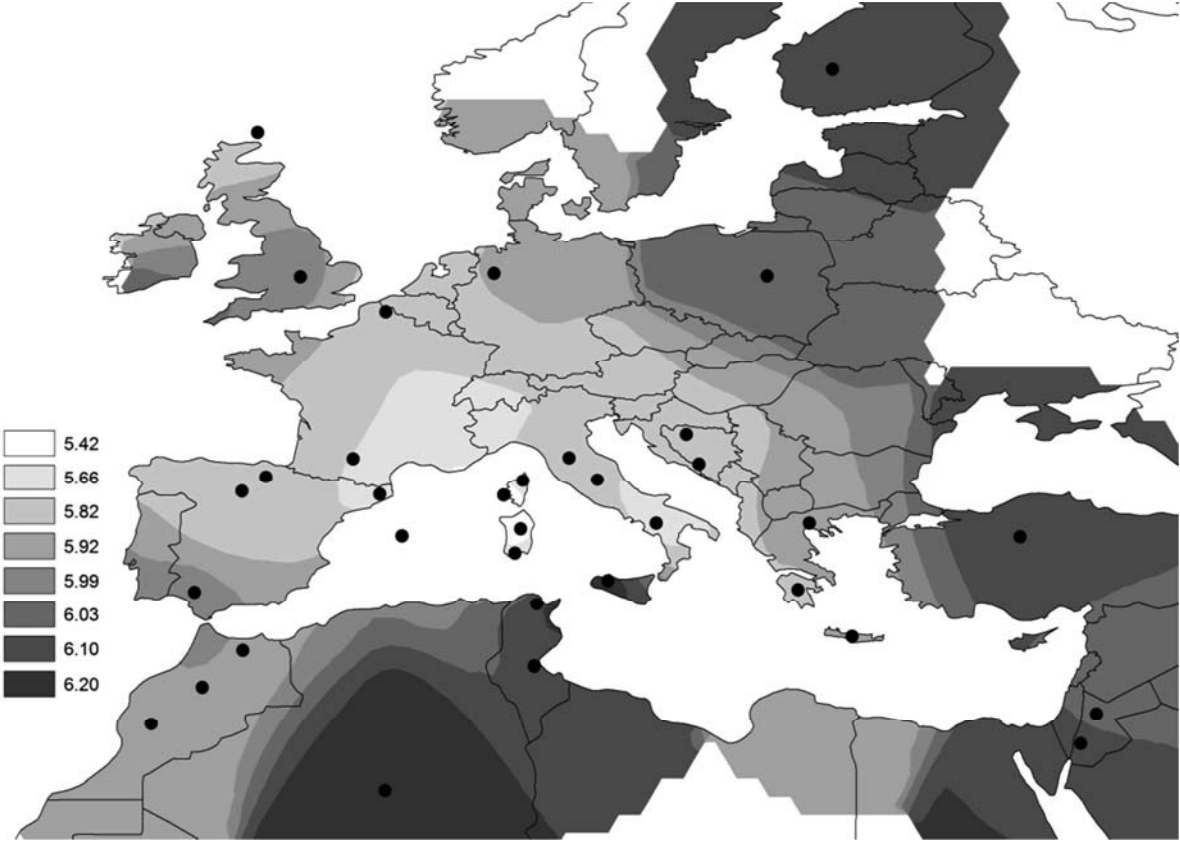


Figure 1 – Contour map of NOS risk score in the European and Mediterranean samples.

Eco-epidemiology of NOS gene variation and cardiovascular events.

The eleven populations with both *NOS* genotype data and MONICA information are shown in ST Tab 10. From the MONICA parameters, only daily smoking rate appeared as slightly correlated with coronary event rates in women ($\rho=0.57$; $p=0.064$) and explained 39% of the population variance of coronary event rates ($p=0.022$).

Mean *NOS* risk score values in the eleven populations considered were positively correlated with coronary event rates in men ($\rho=0.82$; $p<0.01$) (Figure 2A) and women ($\rho=0.76$; $p<0.01$) (Figure 2B).

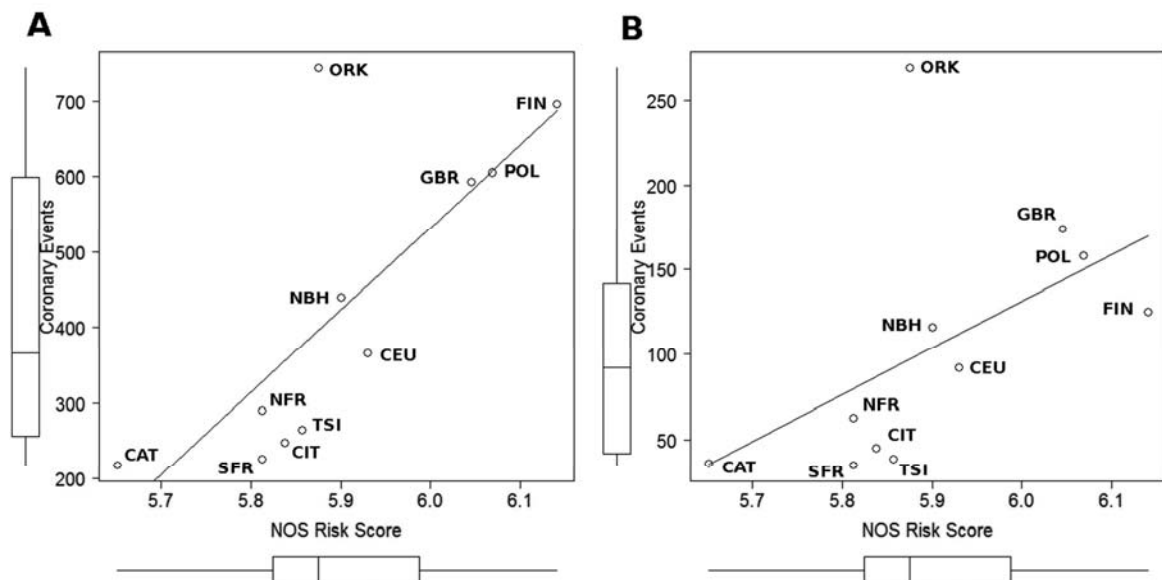


Figure 2 – Correlation plots between average *NOS* risk scores and coronary events in men (A) and women (B). Coronary events: rates per 100,000 people from the MONICA project.

In these figures, the ORK sample appeared as substantially different from the others. In the regression analysis, variation in mean risk scores explained 53% of interpopulation variance in coronary event rates in men and 19% in women (Table 3), and the ORK sample was confirmed as an outlier sample (Bonferroni $p<0.01$ for both men and women). The ORK sample and its

MONICA counterpart were excluded from the regression analyses. The outlier character of the Orkney Islands sample was probably due to an island genetic drift phenomenon when only a few markers are analyzed. Following this trend, any MONICA parameter was correlated with coronary events in the remaining ten MONICA samples, and the *NOS* risk score of the ten continental samples significantly explained 86% of coronary events in men and 67% in women (Table 3). Regarding SBP, 35% for men and 27% for women of the population variance was accounted for by the *NOS* risk score (Table 3). Individually, frequency distributions of 11 SNPs were associated with population coronary event rates or systolic blood pressure levels in men or women as estimated by correlation and univariate regression analyses (Table 3). Eight SNPs were correlated with coronary events in both men and women, explaining a remarkable proportion of the coronary rates variance: 34-67% in men and 36-52% in women. Out of these eight SNPs, 5 belonged to the shortest (42kb) region examined comprising the *NOS3* and *ATG9B* genes. This region was initially tested by nine SNPs, indicating that most of the tested genetic variation of this region had a similar geographic distribution pattern to coronary event rates. Among these SNPs correlating with CAD incidence, only the G risk allele of the rs1799983 was included in the risk score. Moreover, the same allele was the only one to be positively correlated with SBP in both men and women, explaining similar proportions of interpopulation variation, 40% in men ($p<0.05$) and 51% in women ($p<0.01$) (Table 3).

Since north-to-south and east-to-west geographic patterns of variation in genetic and environmental risk factors could underlie the observed associations between polymorphisms in *NOS* regions and coronary events, we then assessed the influence of geography in the distribution of coronary

events. Latitude was strongly correlated in men ($\rho=0.87$; $p<0.01$) and women ($\rho=0.75$; $p<0.05$) explaining a high proportion of population incidence (66% in men and 38% in women).

			Correlation and univariate linear regression analysis in men				Correlation and univariate linear regression analysis in women			
			Coronary event rate		SBP		Coronary event rate		SBP	
			ρ	R ²	ρ	R ²	ρ	R ²	ρ	R ²
NOS GRS			0.82**	0.53**	ns	0.36*	0.76**	0.19'	0.52'	0.28'
Continental NOS GRS			0.93***	0.86***	ns	0.35*	0.85**	0.67**	ns	0.27'
ID reference	Gene region	MAF Allele								
rs11771443	NOS3	T	0.68*	ns	ns	ns	ns	ns	ns	ns
rs1799983	NOS3	T	-0.85**	0.61**	-0.80**	0.40*	-0.79**	0.45*	-0.86**	0.51**
rs753482	NOS3	G	-0.82**	0.65**	ns	ns	-0.71*	0.52*	ns	ns
rs7830	ATG9B	T	0.84**	0.56**	ns	ns	0.74*	0.49*	ns	ns
rs2373929	ATG9B	T	0.78**	0.49*	ns	ns	0.63*	ns	ns	ns
rs13307588	ATG9B	A	-0.83**	0.48*	ns	ns	-0.72*	0.38*	ns	ns
rs10774907	NOS1	A	0.87**	0.67**	ns	ns	0.81**	0.64**	ns	ns
rs2271986	NOS1	T	0.65*	ns	ns	ns	0.71*	ns	ns	ns
rs1889024	NOS2A	G	ns	0.33*	ns	ns	0.63*	ns	ns	ns
rs953527	NOS2A	A	ns	ns	ns	ns	ns	ns	0.65*	ns
rs8072199	NOS2A	T	0.75*	0.34*	ns	ns	0.65*	0.36*	ns	ns

Table 3 - Spearman's Rho and % of explained interpopulation variance (adjusted R²) of coronary event rates and SBP by gender of significantly correlated minor allele frequency (MAF) variants and population NOS risk scores. ': P value<0.1; *: P value<0.05; **: P value<0.01; ***.P value<0.001. ns: non-significant.

Grouping the different parameters that were independently correlated with coronary events rates, a multivariate analysis was performed including the NOS risk score and the latitude in both male and female models. The result

showed latitude was no longer significant, and the proportions of population coronary incidence for both men and women accounted for by the multivariate models were not higher compared with the *NOS* risk score univariate models. For men, the multivariate model explained 85% of population coronary incidence ($p < 0.001$; *NOS* risk score $p = 0.01$, latitude $p = 0.45$). And for women, it explained 62% of the population coronary events rate ($p = 0.01$; *NOS* risk score $p = 0.04$, latitude $p = 0.82$).

Discussion

This study analyzes the role of molecular variation from *NOS* genes in cardiovascular patients and assesses the population distribution of genetic risk scores as an ecological predictor of the CAD burden across the European and Mediterranean landscape for the first time. The *NOS* GRS included the 4 genetically independent SNPs with lower P value associated with MI in a meta-analysis of three European case-control studies. Since *NOS* genes regulate the physiological availability of NO, this GRS constitutes a polygenic approach to the potential contribution of NO to CAD. The interindividual contribution of the GRS was lower than 1%. However, from an ecological perspective, GRS values across Europe were positively correlated with the incidence of coronary events explaining 65-85% of interpopulation variation of CAD incidence. These contrasting contributions and the usefulness of GRSs as ecological predictors are discussed below.

Interindividual (intrapopulation) and interpopulation contribution to CAD of NOS GRS

In the context of case-control studies, the *NOS* GRS was only a significant MI risk factor in the sample with the largest size (ATVB) but not in the other

studies. Also, the AUC indicated that the predictive value of the GRS was null or very limited (AUC of 0.527).

The weak effects of our GRS performed in terms of both association and prediction are in complete accordance with previous reports in the literature. The proportion of variance explained by the relatively large number of loci associated with CAD is lower than 1% [54]. In addition, the improvement in risk prediction provided by genetic markers appeared to be null or insufficient [55], even with the strongest and most replicated CAD risk factor identified in the 9p21.3 locus [7,56]. When we move from single genetic risk variants to a genomic profile, the combined effect of dozens of risk variants generally explains only a small proportion of disease variance [56] and shows a limited predictive ability (AUC of 0.55-0.62) [6,55]. For instance, in the MIGen Consortium, the effect of the nine top-associated loci explained 2.8% of phenotype variance [40]. Even a more comprehensive genetic risk score of 101 SNPs associated with MI and other cardiovascular risk factors explained less than 5% of interindividual variance [57]. In spite of the low genetic contribution of genomic profiles, the genetic basis of CAD is strong as reflected in family aggregation data (40% for women and 60% for men) [6,7,58]. The proportion of heritability that remains unaccounted for (referred to as “missing heritability” elsewhere) would be explained by common genetic variants (MAF > 0.05) having very small effects and rare variants with a larger contribution to the complex phenotype [9,58–60].

The geographic distribution of GRS presented an interesting variation pattern across European and Mediterranean populations. This distribution showed a concentric pattern from a center of lower risk scores in North-Western Mediterranean, specially the Islands of Corsica and Sardinia (Figure

1). The gradual increase towards North (UK) and North-East Europe (Poland and Finland) through the scarcely sampled area of North-Central Europe does not seem unreasonable given the general trends across the European continent and does not suggest any major problems with spurious interpolation. This European south-to-north cline in population GRS explained a large proportion of variance in coronary incidence across 10 MONICA populations, 67% in women and 86% in men (Figure 2). This large contribution contrasts with the intrapopulation (i.e. interindividual) contribution of the GRS (<1%). Looking at single genetic markers to understand this phenomenon, we have identified some genetic variants, mainly in the *NOS3/ATG9B* region, with frequencies correlating with CAD incidence. According to these correlations all these variants would explain a similar proportion of variance in CAD incidence (35-65%), but lower than GRS (Table 3). Out of this group of correlating variants only rs1799983 was included in the GRS. The other genetic variants correlating with CAD incidence were not associated with CAD phenotype. In the literature, empirical data on the ecological applications of GRSs are lacking, but some studies have been done using single markers. Previous studies on the correlation between risk allele frequencies and CAD incidence across MONICA populations did not find conclusive results [25,26]. One of these studies [26] extended the Mediterranean paradox discussion into genetics due to the observed negative correlations between some genetic risk factors and CAD events. These authors concluded that the observed north-to-south cline in the frequency of some genetic risk variants was most probably the result of spatial distribution of the whole genome variation present in the European continent, which has been mainly shaped by the history of populations [61–63]. In this context, the variance explained by the markers

correlating with CAD incidence in this study cannot be considered a specific effect of each variant, but rather the combined effect of many risk variants showing the same distribution. Thus, the apparent high effect (46% in women and 61% in men) of the risk SNP rs1799983 is likely to be a colinearity effect with other risk genetic factors in the regression analysis. In the same way, the estimated contribution of GRS (67% in women and 86% in men) would be the consequence of the joint effect of risk variants from the common frequency spectrum, with similar population patterns as the 4 SNPs included in the GRS. In other words, the contrasting contributions from interindividual variance (<1%) to interpopulation variance (65-85%) suggests that this GRS population approach most probably suffers from colinearity with other genetic as well as environmental risk factors. These results would support that a high proportion of population CAD incidence is determined by common genetic variant distributions because classical risk factors contribute in 30-40% of CAD population incidence, and because rare variants are basically population-specific and are not distributed in population gradients [41,60].

The important genetic contribution to CAD incidence variation suggests some considerations about the role of genetic factors on the individual risk to CAD. High incidence of CAD in a population would be determined by high frequency of genetic risk combinations and, hence, a high proportion of individuals carrying these genetic risk combinations. In this ecological approach, the GRS is capturing the contribution to CAD incidence of the myriad of genetic risk variants with similar geographic distribution. Thus, only the joint inclusion of this myriad of genetic variants would explain a considerable proportion of the genetic contribution to CAD outcome at the individual level. In accordance with these results, previous studies have dealt

with the number and the effect sizes of the genetic variants involved in the genetic architecture of CAD. These simulation models predicted that the number of genetic variants needed to explain the estimated heritability, under a purely additive model, would range from few hundred low frequency variants with large effect to several hundred or few thousand for common variants with small effect [6,64,65].

In addition to the "Finland-to-Spain" axis, the variation in GRS is also clinal between the West and East in the northern Mediterranean shore. Although the correlation with CAD incidence is not easy to demonstrate due to a sizeable lack of epidemiological data in these populations, some partial data [66] point in this direction. This constitutes an additional insight into our working hypothesis on the correlation between CAD incidence and GRS in Europe. So, the population distribution of both our GRS and cardiovascular incidence are deeply influenced by similar demographic processes that have modulated genetic variation in current human groups. Available data in North Africa [67] are too scarce to extend any conclusion to the southern shore of the Mediterranean.

Usefulness of population GRS approach

In the European continent, environmental factors as well as genetic variation seem to be structured in south-to-north clines that can be correlated with observed CAD incidence as reflected by latitude in our study. In previous data such as the original MONICA project, modifiable risk factors explained only 30% of coronary incidence variance [24]. And in this study, multivariate analyses stressed the importance of GRS to explain the distribution of CAD incidence, especially in men. These results reflect that most variance in CAD incidence among populations is accounted for by

genetic background. A priori, this would contrast with the fact that at the interindividual level (within a population) environmental and behavioral factors are involved in 80% of all cardiovascular events. However, they are two complementary sides of the same phenomenon that explain different features of the disease: individual outcome and population incidence. Whereas within a population environmental factors explain a large proportion of individual events, the total amount of coronary events in populations under similar environmental pressures would depend on their genetic predisposition. The potential incidence of CAD in a population would be mainly determined by its genetic risk background but it would be triggered by behavioral and life style factors.

Our results highlight the usefulness of GRSs as population estimates of the genetic burden of disease or as ecological predictors. It has been stated that the assessment of disease risk and its temporal trends is of critical importance to predict incidences of CAD [68]. An ideal GRS would include all CAD risk variants in the genome. Nevertheless, an exploratory strategy could include the design of different GRSs for different pathophysiological processes related to CAD, such as endothelial dysfunction, accelerated atherosclerosis or thrombosis, each one having its own genetic basis [68]. These GRSs could provide a solid basis for developing lifestyle intervention programs to prevent intermediate risk factors (e.g. obesity, high levels of blood pressure, glucose and lipids) in population subgroups before environmental factors trigger a potentially high genetic predisposition for the disease.

Another particular advantage of GRS is that they can be constructed from samples of a few hundreds of individuals per population. High potential benefits, no interventional harms and low cost make this approach very

promising for future public health studies.

Limitations

The analyses of this study have some limitations that can be commented on. A first aspect refers to the genotyping strategy; with the aim of capturing the maximum common genetic variation in *NOS* genes we genotyped 1 SNP every 5 kb in our samples (65 SNPs in total). The accuracy of this strategy was assessed by imputing all variants present in the TSI 1000 Genomes population sample in our general population sample from Lazio, Italy (genotyped with our set of 65 SNPs). From the result, we concluded that this genotyping strategy was representative of >70% of common genetic variation. Therefore, a remaining 30% of common genetic variation of *NOS* genes is not well represented in these analyses. Secondly, in MIGen case-control samples a considerable proportion of genotypes (~60%) were imputed, and, even though imputation quality controls were performed, this fact could have affected the association and meta-analysis results. In any case, this fact does not invalidate the GRS population approach because the distribution of a robust (large number of polymorphisms) GRS does not depend on the distribution of single polymorphisms. Thirdly, the GRS in this study correspond only to a small piece of the genetic basis of cardiovascular diseases jigsaw. So, this initial study should be further developed beyond polymorphisms in *NOS* genes. Finally, the size of the general population samples was robust enough to check the frequency distributions of polymorphisms. However, future studies should include larger sample sizes if less common (<5%) polymorphisms are to be included.

Conclusions

This study of cardiovascular *NOS*-GRS in European populations shows for the first time that GRSs are a powerful way of analyzing the distribution of genetic risk and a promising tool for ecological predictions of disease.

Although the contribution of GRS to CAD at the individual level was lower than 1%, GRS explained a large proportion of interpopulation differences in CAD incidence (65%-85%). This large contribution to CAD incidence across populations might be the result of colinearity with several other common genetic and environmental factors. From the GRS perspective, the so-called cardiovascular Mediterranean paradox would be no longer held and CAD genetic architecture would be mainly based on common genetic polymorphisms. The genetic risk score population approach seems very promising in future public health interventions to develop lifestyle programs and prevent intermediate risk factors in population subgroups with especially high genetic predisposition.

Acknowledgements

This research has been supported by the Ministerio de Ciencia y Tecnología CGL2008-03955 and CGL2011-27866 projects to PM; the Ministerio de Ciencia y Tecnología Predoctoral Grant (BES-2009-016334) and Travel Grant (EEBB-2011-44310) to RCT; and the Beatriu de Pinós postdoctoral grant (2009 BP-B 00274) to MV. We want to thank ACJW Janssens for the supervision of RCT work stay at the University Medical Center of Rotterdam, the MIGen Consortium from NBCI dbGaP for providing the MIGen case-control samples, and the Spanish National DNA Bank for providing cardiovascular samples as well as general Spanish population samples. We also want to thank all our collaborators that provided samples of general populations: N Harich and M Kandil (Morocco), H Chabaani (Tunisia), M Sadiq (Jordan), JM Dougoujon (France), L Varesi (Corsica), F Cruciani and G Destro-Bisol (Italy), C Calo (Sardinia), N Pojskic (Sarajevo), A Kouvatzi (Greece), Ewa Bartnik (Poland), L Murphy (United Kingdom) and The 1000 Genomes Project. All

volunteers are gratefully acknowledged for sample donation. Finally, thanks to the academic editor and the two anonymous reviewers who substantially improved this manuscript.

References

1. World Health Organization (2011) Global status report on noncommunicable diseases 2010. WHO Libr.
2. Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, et al. (1998) Prediction of Coronary Heart Disease Using Risk Factor Categories. *Circulation* 97: 1837–1847.
3. Ridker PM, Buring JE, Rifai N, Cook NR (2007) Development and Validation of Improved Algorithms for the Assessment of Global Cardiovascular Risk in Women The Reynolds Risk Score. *JAMA* 297: 611–620.
4. Assmann G (2002) Simple Scoring Scheme for Calculating the Risk of Acute Coronary Events Based on the 10-Year Follow-Up of the Prospective Cardiovascular Munster (PROCAM) Study. *Circulation* 105: 310–315.
5. Conroy R (2003) Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 24: 987–1003.
6. Van der Net JB, Janssens ACJW, Sijbrands EJG, Steyerberg EW (2009) Value of genetic profiling for the prediction of coronary heart disease. *Am Heart J* 158: 105–110.
7. Thanassoulis G, Vasan RS (2010) Genetic cardiovascular risk prediction: will we get there? *Circulation* 122: 2323–2334.
8. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461: 747–753.
9. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, et al. (2010) Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 11: 446–450.
10. Morrison AC, Bare LA, Chambless LE, Ellis SG, Malloy M, et al. (2007) Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. *Am J Epidemiol* 166: 28–35.
11. Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, et al. (2010) A multilocus genetic risk score for coronary heart disease: case-control and

- prospective cohort analyses. *Lancet* 376: 1393–1400.
12. Brautbar A, Pompeii LA, Dehghan A, Ngwa JS, Nambi V, et al. (2012) A genetic risk score based on direct associations with coronary heart disease improves coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC), but not in the Rotterdam and Framingham Offspring, Studies. *Atherosclerosis* 223: 421–426.
 13. Hughes MF, Saarela O, Stritzke J, Kee F, Silander K, et al. (2012) Genetic markers enhance coronary risk prediction in men: the MORGAM prospective cohorts. *PLoS One* 7: e40922.
 14. Vaarhorst AA, Lu Y, Heijmans BT, Dollé MET, Böhringer S, et al. (2012) Literature-based genetic risk scores for coronary heart disease: the Cardiovascular Registry Maastricht (CAREMA) prospective cohort study. *Circ Cardiovasc Genet* 5: 202–209.
 15. Thanassoulis G, Peloso GM, Pencina MJ, Hoffmann U, Fox CS, et al. (2012) A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium: the Framingham Heart Study. *Circ Cardiovasc Genet* 5: 113–121.
 16. Bolton JL, Stewart MCW, Wilson JF, Anderson N, Price JF (2013) Improvement in prediction of coronary heart disease risk over conventional risk factors using SNPs identified in genome-wide association studies. *PLoS One* 8: e57310.
 17. Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, Arveiler D, Rajakangas a. M, et al. (1994) Myocardial infarction and coronary deaths in the World Health Organization MONICA Project. Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents. *Circulation* 90: 583–612.
 18. Bronte-Stewart B (1956) THE EPIDEMIOLOGY OF ISCHAEMIC HEART DISEASE. *Postgrad Med J* 35: 180–185.
 19. Keys A, Aravanis C, Blackburn H, Buzina R, Djordjevic B, et al. (1980) Seven Countries. A Multivariate Analysis of Death and Coronary Heart Disease. Harvard Univ Press: 1–381.
 20. Tunstall-pedoe H, Kuulasmaa K, Tolonen H, Ruokokoski E, Amouyel P (1999) Contribution of trends in survival and coronar y - event rates to changes in coronary heart disease mortality: 10-year results from 37 WHO MONICA Project populations. *353*: 1547–1557.
 21. Tunstall-pedoe H (1988) *Autres pays, autres moeurs. Theories on why the French*

- have less heart disease than the British. *Br Med J* 297: 1559–1560.
22. Artaud-Wild SM, Connor SL, Sexton G, Connor WE (1993) Differences in coronary mortality can be explained by differences in cholesterol and saturated fat intakes in 40 countries but not in France and Finland. A paradox. *Circulation* 88: 2771–2779.
 23. The WHO MONICA Project (1994) Ecological Analysis of the Association between Mortality and Major Risk Factors of Cardiovascular Disease. *Int J Epidemiol* 23: 505–516.
 24. Riudavets J-B, Ducimetière P, Evans A, Montaye M, Haas B, et al. (2010) Patterns of alcohol consumption and ischaemic heart disease in culturally divergent countries: the Prospective Epidemiological Study of Myocardial Infarction (PRIME). *Br Med J* 341:c6077. doi:10.1136/bmj.c6077.
 25. Stengård JH, Weiss KM, Sing CF (1998) An ecological study of association between coronary heart disease mortality rates in men and the relative frequencies of common allelic variations in the gene coding for apolipoprotein E. *Hum Genet* 103: 234–241.
 26. Lao O, Dupanloup I, Barbujani G, Bertranpetit J, Calafell F (2008) The Mediterranean Paradox for Susceptibility Factors in Coronary Heart Disease Extends to Genetics. *Ann Hum Genet* 72: 48–56.
 27. Relethford JH (2008) Geostatistics and spatial analysis in biological anthropology. *Am J Phys Anthropol* 136: 1–10.
 28. Mungrue IN, Brecht DS, Stewart DJ, Husain M (2003) From molecules to mammals: what's NOS got to do with it? *Acta Physiol Scand* 179: 123–135.
 29. Robb GB, Carson AR, Tai SC, Fish JE, Singh S, et al. (2004) Post-transcriptional regulation of endothelial nitric-oxide synthase by an overlapping antisense mRNA transcript. *J Biol Chem* 279: 37982–37996.
 30. Casas JP, Cavalleri GL, Bautista LE, Smeeth L, Humphries SE, et al. (2006) Endothelial nitric oxide synthase gene polymorphisms and cardiovascular disease: a HuGE review. *Am J Epidemiol* 164: 921–935.
 31. Paternoster L, Martinez-Gonzalez NA, Charleton R, Chung M, Lewis S, et al. (2010) Genetic effects on carotid intima-media thickness: systematic assessment and meta-analyses of candidate gene polymorphisms studied in more than 5000 subjects. *Circ Cardiovasc Genet* 3: 15–21.

32. Li J, Wu X, Li X, Feng G, He L, et al. (2010) The endothelial nitric oxide synthase gene is associated with coronary artery disease: a meta-analysis. *Cardiology* 116: 271–278.
33. Zhang K, Bai P, Shi S, Zhou B, Wang Y, et al. (2012) The G894T polymorphism on endothelial nitric oxide synthase gene is associated with increased coronary heart disease among Asia population: evidence from a Meta analysis. *Thromb Res* 130: 192–197.
34. Zintzaras E, Kitsios G, Stefanidis I (2006) Endothelial NO synthase gene polymorphisms and hypertension: a meta-analysis. *Hypertension* 48: 700–710.
35. Pereira TV, Rudnicki M, Cheung BMY, Baum L, Yamada Y, et al. (2007) Three endothelial nitric oxide (NOS3) gene polymorphisms in hypertensive and normotensive individuals: meta-analysis of 53 studies reveals evidence of publication bias. *J Hypertens* 25: 1763–1774.
36. Morris BJ, Glenn CL, Wilcken DE, Wang XL (2001) Influence of an inducible nitric oxide synthase promoter variant on clinical variables in patients with coronary artery disease. *Clin Sci (Lond)* 100: 551–556.
37. Iwai N, Tago N, Yasui N, Kokubo Y, Inamoto N, et al. (2004) Genetic analysis of 22 candidate genes for hypertension in the Japanese population. *J Hypertens* 22: 1119–1126. doi:10.1097/01.hjh.0000125407.50839.5c.
38. Kunnas TA, Mikkelsen J, Ilveskoski E, Tanner MM, Laippala P, et al. (2003) A functional variant of the iNOS gene flanking region is associated with LAD coronary artery disease: an autopsy study. *Eur J Clin Invest* 33: 1032–1037.
39. Fu L, Zhao Y, Lu J, Shi J, Li C, et al. (2009) Functional single nucleotide polymorphism-1026C/A of inducible nitric oxide synthase gene with increased YY1-binding affinity is associated with hypertension in a Chinese Han population. *J Hypertens* 27: 991–1000.
40. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, et al. (2009) Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* 41: 334–341.
41. Abecasis GR, Auton A, Brooks LD, DePristo M a, Durbin RM, et al. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491: 56–65.
42. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M a R, et al. (2007)

- PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559–575.
43. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 34: 816–834.
 44. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263–265.
 45. Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 5: e1000529.
 46. Kundu S, Aulchenko YS, van Duijn CM, Janssens ACJW (2011) PredictABEL: an R package for the assessment of risk prediction models. *Eur J Epidemiol* 26: 261–264.
 47. Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26: 2190–2191.
 48. De Bakker PIW, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, et al. (2005) Efficiency and power in genetic association studies. *Nat Genet* 37: 1217–1223.
 49. Thioulouse J, Chessel D, Champely S (1995) Multivariate analysis of spatial patterns: a unified approach to local and global structures. *Environ Ecol Stat* 2: 1–14.
 50. Rosenberg MS, Anderson CD (2011) PASSaGE: Pattern Analysis, Spatial Statistics and Geographic Exegesis. Version 2. *Methods Ecol Evol* 2: 229–232.
 51. Chikhi L, Destro-Bisol G, Pascali V, Baravelli V, Dobosz M, et al. (1998) Clinal variation in the nuclear DNA of Europeans. *Hum Biol* 70: 1.
 52. Kuulasmaa K, Tunstall-Pedoe H, Dobson A, Fortmann S, Sans S, et al. (2000) Estimation of contribution of changes in classic risk factors to trends in coronary-event rates across the WHO MONICA Project populations. *Lancet* 355: 675–687.
 53. Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, et al. (2008) Correlation between genetic and geographic structure in Europe. *Curr Biol* 18: 1241–1248.
 54. Ioannidis JP (2009) Prediction of cardiovascular disease outcomes and established cardiovascular risk factors by genome-wide association markers. *Circ Cardiovasc Genet* 2: 7–15.
 55. Janssens ACJW, van Duijn CM (2008) Genome-based prediction of common

- diseases: advances and prospects. *Hum Mol Genet* 17: R166–73.
56. Janssens ACJW, van Duijn CM (2009) Genome-based prediction of common diseases: methodological considerations for future research. *Genome Med* 1: 20.
 57. Paynter NP, Chasman DI, Paré G, Buring JE, Cook NR, et al. (2010) Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* 303: 631–637.
 58. Prins BP, Lagou V, Asselbergs FW, Snieder H, Fu J (2012) Genetics of coronary artery disease: genome-wide association studies and beyond. *Atherosclerosis* 225: 1–10.
 59. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, et al. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9: 356–369.
 60. Gibson G (2011) Rare and common variants: twenty arguments. *Nat Rev Genet* 13: 135–145.
 61. Sokal R, Harding R, Oden N (1989) Spatial patterns of human gene frequencies in Europe. *Am J Phys Anthr* 80: 267–294.
 62. Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *The History and Geography of Human Genes* Princeton University Press.
 63. Barbujani G, Goldstein DB (2004) Africans and Asians abroad: genetic diversity in Europe. *Annu Rev Genomics Hum Genet* 5: 119–150.
 64. Janssens ACJW, Aulchenko YS, Elefante S, Borsboom GJJM, Steyerberg EW, et al. (2006) Predictive testing for complex diseases using multiple genes: Fact or fiction? *Genet Med* 8: 395–400.
 65. Stahl E, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, et al. (2012) Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet* 44: 483–489.
 66. Müller-Nordhorn J, Binting S, Roll S, Willich SN (2008) An update on regional variation in cardiovascular mortality within Europe. *Eur Heart J* 29: 1316–1326.
 67. World Health Organization (2004) *The global burden of disease 2004*. WHO.
 68. Batsis J, Lopez-Jimenez F (2010) Cardiovascular risk assessment--from individual risk prediction to estimation of global risk and change in risk in the population. *BMC Med* 8: 29.

3.4 Resultats III: Carreras-Torres et al., (article III)

Clinal genetic structures from *NOS* genes explain geographical differences in European coronary incidence and fit a dual Neolithic expansion across Europe and the Mediterranean

Robert Carreras-Torres, Albert Ferran, Daniela Zanetti, Esther Esteban, Giovanni Destro-Bisol, Laurent Varesi, Naris Pojskic, Hassen Chaabani, Marc Via, Laurent Excoffier, Pedro Moral.

3.4.1 Resum en Català

Les estructures genètiques en clina observades pels gens *NOS* expliquen les diferències geogràfiques en la incidència coronària a Europa i s'ajusten a una expansió Neolítica doble a Europa i la Mediterrània

La distribució en les poblacions humanes de les bases genètiques de les malalties complexes pot trobar-se darrera la diferent susceptibilitat a aquestes malalties. La distribució poblacional de polimorfismes genètics associats a les malalties coronàries no sembla respondre a efectes de pressió selectiva. Per tant, existeix la possibilitat que l'arquitectura genètica de les malalties coronàries es trobi estructurada entre les poblacions europees de manera similar a la variació genètica global.

En aquest estudi, s'ha analitzat la variació genètica de 114 SNPs i 17 polimorfismes de repeticions en tàndem (TRP, de l'anglès *tandem repeat polymorphism*) localitzats en les regions dels gens de les sintases d'òxid nítric (*NOS*, de l'anglès *Nitric Oxide Synthases*) en 1451 individus corresponents a 35 poblacions majoritàriament d'Europa, nord d'Àfrica i Orient Mitjà. Es va testar si la incidència coronària registrada pel projecte

MONICA de l'OMS en 11 de les poblacions europees genotipades podia ser explicada per les estimacions d'estructura genètica en les poblacions. Com a resultats, es va observar que la diferenciació genètica entre les poblacions d'Europa, nord d'Àfrica i Orient Mitjà era molt baixa però significativa ($F_{ST} = 1.2\%$ per SNPs i $F_{ST} = 1.6\%$ per TRPs). Mitjançant l'anàlisi de components principals espacial (sPCA) en aquestes poblacions, es va determinar que les principals estructures de variació genètica eren gradients de continuïtat genètica distribuïts seguint eixos similars a nord-sud i est-oest. L'estimació de components genètics "ancestrals" va indicar l'existència de 3 components similars en SNPs i TRPs que van ser interpretats com components genètics d'origen Paleolític europeu, Paleolític nord africà i Neolític europeu. La majoria de poblacions estudiades presentaven una contribució elevada (>60%) de la component genètica d'origen paleolític. D'aquest conjunt de resultats destaquen quatre grups de poblacions que presenten les estimacions més extremes dins d'aquest contínuum genètic observat: la població del País Basc, les poblacions jordanes, les poblacions de Bòsnia-Hercegovina i les poblacions nord africanes. En els tests per determinar el possible efecte de la selecció natural a l'hora de modular la distribució genètica en les poblacions, es va observar com els polimorfismes detectats com a associats a la malaltia coronària en estudis anteriors no presentaven cap senyal de selecció. A més, aquests tests tampoc van revelar cap efecte selectiu sobre un grup de polimorfismes localitzats al gen *NOS3* que presenta cert gradient poblacional correlacionat amb la incidència coronària. Finalment, les nostres estimacions d'estructuració genètica poblacional (distàncies genètiques, components principals espacials i proporcions ancestrals) van explicar altes contribucions a la incidència coronària de les 11 poblacions analitzades (36-98%).

Aquests resultats reflecteixen la continuïtat genètica entre les poblacions del continent europeu i la Mediterrània i, per tant, l'absència de barreres geogràfiques i culturals al flux genètic entre les ribes nord i sud de la Mediterrània. L'elevada contribució paleolítica al patrimoni genètic actual contrastaria amb la marcada distribució genètica nord-sud i est-oest fruit de processos més recents. Però aquests resultats s'ajusten a simulacions computacionals on aquesta distribució s'explicaria per un doble procés de Neolització: una expansió neolítica al continent europeu lligada a la cultura ceràmica *Linearbandkeramik* i una expansió neolítica a la Mediterrània lligada a la cultura ceràmica Cardial. El fet que les nostres estimacions d'estructuració genètica expliquin la gairebé totalitat de la incidència cardiovascular indica que la predisposició poblacional a la malaltia cardiovascular es relaciona amb el patrimoni genètic poblacional i que és fruit de la demografia històrica de les poblacions i no a efectes d'adaptació al medi. De fet, el gradient nord-sud en la incidència coronària al continent europeu seria conseqüència d'aquest doble procés de Neolització, ja que s'haurien generat poblacions de baixa diversitat genètica al centre-nord del continent (elevant-ne la predisposició coronària) i poblacions de diversitat genètica més elevada al sud del continent, allà on els patrimonis genètics de les dues expansions neolítiques van convergir (disminuint-ne la predisposició coronària).

3.4.2 Informe dels directors de tesi de la participació del doctorant en l'article i de la singularitat de l'article en la elaboració de la tesi

Dr. Pedro Moral Castrillo, professor del Departament de Biologia Animal de la Universitat de Barcelona, i Dr. Marc Via García, professor del Departament de Psiquiatria i Psicobiologia Clínica de la Universitat de Barcelona, supervisors de la tesi doctoral “**Variació genòmica de les Sintases d’Òxid Nítric i de regions associades amb Risc Cardiovascular en Poblacions Humanes**”, manifesten que l’estudiant de doctorat Robert Carreras Torres ha participat en el treball titulat “*Clinal genetic structures from NOS genes explain geographical differences in European coronary incidence and fit a dual Neolithic expansion across Europe and the Mediterranean*”, i llest per publicació, en les següents tasques:

- Disseny i selecció de les mostres i marcadors analitzats (90%).
- Recol·lecció i preparació de les mostres poblacionals (90%)
- Genotipació dels STRs en el laboratori (90%).
- Creació de las bases de dades genètiques i selecció de resultats disponibles a la bibliografia per tal de realitzar comparacions (100%).
- Anàlisis estadístics (100%).
- Redacció de l’article (95%).

Cap dels coautors de l’article ha utilitzat els resultats d’aquest estudi de manera implícita o explícita per elaborar una altra tesi doctoral. Per tant, aquest article forma part de manera exclusiva de la tesi doctoral de Robert Carreras Torres.

Signat per Dr. Pedro Moral Castrillo i Dr. Marc Via García
Barcelona, 22 Abril 2014.

3.4.3 Article III

Clinal genetic structures from *NOS* genes explain geographical differences in European coronary incidence, and fit a dual Neolithic expansion across Europe and the Mediterranean

Robert Carreras-Torres¹, Albert Ferran¹, Daniela Zanetti¹, Esther Esteban^{1,2}, Giovanni Destro-Bisol³, Laurent Varesi⁴, Naris Pojskic⁵, Hassen Chaabani⁶, Marc Via⁷, Laurent Excoffier⁸, Pedro Moral^{1,2}.

¹Departament Biologia Animal - Antropologia, Universitat de Barcelona, Barcelona, Spain;

²Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain.

³Department of Animal and Human Biology, University La Sapienza, Rome, Italy.

⁴Faculté des Sciences et Techniques, Université de Corse, Corte, France.

⁵Institute for Genetic Engineering and Biotechnology, Sarajevo, Bosnia and Herzegovina.

⁶Laboratory of Human Genetics and Anthropology, Faculty of Pharmacy, University of Monastir, Monastir, Tunisia.

⁷Departament Psiquiatria i Psicobiologia Clínica and Institute for Brain, Cognition and Behaviour (IR3C), Universitat de Barcelona, Barcelona, Spain;

⁸Computational and Molecular Population Genetics Lab, Institute of Ecology and Evolution, University of Bern, Berne, Switzerland;

Abstract

The distribution of the genetic basis of complex diseases across human populations could be behind different population susceptibility to the diseases. Population distribution of genetic polymorphism associated with coronary artery disease (CAD) does not seem to be the consequence of selective pressures. There is the possibility that CAD genetic architecture among European populations is structured similarly to whole genome variation. In this study, inferences of genetic structure were obtained using

114 single nucleotide polymorphisms (SNP) and 17 tandem repeat polymorphisms (TRP) from the nitric oxide synthases (*NOS*) gene regions in 1451 individuals from 35 populations mainly from Europe, North Africa and the Middle East. Moreover, we tested whether coronary events in 11 of these European populations could be explained by population genetic structure estimates. Our results confirmed the lack of a Mediterranean barrier for gene flow and highlighted that the European and North African genetic clines constitute a continuous pattern. Three main ancestral genetic pools were identified and interpreted, with major European and North African Palaeolithic genetic contribution (>60%) for most populations. Surprisingly, these large amounts of Palaeolithic ancestral components depicted a recent population distribution as a consequence of a strong impact from a dual Neolithic expansion: one expansion in the European continent and another in the Mediterranean basin. Finally, the main objective of the survey was achieved when population structure estimates explained large proportions of coronary event rates (39-98%). These results were interpreted as a possible consequence of dual Neolithisation. In the European context, this differentiated Neolithisation generated populations of low genetic diversity and high CAD incidence in the core area of European continental expansion (Central-North Europe) and populations with higher genetic diversity and low CAD incidence on the boundaries of differentiated expansion areas (Southwest Europe) where different genetic backgrounds would converge.

Keywords

NOS genes, European genetic stratification, Neolithic expansion, European CAD incidence

Introduction

Population distribution of coronary artery disease (CAD) traditionally presents a southwest-to-northeast cline of incidence and prevalence in Europe, from the South (Spain and France), with the lowest CAD mortality, towards the North (Finland and UK) [1,2]. It has been assessed that classical risk factors contribute to 30-40% of CAD population incidence [3,4]. Moreover, it has been suggested that geographic genetic differences could also be behind the population gradient related to coronary incidence. A recent companion study [5] has shown that the distribution of a few genetic cardiovascular-associated markers from nitric oxide synthase (*NOS*) genes explained a large proportion (65-85%) of CAD incidence. Given the polygenic nature of cardiovascular disorders, the few *NOS* associated markers explained only a modest contribution to CAD phenotype (<1%), so it is reasonable to think that other cardiovascular-associated markers would have a similar population distribution.

Genetic differences among populations are mainly caused by demographic processes linked to human migration and adaptation to new environments. Phenotype differentiation, as a response to the environment, could imply unusual allele frequency patterns in populations. This seems to be the case in Type II Diabetes and pigmentation [6–8] under different evolutionary pressures [9,10]. However, this has not been observed for CAD or hypertension, for which risk genetic differences among world-wide populations have been explained by random processes [6,11]. Indeed, sustained selection processes that drive alleles near to fixation have been rare during recent human evolution [12]. On the European continent, broad human groups share a demographic history with similar genetic profiles and thus, low genetic differentiation. However, a strong correlation between

genetic and geographic distances has produced a genetic continuum [13,14] shaped by north-to-south and east-to-west clines of genetic diversity [13,15]. Therefore, this well-known phenomenon could parallel the continental north-to-south geographic variation in some complex traits and diseases.

In this study, we hypothesize that the genetic bases of CAD are similarly structured to whole genomic variation in European populations. This general genetic structure would also reflect the geographic differences in CAD incidence and prevalence on the European continent.

To assess this, we have described variation in and around nitric oxide synthases (*NOS*) genomic regions. *NOS* genes are responsible for nitric oxide (NO) availability. The role of NO in regulation of vascular tone homeostasis, tissue perfusion, and platelet aggregation [16] is considered a key piece of the CAD jigsaw. There are three *NOS* genes: *endothelial NOS* (*eNOS* or *NOS3*), *neuronal NOS* (*nNOS* or *NOS1*) and *inducible NOS* (*iNOS* or *NOS2A*) [16]. According to the Human Genome Epidemiology (HuGE) Navigator browser (www.hugenavigator.net), the *NOS3* gene is the second most reported gene for CAD, with 139 related papers, and the fourth most reported for myocardial infarction, with 75 reports (April, 2014). We have analysed single nucleotide polymorphisms (SNP) and tandem repeat polymorphisms (TRP), covering coding and non-coding regions around *NOS* genes. Non-coding regions have been included in the analyses because variation in intergenic regions is under-represented among disease-associated markers [17] and can provide a broader evolutionary perspective. Concerning populations included in this study, some of them were selected taking presumably European ancestral genetic pools into account. The observed genetic diversity patterns in Europe have been interpreted as an indicator of recent demographic events since Palaeolithic times [18–20]. And

according to some authors, European genetic background would be provided by the first Palaeolithic Europeans, Neolithic farmers [19] and migration from North Africa [21].

In order to test our working hypothesis, we have followed two different approaches: 1) we have tested several genetic *NOS* markers for selection, both associated and non-associated with CAD; and 2) we have performed, a systematic correlation analysis of *NOS* molecular population estimates and coronary incidences in European populations for the first time.

Materials and Methods

Ethics Statement

The study has been specifically approved by the Ethical Committee of the University of Barcelona (Institutional Review Board: IRB00003099) and all the participants provided a written informed consent.

Sample description

A total of 35 world-wide population samples from Europe, Middle East, North Africa, South-Saharan Africa, and South America were genetically tested. Blood samples of 1451 individuals were collected for DNA extraction from healthy unrelated individuals of both sexes whose four grandparents had been born in the same geographical region. Population details of all samples, including sample size, geography and ethnic group when available, are recorded in Supporting Table (ST) Tab 1. The geographic distribution of the European and Mediterranean population samples are displayed in Figure S1. The European samples corresponded to populations from UK, Poland, Spain (5 samples), France (4 samples), Italy (5 samples), Bosnia-Herzegovina (2 samples), Greece (3 samples) and Turkey. The

Middle East samples are represented by Jordan (2 samples), and North Africa by Tunisia (2 samples), Morocco (3 samples), Algeria and Egypt. Finally, two samples from Africa (the Ivory Coast and Cameroon) and South America (two Bolivian Andean groups) have been included to give a broader context to our European and Mediterranean analyses. In addition to DNA samples, genetic data from four European 1000 Genomes Project samples (GBR, FIN, CEU and TSI) [22] were included in the analyses (ST Tab1).

Polymorphisms and genotyping

DNA samples were genotyped for a combined set of 144 SNPs, 1 variable number tandem repeat (VNTR) of 27 base pairs (bp), and 16 microsatellites or short tandem repeats (STR) (tri- tetra- and penta-nucleotides; 13 of them were novel STRs). All polymorphisms were located in the three *NOS* genes regions and in three non-coding regions 300-500 kilobases (kb) apart from *NOS* genes. The SNPs were selected as a representative set of the common variation existing in these six genomic regions, according to an average coverage of 1 SNP every 5kb and high heterozygosity in CEU population as reported in the HapMap project (www.hapmap.org). SNPs were genotyped using a Golden Gate Genotyping Assay (Illumina Inc., San Diego, CA). Genomic location of genetic variants is shown in ST Tab 2.

The 27bp VNTR and the 16 STRs, from now on coded as TRP, were selected according to previously reported descriptions [23,24] or as potentially polymorphic loci described by the UCSC Genome Browser website (<http://www.genome.ucsc.edu/>). TRPs were determined by DNA fragment analyses with fluorescent-labeled primers. PCR amplification was carried out in a total volume of 10 µl using 10-20 ng of template DNA, 1 µl

Buffer 10x, 1 mM of MgSO₄, 0.2 mM of each dNTPs, 0.25 μM of each primer and 0.2 μl KOD Hot Start DNA Polymerase (Novagen, Merck KGaA, Darmstadt, Germany). The amplification program consisted of an initial 5 min denaturalization step at 92 °C, followed by 30 cycles of 30 s at 92°C, 30 s at annealing temperature (61-62°C) and 30 s at 72°C, and with a final 5 min extension step at 72°C. PCR products were 1:4 standardly diluted and electrophoresed on ABI PRISM 3700 DNA sequencer (Applied Biosystems, Foster City, CA) using CST ROX 70-500 (Bioventures. Inc, Murfreesboro, TN) as size standard. GeneMapper 3.0 software (Applied Biosystems, Foster City, CA) was used to genotype individuals. Standard positive and negative controls were included and DNA from homozygote individuals was directly sequenced to establish the exact number of repeats and control allele designations. The results from the analysis of the TRPs were submitted to GenBank and are available in dbSNP Build B139. The newly assigned reference IDs, genomic location, repeat motive descriptions, PCR primers descriptions and annealing temperatures are indicated in ST Tab 3.

Statistical Analyses

Genotyping rate was assessed using PLINK v1.07 [25]. Non polymorphic SNPs, or with a genotyping rate lower than 0.75, were removed from the analyses. Since mutation mechanisms and rates are different for SNPs and TRPs, analyses were performed separately.

Population allele frequencies, gene diversities and fitting for Hardy-Weinberg equilibrium were performed with Arlequin v3.5 software [26]. Analyses of molecular variation (AMOVA) using Wright's F-statistics have been conducted by clustering samples according to geographic criteria.

AMOVA indices were calculated in the whole set of populations grouped into South-Saharan Africa, South America, and the European-Mediterranean region; and also in the reduced set of European and Mediterranean populations according to North-Central Europe, South Europe, Middle East, and North Africa. Population genetic relationships were assessed by pairwise F_{ST} genetic distances for the SNP data set and pairwise R_{ST} genetic distances for the TRP data set. Both AMOVAs and genetic distances were carried out using Arlequin v3.5 [26]. Genetic distances were represented by multidimensional scaling (MDS) analyses using *stat* and *graphics* R software packages (R Development Core Team).

Spatial genetic patterns were estimated in the European and Mediterranean populations through the georeferenced analysis known as spatial principal component analysis (sPCA) [27]. This method provides evidence of geographic global or local genetic structures. The neighboring graph of populations was built using Delaunay triangulation. The main results are maps of scores provided by positive and negative eigenvalues. Positive eigenvalues correspond to global patterns (spatial grouping or genetic clines), whereas negative eigenvalues indicate local patterns (considerable genetic differences between neighbors). The significance of the global or local patterns was assessed through Global and Local Monte-Carlo tests with 10000 permutations. The sPCA, and the Global and Local Monte-Carlo tests were performed using *adegenet* R package [28].

Genetic admixture in the Mediterranean and European samples was estimated using STRUCTURE [29]. According to previous studies [19], no more than four parental populations are expected in the analyzed samples. Thus, to assess the number of clusters (K) that best fitted the genetic data studied in this work, successive numbers of clusters were estimated (K from

2 to 7) iterating the estimation five times per K. The optimal K was inferred performing the Evanno's method [30] implemented in the STRUCTURE HARVESTER web-based program [31]. In order to visualize the spatial distribution of estimated ancestral genetic contributions across populations, membership coefficients for each cluster per polymorphism datasets were mapped using the geostatistical method known as kriging from ArcGIS software (ESRI, Redlands, CA, USA). Since anisotropy was not detected in the semivariogram, we used the ordinary spherical interpolation kriging method [32]. To standardize the visualization of ancestral genetic proportions among the different contour maps, eight equal categories were established for all contour maps. Furthermore, the covariation of observed spatial distribution of ancestral components with geography was assessed by Moran's I and Geary's C randomization tests for spatial autocorrelation [33] using *ade4* R package.

Detection of potential loci under selection was assessed from two different perspectives based on the patterns of genetic diversity among populations. The first method is based on the probability of observing locus by locus AMOVA statistics as a function of heterozygosity, given a null distribution generated under a hierarchically island model of population differentiation [34]. This test detects the effects of balancing and positive selection through loci with comparative significant low or large differences between populations. This method was performed using Arlequin v3.5 software [26] on both SNPs and TRPs in European and Mediterranean populations. The second approach was spatial ancestry (SPA) analysis [35]. In this analysis, only SNP data were used to model allele frequency distributions in the geographic space. Applying the SPA approach, polymorphisms showing steep geographic gradients in allele frequencies can

be identified through SPA scores reflecting the steepness of the geographic gradient. Large SPA scores (higher than the 95th percentile) were taken as indicators of potential selection. In order to detect continental adaptation events, analyses were performed in two different groups of samples: continental Europe plus the Middle East; and North Africa plus the Middle East, excluding the Mediterranean Islands.

Finally, in order to assess the relationship between genetic population structure and variation of coronary incidence among European populations, several genetic parameters from our tested populations were correlated with coronary event rates in 11 populations. Average annual coronary event rates over 5 years by gender in middle-aged individuals were compiled from the World Health Organization (WHO) MONICA Project [36] for the European populations. Among the analyzed populations, four of them (POL, NFR, SFR, and CAT) had a MONICA counterpart and seven additional populations (ORK, GBR, CEU, FIN, TSI, CIT and NBH) had a MONICA population within a 200-km radius or from the same country (ST Tab 4). The CEU sample was considered a counterpart of MONICA German-Bremen population according to Lao et al. [13]. For TRP data correlations, the GBR, CEU, FIN and TSI populations were excluded from the analysis because of the lack of genetic data. Three different genetic parameters were correlated with coronary event rates: 1) genetic distances between the above matched populations and those populations in our study that showed the highest genetic differentiation to the whole set; 2) sPCA scores of significant global or local eigenvalues; and 3) ancestral membership coefficients observed in the admixture analyses. Spearman's correlation, and univariate and multivariate linear regression analyses between genetic stratification estimates and coronary event rates were performed and plotted using *stats*

and *graphics* R packages.

Results

Levels of diversity

SNP genotyping rates and polymorphic status are shown in ST Tab 2. Genotyping rates ranged from 81.0 to 91.2%. Twenty-two SNPs were not successfully genotyped, and 8 SNPs were not polymorphic in the tested populations. These 30 SNPs were removed from the study and, hence, subsequent analyses were based on the variation in 114 SNPs. After Bonferroni correction no SNP or TRP showed a significant departure from the Hardy-Weinberg equilibrium in any population sample (data not shown). SNP minor allele frequencies (MAF) in population samples and global heterozygosities per marker and population are presented in ST Tab 5. Global heterozygosities per SNP were moderate-high, ranging from 0.12 to 0.51. Per population, global heterozygosities ranged from 0.30 to 0.43; populations with the lowest overall heterozygosity were the South-Saharan African and Amerindian groups, most likely as a result of ascertainment biased in the SNPs selected for this study.

All tandem repeat markers (TRPs) were successfully genotyped and were polymorphic in the tested populations. Genotyping rates were slightly lower compared to SNPs, ranging from 65.1 to 87.1% (ST Tab 3). Allele frequencies and global heterozygosities per marker and population are described in ST Tab 6. VNTR ss825679095 and STR ss825679089 were biallelic, both presented alleles of 4 and 5 motive repeats and showed similar heterozygosities (0.236 and 0.240, respectively). The other STRs ranged from 6 to 21 alleles, with core repeats ranging from 2 to 28. Global heterozygosities per marker revealed moderate-high diversity (from 0.37 to

0.84). The population with the highest overall TRP diversity was the Egyptian Siwa sample (0.62) whereas the populations with the lowest diversity were the two Bolivian samples (0.42 and 0.41).

Population relationships and genetic structure

Among European, African and Amerindian populations, global AMOVA indices were $F_{ST} = 0.150$ ($P < 0.0001$) and $F_{CT} = 0.140$ ($P < 0.0001$) for SNPs; and $F_{ST} = 0.122$ ($P < 0.0001$) and $F_{CT} = 0.105$ ($P < 0.0001$) for TRPs. As expected, variance among groups explained an important part of the global variance among populations (see MDS plots in Figure 1A for SNPs, and Figure 2A for TRPs).

In the case of the European and Mediterranean populations, the SNP data set revealed a relatively homogeneous cluster, with the only exception being the Egyptian Berber Siwa (SIW) population which was slightly closer to the South-Saharan African samples (Figure 1A). The TRP data set (Figure 2A) also showed a general homogeneous cluster, with a certain separation of some European and Mediterranean populations. AMOVA estimates were $F_{ST} = 0.012$ ($P < 0.0001$) and $F_{CT} = 0.008$ ($P < 0.0001$) for SNPs, and $F_{ST} = 0.016$ ($P < 0.0001$) $F_{CT} = 0.002$ ($P < 0.76$) for TRPs. These values suggested a geographic structure of genetic variation only for SNPs. MDS analyses using SNP data (Figure 1B) represented rough North-South and East-West axes in the first two dimensions. Moroccan, Algerian and Tunisian groups clustered together in the lower right quadrant, while the Egyptian Siwa and Jordan populations were placed in the upper right quadrant. The European populations clustered in the left quadrants, with few exceptions. Interestingly, the Basque Country is located opposite the Middle Eastern group, and the Bosnia-Herzegovinian samples opposite the North African

group. Populations of these four groups show high averaged F_{ST} genetic distances per populations (ST Tab7) compared with the rest.

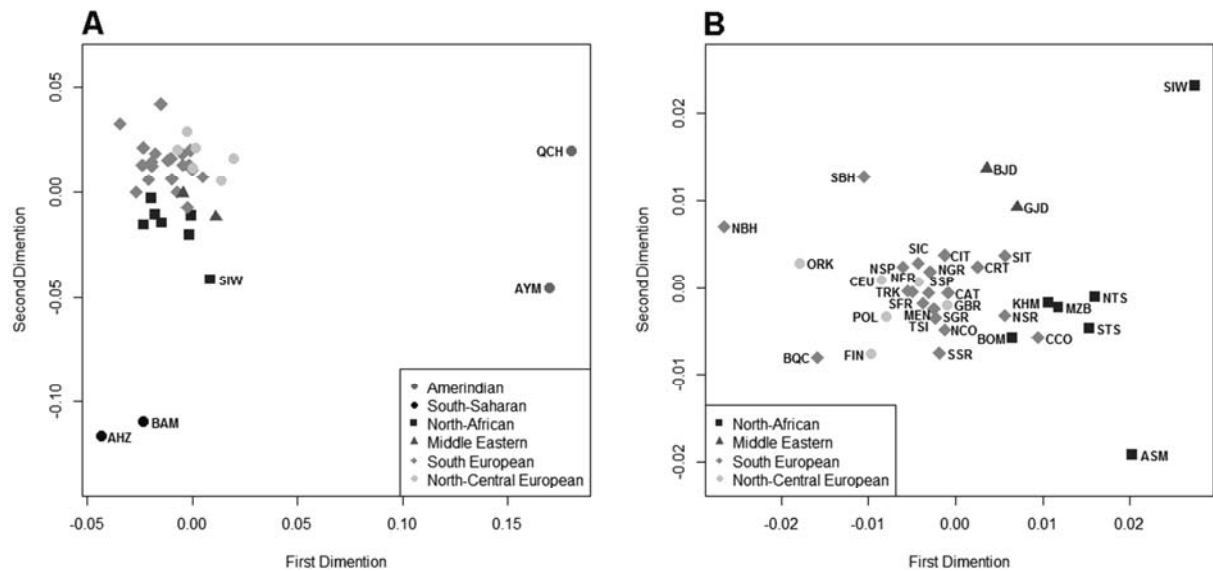


Figure 1- MDS plots for F_{ST} genetic distances between populations for the SNP dataset. A: European-Mediterranean, South-Saharan and Amerindian populations (observed stress 0.052). B: Only European and Mediterranean populations (observed stress 0.186). United Kindom: ORK and GBR; Finland: FIN; Warsaw: POL; Central Europe: CEU; France: NFR and SFR; Corsica: NCO and CCO; Spain: BQC, CAT NSP, SSP and MEN; Italy: TSI, CIT, SIT, NSR, SSR, SIC; Bosnia-Herzegovina: NBH and SBH; Greece: NGR, SGR and CRT; Turkey: TRK; Jordan: GJD and BJD; Morroco: ASM, KHM and BOM; Algeria: MZB; Tunisia: NTS and STS; Egypt: SIW; Coté d'Ivory: AHZ; Cameroon: BAM; Bolivia: AYM and QCH.

Concerning TRP data, most European and Mediterranean samples clustered together in the MDS plot without any geographic structure (Figure 2B), as reflected by the non-significant F_{CT} index. This lack of geographic structure indicates that the TRP dataset is not powerful enough to discriminate between European-Mediterranean geographic groups. However, a similar trend of higher averaged R_{ST} distances per population has also been observed for populations from the four groups mentioned above (ST Tab7).

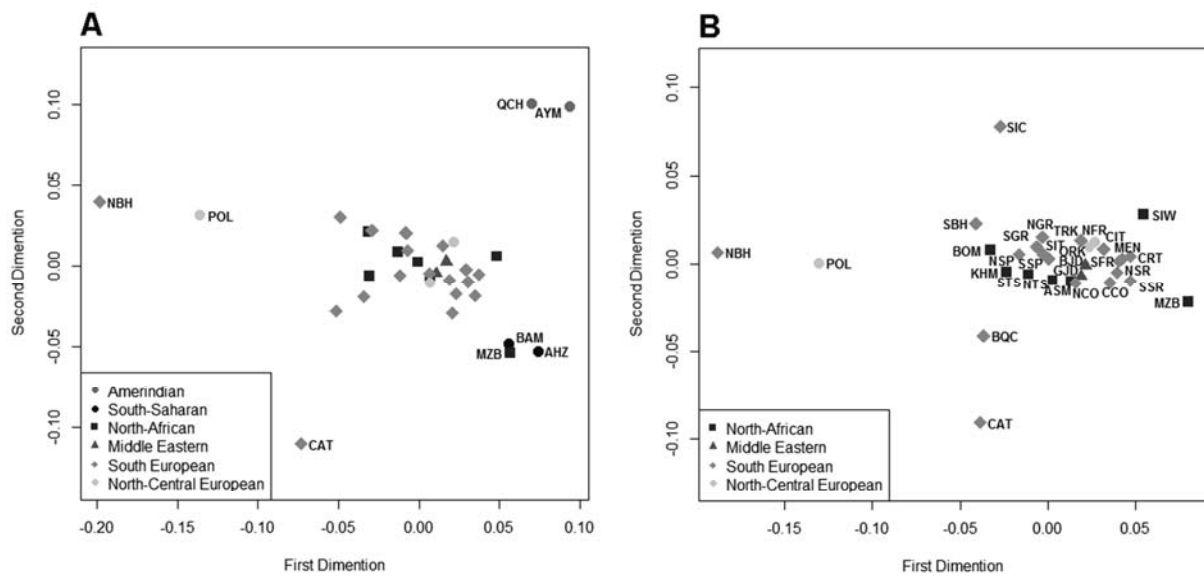


Figure 2- MDS plots for R_{ST} genetic distances between populations for the TRP dataset. A: European-Mediterranean, South-Saharan and Amerindian populations (observed stress 0.169). B: Only European and Mediterranean populations (observed stress 0.244).

Spatial structure

The Egyptian Siwa population was excluded from this analysis due to its characteristics of genetic outlier within the European and Mediterranean context (Figure 1). In the SNPs, sPCA suggested the existence of a global pattern (Global test $\max(t)=0.175$, $P=0.077$), whereas it did not detect any local structure (Local test $\max(t)=0.113$, $P=0.56$). The detected global patterns depicted genetic continuums orientated south-to-north in the first global sPCA score (Figure 3A) and east-to-west in the second (Figure 3B), with relative peaks in the first global score corresponding to the samples of Bosnia-Herzegovina and Italian TSI. For TRPs, sPCA confirmed the presence of a global structure (Global test $\max(t)=0.057$, $P=0.032$); Local test $\max(t)=0.043$, $P=0.83$). The observed global patterns were north-to-south (Figure 4A) and east-to-west in the Mediterranean, or southwest-to-northeast in continental Europe (Figure 4B). Samples from Bosnia-

Herzegovina and the Basque Country slightly contrasted with their neighbours in the first global component. These results indicate the lack of sharp genetic boundaries among the studied populations.

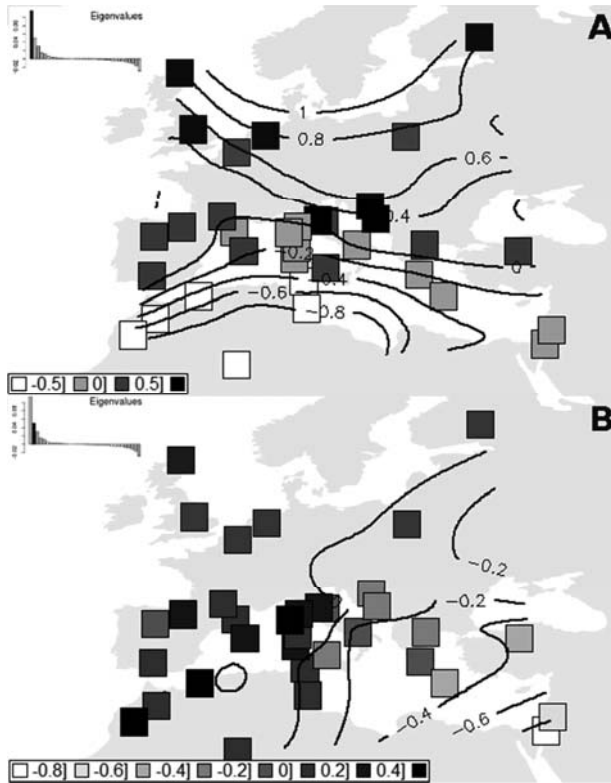


Figure 3- Maps displaying the first (A) and second (B) global eigenvalues scores of sPCA from SNP data.

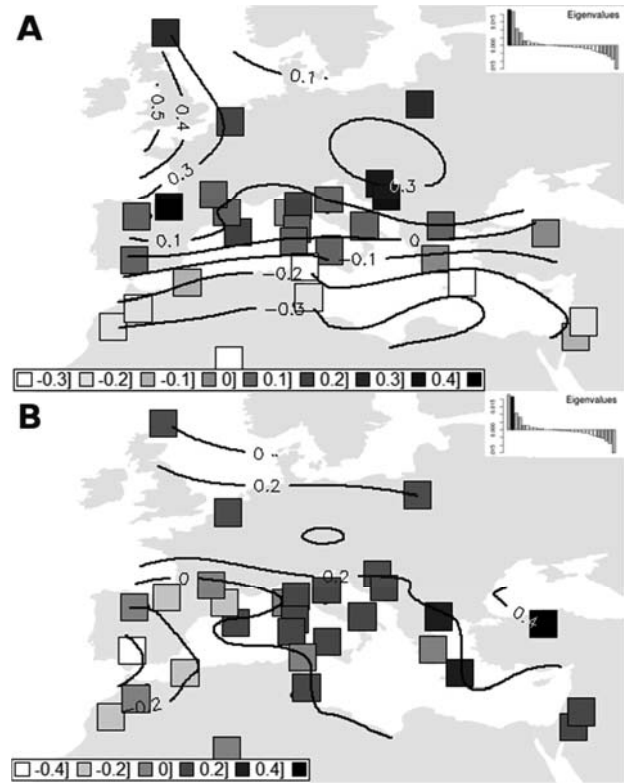


Figure 4- Maps displaying the first (A) and second (B) global eigenvalues scores of sPCA from TRP data.

Genetic admixture

STRUCTURE likelihood estimates and Evanno's method results are summarized in ST Tab8. The highest Delta K estimate corresponded to K=3 for both SNPs and TRPs. The ancestral genetic pools were designed as cluster A, B and C. Contour maps representing the spatial distribution of estimated parental genetic pools across populations are depicted in Figure 5.

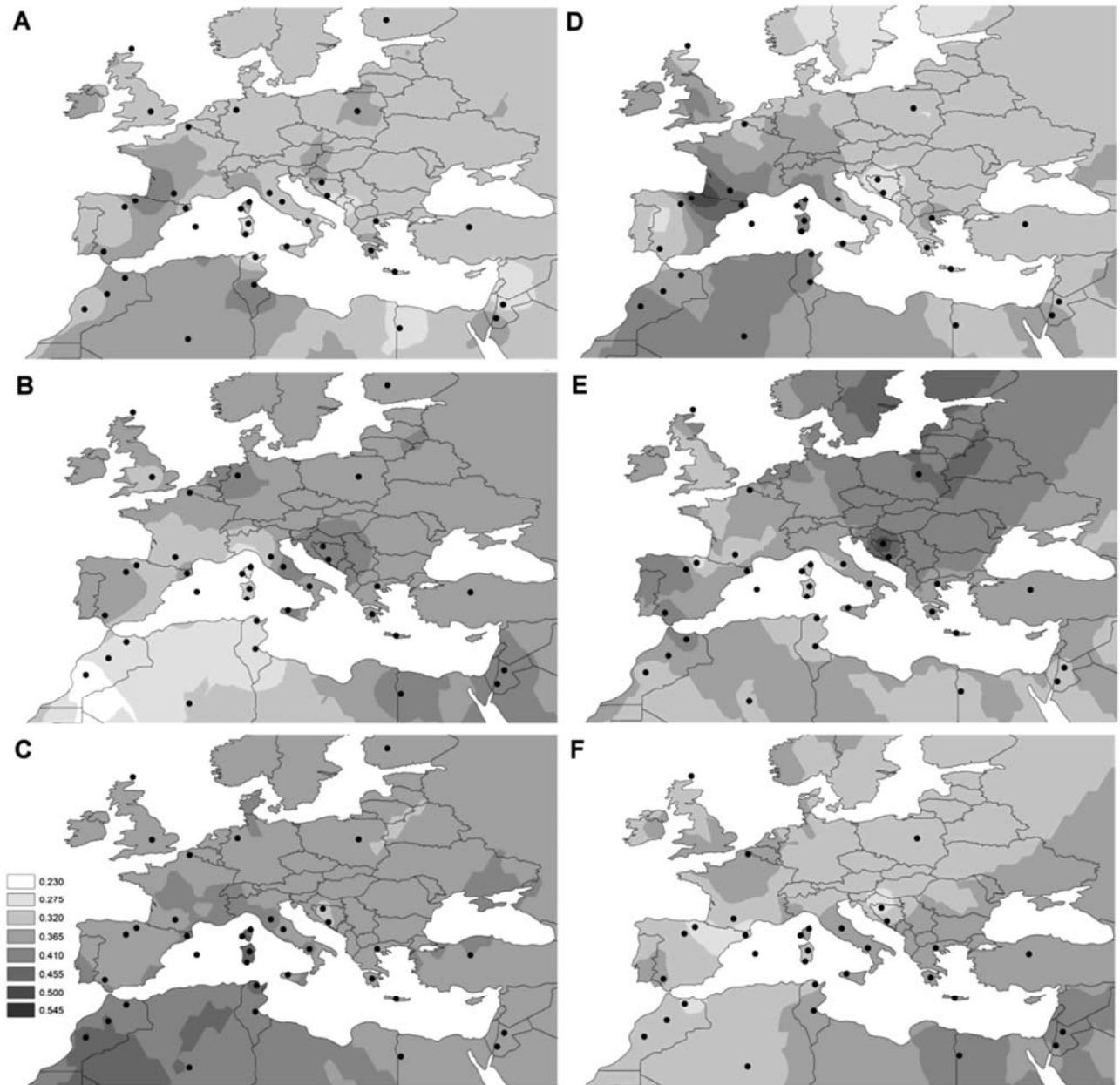


Figure 5- Contour maps of estimated parental genetic pools across populations. A: SNP's A cluster; B: SNP's B cluster; C: SNP's C cluster; D: TRP's A cluster; E: TRP's B cluster; F: TRP's C cluster. The legend indicates the top membership coefficient for each category. See ST Tab9 for population membership coefficients.

Estimated population membership coefficients of cluster A, B and C for SNPs and TRPs showed only slight differences. In both analyses, the population with the highest proportion of cluster A was the Basque County

(0.434 for SNP analyses and 0.525 for TRP analyses) (Figures 5A and 5D and ST Tab9). Regarding cluster B, the populations with the highest proportion were the Bosnian-Herzegovinian samples in both analyses (0.410-0.438 for SNPs and 0.431-0.512 for TRPs) (Figure 5B and 5E). Finally, in the case of cluster C, Mediterranean populations showed the highest proportions. However, for SNP analyses they were the Western North-African samples of Morocco and North Tunisia and the islands of Corsica and Sardinia (0.418-0.502), whereas for STR analyses, they were the Eastern Mediterranean samples of Siwa and General Jordans (0.395-0.409) (Figures 5C and 5F). No population had a membership coefficient close to 1, which would mean that none of the studied populations could be considered as representative of ancestral or parental genetic pools. Moreover, almost all parental genetic pools covariate significantly with geography as reflected by the spatial autocorrelation analysis ($p=0.88$ for SNP cluster A; $p<0.01$ for SNP cluster B; $p<0.05$ for SNP cluster C; $p<0.05$ for TRP cluster A; $p<0.05$ for TRP cluster B; and $p<0.05$ for TRP cluster C).

Detection of potential selection

The results of the locus by locus F_{ST} statistics and their significance as potential result of selection are presented in the ST Tab 10 (SNPs) and ST Tab 11 (TRPs). SNP F_{ST} values ranged from <0.0001 to 0.050 and 14 loci were potentially under selection (highlighted in ST Tab 10). Out of these 14 loci, four had large F_{ST} values (0.039-0.050) indicating potential positive selection, and the other ten showed low F_{ST} values (<0.0001 -0.002) suggesting balancing selection. TRP F_{ST} indices ranged from 0.001 to 0.078 and 5 loci were potentially under selection (highlighted in ST Tab 11). Out of these 5 loci, three presented large F_{ST} values (0.043-0.078), and the other

two had low F_{ST} values (0.001- 0.002).

The second method applied for detection of selection provided SPA scores for SNPs (ST Tab 10). Since no SPA score showed an extreme value, no SNP presented an extreme steep gradient of allele frequency across geography. SPA scores above the 95th percentile (1.55) across Europe identified seven SNPs (highlighted in ST Tab 10), and SPA scores across North-Africa above the 95th percentile (1.52) indicated six SNPs (ST Tab 10).

No genetic marker associated with CAD in our previous study [5] was detected as being under potential selective pressure.

Epidemiological analyses

As an approach to the contribution of population genetic structure to coronary event variation in Europe, mean genetic distances, sPCA scores, and ancestral admixture values were correlated with coronary incidences. Average F_{ST} and R_{ST} genetic distances were those between the population samples with CAD incidence data (see Materials and Methods section) and the four population groups (the Basque Country, Bosnia-Herzegovina, the Middle East, and North Africa) found to be most differentiated from the whole set of populations. The Egyptian Siwa sample was excluded from the North African group due to its outlying position. Correlation and regression results for all population pairs (11 for SNPs and 7 for TRPs) by gender are summarized in Table 1.

Genetic structure parameters		All populations				Populations with low-medium incidence			
		Men CER		Women CER		Men CER		Women CER	
		rho	R ²	rho	R ²	Rho	R ²	rho	R ²
SNP data	sPCA score								
	First Global Com.	0.72*	0.28'	0.61'	0.20'	0.89*	0.65*	0.71'	0.55*
	Second Global Com.	0.05	<0.01	0.02	<0.01	-0.49	0.02	-0.54	0.07
	First*Second Coms.		0.09		0.02		0.97**		0.89*
	F_{ST} distance to								
	Basque Country	0.48	0.04	0.36	0.02	0.61	0.31	0.54	0.23
	Bosnia-Herzegovina	-0.29	<0.01	-0.29	<0.01	-0.78*	0.24	-0.79*	0.29
	Middle East	0.70*	0.39*	0.52	0.42*	0.86*	0.73**	0.68	0.63*
	North Africa	0.63*	0.11	0.49	0.19	0.86*	0.82**	0.79*	0.82**
	Admixture coefficients								
	cluster A	0.13	<0.01	0.02	<0.01	0.04	<0.01	-0.17	<0.01
	cluster B	0.06	<0.01	0.21	<0.01	0.05	0.32	0.75'	0.43'
	cluster C	-0.17	<0.01	-0.18	<0.01	-0.32	0.14	-0.39	0.21
TRP data	sPCA score								
	First Global Com.	0.82*	0.51*	0.75'	0.41'	0.90'	0.98**	0.70	0.97**
	Second Global Com.	0.46	<0.01	0.43	<0.01	0.90'	0.31	0.80	0.29
	First*Second Coms.		0.90*		0.85*		0.98'		0.95
	R_{ST} distance to								
	Basque Country	0.76*	0.61*	0.75'	0.70*	0.70	<0.01	0.60	<0.01
	Bosnia-Herzegovina	-0.46	0.07	-0.57	<0.01	-0.70	0.79*	-0.9'	0.82*
	Middle East	0.28	<0.01	0.36	<0.01	0.00	0.60'	0.10	0.59'
	North Africa	0.32	<0.01	0.39	<0.01	0.10	0.64'	0.20	0.64'
	Admixture coefficients								
	cluster A	-0.89*	0.43'	-0.86*	0.36'	-1.00*	0.65'	-0.9'	0.67'
	cluster B	0.64	0.31	0.75'	0.24	0.40	0.81*	0.70	0.83*
	cluster C	0.04	<0.01	-0.11	<0.01	0.40	<0.01	0.20	<0.01

Table 1- Spearman's rho and % of explained interpopulation variance (adjusted R²) of coronary event rates by genetic structure parameters for SNP data and for TRP data. CER: Coronary event rate. (':P<0.1; *: P<0.05; **:P<0.01.).

For SNPs, the first global sPCA score and genetic distances with the Middle East and North Africa in men yielded significant positive correlations with coronary event rates. Regression analyses of the same data yielded significant results for F_{ST} distances with the Middle East, explaining 39% and 42% of the variance in coronary incidence in men and women, respectively.

For TRPs, coronary event rates were positively correlated with the first global sPCA score and R_{ST} distance to the Basque Country, and negatively with cluster A coefficients. In the regression analyses, the first global scores contributed to 51% of coronary event variance in men, while R_{ST} distance with the Basque Country accounted for 61% of coronary incidence variance in men and 70% in women. Finally, the interaction between the first and second global components explained 90% and 85% of the coronary event variance in men and women.

When correlations of SNP data in Table 1 were plotted, two groups of samples were observed: a group of high coronary incidence and a group of low-medium incidence (Figure 6 for sPCA scores, Figure S2 for F_{ST} distances, and Figure S3 for admixture proportions). For TRPs, samples with a low-medium coronary incidence showed a correlation trend similar to the one observed for SNPs (Figure 7 for sPCA scores, Figure S4 for R_{ST} distances, and Figure S5 for ancestral coefficients). Given this observation, post hoc regression and correlation analyses were carried out in the low-medium coronary incidence group (Table 1). The number of samples with high incidence was too small to test for correlation and regression.

For SNPs, coronary events were positively correlated with the first global sPCA component score (Figure 6) and F_{ST} distances to the Basque Country, the Middle East and North Africa, and negatively with distances to

Bosnia-Herzegovina (Figure S2) in both men and women. However, regression analyses were only significant for the first global sPCA score (contributing in 65% and 55% of coronary events in men and women), distances to Middle East ($R^2 = 73\%$ in men and 63% in women) and North Africa ($R^2 = 82\%$ in both men and women). It is interesting to note that multivariate regression analyses of the first and second global components explained almost all coronary variance (97% in men and 89% in women). Trends observed in the samples from the high incidence group were frequently similar to those described for the low-medium incidence group (Figure 6, Figure S2 and Figure S3).

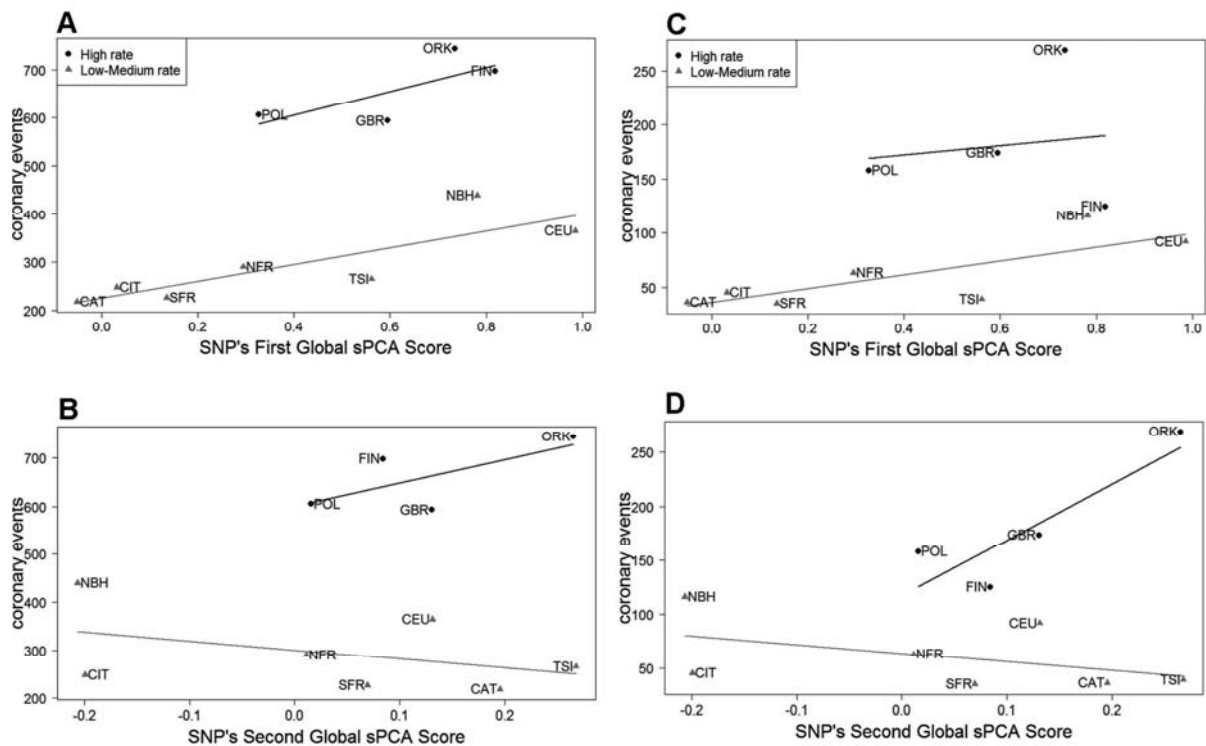


Figure 6- Correlation plots between coronary event rates and SNPs' global sPCA scores by incidence group and by gender (A and B men incidence vs global scores; C and D women incidence vs global scores).

TRP data analysis yielded similar results (Figure 7, Figure S4 and Figure S5). Variance in coronary events in the low-medium incidence group (98% in men and 97% in women) could be almost entirely explained by the first global sPCA component scores (Figure 7). R_{ST} genetic distances to Bosnia-Herzegovina and B ancestral coefficients explained around 80% of coronary incidence (79-83%) in both men and women (Figure S3 and Figure S4).

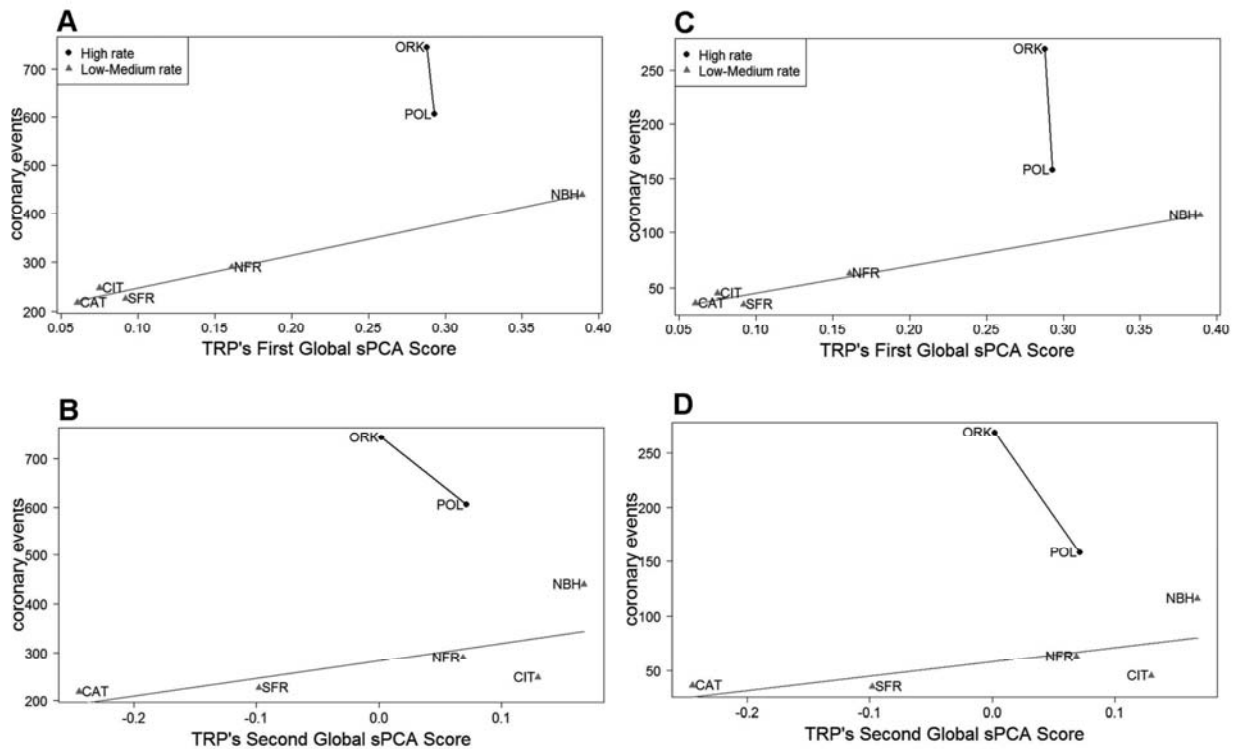


Figure 7- Correlation plots between coronary event rates and TRPs' global sPCA scores by incidence group and by gender (A and B men incidence vs global scores; C and D women incidence vs global scores).

Discussion

Our working hypothesis in this study was that the distribution of CAD genetic variants follows the same pattern as the whole genome across the European continent. To test this hypothesis, we have analyzed the molecular variation of 114 SNPs and 17 TRPs from six genomic regions (the

three *NOS* genes and 3 adjacent non-coding regions) in 35 European and Mediterranean populations and we have compared the population structure revealed by these markers with the distribution of CAD incidence across Europe.

Global population variation

Diversity indices for world-wide populations (F_{ST} : 0.15 SNPs-0.14 TRPs) as well as for the European-Mediterranean group (F_{ST} : 0.016 SNPs-0.012 TRPs) are in accordance with previous publications where other autosomal markers were analyzed in a similar group of populations [37–39]. Interestingly, the proportion of genetic variance attributable to differences among groups is lower for TRPs as compared with SNPs, both for worldwide and European-Mediterranean populations. This phenomenon has also been described by other authors [37] and could be due to the higher mutation rate of tandem repeats, which can produce alleles identical by state but not by descent, thereby rubbing out the effect of genetic drift differentiation in these populations [37].

*Selection signs in *NOS* gene regions*

In a previous survey of *NOS* genes [5], we found that *NOS* regions had low levels of linkage disequilibrium, 4 genetically independent SNPs were found associated with CAD, and several SNPs in *NOS3* correlated with CAD incidence in Europe. In this current study, the CAD-associated SNPs did not show any sign of selection. The CAD-correlated *NOS3* SNPs seemed to present mild allele frequency gradients across Europe, as revealed by the SPA scores, but they are not related to positive selection. However, some signs of selection have been detected in other markers. The VNTR

ss825679095 in the *NOS3* gene, associated in some previous reports with CAD and vessel diseases [40–42], appeared to be related to balancing selection processes. In general, our results are consistent with the absence in the literature of clear signs of selection for genetic variants associated with CAD [6,11] without excluding particular cases. In the *NOS1* gene, cases of positive and balancing selection have been detected. These markers might be related to neuronal pathophysiology since the *NOS1* gene has been associated with genetic predisposition to some mental disorders [43]. Also in the *NOS2A* gene, a few possible positive selection signs have been observed, most probably as a result of the role of this gene in acquired immunity [44]. In non-coding regions, signs of positive as well as balancing selection have been found in *NOS3* and *NOS2A* adjacent regions. In the absence of additional functional information, our results showing slightly differential patterns could be compatible with selection but random processes cannot be discarded.

European and North African genetic continuity

Population relationships provided by genetic distances and their graphical representation reflected an homogeneous genetic cluster for the European and Mediterranean populations with the exception of the well-known outlier population of the Egyptian Siwa Berbers [45] in the SNP set (Figure 1A), and a few samples in the TRP set (Figure 2A) probably due to data missingness. The European and Mediterranean populations showed a genetic continuum in the sPCA characterized by orthogonal north-to-south and east-to-west genetic clines for both kinds of markers. Therefore, the *NOS* genome variation is consistent with genetic variation described in other surveys [13,15], and the lack of genetic barriers between South Europe and

North Africa [21,37,39], especially in the westernmost part. These results highlight that both the genetic continuity around the Mediterranean basin and the well-known European genetic clines are two pieces of the same genetic structure phenomenon.

NOS variation patterns support the dual Neolithisation model.

Our hypothesis-free admixture analyses detected three ancestral components for both kinds of markers. The estimated ancestral contributions ranged between 15-55% and identified the following as populations with the highest ancestral proportions: Basque Country (for cluster A), Bosnia-Herzegovina (for cluster B), North Africa (for cluster C from SNPs), and Jordan (for cluster C from TRPs). The slight discrepancy between SNPs and TRPs on the identification of the populations with the highest cluster C component could be related to different features of the analysed datasets. The limited number of analysed markers, the ascertainment bias for the selected highly polymorphic SNPs, together with the different mutation rate of SNPs and TRPs could influence the results of admixture analysis [46].

These results contrast with classical Palaeolithic vs Neolithic interpretation. Previous admixture analyses [18–20] estimated a large Neolithic contribution of 70% in the Balkan region decreasing westwards to 30% in Iberia, using current Basque Country and Middle East populations as representative of parental populations. However, it was argued that if the chosen parental populations were not representative of the ancestral ones, these results would show a longitudinal genetic variation as a result, not only of admixture events, but also of isolation by distance [20]. Indeed, our results showed the Basque Country and Jordan (second variation axis), and Bosnia-Herzegovina and North Africa (primary variation axis) to be poles of

the observed continuous genetic frame. Since no ancestral genetic proportion was close to 100%, choosing any given population as representative of ancestral pools would not be justified, and they cannot be considered outlier genetic entities as was previously suggested for the Basque Country [47]. Our data suggest that these four groups of samples simply represent the extremes of the European and Mediterranean continuum of genetic variation studied in this survey.

However, our results support a model of dual Neolithic dispersal in Europe and the Mediterranean region. A synthesis of archaeological records and ancient DNA analyses indicates that Neolithic transition in Europe, around 11 thousand years ago (kya), involved two main routes [48]. Early Neolithic man had appeared in Central Europe 7.5 kya associated with the *Linearbandkeramik* (LBK) culture [49], whereas the diffusion in the West Mediterranean (8-7.5 kya) was linked to Cardial culture and attributed to “maritime pioneers” [50]. Recent genetic data based on ancient DNA supported the dual model of Neolithic diffusion for maternal and paternal lineages showing genetic discontinuity among Neolithic samples [48,51–53]. This dual Neolithic expansion has been also described using autosomal data related to the lactase persistence gene in Neolithic samples [52,53]. This context provides a general framework to interpret the ancestral components showed by *NOS* genes. Accordingly, the ancestral cluster A component could be attributed to a component of Palaeolithic European origin since it presents the highest proportion in the Basque Country and the lowest in groups where Neolithic economy set up. Cluster B, with the highest frequency in Bosnia-Herzegovina, could be interpreted as coming from early farmers with LBK culture dwelling along the Danube River. This ancestral pool would have separated from Levant and Middle East farmers 12 kya in consistency with

recent genetic data [54]. Thus, our results highlight the need to properly genetically characterize the populations from the Balkan region, because these populations' genome might carry key information about 8 kya European continental Neolithisation. Finally, cluster C could represent the North African genetic background and its dispersion could be a product of the Mediterranean maritime Neolithisation route. This interpretation is in agreement with the recently described gradient of Palaeolithic “Maghrebi” ancestral component decreasing eastwards [55], also being observed in South European populations [56]. These results would also be consistent with recent spatial simulations in the West Mediterranean showing that current genetic patterns were mostly constituted in pre-Neolithic times [37]. On the other hand, the C ancestral component from TRP data showed the highest frequencies in the Eastern Mediterranean decreasing towards the West. These differences could be related to the intrinsic features of these markers.

European and North-African Palaeolithic components shaped by a Neolithic distribution

Genetic variation in the *NOS* regions shows two main orthogonal axes of variation, North-South and East-West, as revealed by sPCA and genetic distance analyses. Besides, ancestry estimation in *NOS* regions variation is interpreted as a substantial presence of Palaeolithic gene pool in the current genetic variation in Europe and the Mediterranean region (>60%).

These results are explained by recent population expansion simulations. According to these simulations, the east-to-west variation axis would be a result of allele frequency clines generated in range expansions [46,57]. In addition, a high proportion of Palaeolithic component can be

explained by introgression, however small, of Palaeolithic elements during the Neolithic “wave of advance” [46,57]. Thus, North African Palaeolithic introgression in Mediterranean maritime Neolithic expansion led to the spread of the “Maghrebi” ancestral component across the Western Mediterranean. Interestingly, the orthogonality of the main north-to-south axis of genetic variation is explained by the expansion in parallel sectors differentiated by drift aligned along the direction of expansion [58,59]. The dual Neolithic expansion through continental Europe and the Mediterranean area initially generated two highly differentiated regions of low genetic diversity, especially in their westernmost part. Later on, migrations and gene flow related to metal ages and flourishing civilizations might blur the sectorial boundaries, thus genetically enriching these areas and generating current orthogonal patterns of genetic diversity.

In summary, the *NOS* gene variation described in this study could indicate that the current north-to-south and east-to-west patterns of genetic gradient, diversity, and admixture were generated by the deep impact of dual Neolithic expansion, and have not yet been eroded by posterior migrations.

Neolithic expansion and European coronary incidence

Since it was assessed that classical risk factors (e.g. fat consumption, alcohol intake, smoking habits) contribute to 30-40% of CAD population incidence [3,4], the population distribution of CAD genetic architecture may explain the remaining (60-70%) variance. The multivariate analysis of TRPs first and second sPCA scores in the eleven tested populations already explain a larger 85-90% of coronary incidence. As several classical risk factors present a latitudinal distribution (cultural differences in food and drink production), when the north-to-south distributed first global sPCA

component is included in the regression models, a percentage (not higher than 30-40%) of the variance estimations could be accounted for by those factors. In the correlation plots, two groups of populations differentiated by high or low-moderate incidence were observed. They can be interpreted as groups with similar environmental conditions in food and drink production. Thus, considering these groups separately will considerably reduce the amount of coronary incidence variance explained by these risk factors, at the expense of losing statistical power. At this point, it is worth mentioning that the sPCA components of both markers in the low-moderate incidence group of populations explained almost the total amount (97-98%) of coronary events rates. In conclusion, these results suggest that the main parameters of genetic population structure of the European and North African populations tested in this survey explain a substantial proportion of the observed differences in CAD incidence across European populations. This reinforces our working hypothesis that CAD genetic architecture is commonly structured in the European continent and contributes to the geographic differences in CAD incidence.

Correlations between coronary events and genetic structure parameters could at first glance suggest that a high proportion of the “European Neolithic” ancestral cluster (B) and shorter genetic distances from Bosnia-Herzegovina would represent higher coronary incidence, whereas the “Maghrebi” ancestrality (C) would provide European populations with genetic CAD protective factors. However, a more plausible explanation would be that the observed coronary incidence is related to genetic diversity in the populations. Thus, high proportions of both “European Neolithic” and “Maghrebi” ancestral components in the South European populations would constitute protective factors for CAD.

These initial interpretations need to be further assessed in order to be confirmed and refined with studies including both more polymorphisms and samples with a better geographic representation of Central European populations. In any case, the inferred genetic structure and admixture components of this survey, as well as the interpretation of a dual Neolithisation, appear to be reliable insofar as they have been able to explain the distribution pattern of a complex phenotype with a genetic architecture commonly structured across populations.

Conclusions

This survey provides data on the distribution of molecular variation in the *NOS* genes regions and non-coding adjacent regions among European and Mediterranean populations for the first time. No selective forces seem to have modulated the variation of these genetic markers related with CAD: neither in those associated with the coronary phenotype nor in those correlated with the distribution of cardiovascular incidence. On the other hand, geographic differences in European CAD incidence can be explained by parameters of genetic structure. This suggests, for the first time, that CAD genetic bases are commonly structured in the European continent. Regarding population structure inferences, the lack of a Mediterranean genetic barrier has been confirmed. In addition, a major European and North African Palaeolithic genetic contribution, instead of continental European Neolithic ancestry, has been identified for most populations. Surprisingly, these large amounts of Palaeolithic ancestral components depict a recent population distribution that can be explained by dual Neolithic expansion in the European continent and the Mediterranean basin.

Acknowledgements

This research has been supported by the Ministerio de Ciencia y Tecnología CGL2008-03955 and CGL2011-27866 projects to PM; the Ministerio de Ciencia y Tecnología Predoctoral Grant (BES-2009-016334) and Travel Grant (EEBB-I-12-04854) to RCT; and the Beatriu de Pinós postdoctoral grant (2009 BP-B 00274) to MV. We want to thank all our collaborators that provided samples of general populations: Spanish National DNA Bank (Spain), N Harich and M Kandil (Morocco), M Sadiq (Jordan), JM Dougoujon (France), F Cruciani (Italy and Cameroon), C Calo (Sardinia), A Kouvatsi (Greece), Ewa Bartnik (Poland), L Murphy (United Kingdom) and The 1000 Genomes Project. All volunteers are gratefully acknowledged for sample donation.

References

1. Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, Arveiler D, Rajakangas M, et al. (1994) Myocardial infarction and coronary deaths in the World Health Organization MONICA Project. Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents. *Circulation* 90: 583–612.
2. Keys A, Aravanis C, Blackburn H, Buzina R, Djordjevic B, et al. (1980) Seven Countries. A Multivariate Analysis of Death and Coronary Heart Disease. Harvard Univ Press: 1–381.
3. Artaud-Wild SM, Connor SL, Sexton G, Connor WE (1993) Differences in coronary mortality can be explained by differences in cholesterol and saturated fat intakes in 40 countries but not in France and Finland. A paradox. *Circulation* 88: 2771–2779.
4. The WHO MONICA Project (1994) Ecological Analysis of the Association between Mortality and Major Risk Factors of Cardiovascular Disease. *Int J Epidemiol* 23: 505–516.
5. Carreras-Torres R, Kundu S, Zanetti D, Esteban E, Via M, et al. (2014) Genetic risk score of NOS gene variants associated with myocardial infarction correlates with coronary incidence across Europe. *PLoS ONE*. doi:10.1371/journal.pone.0096504
6. Corona E, Dudley JT, Butte AJ (2010) Extreme evolutionary disparities seen in positive selection across seven complex diseases. *PLoS One* 5: e12236.
7. Casto AM, Feldman MW (2011) Genome-wide association study SNPs in the

human genome diversity project populations: does selection affect unlinked SNPs with shared trait associations? *PLoS Genet* 7: e1001266.

8. Chen R, Corona E, Sikora M, Dudley JT, Morgan A, et al. (2012) Type 2 diabetes risk alleles demonstrate extreme directional differentiation among human populations, compared to other diseases. *PLoS Genet* 8: e1002621.
9. Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, et al. (2009) Signals of recent positive selection in a worldwide sample of human populations. *Genome Res* 19: 826–837.
10. Marigorta UM, Lao O, Casals F, Calafell F, Morcillo-Suárez C, et al. (2011) Recent human evolution has shaped geographical differences in susceptibility to disease. *BMC Genomics* 12: 55.
11. Ding K, Kullo IJ (2011) Geographic differences in allele frequencies of susceptibility SNPs for cardiovascular disease. *BMC Med Genet* 12: 55.
12. Coop G, Pickrell JK, Novembre J, Kudaravalli S, Li J, et al. (2009) The role of geography in human adaptation. *PLoS Genet* 5: e1000500.
13. Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, et al. (2008) Correlation between genetic and geographic structure in Europe. *Curr Biol* 18: 1241–1248.
14. Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, et al. (2008) Genes mirror geography within Europe. *Nature* 456: 98–101.
15. Auton A, Bryc K, Boyko AR, Lohmueller KE, Novembre J, et al. (2009) Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res* 19: 795–803.
16. Mungrue IN, Bredt DS, Stewart DJ, Husain M (2003) From molecules to mammals: what's NOS got to do with it? *Acta Physiol Scand* 179: 123–135.
17. Hindorff L, Sethupathy P, Junkins H, Ramos EM, Mehta JP, et al. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 106: 9362–9367.
18. Chikhi L, Nichols R, Barbujani G, Beaumont M (2002) Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci U S A* 99: 11008–11013.
19. Dupanloup I, Bertorelle G, Chikhi L, Barbujani G (2004) Estimating the impact of prehistoric admixture on the genome of Europeans. *Mol Biol Evol* 21: 1361–1372.
20. Belle EMS, Landry P-A, Barbujani G (2006) Origins and evolution of the Europeans' genome: evidence from multiple microsatellite loci. *Proc Biol Sci* 273:

- 1595–1602.
21. Botigué LR, Henn BM, Gravel S, Maples BK, Gignoux CR, et al. (2013) Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. *Proc Natl Acad Sci U S A* 110: 11791–11796.
 22. Abecasis GR, Auton A, Brooks LD, DePristo M, Durbin RM, et al. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491: 56–65.
 23. Via M, González-Pérez E, Esteban E, López-Alomar A, Vacca L, et al. (2003) Molecular variation in endothelial nitric oxide synthase gene (eNOS) in western Mediterranean populations. *Coll Antropol* 27: 117–124.
 24. Asensi V, Montes H, Valle E, Ocaña MG, Astudillo A, et al. (2007) The NOS3 (27-bp repeat, intron 4) polymorphism is associated with susceptibility to osteomyelitis. *Nitric Oxide* 16: 44–53.
 25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MR, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559–575.
 26. Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10: 564–567.
 27. Jombart T, Devillard S, Dufour A-B, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity (Edinb)* 101: 92–103.
 28. Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403–1405.
 29. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
 30. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14: 2611–2620.
 31. Earl D, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4: 359–361.
 32. Relethford JH (2008) Geostatistics and spatial analysis in biological anthropology.

Am J Phys Anthropol 136: 1–10.

33. Thioulouse J, Chessel D, Champely S (1995) Multivariate analysis of spatial patterns: a unified approach to local and global structures. *Environ Ecol Stat* 2: 1–14.
34. Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity (Edinb)* 103: 285–298.
35. Yang W-Y, Novembre J, Eskin E, Halperin E (2012) A model-based approach for analysis of spatial structure in genetic data. *Nat Genet* 44: 725–731.
36. Kuulasmaa K, Tunstall-Pedoe H, Dobson A, Fortmann S, Sans S, et al. (2000) Estimation of contribution of changes in classic risk factors to trends in coronary-event rates across the WHO MONICA Project populations. *Lancet* 355: 675–687.
37. Currat M, Poloni ES, Sanchez-Mazas A (2010) Human genetic differentiation across the Strait of Gibraltar. *BMC Evol Biol* 10: 237.
38. El Moncer W, Esteban E, Bahri R, Gayà-Vidal M, Carreras-Torres R, et al. (2010) Mixed origin of the current Tunisian population from the analysis of Alu and Alu/STR compound systems. *J Hum Genet* 55: 827–833.
39. Athanasiadis G, Moral P (2013) Spatial principal component analysis points at global genetic structure in the Western Mediterranean. *J Hum Genet*: 1–4.
40. Matyar S, Attila G, Acartürk E, Akpınar O, Inal T (2005) eNOS gene intron 4 a/b VNTR polymorphism is a risk factor for coronary artery disease in Southern Turkey. *Clin Chim acta* 354: 153–158.
41. Hassan A, Gormley K, O’Sullivan M, Knight J, Sham P, et al. (2004) Endothelial nitric oxide gene haplotypes and risk of cerebral small-vessel disease. *Stroke* 35: 654–659.
42. Khurana VG, Meissner I, Sohni YR, Bamlet WR, McClelland RL, et al. (2005) The presence of tandem endothelial nitric oxide synthase gene polymorphisms identifying brain aneurysms more prone to rupture. *J Neurosurg* 102: 526–531.
43. Galimberti D, Scarpini E, Venturelli E, Strobel A, Herterich S, et al. (2008) Association of a NOS1 promoter repeat with Alzheimer’s disease. *Neurobiol Aging* 29: 1359–1365.
44. Velez D, Hulme W, Myers J, Stryjewski M, Abbate E, et al. (2010) Association of SLC11A1 with tuberculosis interactions with NOS2A and TLR2 in African-Americans and Caucasians. *Int J Tuberc Lung Dis* 13: 1068–1076.

45. González-Pérez E, Esteban E, Via M, Gayà-Vidal M, Athanasiadis G, et al. (2010) Population relationships in the Mediterranean revealed by autosomal genetic data (Alu and Alu/STR compound systems). *Am J Phys Anthropol* 141: 430–439.
46. Currat M, Excoffier L (2005) The effect of the Neolithic expansion on European molecular diversity. *Proc R Soc* 272: 679–688.
47. Laayouni H, Calafell F, Bertranpetit J (2010) A genome-wide survey does not show the genetic distinctiveness of Basques. *Hum Genet* 127: 455–458.
48. Pinhasi R, Thomas MG, Hofreiter M, Currat M, Burger J (2012) The genetic history of Europeans. *Trends Genet* 28: 496–505.
49. Itan Y, Powell A, Beaumont M, Burger J, Thomas MG (2009) The origins of lactase persistence in Europe. *PLoS Comput Biol* 5: e1000491.
50. Linstädter J, Medved I, Solich M, Weniger G-C (2012) Neolithisation process within the Alboran territory: Models and possible African impact. *Quat Int* 274: 219–232.
51. Sampietro ML, Lao O, Caramelli D, Lari M, Pou R, et al. (2007) Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proc R Soc* 274: 2161–2167.
52. Lacan M, Keyser C, Ricaut F-X, Brucato N, Tarrús J, et al. (2011) Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proc Natl Acad Sci* 108: 18255–18259.
53. Lacan M, Keyser C, Ricaut F, Brucato N, Duranthon F, et al. (2011) Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proc Natl Acad Sci U S A* 108: 9788–9791.
54. Haber M, Gauguier D, Youhanna S, Patterson N, Moorjani P, et al. (2013) Genome-wide diversity in the levant reveals recent structuring by culture. *PLoS Genet* 9: e1003316.
55. Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, et al. (2012) Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS Genet* 8: e1002397.
56. Sánchez-Quinto F, Botigué LR, Civit S, Arenas C, Avila-Arcos MC, et al. (2012) North African populations carry the signature of admixture with Neandertals. *PLoS One* 7: e47765.
57. Klopstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the

wave of a range expansion. *Mol Biol Evol* 23: 482–490.

58. François O, Currat M, Ray N, Han E, Excoffier L, et al. (2010) Principal component analysis under population genetic models of range expansion and admixture. *Mol Biol Evol* 27: 1257–1268.
59. Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol Evol* 23: 347–351.

4. DISCUSSIÓ

4.1 Discussió general

En aquesta tesi diferents tipus de polimorfismes són analitzats en diferents regions genètiques tan a nivell epidemiològic com poblacional en mostres europees i mediterrànies. Aquesta discussió es troba estructurada d'acord amb els objectius proposats en la secció segona.

4.2 Identificació de polimorfismes del gen *F13A1* de risc per l'infart de miocardi en població espanyola i possible interacció genètica de risc dins del gen

L'anàlisi conjunt de diferents polimorfismes localitzats en el gen *F13A1* ha permès obtenir un possible model interactiu entre els marcadors genètics que explicaria les contradiccions obtingudes en estudis anteriors.

4.2.1 Marcadors de risc per infart de miocardi al gen *F13A1* en població espanyola

Inicialment, en aquest treball es va realitzar una aproximació epidemiològica individual als polimorfismes analitzats, on els resultats obtinguts no van destacar cap efecte individual remarcable, tal i com s'havia observat en altres mostres mediterrànies.

Pel que fa als SNPs analitzats, el polimorfisme Val34Leu (rs5985) va ser observat a una freqüència similar a altres poblacions europees (0.2-0.3) i no va mostrar associació per l'infart de miocardi en la nostra mostra de malalts espanyols. En altres estudis realitzats prèviament per altres autors, aquest polimorfisme tampoc s'havia trobat associat a malalties vasculars en poblacions del sud d'Europa [86,87], però sí havia destacat l'al·lel Leu34 com a protector davant trombosis en poblacions del centre-nord d'Europa [86,87]. Aquesta associació s'atribueix a la proximitat de l'aminoàcid

corresponent al centre d'activació del factor de coagulació per part de la trombina produint fibres de fibrina més sensibles a la fibrinòlisi [15]. El polimorfisme Pro564Leu (rs5982) també va ser observat a freqüències similars en poblacions europees (0.1-0.2) sense mostrar associació significativa en el nostre estudi. Per aquest polimorfisme, altres estudis havien destacat el rol de risc de l'al·lel Leu564 en accidents cerebrovasculars hemorràgics [88], atribuïble a l'acció del polimorfisme en la interacció entre la subunitat A i la subunitat B del factor de coagulació i, per tant, en la disponibilitat de factor de coagulació activat [14].

En canvi, cap dels STRs analitzats havia estat estudiat epidemiològicament per cap malaltia vascular. Un d'aquests polimorfismes, el *F13A1*-STR localitzat a 704 pb de l'inici de transcripció, és utilitzat en anàlisi forense i, per tant, àmpliament descrit en moltes poblacions d'arreu del planeta. En el nostre estudi, destaca inicialment l'al·lel més petit en nombre de repeticions (3.2) per presentar un paper protector cap a l'infart de miocardi. L'altre STR, el D6S477 localitzat 3.6 kb enllà del punt final de transcripció, va mostrar freqüències similars a altres estudis [89,90] i no va mostrar cap associació destacada amb l'infart de miocardi.

4.2.2 Interacció de la variació genètica de risc dins del gen *F13A1*

En el segon pas estratègic d'aquest estudi, les anàlisis que tingueren en compte més d'un polimorfisme a la vegada mostraren efectes interactius entre els marcadors de risc.

Es va observar com la combinació al·lèlica o haplotip Val34-Pro564 presentava una sobretransmissió significativa (36 de 45 casos possibles) de pares heterozigots a fills afectats. Aquesta interacció significativa també es va observar en la sobretransmissió (21 casos dels 23 possibles) de l'al·lel

Pro564 del polimorfisme rs5982 quan s'analitzava separatament la submostra de pacients amb genotip Val34Val34 pel polimorfisme rs5985. A la vegada, pel que fa el *F13A1*-STR, la combinació dels al·lells d'aquest STR amb l'al·lel Pro564 mostrava com es mantenia el rol protector de l'al·lel de 3.2 repeticions i com apareixia l'al·lel de 7 repeticions, el de més repeticions entre els freqüents, com a factor de risc.

Aquests resultats s'expliquen des d'un punt de vista epistàtic o interactiu entre els factors de risc. S'estima que en les malalties complexes les interaccions entre factors de risc genètics i ambientals són més importants que els efectes individuals [91]. De fet, aquestes interaccions serien la causa de la manca de replicació de resultats en la majoria de estudis d'associació. De la mateixa manera que interaccions entre diferents gens poden implicar un risc important per una malaltia complexa com la cardiovascular, interaccions de polimorfismes dins d'un mateix gen poden ser els responsables del rol de risc o protector que presenti una determinada versió d'un producte gènic. A més, un altre factor a tenir en compte, és el doble rol que les variants genètiques poden mostrar en les malalties complexes. Pel que fa als polimorfismes estudiats, en similars poblacions nord-centre europees on l'al·lel Leu34 ha destacat com a protector davant trombosis [86,87], aquest mateix al·lel s'ha mostrat de risc davant fenòmens hemorràgics [92,93]. Per tant, probablement, aquest és el cas del nostre estudi on la implicació del factor de coagulació XIII en processos vasculars depèn de la combinació de variants genètiques en el gen *F13A1*, que són les que acabarien conformant una condició hemorràgica o trombòtica del factor de coagulació. Una combinació equilibrada o neutralitzada de variants pro-trombòtiques (com Val34, Pro564 i *F13A1*-STR*7) i variants pro-hemorràgiques (com Leu34, Leu564 i *F13A1*-STR*3.2) en diferents llocs del

gen *F13A1* pot originar resultats negatius d'associació quan aquests marcadors són analitzats individualment.

En el cas del polimorfisme *F13A1*-STR, aquest STR es trova localitzat a la regió promotora del gen entre seqüències reguladores, anomenades GATA-1 i Ets-1, responsables d'estimular la transcripció del gen [94]. Aquest polimorfisme en nombre de repeticions variables provoca un distanciament entre aquestes regions reguladores i, possiblement, podria debilitar l'estabilitat del complex estimulador de la transcripció i acabar afectant als nivells d'expressió del gen. Aquesta alteració funcional per part de polimorfismes STR ha estat prèviament descrita en altres regions genètiques promotores i ha estat proposat com a mecanisme de regulació capaç d'una regulació quantitativa dins d'un espectre continu d'expressió [95]. Per tant, aquest polimorfisme *F13A1*-STR podria mostrar-se com a funcional generant diferents nivells d'expressió de FXIII segons el nombre de repeticions i explicant el rol protector de l'al·lel amb un nombre més petit de repeticions (3.2) i el rol de risc de l'al·lel amb un nombre més elevat de repeticions entre els freqüents (7) davant malalties trombòtiques.

4.2.3 Efecte poblacional en la identificació de marcadors de risc amb rol interactiu

Tal i com s'ha esmentat anteriorment, els estudis previs d'associació on el polimorfisme Val34Leu apareixia relacionat amb malalties vasculars trombòtiques i hemorràgiques van ser realitzats principalment en poblacions del centre-nord d'Europa. En canvi, en la majoria d'estudis on el polimorfisme no mostrava cap efecte significatiu les mostres utilitzades eren provinents de poblacions mediterrànies del sud d'Europa.

Inicialment aquesta contradicció podria ser explicada pel fet que les

malalties complexes presenten diferents detonants genètics segons la regió geogràfica de la població on apareixen. Per tant, estariem davant d'una especificitat de factors genètics associats segons l'origen poblacional. Però els nostres resultats no apunten en aquesta direcció. El fet que en la nostra mostra sud europea els polimorfismes estudiats no presentin associació individual però sí en combinació dóna peu a proposar un model interactiu entre marcadors genètics de risc que són detectats segons les característiques intrínseques d'aquests marcadors en la població on es troben. Per exemple, ha estat suggerit que, sota un model interactiu, factors independents de risc només poden ser estadísticament detectats depenent de les freqüències al·lèliques en què es presentin en la població [91]. Alhora, el desequilibri de lligament entre els marcadors de risc pot contribuir a la seva detecció quan la interacció es produeix dins d'un mateix gen. Al continent europeu és àmpliament coneguda la baixa diferenciació genètica entre poblacions que alhora es correlaciona amb la geografia [75]. Les característiques que en destaquen d'aquesta distribució geogràfica són una major heterozigositat global i un menor desequilibri de lligament global de les poblacions del sud d'Europa comparades amb les del nord-centre d'Europa [75]. Per tant, en un model interactiu, estudis realitzats en poblacions nord-centre europees tindran una tendència major a assolir significació estadística que estudis realitzats en poblacions sud europees. Aquest és precisament el cas del nostre estudi realitzat amb polimorfismes del gen *F13A1* en una mostra epidemiològica mediterrània on l'associació estadística no apareix de manera individual i, per contra, sí sota un model interactiu entre els marcadors genètics.

4.3 Identificació de factors de risc en els gens *NOS* i distribució poblacional de risc a Europa i la Mediterrània en relació a la incidència coronària

La incorporació de l'ús de *scores* de risc genètic en l'estudi de la distribució geogràfica dels factors de risc dels gens *NOS* ha permès interpretar la diferent incidència cardiovascular al continent europeu degut a una possible diferent susceptibilitat genètica de les poblacions.

4.3.1 Identificació de marcadors de risc genètic dels gens *NOS* per infart de miocardi en població europea

Del conjunt de 61 SNPs analitzats en els tres gens *NOS* (*NOS3*, *NOS1* i *NOS2A*) 5 polimorfismes van presentar indicis d'una certa associació ($p < 0.1$) amb el fenotip infart de miocardi. El nombre final de marcadors associats un cop es va comprovar la independència genètica (desequilibri de lligament $r^2 < 0.2$) entre aquests marcadors va ser de 4 SNPs, 3 en el gen *NOS3* i un en el gen *NOS1*.

Segons la literatura disponible, el gen més associat a les malalties cardiovasculars del tres gens *NOS* és el gen *NOS3*, ja que és el màxim responsable de la disponibilitat d'òxid nítric a l'endoteli de la paret vascular, i per tant, de la vasodilatació i vasocontracció. Així, no és sorprenent que la majoria de variació genètica associada al nostre estudi es trobi en aquest gen. Un dels polimorfismes detectats en el nostre estudi (rs1799983) ha estat àmpliament estudiat per malalties cardiovasculars i hipertensió, però amb resultats ben contradictoris. La majoria de publicacions en què aquest polimorfisme es troba associat són treballs amb un nombre de mostres baix, alhora que sembla trobar-se associat amb més facilitat en mostres asiàtiques que europees [18–23]. Fins i tot, s'han observat els dos al·lels com de risc

cardiovascular, fenomen comunament anomenat de “flip-flop” [96]. L'al·lel determinat com de risc en aquest estudi és l'al·lel G. En suport a aquest resultat tenim que el nombre de mostres europees utilitzades s'aproxima a les 5.000, i en contra, que aquest polimorfisme presenta uns dels índex de qualitat d'imputació més baixos (0.6-0.7) entre els marcadors utilitzats en l'estudi. Per tant, la determinació de factor de risc per part d'aquest polimorfisme no es troba ni de bon tros resolta. Futurs estudis que explorin el paper que juga aquest polimorfisme en la malaltia coronària són encara necessaris.

4.3.2 Habilitat predictiva de la puntuació de risc genètic dels gens *NOS*

Un model de *score* de risc genètic (GRS) va ser obtingut per les mostres cas-control calculant el número d'al·lells de risc dels 4 polimorfismes dels gens *NOS* associats que presentava cada individu. El valor mig de GRS en les mostres era d'uns 6 al·lells de risc, oscil·lant entre 3 i 8 al·lells de risc. Aquest GRS, com a nou paràmetre quantitatiu, va mostrar associació significativa amb la malaltia només en la mostra cas-control italiana ATVB, però explicant menys del 1% del fenotip infart de miocardi. De fet, aquesta mostra cas-control era la més nombrosa de les utilitzades (~3.000 mostres) i, per tant, la que més pes va tenir a l'hora d'identificar els marcadors de risc. En la resta de mostres cas-control, el GRS no es va mostrar com un paràmetre significativament associat. Alhora, la capacitat del GRS de discernir entre afectats i no afectats per la malaltia va ser nul·la o molt limitada, tal i com mostrava el paràmetre AUC (0.45-0.53).

La dèbil contribució del nostre GRS dels gens *NOS* en termes d'associació i predicció es troba en plena coherència amb l'efecte mostrat per altres variants genètiques associades a les malalties complexes. Tal i com

s'ha observat en la secció 1.4.2 d'aquesta tesi, la proporció de variància fenotípica explicada (contribució a la malaltia) per la gran majoria de polimorfismes associats a la malaltia cardiovascular es inferior al 1% [38]. Generalment, la millora en la capacitat predictiva del risc per part dels factors de risc genètic sol ser nul·la o insuficient [46], inclús per l'àmpliament replicat polimorfisme del locus 9p21.3 [7,47]. Pel que fa a puntuacions de risc utilitzades en altres estudis, la variància explicada per dotzenes de marcadors de risc no sol superar el 5% del fenotip [48,97], bastant lluny del 40-60% estimat a partir d'altres mètodes, i donant una visió de la complexitat genètica de la malaltia coronària.

4.3.3 Distribució poblacional del GRS a través d'Europa i la Mediterrània

El *score* de risc genètic dels gens *NOS* també va ser calculat en els individus representants de les poblacions, i la distribució geogràfica de la mitjana poblacional va ser representada mitjançant un mapa de contorns (Figura 1 de la secció 3.3.3). Es va observar la regió de les illes de Còrsega i Sardenya i les regions continentals nord mediterrànies properes com les poblacions amb menor risc genètic *NOS* (5.4 - 5.7 al·lels de risc). A mesura que augmentava la distància a aquesta regió geogràfica, el risc genètic *NOS* augmentava de manera esglaonada en totes direccions, covariant significativament amb la geografia.

Aquesta covariació geogràfica de la variació genètica que s'observa pel *score* de risc (GRS) dels gens *NOS* respon a un patró similar al patró d'estructuració genètica global que mostren les poblacions europees i mediterrànies [69,75]. Aquesta estructuració genètica àmpliament descrita a Europa i a la Mediterrània on les diferències genètiques són poques, però

significatives, i es mostren en un gradient poblacional, és precisament el que s'observa per aquests polimorfismes detectats de risc en els gens *NOS*, tal i com es discutirà en la secció 4.4.2 juntament amb la resta de variació genètica analitzada.

4.3.4 Correlació del GRS i de la variació genètica individual amb la incidència coronària descrita a Europa

La variació geogràfica del GRS dels gens *NOS* al continent europeu destacava per presentar els valors més baixos a l'est dels Pirineus i els valors més alts a les poblacions de Gran Bretanya, Polònia i Finlàndia. Aquest increment gradual cap al nord i nord-est correlacionava considerablement amb la incidència coronària registrada pel projecte MONICA de la OMS al continent europeu (Figura 2 de la secció 3.3.3) i arribava a explicar-ne una gran proporció (67% en dones i 86% en homes). Pel que fa a la variació geogràfica de les freqüències dels polimorfismes analitzats individualment, un grup de polimorfismes, bàsicament localitzats a la regió del gen *NOS3*, presentava freqüències poblacionals correlacionades amb la incidència coronària de les poblacions MONICA, explicant-ne una proporció considerable (35-65%), tot i que inferior al GRS dels gens *NOS* (Taula 3 secció 3.3.3). D'aquest grup de SNPs la freqüència dels quals correlaciona amb la incidència, només un d'ells (rs179993) es va trobar associat a la malaltia coronària i, per tant, inclòs al GRS.

Inicialment la gran contribució del GRS a la incidència coronària poblacional va contrastar amb la minúscula contribució que el GRS presentava per la malaltia a nivell individual. No existeixen dades prèvies en la literatura on s'hagi analitzat la distribució poblacional d'altres GRS i, per tant, que es puguin comparar amb aquest estudi. Però sí es troben estudis on

s'ha descrit la variació poblacional de marcadors de risc genètic i la seva correlació amb la incidència coronària. Aquests estudis van revelar una correlació significativa de la isoforma 4 del gen de la *APOE* amb la incidència coronària a Europa, però de cap altre marcador dels estudiats [57,58]. En un d'aquests estudis fins i tot es van detectar clines genètiques correlacionades negatives amb la incidència, on a major freqüència poblacional de risc genètic menor incidència observada [58]. Aquest fet va fer que la idea de la paradoxa mediterrània, on els factors de risc tradicionals no explicaven la incidència observada (només un 30%), s'ampliés també als factors de risc genètic. La interpretació d'aquests autors a les clines genètiques observades de nord a sud fou que amb molta probabilitat eren el reflex de la distribució i estructuració genètica global al continent europeu fruit de la història demogràfica d'aquestes poblacions [78,98,99]. Per tant, en aquest context d'estructuració genètica de les poblacions europees, la variància de la incidència coronària explicada per cada polimorfisme correlacionat no pot ser considerada com a específica d'aquell polimorfisme ja que molt probablement la proporció d'incidència explicada és fruit de l'efecte combinat o additiu de moltes variants de risc genètic amb la mateixa distribució poblacional. Per exemple, l'aparent elevada contribució (46% en dones i 61% en homes) del polimorfisme de risc rs1799983 es deu molt probablement a un efecte de col·linealitat d'aquest factor amb molts altres factors de risc amb el mateix patró geogràfic. De la mateixa manera, la gran contribució del GRS (67% en dones i 86% en homes) deu reflectir l'efecte conjunt de totes les variants genètiques de risc, i també de possibles factors de risc ambiental, que presenten un patró poblacional similar als 4 SNPs inclosos en la puntuació. Així, els nostres resultats donen peu a proposar que una elevada proporció de la incidència coronària de les poblacions europees

es troba efectivament determinada per la distribució geogràfica de variants genètiques de risc, ja que els factors de risc tradicionals només són capaços d'explicar entre un 30 i un 40% de la incidència. A més, aquests factors genètics de risc serien comuns en la població (freqüències superiors a 5%), ja que les variants rares solen ser específiques de cada població i no es presenten distribuïdes en gradients geogràfics [10,30].

Pel que fa la variació geogràfica del GRS en la costa nord de la Mediterrània, també s'observa una gradació esglaonada amb valors baixos de risc genètic *NOS* a l'oest i valors de risc elevats a l'est de la Mediterrània. Existeixen poques dades epidemiològiques que cobreixin a la vegada aquestes regions, però aquests estudis també apunten a una incidència coronària més elevada a la regió est de la Mediterrània [100]. Per tant, també hi ha correlació entre el nostre GRS dels gens *NOS* i la incidència coronària, reforçant la interpretació de que tan el GRS com la incidència es troben profundament influenciats per l'estructuració genètica global fruit de la història demogràfica de les poblacions actuals.

Pel que fa la distribució del GRS per la costa del nord d'Àfrica, es fa difícil estimar si l'estructuració genètica poblacional reflecteix una diferent incidència cardiovascular ja que les dades epidemiològiques disponibles són excessivament minses [101]. Tot i això, els elevats valors del *score* poblacional observats en aquesta àrea geogràfica estimulen futures aproximacions en aquest àmbit d'estudi eco-epidemiològic.

4.3.5 Incidència coronària poblacional, predisposició genètica i hàbits i estil de vida

La important contribució genètica observada en el nostre estudi a la incidència coronària suggereix algunes consideracions sobre el rol que

aquests factors de risc genètic tenen sobre el risc individual de la malaltia coronària. Elevades incidències poblacionals estarien determinades per elevades freqüències de combinacions genètiques de risc i, per tant, d'una proporció considerable d'individus portadors d'aquestes combinacions genètiques de risc.

En aquest estudi d'epidemiologia ecològica, es podia interpretar que el GRS dels gens *NOS* capturava la contribució a la malaltia de la miríada de variants genètiques de risc que tenen una distribució geogràfica similar als 4 SNPs utilitzats. Per tant, només la inclusió d'aquesta miríada de variants genètiques en les anàlisis epidemiològiques podria explicar una considerable proporció de la variabilitat en malaltia coronària a nivell individual. En aquest context, estudis previs han intentat obtenir un model simulat de l'arquitectura genètica de les malalties complexes, per tal de determinar el nombre de polimorfismes genètics necessaris per a assolir una contribució significativa. Les prediccions d'aquests models de simulació són que, sota un model purament additiu entre marcadors, el nombre de marcadors genètics necessaris per tal d'explicar l'heretabilitat observada variaria entre pocs centenars de marcadors de baixa freqüència i contribució elevada i bastants centenars o pocs milers de marcadors d'elevada freqüència i baixa contribució [6,11,12]. Tenint tot això en compte, les nostres dades apunten cap a un model d'arquitectura genètica bastant complexa on els polimorfismes genètics implicats són majoritàriament d'elevada freqüència poblacional, ja que es distribueixen en gradients geogràfics, i de baixa contribució al fenotip.

Dins la contribució de factors de risc al desenvolupament de la malaltia cardiovascular, també tenen molt de pes els factors ambientals. Al continent europeu, la majoria de factors ambientals també es troben

distribuïts en clines nord-sud correlacionant-se amb la incidència coronària, però la seva contribució a la incidència poblacional no és superior al 30-40%, tal i com ja va reflectir l'estudi MONICA de la OMS [55]. A nivell individual, diferents estudis epidemiològics han determinat que els factors ambientals de risc es troben relacionats en el 80% del casos de malaltia coronària en una població. Així, existeix una discrepància on els factors ambientals semblarien ser més determinants a nivell individual i, en canvi, a nivell ecològic el factor més determinant en la malaltia és la predisposició genètica. De fet, aquestes observacions conformarien dues cares del mateix fenomen. Tot i que els factors ambientals expliquen una gran proporció dels casos de la malaltia en una població, el nombre total de casos en aquesta població depèn de la predisposició genètica dels seus individus. Per tan, dues poblacions sota la mateixa pressió ambiental de risc mostraran incidències poblacionals en funció de la seva predisposició genètica. La incidència coronària potencial d'una població estaria determinada per la predisposició genètica dels seus habitants però seria disparada pels hàbits i estils de vida dels mateixos.

4.3.6 Utilitat dels GRS per futurs estudis eco-epidemiològics

Els resultats d'aquesta tesi subratllen la utilitat que les puntuacions de risc genètic poden tenir a l'hora de determinar la càrrega o predisposició genètica de les poblacions i, per tant, com a predictors ecològics. La valoració dels riscos poblacionals cap a les malalties és de gran utilitat en salut pública, així com conèixer la seva variació al llarg del temps per tal de predir possibles brots no desitjats d'aquestes malalties [102]. El GRS ideal seria aquell que inclogués totes les variants genètiques de risc coronari que existeixen en el genoma, tasca difícil d'aconseguir actualment i segurament

durant els propers anys. No obstant, en una estratègia exploratòria es podrien dissenyar diversos GRS per diferents processos patofisiològics relacionats amb les malalties cardiovasculars, com per exemple per la disfunció endotelial, la aterosclerosi accelerada, l'hemostàsia en coagulació sanguínia, etc.. cadascuna amb la seva pròpia base genètica [102]. Aquests GRS podrien aportar una base sòlida per tal de desenvolupar programes d'intervenció en els estils de vida dels individus, prevenint factors de risc intermedis (com l'obesitat, la hipertensió, la hiperglucèmia, la hipercolesterolèmia) en subgrups de poblacions amb elevada predisposició. Així, la càrrega sanitària associada a aquestes patologies podria disminuir-se considerablement abans que els factors ambientals provoquessin la seva aparició degut a la predisposició genètica dels individus segons la seva població d'origen. Cal afegir, que una gran avantatge d'aquesta estratègia seria la necessitat de poques mostres representatives per població, i per tan un baix cost per tal de dur-la a terme. Tot això fa pensar que els GRS poden ser una gran eina per futurs estudis en salut pública.

4.4 Variació dels gens *NOS* a Europa i la Mediterrània: identificació dels processos demogràfics i/o adaptatius en relació a la incidència coronària

La distribució poblacional de la variació analitzada a les regions cromosòmiques dels gens *NOS* ha permès identificar el tipus d'estructures genètiques i processos adaptatius i demogràfics associats als gens *NOS* en les poblacions d'Europa i la Mediterrània, alhora que s'han inferit models poblacionals que ens expliquen la diferent incidència cardiovascular al continent europeu.

4.4.1 Estimació del grau de diferenciació genètica entre les poblacions analitzades, especialment d'Europa, nord d'Àfrica i Orient Mitjà

L'anàlisi dels conjunts de 114 SNPs i 17 polimorfismes de repeticions en tàndem (TRPs) de les sis regions genòmiques estudiades (tres d'elles inclouen els gens *NOS3*, *NOS1* i *NOS2A*) van revelar magnituds de diferenciació similar en les 39 poblacions analitzades (F_{ST} : 15% SNPs – 12.2% TRPs). En la jerarquització d'aquesta diferenciació poblacional, la major contribució l'aportava la diferenciació entre poblacions de diferents continents (entre continents: F_{CT} : 14% SNPs – 10.5% TRPs). En canvi, la magnitud de diferenciació observada entre poblacions d'Europa i la Mediterrània va ser bastant reduïda (F_{ST} : 1.2% SNPs – 1.6% TRPs) però similar a la detectada utilitzant altres marcadors autosòmics en un grup similar de poblacions [84,85,103]. Alhora, la proporció de la variància atribuïble a diferències entre grups fou més baixa en els TRPs que en els SNPs tant pel conjunt total de poblacions analitzades com pel grup de poblacions europees i mediterrànies. Aquest fenomen ha estat descrit prèviament per altres autors [85] i es pot deure a l'elevada taxa de mutació que presenten les repeticions en tàndem, on es poden generar al·lels idèntics

per estat però no per descendència, que contribuirien a difuminar la diferenciació entre-poblacions [85].

4.4.2 Processos d'adaptació genètica al medi reflectits en els gens *NOS*

L'anàlisi de possibles senyals de selecció en la variació genètica estudiada va aportar informació molt rellevant pel que fa la funcionalitat dels gens *NOS* i la seva relació amb la malaltia cardiovascular.

Primer de tot, cal destacar que cap dels 4 marcadors genètics associats al fenotip malaltia coronària (inclosos al GRS) van mostrar patrons de diferenciació poblacional (F_{ST}) diferents a la majoria de variació genètica estudiada ni, per tant, senyals d'adaptació al medi. Alhora, la variació genètica correlacionada geogràficament amb la incidència coronària (majoritàriament en el gen *NOS3*) presentava lleugers gradients de freqüència al·lèlica al continent però no associats a cap senyal de selecció positiva (alta diferenciació). Per tant, la variació geogràfica del conjunt polimorfismes de risc (GRS) dels gens *NOS* respon a un patró similar al patró d'estructuració genètica global que mostren les poblacions europees i mediterrànies. Tot i així, alguns senyals de selecció van ser detectats en altres marcadors estudiats. El polimorfisme de repetició en tàndem ss825679095 del gen *NOS3*, associat en altres estudis a malalties vasculares [104–106], aparegué relacionat amb processos de selecció equilibradora (baixa diferenciació poblacional). Per tant, en general, els nostres resultats foren consistents amb l'absència d'una clara detecció de senyals de selecció en variants genètiques associades a malalties cardiovasculars, tal i com s'havia observat anteriorment [107,108].

Altrament, senyals de selecció positiva i selecció equilibradora van ser detectats en els gens *NOS1* i *NOS2A*, així com en regions no codificants.

Les variacions genètiques amb patró poblacional diferenciat del gens *NOS1* i *NOS2A* poden trobar-se relacionades amb diferents processos biològics com la formació i funcionament de xarxes neuronals [109], o com processos immunològics adquirits i, per tant, d'adaptació a nous patògens [110]. Els casos detectats en les regions no codificants no poden ser descartats com a fruit de processos estocàstics a falta d'informació sobre la seva funcionalitat, d'igual manera que pels casos de les regions codificants.

4.4.3 Estructuració genètica observada a Europa i la Mediterrània per les regions genòmiques dels gens *NOS*

Les relacions poblacionals aportades per les distàncies genètiques i la seva representació geogràfica mostraren un clúster genètic homogeni per les poblacions europees i mediterrànies, amb l'excepció de la població aïllada de l'oasi de Siwa a Egipte [111] pel conjunt de SNPs (Figura 1A secció 3.4.3), i d'algunes poblacions pel conjunt de TRPs (Figura 2A secció 3.4.3), probablement degut a la baixa taxa de genotipació en aquestes mostres. Les poblacions europees i mediterrànies mostraren una estructura de gradient genètic continu, tal i com va revelar l'anàlisi de components principals espacial (sPCA), caracteritzada per dos eixos de variació ortogonals simulant els eixos nord-sud i est-oest (Figures 3 i 4 secció 3.4.3). Per tant, el patró de variació observat pels gens *NOS* és similar a l'observat per altres marcadors on la diferenciació poblacional és baixa, però significativa, i estructurada en gradients longitudinals i latitudinals [74,75]. Alhora, aquest patró observat pels gens *NOS* reforça la percepció d'absència de barreres genètiques entre el sud d'Europa i el nord d'Àfrica [76,84,85], especialment en la regió de l'estret de Gibraltar. Així, els nostres resultats subratllen que ambdós fenòmens de clines genètiques al continent europeu i continuïtat genètica

entre les dues costes mediterrànies són dos peces de la mateixa fotografia d'estructuració genètica.

4.4.4 Components “ancestrals” estimats del patrimoni genètic actual de les poblacions d'Europa i la Mediterrània

Els resultats d'estimació de components genètics “ancestrals” detectaren que la variació genètica observada en les poblacions era més òptimament explicada per la presència de 3 components genètics. Les proporcions d'aquests components genètics oscil·laven segons la població entre el 15 i el 55% del patrimoni genètic actual. Les poblacions identificades per mostrar les proporcions més altes de cada component foren el País Basc pel component anomenat A, les poblacions de Bòsnia-Hercegovina pel component anomenat B, i les poblacions del nord d'Àfrica pel component anomenat C dels SNPs i les poblacions de Jordània pel component C dels TRPs. Aquesta lleugera discrepància en la identificació de les poblacions amb major proporció de C entre SNPs i STRs pot trobar-se relacionada amb el limitat nombre de polimorfismes utilitzats i, alhora, amb les diferents característiques dels marcadors, on la diferent taxa de mutació pot influir en l'anàlisi donant més pes a uns processos demogràfics o a uns altres.

Aquest resultat contrasten amb la visió clàssica de dos components genètics ancestrals a Europa: un provinent de les poblacions paleolítiques que colonitzaren per primer cop el continent i l'altre provinent de les poblacions neolítiques que difongueren la nova economia sedentària. Estudis previs, que consideraren la població del País Basc i poblacions del Pròxim Orient com a poblacions ancestrals (representants del Paleolític i del Neolític respectivament), estimaren una considerable contribució neolítica del 70%

per les poblacions de la regió dels Balcans que decreixia en direcció oest fins al 30% en la Península Ibèrica [79–81]. Aquests mateixos estudis, argumentaren que si l'elecció de les poblacions ancestrals fos errònia, els seus resultats estarien reflectint un gradient de variació genètica longitudinal no només compatible amb barreja poblacional ancestral, sinó també amb diferenciació per distància geogràfica [81]. De fet, els resultats d'aquesta tesi indiquen que les poblacions del País Basc i Jordània (segon eix de variació genètic observat), i Bòsnia-Hercegovina i nord d'Àfrica (primer eix de variació) representen els pols o extrems del marc genètic continu observat. I pel fet que no mostren proporcions de components “ancestrals” propers al 100%, aquestes poblacions no poden ser escollides com a representants de cap component genètic parental o ancestral ni tampoc considerades com aïllats genètics poblacionals, tal i com ja s'ha suggerit prèviament en el cas de la població del País Basc [112]. Les nostres dades suggereixen que aquests quatre grups poblacionals simplement representen els extrems o pols del contínuum de variació genètica de les poblacions europees i mediterrànies seleccionades en aquest estudi.

Per contra, els nostres resultats són consistents amb el model de doble dispersió del Neolític a Europa i la Mediterrània exposat en la secció 1.5.2.1. En resum, la transició neolítica a Europa es va dispersar a través de dues rutes principals [82], una ruta terrestre a través de les valls fluvials centre europees (Danubi-Rin) fa 7.500 anys lligada a la cultura ceràmica *Linearbandkeramik* (LBK), i una ruta marítima a través de les costes nord i sud de la Mediterrània fa 8-7.500 anys lligada a la cultura ceràmica Cardial [83]. Dades paleogenètiques recents, basades en DNA antic, donen suport a aquest doble model de difusió del Neolític degut a la discontinuïtat genètica observada en les línies haplotípiques parentals (mtDNA i ChrY) entre les

mostres neolítiques [82,113–115]. En mostres neolítiques que se'ls ha pogut extreure DNA, el llinatge uniparentals mitocondrials i del cromosoma Y no coincideixen entre les mostres centre i sud europees, indicant que els grups poblacionals no eren genèticament idèntics. Alhora, aquesta diferenciació genètica poblacional de les mostres del Neolític també es va observar pel gen de la lactasa [114,115]. Les mostres centre europees presentaven la variant genètica de persistència a la lactasa que permet digerir lactosa en edat adulta, per contra, les mostres mediterrànies presentaven intolerància total a la lactosa. Aquest context de dades arqueològiques i genètiques ens permet interpretar els components genètics observats en la variació analitzada dels gens *NOS*. D'acord amb aquest model, el component “ancestral” o clúster anomenat A es podria atribuir a un component paleolític europeu pel fet de presentar-se en major proporció en la mostra del País Basc i en menor proporció en les poblacions on el Neolític es va originar. El clúster B, amb major proporció en les poblacions de Bòsnia-Hercegovina, es podria interpretar provinent dels primers agricultors de cultura LBK habitants de la vall del Danubi. Segons estudis recents, aquesta porció ancestral centre europea s'hauria diferenciat de les poblacions del Llevant mediterrani i Pròxim Orient fa 12.000 anys [116], just abans d'adquirir la economia neolítica. Tot i així, els seus portadors ancestrals haurien estat els principals responsables de difondre el Neolític al centre d'Europa. Les nostres dades subratllen la necessitat de caracteritzar genèticament les poblacions dels Balcans i el Mar Negre d'una forma més detallada per tal d'obtenir més informació sobre aquest component genètic “ancestral”, els portadors del qual van difondre el Neolític per l'Europa continental. Finalment, el clúster C podria estar representant un component genètic nord-africà i la seva dispersió seria producte de la ruta de dispersió marítima del Neolític per la

Mediterrània. Aquesta interpretació encaixaria amb la recent descripció d'un component genètic "ancestral" característic de les poblacions actuals del Magrib procedent de poblacions paleolítiques que decreix en proporció d'oest a est en les poblacions del nord d'Àfrica [117]. Aquest component paleolític "magribí" també ha estat observat en poblacions del sud d'Europa [118]. Aquesta interpretació casaria amb dades provinents de simulacions de difusió espacial de les poblacions de l'oest de la Mediterrània, on s'observa que el patró de variació genètic observat en la actualitat per diferents tipus de marcadors genètics hauria d'haver-se generat en èpoques anteriors a la difusió neolítica [85]. D'altra banda, el component ancestral C obtingut amb els marcadors de repeticions en tàndem mostra les proporcions més elevades en les poblacions de l'est del Mediterrani i un gradient decreixent cap a l'oest. Aquestes diferències es podrien deure a les diferents característiques dels marcadors genètics que accentuarien de manera diferent la importància de determinades poblacions o porcions genètiques en la composició de les poblacions actuals.

4.4.5 Interpretació dels processos poblacionals que originaren l'estructuració genètica en gradients ortogonals i la distribució de components ancestrals observats

Les dades genètiques analitzades per les regions dels gens *NOS* presenten dos eixos ortogonals principals de variació, simulant els eixos Nord-Sud i Est-Oest, tal i com revelen les distàncies genètiques i l'anàlisi de components principals espacial (sPCA). A més, els resultats de l'estimació de barreja genètica ancestral són interpretats amb una presència considerable de porció genètica d'origen paleolític (>60%) en les poblacions actuals d'Europa i la Mediterrània.

Aquests resultats encaixen amb resultats provinents de simulacions computacionals d'expansions poblacionals. Aquestes simulacions indicarien que l'eix de variació est-oest és degut a la formació de clines geogràfiques de freqüències al·lèliques a mesura que les poblacions es van expandint i colonitzant nous territoris [119,120]. Alhora, aquestes simulacions mostren com petites incorporacions o introgressions de porció genètica d'origen paleolític en les poblacions difusores del Neolític, en l'anomenat límit de "l'onada d'avanç", provoquen que poblacions inicialment genèticament idèntiques a les neolítiques originàries incorporin progressiva i ràpidament component genètic d'origen paleolític [119,120]. Aquest procés d'introgessió genètica d'origen paleolític en l'onada poblacional d'avanç neolítica, fa que al final les poblacions que difonen el Neolític continguin majoritàriament patrimoni genètic paleolític. Per tant, el fort impacte del creixement demogràfic associat a l'adquisició de la tècnica neolítica redistribueix en clines geogràfiques tant el patrimoni genètic nou com també el ja existent, seguint la direcció d'expansió. Aquest seria precisament el cas de la introgessió de patrimoni genètic de les poblacions paleolítiques del nord d'Àfrica en la ruta de dispersió marítima del Neolític en la Mediterrània que explicaria la distribució observada del component ancestral "magribí" per tot l'oest de la Mediterrània.

Un fet molt interessant és l'observació de que el principal eix de variació genètica en les poblacions europees observat en aquest i altres estudis és nord-sud, de forma ortogonal a la direcció d'expansió i diferenciació poblacional est-oest típica del continent europeu. Aquest fet tan interessant també és explicat per models de simulació, on en el procés d'expansió poblacional es generen diferents sectors geogràfics en paral·lel a l'eix d'expansió que es diferencien de l'origen a mesura que avança la

colonització [121,122]. Aquesta diferenciació poblacional dins de cada sector és completament a l'atzar, degut al nombre reduït del grup que facilita la deriva genètica, i independent dels altres sectors. Per tant, als extrems d'expansió de cada sector es trobarien poblacions que són considerablement diferents genèticament. En el nostre cas, en la doble expansió neolítica a través de l'Europa continental i la Mediterrània, inicialment s'haurien generat dues regions geogràfiques amb baixa diversitat genètica i altament diferenciades entre elles, especialment a l'extrem oest. Posteriorment, altres moviments migratoris i de flux genètic entre aquestes dues regions, originats per expansions poblacionals en l'era dels metalls i de civilitzacions més modernes, haurien difuminat els contrastos genètics entre els diferents grups neolítics enriquint genèticament aquestes poblacions frontereres i generant els principals patrons de variació genètica nord-sud, ortogonals a l'eix d'expansió inicial.

En resum, la variació genètica en les regions *NOS* observada en aquest estudi indicaria que els actuals patrons genètics est-oest i nord-sud de gradient de freqüències, diversitat i barreja poblacional podrien explicar-se com a generats per l'impacte d'un doble procés de Neolització a Europa i la Mediterrània que encara no han estat esborrats per moviments poblacionals posteriors.

4.4.6 Estimació de l'estructuració poblacional de l'arquitectura genètica de les malalties coronàries i processos poblacionals que determinarien la diversa incidència coronària a les poblacions europees

En l'estudi de la distribució geogràfica de la incidència coronària en les poblacions europees, s'ha determinat freqüentment que els factors de risc considerats com a clàssics (per exemple, consum de greixos, ingesta

d'alcohol, tabaquisme) contribueixen tan sols en un 30-40% en la diferent incidència coronària observada en els poblacions [54,55]. Per tant, s'espera que la porció restant (60-70%) sigui explicada per la distribució poblacional dels factors genètics que originen les malalties.

En els nostres resultats de correlació entre paràmetres genètics poblacionals i la incidència coronària, destacava l'anàlisi multivariant dels TRPs que incloïa el primer i el segon component (eix) de variació espacial i explicava una gran proporció (85-90%) de la variació poblacional de la incidència coronària. S'ha de tenir en compte que, pel fet que molts factors de risc clàssics presenten una distribució latitudinal (lligat a diferències culturals en producció alimentària i consum de begudes alcohòliques), quan el component de variació genètica nord-sud és inclòs en un model de regressió, un percentatge de la variació estimada (no superior al 30-40%) pot ser deguda a aquest factors clàssics i no pròpiament a la variació genètica. En els nostres gràfics de correlacions (Figures 6 i 7 secció 3.4.3), es van observar dos grups de poblacions diferenciades segons la incidència que mostraven, alta o baixa-moderada. Aquests grups de poblacions es van interpretar com grups amb condicions ambientals i de producció alimentària similars. Per tant, l'anàlisi per separat d'aquests grups redueix considerablement la variació en la incidència lligada als factors de risc clàssics, amb l'inconvenient de la pèrdua de poder estadístic de l'anàlisi. En el grup de poblacions amb incidència baixa-moderada, és remarcable que els dos principals components (eixos) de variació genètica espacial dels dos tipus de marcadors genètics (SNPs i TRPs) explicaven gairebé la totalitat (97-98%) de la variació en la incidència coronària d'aquestes poblacions. Per tant, les estimacions d'estratificació poblacional dels gens *NOS* expliquen les diferències en incidència coronària de les poblacions estudiades. Aquest

conjunt de resultats indicarien que la variació en la incidència cardiovascular és deguda a la distribució dels factors genètics de risc que predisposen per malalties cardiovasculars i que aquests es troben estructurats en les poblacions de la mateixa manera que s'hi troba la variació dels gens *NOS* estudiada.

Les correlacions entre la incidència coronària i els altres paràmetres d'estructuració analitzats podrien donar una falsa visió si fossin interpretades sense cura. Per exemple, es podria interpretar que una elevada proporció de component ancestral provinent del Neolític europeu (B) i una proximitat genètica a les poblacions de Bòsnia-Hercegovina implicaria una elevada incidència coronària en les poblacions, mentre que el component ancestral “magribí” (C) aportaria protecció genètica enfront de malalties coronàries. Però una interpretació més plausible relacionaria la incidència coronària amb la diversitat genètica que presenten les poblacions de manera que poblacions amb elevada contribució dels dos components ancestrals “Neolític europeu” i “magribí”, enriquides genèticament per la confluència d'aquests dos components, mostrarien diversitat poblacional més elevada, menor predisposició genètica cardiovascular (degut als efectes equilibradors discutits a la secció 4.2.2), i per tant, menor incidència coronària en la població (tal i com s'ha discutit a la secció 4.3.5).

Aquesta interpretació necessita ser valorada més endavant per tal de ser rebutjada o refinada amb estudis que incloguin més marcadors genètics i més poblacions. De totes maneres, l'estructuració i barreja genètica estimada en aquest estudi, així com la interpretació d'un doble procés de Neolització, semblen plausibles ja que són capaços d'explicar la distribució poblacional d'una malaltia complexa amb una base genètica comunament estructurada en les poblacions.

5. CONCLUSIONS

5.1 Conclusions d'aquesta tesi

Les principals conclusions obtingudes en aquest treball són les següents:

1. La variació genètica analitzada al gen *F13A1* no presentà una associació individual amb l'infart de miocardi en la població espanyola, però sí un efecte interactiu entre les variants dins del gen *F13A1*.
2. Característiques genètiques de les poblacions, com l'heterozigositat i el desequilibri de lligament, podrien ser les responsables de la detecció d'efectes individuals en variants genètiques que formen part de models interactius.
3. L'habilitat predictiva del score de risc genètic dels gens *NOS*, a partir de 4 SNPs genèticament independents, va presentar una capacitat nul·la o molt limitada de discernir entre malalts i no malalts.
4. Tot i que, la distribució poblacional del GRS covaria amb la geografia i explica un percentatge molt elevat (65-85%) de la incidència coronària a Europa.
5. L'elevada proporció d'incidència coronària explicada pel GRS deu ser degut a l'efecte col·lineal de la miriada de variants genètiques de risc amb el mateix patró de variació poblacional que els SNPs que conformen el GRS.
6. L'ús de GRS per tal d'estimar la predisposició genètica de les poblacions a determinades malalties complexes pot ser de gran utilitat en salut pública.
7. La major part de variació en els gens *NOS*, inclosa la variació relacionada amb la malaltia coronària, no van presentar senyals de selecció o adaptació al medi en el conjunt de poblacions analitzades d'Europa i la Mediterrània.

8. Els principals eixos de variació genètica poblacional registrats pels gens *NOS* descriuen un patró de variació continu que emula els eixos nord-sud i est-oest.
9. Tres components genètics ancestrals foren detectats en les poblacions estudiades i interpretables com patrimoni genètic “Paleolític europeu”, “Paleolític nord-africà” i “Neolític europeu”.
10. El patró d'estructuració genètica i barreja ancestral s'ajusta a un model de doble procés de Neolització seguint dues rutes principals: una de terrestre a través de les valls fluvials centre-europees (Danubi-Rin), i una de marítima a través de les costes nord i sud de la Mediterrània.
11. Els paràmetres d'estructuració genètica espacial dels gens *NOS* expliquen gairebé tota la variabilitat en incidència coronària entre poblacions europees, posant de relleu que la miriada de variants genètiques de risc per la malaltia coronària es trobaria estructurada en les poblacions europees d'igual manera que la variació genètica global.
12. La predisposició genètica de risc poblacional pot estar lligada a la diversitat genètica de la població aportada per la confluència de diferents components ancestrals com el “Neolític europeu” i “Paleolític nord-africà”.

5.2 Conclusions of this thesis

The main conclusions obtained in this work are the following:

1. The genetic variation analysed in the *F13A1* gene did not show any individual association with myocardial infarction in our Spanish sample, however allele-allele interactions within *F13A1* were detected.
2. Population genetic features, such as heterozygosity and linkage disequilibrium, could be involved in the detection of individual effects in markers under interactive models.
3. The predictive ability of the *NOS* genes risk score, computed from 4 genetically independent SNPs, showed no or very limited ability to discriminate between cases and controls.
4. However, the population distribution of the GRS covaried with geography and explained a substantial proportion (65-85%) of coronary incidence in Europe.
5. The proportion of coronary incidence explained by the GRS can be due to the colinear effect of the myriad of genetic risk variants with the same population distribution than the SNPs of the GRS.
6. GRSs can be very useful in public health studies to estimate population genetic risk predisposition to complex diseases.
7. Most of the *NOS* gene variation, included CAD-related markers, showed no signs of selection or adaptation to new environment in the analyzed European and Mediterranean populations.
8. The main population genetic variation axes of *NOS* gene markers showed a continuous genetic gradient following North-South and East-West patterns.
9. Three ancestral genetic components were detected in the studied

populations and interpreted as “European Palaeolithic”, “North African Palaeolithic”, and “European Neolithic” genetic components.

10. The pattern of population genetic structure and the ancestral admixture detected fitted a model of dual Neolithization following two main routes: one expansion in the European continent along the Danube River, and another in the Mediterranean basin attributed to the “maritime pioneers”.
11. Population genetic spatial components of *NOS* genes explained almost all variability in coronary incidences across European populations, highlighting that the myriad of coronary risk genetic variants would be structured among populations similarly to genome-wide variation.
12. Population genetic risk predisposition can be linked to population genetic diversity as a result of ancestral components convergence, such as “European Neolithic” and “North African Paleolithic” genetic components.

6. REFERÈNCIES

1. Watkins H, Farrall M (2006) Genetic susceptibility to coronary artery disease: from promise to progress. *Nat Rev Genet* 7: 163–173.
2. Nabel EG (2003) Cardiovascular Disease. *N Engl J Med* 349: 60–72.
3. Adams RLC, Bird RJ (2009) Review article: Coagulation cascade and therapeutics update: relevance to nephrology. Part 1: Overview of coagulation, thrombophilias and history of anticoagulants. *Nephrology* 14: 462–470.
4. Dias RG, Gowdak MM, Pereira AC (2011) Genetics and cardiovascular system: influence of human genetic variants on vascular function. *Genes Nutr* 6: 55–62.
5. World Health Organization (2011) Global status report on noncommunicable diseases 2010. WHO Libr.
6. Van der Net JB, Janssens ACJW, Sijbrands EJG, Steyerberg EW (2009) Value of genetic profiling for the prediction of coronary heart disease. *Am Heart J* 158: 105–110.
7. Thanassoulis G, Vasan RS (2010) Genetic cardiovascular risk prediction: will we get there? *Circulation* 122: 2323–2334.
8. Prins BP, Lagou V, Asselbergs FW, Snieder H, Fu J (2012) Genetics of coronary artery disease: genome-wide association studies and beyond. *Atherosclerosis* 225: 1–10.
9. Lusis AJ, Mar R, Pajukanta P (2004) Genetics of atherosclerosis. *Annu Rev Genomics Hum Genet* 5: 189–218.
10. Gibson G (2011) Rare and common variants: twenty arguments. *Nat Rev Genet* 13: 135–145.
11. Janssens A CJW, Aulchenko YS, Elefante S, Borsboom GJJM, Steyerberg EW, et al. (2006) Predictive testing for complex diseases using multiple genes: Fact or fiction? *Genet Med* 8: 395–400.
12. Stahl E, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, et al. (2012) Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet* 44: 483–489.
13. Muszbek L, Yee VC, Hevessy Z (1999) Blood coagulation factor XIII: structure and function. *Thromb Res* 94: 271–305.
14. Gallivan L, Markham F, Anwar R (1999) The Leu564 factor XIIIa variant results in significantly lower plasma factor XIII levels than the Pro564 variant. *Thromb Haemost* 82: 1368–1370.
15. Lim BCB, Ariëns RAS, Carter AM, Weisel JW, Grant PJ (2003) Genetic regulation of fibrin structure and function : complex gene- environment interactions may modulate vascular risk. *Lancet* 361: 1424–1431.
16. Mungrue IN, Brecht DS, Stewart DJ, Husain M (2003) From molecules to mammals: what's NOS got to do with it? *Acta Physiol Scand* 179: 123–135.
17. Robb GB, Carson AR, Tai SC, Fish JE, Singh S, et al. (2004) Post-transcriptional regulation of endothelial nitric-oxide synthase by an overlapping antisense mRNA transcript. *J Biol Chem* 279: 37982–37996.

18. Casas JP, Cavalleri GL, Bautista LE, Smeeth L, Humphries SE, et al. (2006) Endothelial nitric oxide synthase gene polymorphisms and cardiovascular disease: a HuGE review. *Am J Epidemiol* 164: 921–935.
19. Paternoster L, Martinez-Gonzalez N, Charleton R, Chung M, Lewis S, et al. (2010) Genetic effects on carotid intima-media thickness: systematic assessment and meta-analyses of candidate gene polymorphisms studied in more than 5000 subjects. *Circ Cardiovasc Genet* 3: 15–21.
20. Zhang K, Bai P, Shi S, Zhou B, Wang Y, et al. (2012) The G894T polymorphism on endothelial nitric oxide synthase gene is associated with increased coronary heart disease among Asia population: evidence from a Meta analysis. *Thromb Res* 130: 192–197.
21. Zintzaras E, Kitsios G, Stefanidis I (2006) Endothelial NO synthase gene polymorphisms and hypertension: a meta-analysis. *Hypertension* 48: 700–710.
22. Pereira TV, Rudnicki M, Cheung BMY, Baum L, Yamada Y, et al. (2007) Three endothelial nitric oxide (NOS3) gene polymorphisms in hypertensive and normotensive individuals: meta-analysis of 53 studies reveals evidence of publication bias. *J Hypertens* 25: 1763–1774.
23. Li J, Wu X, Li X, Feng G, He L, et al. (2010) The endothelial nitric oxide synthase gene is associated with coronary artery disease: a meta-analysis. *Cardiology* 116: 271–278.
24. Morris BJ, Glenn CL, Wilcken DE, Wang XL (2001) Influence of an inducible nitric oxide synthase promoter variant on clinical variables in patients with coronary artery disease. *Clin Sci (Lond)* 100: 551–556.
25. Iwai N, Tago N, Yasui N, Kokubo Y, Inamoto N, et al. (2004) Genetic analysis of 22 candidate genes for hypertension in the Japanese population. *J Hypertens* 22: 1119–1126.
26. Kunnas T, Mikkelsen J, Ilveskoski E, Tanner MM, Laippala P, et al. (2003) A functional variant of the iNOS gene flanking region is associated with LAD coronary artery disease: an autopsy study. *Eur J Clin Invest* 33: 1032–1037.
27. Fu L, Zhao Y, Lu J, Shi J, Li C, et al. (2009) Functional single nucleotide polymorphism-1026C/A of inducible nitric oxide synthase gene with increased YY1-binding affinity is associated with hypertension in a Chinese Han population. *J Hypertens* 27: 991–1000.
28. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.
29. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. *Science* (80-) 291: 1304–1351.
30. Abecasis GR, Auton A, Brooks LD, DePristo M, Durbin RM, et al. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491: 56–65.
31. Porta M (2008) *A dictionary of Epidemiology*. 5th ed. Oxf Univ Press, editor.
32. Wilson PWF, D'Agostino RB, Levy D, Belanger M, Silbershatz H, et al. (1998) Prediction of Coronary Heart Disease Using Risk Factor Categories. *Circulation* 97: 1837–1847.

33. Ridker PM, Buring JE, Rifai N, Cook NR (2007) Development and Validation of Improved Algorithms for the Assessment of Global Cardiovascular Risk in Women The Reynolds Risk Score. *JAMA* 297: 611–620.
34. Assmann G (2002) Simple Scoring Scheme for Calculating the Risk of Acute Coronary Events Based on the 10-Year Follow-Up of the Prospective Cardiovascular Munster (PROCAM) Study. *Circulation* 105: 310–315.
35. Conroy R (2003) Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 24: 987–1003.
36. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, et al. (2010) Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 11: 446–450.
37. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff L, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461: 747–753.
38. Ioannidis JP (2009) Prediction of cardiovascular disease outcomes and established cardiovascular risk factors by genome-wide association markers. *Circ Cardiovasc Genet* 2: 7–15.
39. Morrison AC, Bare L, Chambless LE, Ellis SG, Malloy M, et al. (2007) Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. *Am J Epidemiol* 166: 28–35.
40. Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, et al. (2010) A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet* 376: 1393–1400.
41. Brautbar A, Pompeii L, Dehghan A, Ngwa JS, Nambi V, et al. (2012) A genetic risk score based on direct associations with coronary heart disease improves coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC), but not in the Rotterdam and Framingham Offspring, Studies. *Atherosclerosis* 223: 421–426.
42. Hughes MF, Saarela O, Stritzke J, Kee F, Silander K, et al. (2012) Genetic markers enhance coronary risk prediction in men: the MORGAM prospective cohorts. *PLoS One* 7: e40922.
43. Vaarhorst AA, Lu Y, Heijmans BT, Dollé MET, Böhringer S, et al. (2012) Literature-based genetic risk scores for coronary heart disease: the Cardiovascular Registry Maastricht (CAREMA) prospective cohort study. *Circ Cardiovasc Genet* 5: 202–209.
44. Thanassoulis G, Peloso GM, Pencina MJ, Hoffmann U, Fox CS, et al. (2012) A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium: the Framingham Heart Study. *Circ Cardiovasc Genet* 5: 113–121.
45. Bolton JL, Stewart MCW, Wilson JF, Anderson N, Price JF (2013) Improvement in prediction of coronary heart disease risk over conventional risk factors using SNPs identified in genome-wide association studies. *PLoS One* 8: e57310.
46. Janssens ACJW, van Duijn CM (2008) Genome-based prediction of common diseases: advances and prospects. *Hum Mol Genet* 17: R166–73.

47. Janssens ACJW, van Duijn CM (2009) Genome-based prediction of common diseases: methodological considerations for future research. *Genome Med* 1: 20.
48. Paynter NP, Chasman DI, Paré G, Buring JE, Cook NR, et al. (2010) Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* 303: 631–637.
49. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, et al. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9: 356–369.
50. Keys A, Aravanis C, Blackburn H, Buzina R, Djordjevic B, et al. (1980) Seven Countries. A Multivariate Analysis of Death and Coronary Heart Disease. Harvard Univ Press: 1–381.
51. Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, Arveiler D, Rajakangas M, et al. (1994) Myocardial infarction and coronary deaths in the World Health Organization MONICA Project. Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents. *Circulation* 90: 583–612.
52. Tunstall-pedoe H, Kuulasmaa K, Tolonen H, Ruokokoski E, Amouyel P (1999) Contribution of trends in survival and coronar y - event rates to changes in coronary heart disease mortality : 10-year results from 37 WHO MONICA Project populations. 353: 1547–1557.
53. Tunstall-pedoe H (1988) Autres pays, autres moeurs. Theories on why the French have less heart disease than the British. *Br Med J* 297: 1559–1560.
54. Artaud-Wild SM, Connor SL, Sexton G, Connor WE (1993) Differences in coronary mortality can be explained by differences in cholesterol and saturated fat intakes in 40 countries but not in France and Finland. A paradox. *Circulation* 88: 2771–2779.
55. The WHO MONICA Project (1994) Ecological Analysis of the Association between Mortality and Major Risk Factors of Cardiovascular Disease. *Int J Epidemiol* 23: 505–516.
56. Riudavets J-B, Ducimetière P, Evans A, Montaye M, Haas B, et al. (2010) Patterns of alcohol consumption and ischaemic heart disease in culturally divergent countries: the Prospective Epidemiological Study of Myocardial Infarction (PRIME). *Br Med J* 341:c6077.
57. Stengård JH, Weiss KM, Sing CF (1998) An ecological study of association between coronary heart disease mortality rates in men and the relative frequencies of common allelic variations in the gene coding for apolipoprotein E. *Hum Genet* 103: 234–241.
58. Lao O, Dupanloup I, Barbujani G, Bertranpetit J, Calafell F (2008) The Mediterranean Paradox for Susceptibility Factors in Coronary Heart Disease Extends to Genetics. *Ann Hum Genet* 72: 48–56.
59. Barbujani G, Colonna V (2010) Human genome diversity: frequently asked questions. *Trends Genet* 26: 285–295.
60. Jobling M (2012) The impact of recent events on human genetic diversity. *Philos Trans R Soc Lond B Biol Sci* 367: 793–799.

61. Cann H, de Toma C, Cazes L, Legrand M, Morel V, et al. (2002) A human genome diversity cell line panel. *Science* (80-) 296: 261–262.
62. Ramachandran S, Deshpande O, Roseman CC, Rosenberg N, Feldman MW, et al. (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci* 102: 15942–15947.
63. Ashraf Q, Galor O (2013) The “Out of Africa” Hypothesis, Human Genetic Diversity, and Comparative Economic Development. *Am Econ Rev* 103: 1.46.
64. Jombart T, Pontier D, Dufour B (2009) Genetic markers in the playground of multivariate analysis. *Heredity* (Edinb) 102: 330–341.
65. Liu Y, Nyunoya T, Leng S, Belinsky S, Tesfaigzi Y, et al. (2013) Softwares and methods for estimating genetic ancestry in human populations. *Hum Genomics* 7: 1.
66. Rosenberg N, Pritchard JK, Weber JL, Cann HM, Kidd KK, et al. (2002) Genetic structure of human populations. *Science* 298: 2381–2385.
67. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319: 1100–1104.
68. Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, et al. (2003) Genetic Variation Among World Populations : Inferences From 100 Alu Insertion Polymorphisms: 1607–1618.
69. Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, et al. (2008) Genes mirror geography within Europe. *Nature* 456: 98–101.
70. Coop G, Pickrell JK, Novembre J, Kudaravalli S, Li J, et al. (2009) The role of geography in human adaptation. *PLoS Genet* 5: e1000500.
71. Yu N, Chen F, Ota S, Jorde LB, Pamilo P, et al. (2002) Larger genetic differences within africans than between Africans and Eurasians. *Genetics* 274: 269–274.
72. Cavalli-Sforza LL (2007) Human evolution and its relevance for genetic epidemiology. *Annu Rev Genomics Hum Genet* 8: 1–15.
73. Gignoux CR, Henn BM, Mountain JL (2011) Rapid, global demographic expansions after the origins of agriculture. *Proc Natl Acad Sci U S A* 108: 6044–6049.
74. Auton A, Bryc K, Boyko AR, Lohmueller KE, Novembre J, et al. (2009) Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res* 19: 795–803.
75. Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, et al. (2008) Correlation between genetic and geographic structure in Europe. *Curr Biol* 18: 1241–1248.
76. Botigué LR, Henn BM, Gravel S, Maples BK, Gignoux CR, et al. (2013) Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. *Proc Natl Acad Sci U S A* 110: 11791–11796.
77. Fadhloui-Zid K, Haber M, Martínez-Cruz B, Zalloua P, Benammar Elgaaied A, et al. (2013) Genome-wide and paternal diversity reveal a recent origin of human populations in North Africa. *PLoS One* 8: e80293.

78. Barbujani G, Goldstein DB (2004) Africans and Asians abroad: genetic diversity in Europe. *Annu Rev Genomics Hum Genet* 5: 119–150.
79. Chikhi L, Nichols R, Barbujani G, Beaumont M (2002) Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci U S A* 99: 11008–11013.
80. Dupanloup I, Bertorelle G, Chikhi L, Barbujani G (2004) Estimating the impact of prehistoric admixture on the genome of Europeans. *Mol Biol Evol* 21: 1361–1372.
81. Belle EMS, Landry P-A, Barbujani G (2006) Origins and evolution of the Europeans' genome: evidence from multiple microsatellite loci. *Proc Biol Sci* 273: 1595–1602.
82. Pinhasi R, Thomas MG, Hofreiter M, Currat M, Burger J (2012) The genetic history of Europeans. *Trends Genet* 28: 496–505.
83. Linstädter J, Medved I, Solich M, Weniger G-C (2012) Neolithisation process within the Alboran territory: Models and possible African impact. *Quat Int* 274: 219–232.
84. Athanasiadis G, Moral P (2013) Spatial principal component analysis points at global genetic structure in the Western Mediterranean. *J Hum Genet*: 1–4.
85. Currat M, Poloni ES, Sanchez-Mazas A (2010) Human genetic differentiation across the Strait of Gibraltar. *BMC Evol Biol* 10: 237.
86. Vokó Z, Bereczky Z, Katona É, Ádány R, Muszbek L (2007) Factor XIII Val34Leu variant protects against coronary artery disease. *Thromb Haemost* 97: 458–463.
87. Wells PS, Anderson JL, Scarvelis DK, Doucette SP, Gagnon F (2006) Factor XIII Val34Leu variant is protective against venous thromboembolism: a HuGE review and meta-analysis. *Am J Epidemiol* 164: 101–109.
88. Reiner AP, Schwartz SM, Frank MB, Longstreth WT, Hindorff L., et al. (2001) Polymorphisms of Coagulation Factor XIII Subunit A and Risk of Nonfatal Hemorrhagic Stroke in Young White Women Editorial Comment. *Stroke* 32: 2580–2587.
89. Barral S, Lareu M V, Salas A, Carracedo A (2000) Sequence variation of two hypervariable short tandem repeats at the D22S683 and D6S477 loci. *Int J Legal Med* 113: 146–149.
90. Garofano L, Pizzamiglio M, Bizzaro GP, Donato F, Rossetti M, et al. (1999) Italian population data on two new short tandem repeat loci: D6S477 and D19S433. *Forensic Sci Int* 101: 203–208.
91. Moore JH (2003) The Ubiquitous Nature of Epistasis in Determining Susceptibility to Common Human Diseases. *Hum Hered* 56: 73–82.
92. Catto AJ, Kohler HP, Bannan S, Stickland M, Carter A, et al. (1998) Factor XIII Val 34 Leu : A Novel Association With Primary Intracerebral Hemorrhage. *Stroke* 29: 813–816.
93. Gemmati D, Serino M, Ongaro A, Tognazzo S, Moratelli S, et al. (2001) A common mutation in the gene for coagulation factor XIII-A (VAL34Leu): a risk factor for primary intracerebral hemorrhage is protective against atherothrombotic diseases. *Am J Hematol* 67: 183–188.

94. Kida M, Souri M, Yamamoto M, Saito H, Ichinose A (1999) Transcriptional regulation of cell type-specific expression of the TATA-less A subunit gene for human coagulation factor XIII. *J Biol Chem* 274: 6138–6147.
95. Arnosti DN, Kulkarni MM (2005) Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J Cell Biochem* 94: 890–898.
96. Lin P-I, Vance JM, Pericak-Vance M, Martin ER (2007) No gene is an island: the flip-flop phenomenon. *Am J Hum Genet* 80: 531–538.
97. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, et al. (2009) Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* 41: 334–341.
98. Sokal R, Harding R, Oden N (1989) Spatial patterns of human gene frequencies in Europe. *Am J Phys Anthr* 80: 267–294.
99. Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *The History and Geography of Human Genes* Princeton University Press.
100. Müller-Nordhorn J, Binting S, Roll S, Willich SN (2008) An update on regional variation in cardiovascular mortality within Europe. *Eur Heart J* 29: 1316–1326.
101. World Health Organization (2004) *The global burden of disease 2004*. WHO.
102. Batsis J, Lopez-Jimenez F (2010) Cardiovascular risk assessment--from individual risk prediction to estimation of global risk and change in risk in the population. *BMC Med* 8: 29.
103. El Moncer W, Esteban E, Bahri R, Gayà-Vidal M, Carreras-Torres R, et al. (2010) Mixed origin of the current Tunisian population from the analysis of Alu and Alu/STR compound systems. *J Hum Genet* 55: 827–833.
104. Matyar S, Attila G, Acartürk E, Akpınar O, Inal T (2005) eNOS gene intron 4 a/b VNTR polymorphism is a risk factor for coronary artery disease in Southern Turkey. *Clin Chim acta* 354: 153–158.
105. Hassan A, Gormley K, O’Sullivan M, Knight J, Sham P, et al. (2004) Endothelial nitric oxide gene haplotypes and risk of cerebral small-vessel disease. *Stroke* 35: 654–659.
106. Khurana VG, Meissner I, Sohni YR, Bamlet WR, McClelland RL, et al. (2005) The presence of tandem endothelial nitric oxide synthase gene polymorphisms identifying brain aneurysms more prone to rupture. *J Neurosurg* 102: 526–531.
107. Corona E, Dudley JT, Butte AJ (2010) Extreme evolutionary disparities seen in positive selection across seven complex diseases. *PLoS One* 5: e12236.
108. Ding K, Kullo IJ (2011) Geographic differences in allele frequencies of susceptibility SNPs for cardiovascular disease. *BMC Med Genet* 12: 55
109. Galimberti D, Scarpini E, Venturelli E, Strobel A, Herterich S, et al. (2008) Association of a NOS1 promoter repeat with Alzheimer’s disease. *Neurobiol Aging* 29: 1359–1365.
110. Velez D, Hulme W, Myers J, Stryjewski M, Abbate E, et al. (2010) Association of SLC11A1 with tuberculosis interactions with NOS2A and TLR2 in African-Americans and Caucasians. *Int J Tuberc Lung Dis* 13: 1068–1076.

111. González-Pérez E, Esteban E, Via M, Gayà-Vidal M, Athanasiadis G, et al. (2010) Population relationships in the Mediterranean revealed by autosomal genetic data (Alu and Alu/STR compound systems). *Am J Phys Anthropol* 141: 430–439.
112. Laayouni H, Calafell F, Bertranpetit J (2010) A genome-wide survey does not show the genetic distinctiveness of Basques. *Hum Genet* 127: 455–458.
113. Sampietro ML, Lao O, Caramelli D, Lari M, Pou R, et al. (2007) Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proc R Soc* 274: 2161–2167.
114. Lacan M, Keyser C, Ricaut F, Brucato N, Duranthon F, et al. (2011) Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proc Natl Acad Sci U S A* 108: 9788–9791.
115. Lacan M, Keyser C, Ricaut F-X, Brucato N, Tarrús J, et al. (2011) Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proc Natl Acad Sci* 108: 18255–18259.
116. Haber M, Gauguier D, Youhanna S, Patterson N, Moorjani P, et al. (2013) Genome-wide diversity in the levant reveals recent structuring by culture. *PLoS Genet* 9: e1003316.
117. Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, et al. (2012) Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS Genet* 8: e1002397.
118. Sánchez-Quinto F, Botigué LR, Civit S, Arenas C, Avila-Arcos MC, et al. (2012) North African populations carry the signature of admixture with Neandertals. *PLoS One* 7: e47765.
119. Currat M, Excoffier L (2005) The effect of the Neolithic expansion on European molecular diversity. *Proc R Soc* 272: 679–688.
120. Klopstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol* 23: 482–490.
121. Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol Evol* 23: 347–351.
122. François O, Currat M, Ray N, Han E, Excoffier L, et al. (2010) Principal component analysis under population genetic models of range expansion and admixture. *Mol Biol Evol* 27: 1257–1268.

7. APÈNDIX

Apèndix 1. Material suplementari de Carreras-Torres et al., 2014

Figure S1 – Geographic population distribution of European and Mediterranean samples. See ST Tab 3 for abbreviation codes.



Figure S2 – Plot of linkage disequilibrium (r^2) between tested genetic markers from *NOS3* / *ATG9B* region in CEU sample.

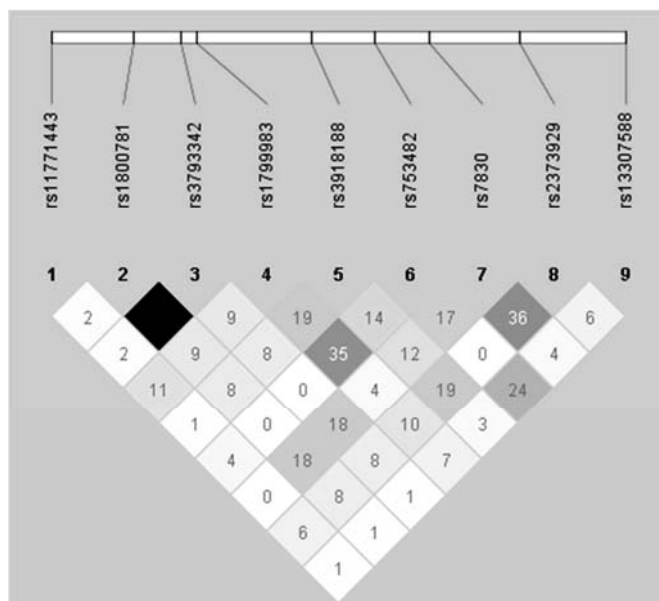


Figure S3 – Plot of linkage disequilibrium (r^2) between tested genetic markers from *NOS1* region in CEU sample.

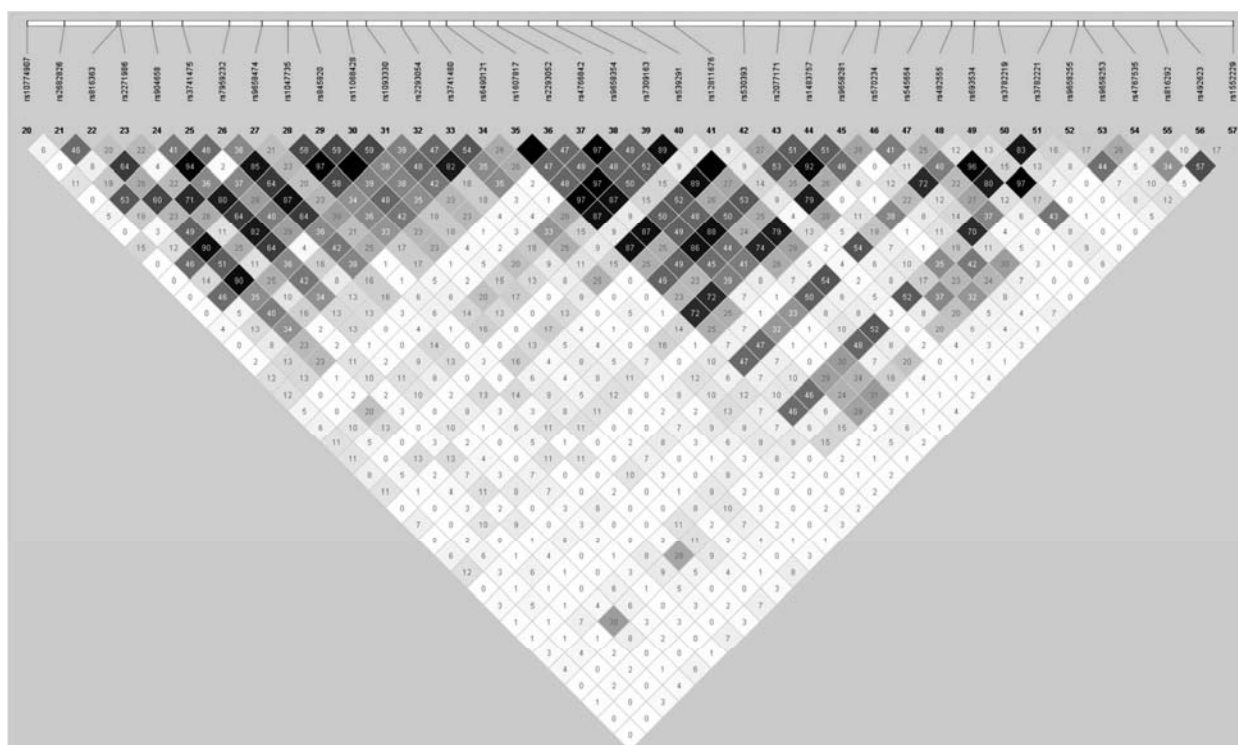


Figure S4 – Plot of linkage disequilibrium (r^2) between tested genetic markers from *NOS2A* region in CEU sample.

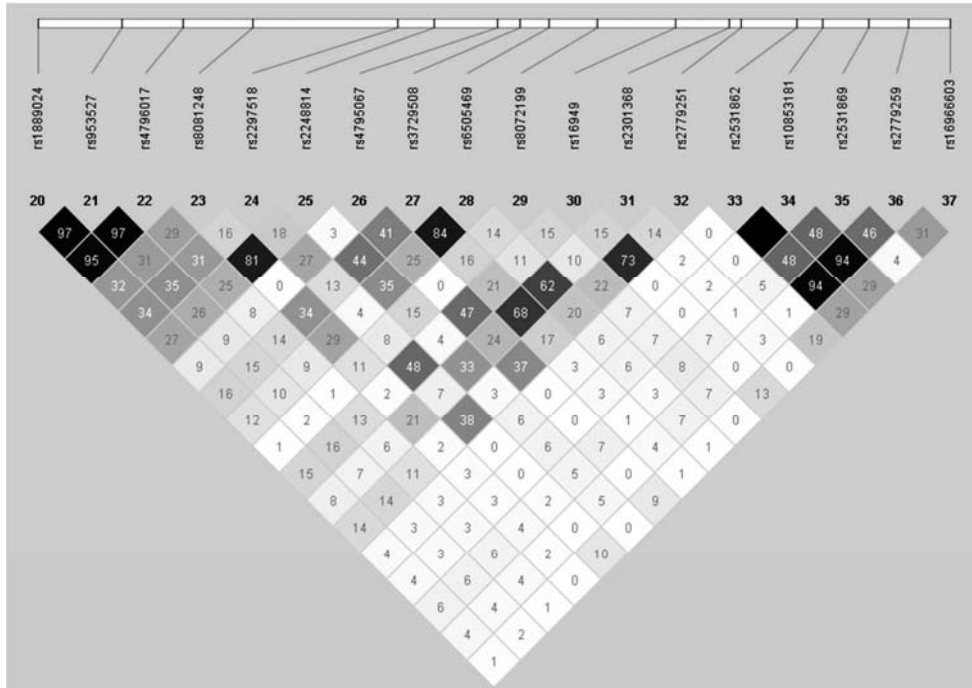
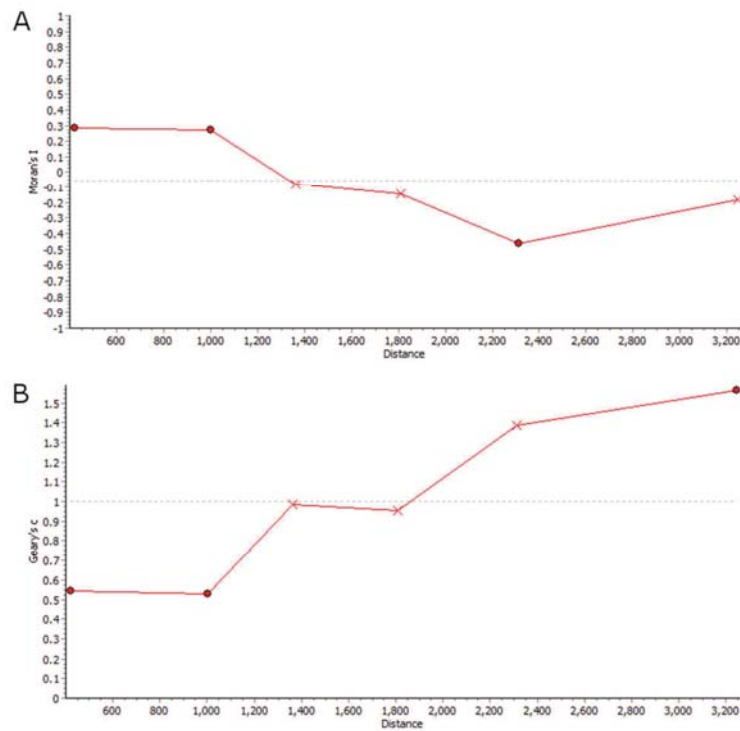


Figure S5- Correlograms of Moran's I (A) and Geary's C (B) autocorrelation coefficients for different distance classes of population pairs. Distances in kilometers. Full circles mean significant coefficients.



Supporting Table (ST) Tab 1 –Genomic location of the genetic variants, genotyping and imputation details. Chromosome positions from Genome Reference Consortium human build 37 (GRCh37). GEN: Genotyped; IMP: Imputed. r^2 : MACH quality metric. i: IMPUTE2 quality metric.

Genetic variants chromosome information						GoldenGate Genotyping Assay	
ID reference	Chr	Position	Alleles	Region	Gene	Genotyped %	Status
rs6951150	7	150681914	C/T	5UTR	NOS3	0.00	Not GEN
rs11771443	7	150687687	C/T	5UTR	NOS3	90.78	GEN
rs1800781	7	150692444	A/G	intron	NOS3	87.55	GEN
rs3793342	7	150695195	A/G	intron	NOS3	90.56	GEN
rs1799983	7	150696111	T/G	coding	NOS3	83.33	GEN
rs3918182	7	150702071	A/G	intron	NOS3	0.00	Not GEN
rs3918188	7	150702781	A/C	intron	NOS3	90.78	GEN
rs753482	7	150706383	T/G	intron	NOS3	89.32	GEN
rs7830	7	150709571	T/G	3UTR	ATG9B	84.93	GEN
rs2373929	7	150714812	T/C	intron	ATG9B	90.34	GEN
rs13307588	7	150720902	A/G	intron	ATG9B	90.99	GEN
rs34459931	7	150721614	-/ACA	5UTR	ATG9B	0.00	Not GEN
rs3763486	7	150723265	A/G	5UTR	ATG9B	0.00	Not GEN
rs10774907	12	117647403	A/G	3UTR	NOS1	90.99	GEN
rs2682826	12	117652838	T/C	3UTR	NOS1	90.92	GEN
rs816362	12	117657345	A/G	intron	NOS1	91.21	Not polymorphic
rs816363	12	117660467	G/C	intron	NOS1	91.14	GEN
rs2271986	12	117660873	T/C	intron	NOS1	88.45	GEN
rs904658	12	117665544	A/C	intron	NOS1	85.73	GEN
rs3741475	12	117669914	A/G	coding	NOS1	90.78	GEN
rs13377860	12	117672539	C/G	coding	NOS1	0.00	Not GEN
rs7959232	12	117676698	A/G	intron	NOS1	89.87	GEN
rs9658474	12	117681509	G/C	intron	NOS1	90.99	GEN
rs1047735	12	117685270	T/C	coding	NOS1	81.01	GEN
rs845820	12	117688699	T/C	intron	NOS1	90.89	GEN
rs11068428	12	117693817	A/G	intron	NOS1	91.10	GEN
rs1093330	12	117696534	A/G	intron	NOS1	90.89	GEN
rs2293054	12	117701714	T/C	coding	NOS1	90.56	GEN
rs3741480	12	117705774	T/C	intron	NOS1	91.07	GEN
rs6490121	12	117708195	A/G	intron	NOS1	90.96	GEN
rs1607817	12	117712186	A/C	intron	NOS1	90.63	GEN
rs2293052	12	117715620	T/C	intron	NOS1	91.07	GEN
rs4766842	12	117720130	G/C	intron	NOS1	91.10	GEN
rs9658354	12	117724225	A/T	intron	NOS1	89.94	GEN
rs7309163	12	117729274	T/C	intron	NOS1	90.96	GEN

MIGen Consortium									
Status	MACH imputation				IMPUTE2 imputation				Imputation Quality
	Finrisk r ²	ATVB r ²	Regicor r ²	MDCS r ²	Finrisk i	ATVB i	Regicor i	MDCS i	
IMP	0.60	0.38	0.35	0.26	0.57	0.44	0.45	0.35	LOW
IMP	0.82	0.84	0.74	0.76	0.90	0.97	0.96	0.87	
GEN	1.00	1.00	0.99	0.99	0.99	1.00	0.98	0.97	
IMP	0.72	0.65	0.61	0.56	0.63	0.66	0.67	0.57	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
IMP	0.87	0.94	0.95	0.87	0.82	0.86	0.88	0.83	
IMP	0.83	0.60	0.60	0.61	0.73	0.57	0.56	0.55	LOW
IMP	0.75	0.49	0.46	0.53	0.71	0.51	0.50	0.54	LOW
IMP	0.88	0.63	0.62	0.55	0.81	0.62	0.67	0.54	
GEN									
IMP	0.72	0.50	0.42	0.59	0.65	0.48	0.49	0.46	LOW
IMP	0.97	0.97	0.97	0.97	0.99	0.97	0.97	0.99	
IMP	0.99	0.97	0.97	0.97	0.98	0.96	0.95	0.98	
IMP	0.96	0.95	0.96	0.99	0.95	0.94	0.95	0.99	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
IMP	1.00	0.99	0.99	0.98	0.99	0.99	0.99	0.99	
IMP	1.00	0.98	0.98	0.98	0.99	0.97	0.98	0.98	
IMP	1.00	0.99	0.99	0.95	0.98	0.98	0.97	0.97	
IMP	1.00	0.99	0.99	0.99	1.00	0.98	0.99	0.98	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
IMP	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99	
IMP	1.00	0.98	0.99	0.98	0.99	0.97	0.98	0.98	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
IMP	0.94	0.87	0.82	0.87	0.94	0.84	0.83	0.88	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
IMP	1.00	1.00	0.99	1.00	1.00	0.99	0.99	1.00	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
IMP	1.00	0.98	0.99	0.99	1.00	0.98	0.99	0.99	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	

rs539291	12	117735141	A/G	intron	NOS1	90.56	GEN
rs12811676	12	117741229	A/G	intron	NOS1	91.03	GEN
rs7295972	12	117747368	A/G	intron	NOS1	0.00	Not GEN
rs530393	12	117751258	T/C	intron	NOS1	89.58	GEN
rs2077171	12	117756502	A/G	intron	NOS1	90.74	GEN
rs1483757	12	117761540	A/G	intron	NOS1	91.03	GEN
rs9658281	12	117767578	T/C	intron	NOS1	90.81	GEN
rs570234	12	117770982	A/C	intron	NOS1	90.20	GEN
rs545654	12	117777049	T/C	intron	NOS1	90.81	GEN
rs482555	12	117781391	T/C	intron	NOS1	90.99	GEN
rs693534	12	117784718	A/G	intron	NOS1	91.10	GEN
rs3782219	12	117788240	T/C	intron	NOS1	89.36	GEN
rs4767533	12	117793530	T/C	intron	NOS1	0.00	Not GEN
rs3782221	12	117795881	A/G	intron	NOS1	90.74	GEN
rs9658255	12	117799756	G/C	5UTR	NOS1	90.81	GEN
rs9658253	12	117800626	T/C	5UTR	NOS1	90.89	GEN
rs4767535	12	117804875	T/C	5UTR	NOS1	90.96	GEN
rs816292	12	117811409	T/C	5UTR	NOS1	90.16	GEN
rs492623	12	117813981	T/C	5UTR	NOS1	90.89	GEN
rs1552229	12	117822187	A/G	5UTR	NOS1	88.93	GEN
rs2091	12	117824885	A/C	5UTR	NOS1	89.54	Not polymorphic
rs1889024	17	26060246	A/G	3UTR	NOS2A	85.08	GEN
rs953527	17	26068742	A/G	3UTR	NOS2A	90.63	GEN
rs4796017	17	26074991	A/G	3UTR	NOS2A	87.76	GEN
rs8081248	17	26081964	A/G	3UTR	NOS2A	89.76	GEN
rs28944201	17	26087190	T/C	coding	NOS2A	90.99	Not polymorphic
rs2255929	17	26087967	T/A	intron	NOS2A	0.00	Not GEN
rs2297514	17	26093315	T/C	intron	NOS2A	0.00	Not GEN
rs2297518	17	26096597	A/G	coding	NOS2A	90.92	GEN
rs2248814	17	26100321	A/G	intron	NOS2A	89.91	GEN
rs4795067	17	26106675	A/G	intron	NOS2A	91.18	GEN
rs3729508	17	26109030	T/C	intron	NOS2A	87.00	GEN
rs6505469	17	26111886	A/T	intron	NOS2A	89.91	GEN
rs8072199	17	26116848	T/C	intron	NOS2A	91.14	GEN
rs16949	17	26124699	T/C	intron	NOS2A	90.23	GEN
rs2779248	17	26127832	A/G	5UTR	NOS2A	0.00	Not GEN
rs2301368	17	26130118	A/G	5UTR	NOS2A	90.70	GEN
rs2779251	17	26131326	A/G	5UTR	NOS2A	85.66	GEN
rs2531862	17	26136991	A/G	5UTR	NOS2A	91.18	GEN
rs10853181	17	26139618	A/G	5UTR	NOS2A	90.85	GEN
rs2531869	17	26144284	A/G	5UTR	NOS2A	90.74	GEN
rs2779259	17	26148284	G/C	5UTR	NOS2A	90.63	GEN
rs16966603	17	26152445	A/T	5UTR	NOS2A	91.10	GEN

IMP	1.00	0.93	0.90	1.00	0.98	0.98	0.98	0.98	
IMP	1.00	0.99	0.99	1.00	1.00	0.99	0.99	1.00	
IMP	1.00	1.00	1.00	1.00	1.00	0.97	0.97	0.99	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
GEN	1.00	1.00	1.00	1.00	0.99	0.99	1.00	0.99	
IMP	0.99	0.94	0.94	0.97	0.98	0.93	0.94	0.96	
IMP	0.97	0.94	0.93	0.94	0.98	0.92	0.93	0.92	
IMP	0.99	0.95	0.96	0.98	0.99	0.95	0.96	0.95	
GEN	1.00	1.00	1.00	0.96	1.00	1.00	1.00	0.74	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
IMP	0.97	0.94	0.97	0.95	0.96	0.92	0.93	0.96	
IMP	1.00	1.00	1.00	1.00	0.99	1.00	0.99	0.98	
IMP	0.98	0.96	0.95	0.98	0.94	0.89	0.88	0.93	
GEN	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
IMP	0.96	0.93	0.93	0.97	0.98	0.88	0.89	0.90	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
IMP	0.99	0.98	0.99	0.99	1.00	0.99	0.99	0.99	
IMP	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.98	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
IMP	0.98	0.96	0.96	0.96	0.99	0.96	0.96	0.96	
GEN									
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
IMP	0.97	0.98	0.97	0.98	0.95	0.95	0.95	0.96	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
GEN	1.00	1.00	1.00	1.00	0.99	0.98	0.99	0.99	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
IMP	0.96	0.76	0.73	0.74	0.87	0.72	0.71	0.73	
IMP	0.95	0.73	0.77	0.85	0.86	0.74	0.71	0.76	
IMP	0.98	0.98	0.99	0.97	0.98	0.99	0.99	0.97	
IMP	0.99	0.99	0.99	0.97	1.00	0.99	0.99	0.98	
IMP	0.98	0.98	0.98	0.97	0.97	0.99	0.99	0.98	
GEN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
IMP	0.98	0.97	0.98	0.97	0.99	0.99	0.99	0.98	

ST Tab 2- Original project, sample size, gender ratio and age (mean, standard deviation and range) of case-control samples. * NDBCS sample is not sex and age matched; upper gender and age description correspond to cases and lower gender and age description to controls.

Case-control sample	Country	Original project	N cases	N controls	% male	age (years)
FINRISK	Finland	MIGen Consortium	167	172	0.69	47.1 ± 6.1 [30.0-60.0]
ATVB	Italy	MIGen Consortium	1688	1664	0.89	39.4 ± 4.9 [16.0-55.0]
Regicor	Spain	MIGen Consortium	310	317	0.79	46.0 ± 5.6 [34.0-60.0]
MDCS	Sweden	MIGen Consortium	86	98	0.57	48.4 ± 4.6 [35.1-56.8]
NDBCS *	Spain	National DNA Bank	291	303	0.81	57.0 ± 10.5 [27.0-80.0]
					0.55	46.0 ± 11.0 [18.0-73.0]

ST Tab 3- Geographic origin, population codification, sample size and geographic coordinates in decimal degrees for the population samples.

Geographic origin	Country	Population	Code	N	Latitude	Longitude
Europe	United Kingdom	Orcadians	ORK	48	59.00	-3.00
	Poland	Warsaw	POL	44	52.25	21.00
	France	North France - Lille	NFR	48	50.60	3.05
		South France - Toulouse	SFR	48	43.65	1.50
		North Corsica	NCO	33	42.60	9.50
		Central Corsica	CCO	45	41.90	8.60
	Spain	Basque Country -Guipuzcoa	BQC	41	42.80	-2.60
		Catalonia - Girona	CAT	43	41.95	2.75
		East Castilla-Leon - Burgos/Valladolid	NSP	39	42.10	-3.73
		West Andalucia - Seville/Cadiz	SSP	46	37.38	-5.97
		Balearic island - Menorca	MEN	49	40.00	3.80
	Italy	Central Italy - Lazio	CIT	37	43.15	13.40
		South Italy - Naples	SIT	39	40.60	15.78
		North Sardinia	NSR	43	40.33	9.40
		South Sardinia	SSR	41	39.20	9.12
		Sicily	SIC	32	38.20	12.50
	Bosnia-Herzegovina	North Bosnia-Herzegovina - Krajina	NBH	40	44.80	17.20
		South Bosnia-Herzegovina - Adriatic coast	SBH	37	43.40	17.80
	Greece	North Greece - Tesaloniki	NGR	48	40.60	22.98
		South Greece - Athens	SGR	43	37.50	22.45
Crete		CRT	44	35.25	25.00	
Turkey	General population	TRK	48	39.95	32.90	
Middle East	Jordania	General Jordanian	GJD	48	31.65	36.50
		Bedouins of Jordania	BJD	42	30.20	35.75
North Africa	Morocco	High Atlas - Asni	ASM	48	31.15	-8.00
		High Atlas - Khenifra	KHM	43	32.90	-5.60
		Low Atlas - Bouhria	BOM	35	34.60	-3.69
	Algeria	M'zab	MZB	50	28.00	3.00
	Tunisia	North-Central Tunisia	NTZ	47	36.80	10.16
		South Tunisia	STZ	49	33.88	10.03
Total DNA samples				1451		
European 1000 Genomes Project	United Kingdom	GBR	GBR	89	52.22	-0.97
	Central Europe	CEU	CEU	85	52.38	8.14
	Finland	FIN	FIN	93	62.02	24.07
	Italy	TSI	TSI	98	43.20	11.67

ST Tab 4 – Case-control allele frequencies and association parameters of MAF variants from MACH imputed data. LIQ: Low imputation quality. LD: Excess linkage disequilibrium. SE: Standard Error. ‘: P value<0.1; *: P value<0.05; **: P value<0.01

Marker information				FINRISK				ATVB			
				Allele frequencies		Association test		Allele frequencies		Association test	
ID reference	Region	Exclusion criteria	MAF Allele	Controls	Cases	Effect	SE	Controls	Cases	Effect	SE
rs11771443	NOS3	LIQ	T	0.158	0.207			0.127	0.134		
rs1800781	NOS3	LD	A	0.170	0.177			0.155	0.163		
rs3793342	NOS3		A	0.166	0.170	-0.13	0.29	0.159	0.169	-0.16	0.11
rs1799983	NOS3		T	0.244	0.268	0.07	0.39	0.345	0.338	-0.46*	0.19
rs3918188	NOS3		A	0.294	0.216	-0.44'	0.24	0.349	0.336	-0.25*	0.10
rs753482	NOS3		G	0.136	0.139	-0.35	0.40	0.266	0.266	0.19'	0.11
rs7830	ATG9B	LIQ	T	0.422	0.471			0.266	0.271		
rs2373929	ATG9B	LIQ	T	0.508	0.560			0.401	0.398		
rs13307588	ATG9B		A	0.053	0.076	0.57	0.42	0.082	0.080	-0.02	0.13
rs10774907	NOS1	LIQ	A	0.313	0.324			0.170	0.185		
rs2682826	NOS1		T	0.319	0.305	0.04	1.00	0.312	0.308	0.08	0.16
rs816363	NOS1	LD	G	0.365	0.395			0.402	0.410		
rs2271986	NOS1		T	0.054	0.096	4.68	4.59	0.062	0.068	0.10	0.23
rs904658	NOS1	LD	C	0.203	0.257			0.347	0.347		
rs3741475	NOS1		A	0.151	0.165	-1.03	1.02	0.260	0.251	0.02	0.20
rs7959232	NOS1	LD	G	0.204	0.267			0.350	0.351		
rs9658474	NOS1		G	0.052	0.093	-5.25	4.65	0.097	0.111	0.13	0.16
rs1047735	NOS1	LD	T	0.198	0.260			0.323	0.316		
rs845820	NOS1	LD	C	0.366	0.401			0.414	0.419		
rs11068428	NOS1		A	0.198	0.259	1.16	0.98	0.322	0.315	-0.18	0.18
rs1093330	NOS1		C	0.366	0.400	0.12	1.08	0.410	0.416	0.10	0.16
rs2293054	NOS1		T	0.299	0.302	-0.27	0.70	0.270	0.264	-0.12	0.13
rs3741480	NOS1		C	0.378	0.425	0.17	1.05	0.422	0.419	-0.07	0.14
rs6490121	NOS1		G	0.302	0.314	-0.17	1.04	0.306	0.310	0.10	0.13
rs1607817	NOS1	LD	A	0.433	0.377			0.382	0.382		
rs2293052	NOS1	LD	T	0.433	0.377			0.379	0.378		
rs4766842	NOS1		C	0.311	0.401	-0.42	0.82	0.397	0.399	-0.08	0.14
rs9658354	NOS1	LD	T	0.311	0.401			0.388	0.393		
rs7309163	NOS1	LD	T	0.433	0.377			0.386	0.386		
rs539291	NOS1	LD	A	0.479	0.398			0.426	0.424		
rs12811676	NOS1		A	0.091	0.144	0.78	1.21	0.024	0.025	0.10	0.24
rs530393	NOS1		T	0.474	0.395	-0.14	0.57	0.392	0.395	-0.08	0.13
rs2077171	NOS1		A	0.224	0.269	0.81	0.95	0.279	0.291	-0.05	0.12
rs1483757	NOS1	LD	G	0.317	0.410			0.400	0.409		
rs9658281	NOS1	LD	T	0.221	0.265			0.268	0.281		

Regicor				Meta-analysis		MDCS		NDBC	
Allele frequencies		Association test				Allele frequencies		Allele frequencies	
Controls	Cases	Effect	SE	Effect	SE	Controls	Cases	Controls	Cases
0.131	0.127					0.132	0.130	0.150	0.120
0.125	0.129					0.113	0.170	0.130	0.160
0.123	0.128	-0.33	0.29	-0.19'	0.10	0.117	0.180	0.120	0.160
0.352	0.350	-1.17*	0.52	-0.44**	0.16	0.305	0.324	0.370	0.370
0.364	0.367	-0.31	0.26	-0.29**	0.09	0.418	0.337	0.340	0.320
0.248	0.277	0.79*	0.31	0.22*	0.10	0.175	0.197	0.290	0.290
0.255	0.242					0.379	0.382	0.330	0.330
0.400	0.392					0.471	0.470	0.420	0.400
0.075	0.084	0.06	0.34	0.04	0.12	0.065	0.070	0.060	0.080
0.197	0.202					0.269	0.281	0.190	0.230
0.276	0.330	-0.32	0.47	0.04	0.15	0.307	0.326	0.310	0.320
0.385	0.443					0.401	0.451	0.410	0.460
0.076	0.079	0.09	0.62	0.11	0.22	0.092	0.129	0.070	0.100
0.309	0.374					0.301	0.343	0.330	0.360
0.199	0.262	0.08	0.58	0.00	0.18	0.213	0.212	0.240	0.240
0.314	0.378					0.297	0.344	0.330	0.370
0.119	0.116	-0.37	0.47	0.07	0.15	0.096	0.151	0.100	0.150
0.275	0.343					0.298	0.324	0.290	0.330
0.394	0.453					0.398	0.448	0.410	0.460
0.274	0.343	0.26	0.55	-0.10	0.17	0.298	0.320	0.290	0.320
0.394	0.452	0.10	0.46	0.10	0.15	0.397	0.449	0.410	0.460
0.248	0.305	-0.03	0.31	-0.11	0.12	0.260	0.279	0.250	0.310
0.431	0.485	0.18	0.38	-0.03	0.13	0.384	0.430	0.450	0.490
0.289	0.358	0.04	0.32	0.09	0.12	0.286	0.302	0.300	0.350
0.375	0.344					0.378	0.343	0.340	0.310
0.377	0.335					0.378	0.343	0.340	0.310
0.418	0.394	-0.21	0.36	-0.10	0.13	0.434	0.494	0.450	0.440
0.416	0.387					0.437	0.495	0.440	0.440
0.379	0.350					0.378	0.343	0.340	0.320
0.447	0.406					0.393	0.355	0.360	0.320
0.041	0.021	-1.39*	0.59	-0.08	0.22	0.111	0.145	0.170	0.150
0.402	0.355	-0.08	0.32	-0.08	0.12	0.398	0.355	0.350	0.320
0.284	0.273	0.11	0.29	-0.02	0.11	0.332	0.355	0.290	0.310
0.437	0.416					0.439	0.506	0.470	0.450
0.252	0.279					0.309	0.337	0.270	0.270

rs570234	NOS1	LD	C	0.482	0.381			0.389	0.397		
rs545654	NOS1		C	0.497	0.546	0.15	0.50	0.510	0.493	-0.04	0.11
rs482555	NOS1	LD	C	0.247	0.254			0.241	0.211		
rs693534	NOS1		A	0.427	0.341	0.66	0.59	0.395	0.398	-0.15	0.12
rs3782219	NOS1		T	0.274	0.292	0.67	0.54	0.244	0.219	-0.24*	0.10
rs3782221	NOS1	LD	A	0.281	0.296			0.266	0.245		
rs9658255	NOS1	LD	C	0.432	0.349			0.399	0.404		
rs9658253	NOS1		T	0.119	0.195	0.64	0.45	0.154	0.170	0.08	0.11
rs4767535	NOS1		C	0.328	0.413	0.03	0.30	0.443	0.433	-0.11	0.09
rs816292	NOS1		T	0.343	0.317	-0.43	0.50	0.277	0.283	0.00	0.12
rs492623	NOS1		T	0.295	0.263	0.08	0.33	0.264	0.260	-0.02	0.09
rs1552229	NOS1		A	0.343	0.320	0.53	0.46	0.364	0.376	0.09	0.10
rs1889024	NOS2A	LD	G	0.486	0.403			0.464	0.462		
rs953527	NOS2A		A	0.452	0.369	-0.17	0.30	0.459	0.457	-0.09	0.07
rs4796017	NOS2A	LD	G	0.453	0.371			0.455	0.456		
rs8081248	NOS2A	LD	A	0.370	0.475			0.377	0.371		
rs2297518	NOS2A		A	0.206	0.159	0.55	0.47	0.220	0.227	0.05	0.11
rs2248814	NOS2A		A	0.364	0.460	0.50	0.31	0.344	0.333	-0.18'	0.09
rs4795067	NOS2A		G	0.422	0.353	-0.54	0.33	0.339	0.353	0.17'	0.09
rs3729508	NOS2A		T	0.334	0.392	-0.97	0.90	0.389	0.391	0.21	0.20
rs6505469	NOS2A		T	0.407	0.437	0.33	0.69	0.452	0.455	0.00	0.18
rs8072199	NOS2A		T	0.410	0.443	0.53	0.35	0.410	0.401	-0.09	0.08
rs16949	NOS2A		C	0.279	0.213	-0.85	0.73	0.248	0.251	-0.06	0.18
rs2301368	NOS2A	LD	G	0.373	0.413			0.405	0.410		
rs2779251	NOS2A		A	0.176	0.150	0.17	0.61	0.172	0.174	0.01	0.24
rs2531862	NOS2A	LD	A	0.392	0.364			0.424	0.432		
rs10853181	NOS2A		G	0.390	0.365	-0.87	5.67	0.424	0.432	0.66	0.66
rs2531869	NOS2A		A	0.154	0.172	0.59	5.64	0.259	0.255	-0.68	0.66
rs2779259	NOS2A	LD	C	0.390	0.365			0.426	0.432		
rs16966603	NOS2A		A	0.236	0.193	0.67	5.67	0.168	0.178	-0.59	0.66

0.400	0.348					0.395	0.320	0.370	0.330
0.500	0.540	0.10	0.26	-0.01	0.10	0.481	0.550	0.520	0.510
0.227	0.256					0.185	0.242	0.240	0.270
0.388	0.346	0.14	0.29	-0.08	0.11	0.418	0.320	0.350	0.350
0.222	0.259	0.17	0.23	-0.15'	0.09	0.179	0.239	0.240	0.250
0.239	0.278					0.209	0.279	0.260	0.270
0.402	0.353					0.429	0.321	0.360	0.360
0.189	0.184	0.01	0.25	0.10	0.10	0.168	0.186	0.160	0.180
0.438	0.485	0.24	0.21	-0.06	0.08	0.388	0.523	0.460	0.440
0.284	0.256	-0.19	0.27	-0.05	0.10	0.327	0.297	0.300	0.290
0.255	0.240	0.06	0.24	0.00	0.08	0.271	0.200	0.240	0.260
0.371	0.368	0.24	0.24	0.13	0.09	0.418	0.459	0.410	0.400
0.402	0.456					0.435	0.413	0.430	0.420
0.397	0.448	0.34	0.20	-0.05	0.07	0.438	0.411	0.430	0.410
0.396	0.445					0.398	0.401	0.430	0.410
0.443	0.419					0.439	0.465	0.440	0.450
0.183	0.195	-0.28	0.26	0.02	0.10	0.189	0.169	0.240	0.190
0.410	0.394	0.21	0.23	-0.08	0.08	0.393	0.458	0.410	0.400
0.301	0.324	0.11	0.23	0.11	0.08	0.352	0.291	0.340	0.310
0.442	0.421	-0.50	0.47	0.06	0.18	0.393	0.471	0.440	0.440
0.500	0.511	0.35	0.43	0.07	0.16	0.469	0.506	0.500	0.500
0.412	0.429	0.19	0.18	-0.03	0.07	0.459	0.506	0.410	0.410
0.202	0.218	0.01	0.42	-0.09	0.16	0.240	0.198	0.250	0.230
0.446	0.444					0.415	0.494	0.410	0.430
0.151	0.146	-0.25	0.63	0.00	0.21	0.171	0.166	0.190	0.170
0.392	0.366					0.347	0.442	0.390	0.390
0.392	0.366	3.82	4.38	0.71	0.65	0.346	0.443	0.390	0.390
0.235	0.234	-3.73	4.35	-0.73	0.64	0.214	0.266	0.190	0.220
0.393	0.366					0.342	0.442	0.390	0.390
0.158	0.133	-3.90	4.39	-0.65	0.65	0.129	0.176	0.200	0.170

ST Tab 5 – Case-control allele frequencies and association parameters of MAF variants from IMPUTE2 imputed data. LIQ: Low imputation quality. LD: Excess linkage disequilibrium. SE: Standard Error. ‘: P value<0.1; *: P value<0.05; **: P value<0.01

Marker information				FINRISK				ATVB			
				Allele frequencies		Association test		Allele frequencies		Association test	
ID reference	Region	Exclusion criteria	MAF Allele	Controls	Cases	Effect	SE	Controls	Cases	Effect	SE
rs11771443	NOS3	LIQ	T	0.158	0.199			0.134	0.146		
rs1800781	NOS3	LD	A	0.163	0.166			0.159	0.169		
rs3793342	NOS3		A	0.166	0.171	-0.19	0.30	0.160	0.169	-0.18	0.12
rs1799983	NOS3		T	0.247	0.271	0.00	0.45	0.342	0.335	-0.51*	0.22
rs3918188	NOS3		A	0.294	0.216	-0.51	0.25	0.349	0.336	-0.25*	0.10
rs753482	NOS3		G	0.145	0.153	-0.37	0.44	0.271	0.268	0.23	0.15
rs7830	ATG9B	LIQ	T	0.425	0.470			0.270	0.276		
rs2373929	ATG9B	LIQ	T	0.507	0.557			0.426	0.420		
rs13307588	ATG9B		A	0.049	0.073	0.78	0.50	0.096	0.095	0.00	0.12
rs10774907	NOS1	LIQ	A	0.330	0.342			0.175	0.182		
rs2682826	NOS1		T	0.318	0.301	-0.14	1.03	0.313	0.310	0.11	0.17
rs816363	NOS1	LD	G	0.371	0.400			0.401	0.409		
rs2271986	NOS1		T	0.056	0.099	1.78	2.37	0.062	0.068	0.10	0.23
rs904658	NOS1	LD	C	0.203	0.257			0.348	0.347		
rs3741475	NOS1		A	0.151	0.165	-1.20	1.05	0.260	0.250	-0.03	0.20
rs7959232	NOS1	LD	G	0.205	0.266			0.351	0.352		
rs9658474	NOS1		G	0.053	0.093	-2.58	2.68	0.096	0.110	0.12	0.16
rs1047735	NOS1	LD	T	0.198	0.260			0.322	0.317		
rs845820	NOS1	LD	C	0.366	0.401			0.413	0.419		
rs11068428	NOS1		A	0.198	0.258	1.30	1.00	0.320	0.315	-0.14	0.18
rs1093330	NOS1		C	0.367	0.400	0.36	1.10	0.409	0.417	0.10	0.17
rs2293054	NOS1		T	0.299	0.302	-0.24	0.73	0.270	0.264	-0.12	0.13
rs3741480	NOS1		C	0.383	0.435	0.00	1.50	0.421	0.419	-0.10	0.17
rs6490121	NOS1		G	0.302	0.314	-0.11	1.36	0.306	0.310	0.15	0.15
rs1607817	NOS1	LD	A	0.433	0.377			0.382	0.382		
rs2293052	NOS1	LD	T	0.433	0.377			0.379	0.379		
rs4766842	NOS1		C	0.311	0.401	-0.46	0.85	0.397	0.399	-0.09	0.14
rs9658354	NOS1	LD	T	0.312	0.401			0.389	0.394		
rs7309163	NOS1	LD	T	0.433	0.377			0.386	0.386		
rs539291	NOS1	LD	A	0.480	0.399			0.397	0.397		
rs12811676	NOS1		A	0.090	0.144	0.84	1.43	0.119	0.113	0.10	0.17
rs530393	NOS1		T	0.474	0.395	-0.18	0.58	0.401	0.404	-0.05	0.13
rs2077171	NOS1		A	0.224	0.269	0.71	0.96	0.279	0.291	-0.02	0.13
rs1483757	NOS1	LD	G	0.314	0.410			0.400	0.408		
rs9658281	NOS1	LD	T	0.221	0.266			0.267	0.280		

Regicor				Meta-analysis		MDCS		NDBC	
Allele frequencies		Association test				Allele frequencies		Allele frequencies	
Controls	Cases	Effect	SE	Effect	SE	Controls	Cases	Controls	Cases
0.135	0.135					0.143	0.141	0.150	0.120
0.124	0.128					0.124	0.181	0.130	0.160
0.123	0.131	-0.36	0.28	-0.20*	0.10	0.117	0.186	0.120	0.160
0.359	0.352	-1.39**	0.53	-0.53**	0.18	0.319	0.327	0.370	0.370
0.366	0.368	-0.37	0.24	-0.30**	0.09	0.418	0.337	0.340	0.320
0.260	0.282	0.94*	0.38	0.26*	0.13	0.821	0.802	0.290	0.290
0.255	0.250					0.369	0.373	0.330	0.330
0.429	0.406					0.526	0.548	0.420	0.400
0.087	0.101	0.14	0.30	0.06	0.11	0.069	0.083	0.060	0.080
0.193	0.197					0.294	0.312	0.190	0.230
0.278	0.333	-0.36	0.46	0.05	0.16	0.698	0.699	0.310	0.320
0.383	0.438					0.404	0.451	0.410	0.460
0.076	0.079	0.70	0.62	0.19	0.22	0.092	0.129	0.070	0.100
0.309	0.374					0.301	0.343	0.330	0.360
0.203	0.265	0.28	0.51	-0.03	0.18	0.213	0.193	0.240	0.240
0.317	0.378					0.294	0.348	0.330	0.370
0.119	0.114	-0.65	0.49	0.03	0.15	0.095	0.156	0.100	0.150
0.274	0.340					0.705	0.681	0.290	0.330
0.394	0.453					0.602	0.552	0.410	0.460
0.274	0.341	0.05	0.49	-0.08	0.17	0.293	0.317	0.290	0.320
0.393	0.452	0.19	0.47	0.12	0.16	0.397	0.447	0.410	0.460
0.249	0.305	-0.26	0.31	-0.15	0.12	0.260	0.279	0.250	0.310
0.437	0.489	0.53	0.41	-0.01	0.16	0.386	0.430	0.450	0.490
0.289	0.360	-0.27	0.36	0.09	0.14	0.286	0.302	0.300	0.350
0.375	0.344					0.378	0.343	0.340	0.310
0.376	0.335					0.622	0.657	0.340	0.310
0.418	0.394	0.07	0.36	-0.08	0.13	0.434	0.494	0.450	0.440
0.417	0.389					0.439	0.496	0.440	0.440
0.379	0.350					0.378	0.343	0.340	0.320
0.393	0.353					0.396	0.355	0.360	0.320
0.152	0.122	-1.19**	0.44	-0.06	0.16	0.113	0.146	0.170	0.150
0.413	0.362	-0.30	0.32	-0.09	0.12	0.398	0.355	0.350	0.320
0.284	0.274	-0.50	0.34	-0.07	0.12	0.668	0.645	0.290	0.310
0.437	0.416					0.439	0.500	0.470	0.450
0.249	0.273					0.306	0.337	0.270	0.270

rs570234	NOS1	LD	C	0.485	0.383			0.390	0.398		
rs545654	NOS1		C	0.495	0.547	0.20	0.51	0.512	0.496	-0.01	0.11
rs482555	NOS1	LD	C	0.247	0.254			0.240	0.211		
rs693534	NOS1		A	0.427	0.341	0.66	0.59	0.395	0.398	-0.15	0.12
rs3782219	NOS1		T	0.275	0.293	0.62	0.53	0.251	0.224	-0.25*	0.10
rs3782221	NOS1	LD	A	0.285	0.296			0.267	0.246		
rs9658255	NOS1	LD	C	0.434	0.348			0.409	0.411		
rs9658253	NOS1		T	0.122	0.195	0.68	0.44	0.154	0.170	0.07	0.11
rs4767535	NOS1		C	0.328	0.413	0.13	0.30	0.443	0.433	-0.13	0.08
rs816292	NOS1		T	0.343	0.317	-0.33	0.50	0.277	0.283	0.00	0.12
rs492623	NOS1		T	0.291	0.265	0.25	0.32	0.287	0.281	-0.03	0.09
rs1552229	NOS1		A	0.343	0.320	0.52	0.47	0.364	0.376	0.08	0.10
rs1889024	NOS2A	LD	G	0.486	0.405			0.476	0.475		
rs953527	NOS2A		A	0.453	0.370	-0.06	0.31	0.460	0.458	-0.10	0.07
rs4796017	NOS2A	LD	G	0.453	0.371			0.455	0.455		
rs8081248	NOS2A	LD	A	0.369	0.477			0.380	0.372		
rs2297518	NOS2A		A	0.206	0.159	0.45	0.47	0.221	0.228	0.06	0.11
rs2248814	NOS2A		A	0.361	0.462	0.62*	0.31	0.342	0.333	-0.19*	0.10
rs4795067	NOS2A		G	0.422	0.353	-0.56'	0.32	0.340	0.353	0.15	0.09
rs3729508	NOS2A		T	0.337	0.389	-0.99	0.83	0.392	0.393	0.10	0.18
rs6505469	NOS2A		T	0.407	0.437	0.24	0.62	0.452	0.455	0.13	0.17
rs8072199	NOS2A		T	0.410	0.443	0.55	0.34	0.410	0.401	-0.12	0.08
rs16949	NOS2A		C	0.279	0.213	-0.66	0.68	0.248	0.252	-0.21	0.18
rs2301368	NOS2A	LD	G	0.383	0.423			0.403	0.402		
rs2779251	NOS2A		A	0.178	0.150	-0.07	0.60	0.176	0.180	0.28	0.25
rs2531862	NOS2A	LD	A	0.403	0.371			0.427	0.433		
rs10853181	NOS2A		G	0.389	0.364	0.62	1.04	0.426	0.432	-0.14	0.67
rs2531869	NOS2A		A	0.168	0.180	-0.87	1.00	0.261	0.256	0.10	0.67
rs2779259	NOS2A	LD	C	0.390	0.365			0.426	0.432		
rs16966603	NOS2A		A	0.236	0.193	-0.73	1.06	0.167	0.178	0.16	0.67

0.403	0.343					0.388	0.316	0.370	0.330
0.503	0.545	0.15	0.26	0.02	0.10	0.488	0.548	0.520	0.510
0.229	0.256					0.230	0.279	0.240	0.270
0.388	0.347	0.04	0.28	-0.09	0.11	0.418	0.320	0.350	0.350
0.237	0.266	0.05	0.25	-0.18'	0.09	0.189	0.247	0.240	0.250
0.241	0.279					0.214	0.279	0.260	0.270
0.408	0.360					0.435	0.324	0.360	0.360
0.189	0.184	-0.05	0.27	0.09	0.10	0.168	0.186	0.160	0.180
0.438	0.485	0.32	0.21	-0.05	0.07	0.388	0.523	0.460	0.440
0.284	0.256	-0.02	0.27	-0.02	0.10	0.327	0.297	0.300	0.290
0.265	0.252	0.11	0.24	0.00	0.08	0.283	0.207	0.240	0.260
0.371	0.368	0.13	0.24	0.10	0.09	0.418	0.459	0.410	0.400
0.409	0.474					0.440	0.413	0.430	0.420
0.398	0.450	0.32	0.20	-0.05	0.07	0.433	0.412	0.430	0.410
0.396	0.444					0.398	0.401	0.430	0.410
0.445	0.422					0.443	0.468	0.440	0.450
0.183	0.195	-0.29	0.26	0.03	0.10	0.189	0.169	0.240	0.190
0.409	0.390	0.24	0.25	-0.08	0.09	0.390	0.459	0.410	0.400
0.301	0.324	0.06	0.22	0.09	0.08	0.352	0.291	0.340	0.310
0.446	0.419	-0.64	0.42	-0.05	0.16	0.398	0.471	0.440	0.440
0.502	0.511	0.34	0.38	0.17	0.15	0.469	0.506	0.500	0.500
0.412	0.429	0.27	0.19	-0.04	0.07	0.459	0.506	0.410	0.410
0.202	0.218	0.12	0.41	-0.19	0.16	0.240	0.198	0.250	0.230
0.448	0.463					0.429	0.485	0.410	0.430
0.153	0.143	-0.55	0.66	0.14	0.22	0.171	0.168	0.190	0.170
0.394	0.368					0.357	0.445	0.390	0.390
0.393	0.366	3.22	5.50	0.12	0.56	0.352	0.445	0.390	0.390
0.236	0.234	-3.16	5.47	-0.23	0.55	0.224	0.268	0.190	0.220
0.393	0.366					0.342	0.442	0.390	0.390
0.157	0.133	-3.22	5.49	-0.13	0.56	0.121	0.175	0.200	0.170

ST Tab 6. Minor allele frequencies (MAF) for population samples.

Marker information		Populations															
ID reference	MAF Allele	ORK	GBR	CEU	POL	FIN	NFR	SFR	NCO	CCO	BQC	CAT	NSP	SSP	MEN	TSI	CIT
rs11771443	T	0.125	0.157	0.135	0.159	0.161	0.094	0.146	0.076	0.089	0.073	0.093	0.218	0.152	0.224	0.148	0.108
rs1800781	A	0.198	0.163	0.147	0.170	0.129	0.135	0.198	0.182	0.089	0.061	0.081	0.141	0.130	0.163	0.148	0.149
rs3793342	A	0.229	0.157	0.147	0.170	0.129	0.135	0.198	0.182	0.089	0.061	0.081	0.141	0.130	0.163	0.148	0.149
rs1799983	T	0.302	0.326	0.359	0.239	0.220	0.313	0.396	0.379	0.433	0.293	0.395	0.346	0.359	0.367	0.388	0.338
rs3918188	A	0.365	0.331	0.341	0.330	0.296	0.396	0.271	0.333	0.356	0.476	0.419	0.410	0.293	0.235	0.296	0.338
rs753482	G	0.208	0.230	0.218	0.159	0.124	0.281	0.260	0.273	0.356	0.232	0.314	0.256	0.283	0.255	0.327	0.297
rs7830	T	0.417	0.365	0.382	0.455	0.446	0.271	0.354	0.333	0.189	0.341	0.209	0.282	0.348	0.255	0.245	0.324
rs2373929	T	0.490	0.455	0.482	0.455	0.538	0.323	0.438	0.394	0.289	0.293	0.267	0.346	0.424	0.337	0.398	0.311
rs13307588	A	0.052	0.062	0.065	0.080	0.048	0.094	0.094	0.121	0.144	0.085	0.140	0.064	0.065	0.041	0.077	0.108
rs10774907	A	0.313	0.331	0.306	0.318	0.333	0.240	0.219	0.258	0.244	0.220	0.209	0.167	0.250	0.286	0.209	0.135
rs2682826	T	0.219	0.309	0.306	0.227	0.306	0.250	0.271	0.318	0.467	0.220	0.360	0.269	0.283	0.235	0.316	0.392
rs816363	G	0.302	0.466	0.465	0.364	0.392	0.396	0.365	0.409	0.556	0.317	0.442	0.397	0.435	0.357	0.434	0.459
rs2271986	T	0.104	0.118	0.153	0.125	0.081	0.115	0.052	0.061	0.011	0.098	0.023	0.115	0.109	0.071	0.082	0.054
rs904658	C	0.250	0.393	0.359	0.307	0.258	0.281	0.292	0.318	0.489	0.293	0.360	0.321	0.304	0.265	0.372	0.392
rs3741475	A	0.135	0.230	0.206	0.182	0.172	0.156	0.219	0.258	0.444	0.195	0.267	0.205	0.185	0.153	0.255	0.297
rs7959232	G	0.260	0.404	0.359	0.330	0.274	0.302	0.281	0.318	0.489	0.305	0.372	0.295	0.304	0.276	0.378	0.392
rs9658474	G	0.125	0.174	0.171	0.136	0.086	0.146	0.083	0.106	0.056	0.134	0.128	0.141	0.141	0.122	0.128	0.108
rs1047735	T	0.292	0.360	0.335	0.307	0.269	0.271	0.260	0.303	0.400	0.317	0.314	0.295	0.283	0.235	0.347	0.351
rs845820	C	0.344	0.478	0.465	0.364	0.409	0.406	0.375	0.379	0.556	0.329	0.453	0.385	0.424	0.388	0.449	0.459
rs11068428	A	0.271	0.360	0.341	0.307	0.269	0.271	0.260	0.303	0.400	0.305	0.302	0.295	0.272	0.235	0.347	0.351
rs1093330	C	0.344	0.478	0.465	0.364	0.409	0.396	0.385	0.379	0.556	0.329	0.453	0.385	0.424	0.388	0.444	0.459
rs2293054	T	0.240	0.326	0.312	0.239	0.317	0.271	0.250	0.182	0.356	0.195	0.244	0.295	0.293	0.265	0.311	0.284
rs3741480	C	0.427	0.449	0.488	0.386	0.430	0.521	0.344	0.303	0.556	0.427	0.500	0.449	0.467	0.459	0.459	0.473
rs6490121	G	0.260	0.354	0.341	0.295	0.344	0.271	0.281	0.258	0.389	0.244	0.326	0.410	0.326	0.286	0.372	0.297
rs1607817	A	0.354	0.354	0.335	0.409	0.371	0.333	0.448	0.409	0.322	0.488	0.372	0.321	0.370	0.408	0.327	0.392
rs2293052	T	0.354	0.354	0.335	0.409	0.371	0.333	0.448	0.409	0.322	0.488	0.360	0.321	0.370	0.408	0.321	0.405
rs4766842	C	0.500	0.438	0.482	0.409	0.414	0.500	0.375	0.409	0.356	0.402	0.384	0.436	0.380	0.378	0.423	0.365
rs9658354	T	0.500	0.421	0.488	0.591	0.414	0.500	0.625	0.591	0.644	0.598	0.628	0.577	0.620	0.622	0.408	0.635
rs7309163	T	0.396	0.360	0.341	0.432	0.371	0.323	0.448	0.394	0.311	0.488	0.360	0.321	0.380	0.429	0.327	0.405
rs539291	A	0.385	0.382	0.353	0.443	0.376	0.344	0.458	0.409	0.322	0.476	0.360	0.346	0.391	0.459	0.357	0.392
rs12811676	A	0.198	0.101	0.153	0.114	0.108	0.240	0.063	0.106	0.156	0.195	0.140	0.090	0.152	0.112	0.077	0.122
rs530393	T	0.385	0.382	0.353	0.432	0.371	0.354	0.469	0.409	0.322	0.476	0.360	0.346	0.391	0.459	0.357	0.405
rs2077171	A	0.313	0.331	0.335	0.273	0.323	0.292	0.292	0.303	0.200	0.244	0.256	0.333	0.250	0.235	0.332	0.243
rs1483757	G	0.542	0.433	0.494	0.386	0.430	0.510	0.385	0.409	0.378	0.378	0.453	0.462	0.391	0.347	0.408	0.378
rs9658281	T	0.292	0.315	0.353	0.261	0.301	0.240	0.250	0.258	0.178	0.220	0.291	0.359	0.250	0.194	0.296	0.257
rs570234	C	0.354	0.388	0.341	0.455	0.366	0.385	0.469	0.424	0.278	0.524	0.326	0.372	0.391	0.418	0.332	0.432
rs545654	C	0.500	0.478	0.465	0.466	0.565	0.510	0.469	0.515	0.611	0.329	0.547	0.462	0.467	0.449	0.536	0.486

SIT	NSR	SSR	SIC	NBH	SBH	NGR	SGR	CRT	TRK	GJD	BJD	ASM	KHM	BOM	MZB	NTZ	STZ
0.167	0.163	0.146	0.250	0.100	0.149	0.156	0.151	0.205	0.250	0.260	0.167	0.146	0.105	0.086	0.050	0.149	0.143
0.167	0.105	0.110	0.141	0.225	0.162	0.146	0.186	0.216	0.063	0.104	0.107	0.146	0.174	0.057	0.100	0.085	0.143
0.167	0.116	0.122	0.141	0.225	0.162	0.135	0.186	0.205	0.146	0.146	0.107	0.156	0.198	0.057	0.140	0.085	0.153
0.372	0.523	0.451	0.219	0.238	0.297	0.281	0.407	0.250	0.208	0.250	0.179	0.260	0.326	0.357	0.350	0.255	0.224
0.295	0.314	0.280	0.344	0.375	0.378	0.396	0.244	0.352	0.323	0.406	0.464	0.313	0.198	0.357	0.240	0.319	0.367
0.308	0.395	0.268	0.156	0.138	0.203	0.167	0.349	0.216	0.146	0.240	0.250	0.240	0.372	0.400	0.460	0.266	0.235
0.295	0.186	0.354	0.375	0.425	0.338	0.313	0.349	0.364	0.260	0.271	0.226	0.354	0.407	0.229	0.460	0.287	0.296
0.346	0.360	0.451	0.344	0.413	0.338	0.406	0.326	0.409	0.385	0.271	0.321	0.458	0.674	0.371	0.460	0.479	0.357
0.103	0.105	0.098	0.016	0.063	0.068	0.031	0.058	0.045	0.042	0.042	0.024	0.021	0.012	0.043	0.460	0.053	0.020
0.231	0.198	0.244	0.125	0.238	0.230	0.188	0.221	0.125	0.135	0.198	0.202	0.333	0.314	0.257	0.460	0.362	0.276
0.372	0.465	0.378	0.375	0.175	0.270	0.302	0.337	0.341	0.313	0.396	0.274	0.396	0.372	0.286	0.460	0.457	0.378
0.513	0.558	0.451	0.438	0.288	0.365	0.406	0.395	0.443	0.406	0.500	0.393	0.583	0.523	0.457	0.460	0.553	0.480
0.090	0.023	0.024	0.031	0.088	0.068	0.115	0.035	0.011	0.063	0.073	0.036	0.031	0.035	0.057	0.460	0.032	0.031
0.449	0.500	0.415	0.328	0.213	0.324	0.365	0.291	0.420	0.229	0.479	0.393	0.563	0.453	0.457	0.460	0.489	0.429
0.321	0.430	0.329	0.281	0.125	0.216	0.250	0.244	0.352	0.281	0.396	0.321	0.375	0.314	0.286	0.460	0.394	0.337
0.462	0.500	0.415	0.328	0.225	0.324	0.354	0.314	0.432	0.365	0.490	0.417	0.573	0.442	0.486	0.460	0.553	0.449
0.141	0.070	0.098	0.078	0.100	0.108	0.125	0.058	0.091	0.115	0.104	0.083	0.219	0.174	0.200	0.460	0.149	0.133
0.397	0.442	0.329	0.328	0.213	0.297	0.333	0.291	0.375	0.365	0.469	0.321	0.344	0.395	0.414	0.460	0.415	0.327
0.487	0.547	0.451	0.453	0.263	0.378	0.427	0.419	0.477	0.396	0.531	0.440	0.667	0.570	0.557	0.460	0.628	0.490
0.397	0.442	0.329	0.328	0.213	0.297	0.333	0.291	0.375	0.333	0.469	0.333	0.333	0.395	0.414	0.460	0.415	0.327
0.500	0.547	0.439	0.453	0.263	0.378	0.406	0.430	0.477	0.396	0.521	0.440	0.656	0.570	0.557	0.460	0.628	0.490
0.397	0.302	0.207	0.266	0.175	0.243	0.229	0.233	0.227	0.260	0.313	0.274	0.323	0.326	0.129	0.460	0.351	0.337
0.577	0.500	0.402	0.359	0.313	0.405	0.375	0.349	0.455	0.417	0.490	0.381	0.573	0.558	0.414	0.460	0.649	0.449
0.462	0.372	0.232	0.313	0.250	0.284	0.271	0.279	0.341	0.271	0.427	0.298	0.365	0.349	0.171	0.460	0.426	0.357
0.321	0.314	0.341	0.375	0.425	0.338	0.406	0.395	0.341	0.354	0.344	0.476	0.406	0.337	0.443	0.460	0.266	0.459
0.321	0.314	0.341	0.375	0.425	0.338	0.385	0.395	0.341	0.344	0.344	0.440	0.333	0.291	0.386	0.460	0.223	0.378
0.346	0.419	0.488	0.375	0.438	0.473	0.438	0.337	0.409	0.479	0.365	0.250	0.240	0.291	0.443	0.460	0.319	0.194
0.654	0.581	0.537	0.625	0.575	0.527	0.583	0.663	0.614	0.510	0.635	0.750	0.760	0.709	0.557	0.460	0.681	0.806
0.321	0.302	0.317	0.375	0.425	0.311	0.406	0.419	0.330	0.333	0.323	0.524	0.448	0.372	0.429	0.460	0.319	0.520
0.321	0.326	0.354	0.375	0.425	0.324	0.385	0.419	0.341	0.323	0.313	0.500	0.365	0.349	0.429	0.460	0.277	0.459
0.115	0.105	0.159	0.078	0.125	0.108	0.135	0.093	0.091	0.188	0.073	0.095	0.156	0.140	0.257	0.460	0.191	0.020
0.321	0.326	0.354	0.359	0.425	0.324	0.396	0.419	0.341	0.323	0.313	0.512	0.354	0.349	0.414	0.460	0.266	0.439
0.256	0.314	0.366	0.344	0.325	0.392	0.313	0.256	0.330	0.292	0.302	0.167	0.104	0.140	0.171	0.460	0.128	0.173
0.372	0.465	0.537	0.422	0.450	0.500	0.417	0.337	0.386	0.448	0.385	0.262	0.260	0.279	0.457	0.460	0.340	0.235
0.231	0.256	0.268	0.313	0.300	0.378	0.333	0.244	0.330	0.260	0.323	0.179	0.094	0.128	0.086	0.460	0.128	0.153
0.295	0.384	0.402	0.406	0.388	0.365	0.354	0.430	0.330	0.354	0.333	0.464	0.313	0.302	0.371	0.460	0.298	0.388
0.551	0.547	0.573	0.469	0.538	0.554	0.458	0.535	0.500	0.573	0.448	0.369	0.490	0.453	0.486	0.460	0.543	0.408

rs482555	C	0.229	0.180	0.182	0.227	0.269	0.250	0.229	0.167	0.411	0.171	0.279	0.231	0.196	0.184	0.240	0.257
rs693534	A	0.354	0.399	0.359	0.409	0.328	0.354	0.427	0.455	0.267	0.573	0.349	0.333	0.435	0.500	0.337	0.405
rs3782219	T	0.167	0.163	0.188	0.227	0.285	0.250	0.229	0.167	0.411	0.171	0.279	0.218	0.196	0.173	0.240	0.257
rs3782221	A	0.198	0.169	0.218	0.250	0.285	0.281	0.250	0.182	0.433	0.171	0.302	0.231	0.207	0.194	0.276	0.257
rs9658255	C	0.365	0.416	0.365	0.409	0.328	0.365	0.427	0.439	0.300	0.561	0.349	0.359	0.435	0.490	0.347	0.405
rs9658253	T	0.208	0.225	0.235	0.193	0.188	0.177	0.167	0.106	0.156	0.195	0.163	0.218	0.130	0.173	0.219	0.149
rs4767535	C	0.552	0.466	0.535	0.466	0.409	0.438	0.417	0.424	0.467	0.341	0.512	0.449	0.391	0.347	0.444	0.419
rs816292	T	0.313	0.202	0.229	0.284	0.237	0.344	0.281	0.379	0.333	0.354	0.267	0.256	0.391	0.439	0.245	0.405
rs492623	T	0.188	0.298	0.253	0.205	0.317	0.208	0.292	0.152	0.211	0.293	0.198	0.269	0.250	0.224	0.260	0.162
rs1552229	A	0.313	0.326	0.341	0.409	0.263	0.396	0.354	0.485	0.389	0.427	0.419	0.385	0.500	0.480	0.291	0.527
rs1889024	G	0.292	0.483	0.394	0.511	0.511	0.469	0.417	0.545	0.467	0.366	0.465	0.474	0.435	0.469	0.439	0.473
rs953527	A	0.292	0.455	0.388	0.489	0.489	0.469	0.406	0.530	0.422	0.390	0.442	0.474	0.424	0.459	0.423	0.459
rs4796017	G	0.292	0.449	0.382	0.489	0.489	0.458	0.406	0.530	0.422	0.390	0.453	0.474	0.424	0.449	0.429	0.432
rs8081248	A	0.542	0.399	0.482	0.432	0.409	0.333	0.406	0.348	0.378	0.451	0.384	0.436	0.457	0.459	0.439	0.324
rs2297518	A	0.083	0.247	0.182	0.295	0.269	0.177	0.188	0.258	0.256	0.232	0.186	0.269	0.217	0.224	0.209	0.243
rs2248814	A	0.479	0.376	0.459	0.398	0.392	0.302	0.417	0.348	0.344	0.476	0.349	0.436	0.435	0.367	0.393	0.324
rs4795067	G	0.271	0.393	0.347	0.432	0.376	0.427	0.313	0.394	0.389	0.329	0.337	0.359	0.315	0.337	0.321	0.284
rs3729508	T	0.490	0.371	0.435	0.432	0.382	0.344	0.417	0.333	0.333	0.500	0.384	0.449	0.435	0.429	0.444	0.378
rs6505469	T	0.521	0.466	0.476	0.489	0.441	0.396	0.490	0.394	0.444	0.561	0.500	0.526	0.500	0.490	0.495	0.432
rs8072199	T	0.479	0.427	0.500	0.602	0.484	0.375	0.396	0.470	0.289	0.549	0.267	0.474	0.391	0.439	0.434	0.419
rs16949	C	0.188	0.270	0.218	0.318	0.285	0.260	0.219	0.333	0.256	0.232	0.221	0.256	0.217	0.255	0.245	0.243
rs2301368	G	0.510	0.399	0.447	0.409	0.414	0.313	0.427	0.318	0.367	0.512	0.395	0.436	0.391	0.459	0.464	0.351
rs2779251	A	0.146	0.169	0.182	0.216	0.194	0.219	0.156	0.258	0.133	0.171	0.093	0.205	0.141	0.173	0.194	0.176
rs2531862	A	0.344	0.360	0.347	0.386	0.414	0.396	0.448	0.394	0.467	0.341	0.349	0.385	0.370	0.429	0.418	0.378
rs10853181	G	0.344	0.360	0.347	0.386	0.414	0.396	0.448	0.394	0.467	0.341	0.360	0.385	0.370	0.429	0.418	0.378
rs2531869	A	0.229	0.213	0.206	0.159	0.183	0.188	0.240	0.212	0.367	0.159	0.174	0.179	0.174	0.255	0.224	0.243
rs2779259	C	0.354	0.365	0.335	0.386	0.409	0.396	0.448	0.394	0.467	0.341	0.372	0.385	0.370	0.418	0.423	0.378
rs16966603	A	0.125	0.146	0.135	0.227	0.226	0.208	0.208	0.182	0.100	0.183	0.186	0.205	0.174	0.163	0.199	0.122

0.321	0.302	0.256	0.172	0.213	0.257	0.167	0.233	0.193	0.271	0.188	0.119	0.292	0.279	0.257	0.460	0.266	0.214
0.333	0.360	0.329	0.375	0.388	0.324	0.427	0.465	0.420	0.333	0.417	0.476	0.344	0.477	0.343	0.460	0.383	0.469
0.321	0.302	0.256	0.141	0.213	0.230	0.188	0.221	0.205	0.229	0.188	0.131	0.302	0.291	0.243	0.460	0.234	0.194
0.372	0.326	0.280	0.125	0.225	0.230	0.219	0.233	0.250	0.292	0.188	0.179	0.323	0.302	0.286	0.460	0.277	0.224
0.333	0.360	0.329	0.406	0.388	0.324	0.427	0.465	0.432	0.354	0.427	0.488	0.344	0.477	0.414	0.460	0.394	0.469
0.154	0.128	0.171	0.125	0.175	0.203	0.167	0.116	0.136	0.167	0.167	0.107	0.167	0.081	0.100	0.460	0.117	0.102
0.526	0.477	0.500	0.438	0.438	0.486	0.479	0.302	0.409	0.458	0.396	0.405	0.521	0.419	0.557	0.460	0.511	0.398
0.256	0.291	0.354	0.219	0.288	0.257	0.240	0.395	0.170	0.198	0.229	0.143	0.313	0.314	0.343	0.460	0.223	0.327
0.295	0.174	0.171	0.281	0.263	0.216	0.302	0.233	0.432	0.323	0.333	0.440	0.208	0.302	0.171	0.460	0.202	0.306
0.321	0.395	0.524	0.313	0.325	0.311	0.313	0.477	0.307	0.240	0.281	0.167	0.406	0.360	0.457	0.460	0.330	0.408
0.526	0.535	0.561	0.531	0.450	0.405	0.427	0.570	0.580	0.448	0.583	0.536	0.656	0.605	0.629	0.460	0.585	0.592
0.526	0.535	0.561	0.531	0.463	0.392	0.427	0.570	0.545	0.458	0.573	0.512	0.635	0.547	0.543	0.460	0.521	0.531
0.500	0.535	0.561	0.531	0.463	0.378	0.417	0.547	0.534	0.458	0.573	0.512	0.635	0.535	0.543	0.460	0.521	0.520
0.321	0.314	0.305	0.359	0.500	0.419	0.385	0.244	0.239	0.323	0.240	0.262	0.260	0.326	0.386	0.460	0.287	0.235
0.231	0.279	0.341	0.141	0.100	0.149	0.240	0.256	0.261	0.219	0.302	0.345	0.354	0.267	0.314	0.460	0.170	0.235
0.269	0.279	0.329	0.344	0.488	0.392	0.344	0.221	0.239	0.323	0.219	0.202	0.271	0.302	0.257	0.460	0.266	0.184
0.385	0.407	0.402	0.266	0.188	0.243	0.406	0.372	0.318	0.333	0.375	0.393	0.396	0.349	0.457	0.460	0.309	0.327
0.385	0.372	0.402	0.516	0.563	0.446	0.344	0.326	0.341	0.406	0.271	0.202	0.344	0.314	0.314	0.460	0.319	0.255
0.462	0.442	0.476	0.563	0.600	0.459	0.385	0.419	0.352	0.448	0.385	0.405	0.406	0.349	0.414	0.460	0.362	0.337
0.372	0.477	0.500	0.406	0.538	0.527	0.427	0.442	0.466	0.354	0.240	0.202	0.333	0.279	0.286	0.460	0.309	0.214
0.295	0.337	0.317	0.203	0.100	0.216	0.271	0.302	0.273	0.229	0.292	0.310	0.292	0.244	0.329	0.460	0.213	0.245
0.385	0.279	0.390	0.438	0.550	0.459	0.365	0.337	0.375	0.427	0.271	0.179	0.385	0.407	0.386	0.460	0.362	0.276
0.205	0.314	0.232	0.156	0.063	0.203	0.229	0.221	0.261	0.156	0.167	0.119	0.219	0.151	0.214	0.460	0.191	0.173
0.474	0.488	0.500	0.375	0.375	0.500	0.354	0.395	0.455	0.438	0.521	0.488	0.333	0.372	0.371	0.460	0.415	0.398
0.474	0.488	0.512	0.375	0.375	0.500	0.354	0.395	0.466	0.438	0.521	0.488	0.333	0.372	0.371	0.460	0.415	0.398
0.333	0.151	0.244	0.266	0.213	0.365	0.229	0.279	0.239	0.271	0.375	0.321	0.219	0.291	0.186	0.460	0.245	0.276
0.474	0.500	0.488	0.375	0.375	0.500	0.354	0.395	0.455	0.438	0.521	0.488	0.323	0.360	0.371	0.460	0.415	0.398
0.128	0.337	0.244	0.109	0.163	0.135	0.125	0.128	0.227	0.177	0.156	0.179	0.094	0.058	0.157	0.460	0.138	0.102

ST Tab 7. Total number of polymorphisms presented in TSI 1000 Genomes Project sample (N polymorphisms) and rates of high accurately imputed (MACH $r^2 > 0.75$) polymorphisms in CIT population sample.

		Minor Allele Frequency spectrum		
		0-0.05	0.05-0.1	0.1-0.5
NOS3 region	N polymorphisms	306	23	179
	N accurately imputed	2	0	108
	% accurately imputed	0.007	0	0.603
NOS1 region	N polymorphisms	776	36	405
	N accurately imputed	13	4	304
	% accurately imputed	0.017	0.111	0.751
NOS2A region	N polymorphisms	1046	135	532
	N accurately imputed	18	47	381
	% accurately imputed	0.017	0.348	0.716

ST Tab 8 - NOS genetic risk score (GRS) distribution for cases and controls and discrimination accuracy for IMPUTE2 imputed dataset. SD: Standard Deviation; Nagelkerke's R^2 : explained interindividual variance of MI by NOS risk score predictive model; AUC: Area Under the receiver operating characteristic (ROC) curve; CI: Confidence Interval.

Case-control sample	Mean GRS \pm SD [min - max]		Nagelkerke's R^2	AUC [95%CI]
	cases	controls		
Finrisk (Finland)	6.10 \pm 0.80 [4.00 - 7.90]	6.04 \pm 0.82 [3.96 - 7.89]	<0.01	0.515 [0.454 - 0.577]
ATVB (Italy)	5.87 \pm 0.69 [3.24 - 7.88]	5.80 \pm 0.71 [3.90 - 7.84]	<0.01	0.527 [0.508 - 0.547]
Regicor (Spain)	5.77 \pm 0.72 [3.92 - 7.49]	5.83 \pm 0.73 [3.95 - 7.80]	<0.01	0.476 [0.431 - 0.521]
MDCS (Sweden)	5.81 \pm 0.71 [4.07 - 7.69]	5.91 \pm 0.68 [3.94 - 7.58]	<0.01	0.444 [0.361 - 0.528]
NDBC (Spain)	5.79 \pm 0.84 [4.00 - 8.00]	5.85 \pm 0.88 [3.00- 8.00]	<0.01	0.469 [0.426 - 0.513]

ST Tab 9 - Mean NOS risk score with standard deviation, minimum and maximum for population samples.

Population	N	NOS risk score
ORK	48	5.88 ± 0.89 [4.00 - 8.00]
GBR	89	6.04 ± 0.78 [4.00 - 8.00]
CEU	85	5.93 ± 0.78 [3.00 - 7.00]
POL	44	6.07 ± 0.90 [4.00 - 8.00]
FIN	93	6.14 ± 0.95 [4.00 - 8.00]
NFR	48	5.81 ± 0.67 [4.00 - 7.00]
SFR	48	5.81 ± 0.70 [5.00 - 8.00]
NCO	33	5.88 ± 0.86 [4.00 - 7.00]
CCO	45	5.42 ± 0.75 [4.00 - 7.00]
BQC	41	5.99 ± 1.03 [3.00 - 8.00]
CAT	43	5.65 ± 0.81 [4.00 - 7.00]
NSP	39	5.77 ± 1.08 [3.00 - 7.00]
SSP	46	6.04 ± 0.84 [4.00 - 7.00]
MEN	49	6.12 ± 0.75 [5.00 - 8.00]
TSI	98	5.86 ± 0.82 [4.00 - 7.00]
CIT	37	5.84 ± 0.96 [4.00 - 8.00]
SIT	39	5.69 ± 0.86 [4.00 - 8.00]
NSR	43	5.49 ± 0.77 [4.00 - 7.00]
SSR	41	5.78 ± 0.88 [4.00 - 7.00]
SIC	32	6.31 ± 0.86 [5.00 - 8.00]
NBH	40	5.90 ± 0.87 [4.00 - 7.00]
SBH	37	5.86 ± 0.75 [4.00 - 7.00]
NGR	48	6.00 ± 0.90 [4.00 - 8.00]
SGR	43	5.88 ± 0.70 [5.00 - 8.00]
CRT	44	5.97 ± 0.78 [4.00 - 7.82]
TRK	48	6.14 ± 1.00 [4.00 - 8.00]
GJD	48	6.02 ± 0.70 [5.00 - 7.02]
BJD	42	6.22 ± 0.81 [5.00 - 8.00]
ASM	48	5.94 ± 1.02 [4.00 - 8.00]
KHM	43	5.98 ± 0.86 [4.00 - 8.00]
BOM	35	5.97 ± 0.92 [4.00 - 8.00]
MZB	50	6.36 ± 0.60 [5.00 - 8.00]
NTS	47	6.19 ± 1.08 [4.00 - 8.00]
STS	49	6.12 ± 0.86 [4.00 - 8.00]

Tab 10 - Present study/MONICA population pairs with coronary event rate and mean levels of traditional risk factors separated by gender. CER: Mean coronary event rates per 100,000 people over 5 years; SBP: Systolic blood pressure (mm Hg); TCH: Total cholesterol (mmol/L); BMI: Body-mass index (Kg/m²); SMK: Daily smoking rate (%).

NOS study populations		MONICA populations		Mean MONICA levels in men					Mean MONICA levels in women				
Population	Code	Population	Code	CER	SBP	TCH	BMI	SMK	CER	SBP	TCH	BMI	SMK
UK – Orcadians	ORK	UK - Glasgow	GLA	744	133	6.1	26.8	41	269	126	6.1	26.9	41
UK – GBR	GBR	UK - Belfast	BEL	593	135	5.9	26.3	29	174	129	5.9	25.6	25
Central Europe - CEU	CEU	Germany - Bremen	BRE	365	132	6.2	26.8	45	92	128	6.2	26.3	30
Poland - Warsaw	POL	Poland - Warsaw	WAR	605	132	5.7	27.1	52	158	128	5.6	27.5	34
Finland - FIN	FIN	Finland – North Karelia	NKA	697	142	6	27.5	27	125	137	5.7	27.1	11
North France – Lille	NFR	France - Lille	LIL	290	135	5.8	26.4	33	63	129	5.8	26.4	17
South France – Toulouse	SFR	France - Toulouse	TOU	226	125	5.8	26.1	24	35	117	5.6	24.5	22
North-West Spain – Girona	CAT	Spain - Catalonia	CAT	218	121	5.6	26.7	41	36	118	5.5	27.4	15
Italy - TSI	TSI	Italy - Area Brianza	BRI	265	131	5.9	26.4	34	39	127	5.9	25.5	23
Central Italy – Lazio	CIT	Italy - Friuli	FRI	248	140	5.9	26.9	29	45	134	5.7	25.8	22
North Bosnia and Herzegovina – Krajina	NBH	Serbia - Novi Sad	NOS	440	136	6.4	27.3	49	116	137	6.2	27.8	30

Apèndix 2. Material suplementari de Carreras-Torres et al., (article III)

Figure S1- Geographic distribution of European and Mediterranean samples. See Supporting Table (ST) Tab 1 for abbreviation codes.



Figure S2- Correlation plots between coronary events rates and F_{ST} genetic distances with Basque Country, Bosnia-Herzegovina, Middle East, and North Africa by incidence group and by gender (A, B, C and D men incidence vs distances; E, F, G and H women incidence vs distances).

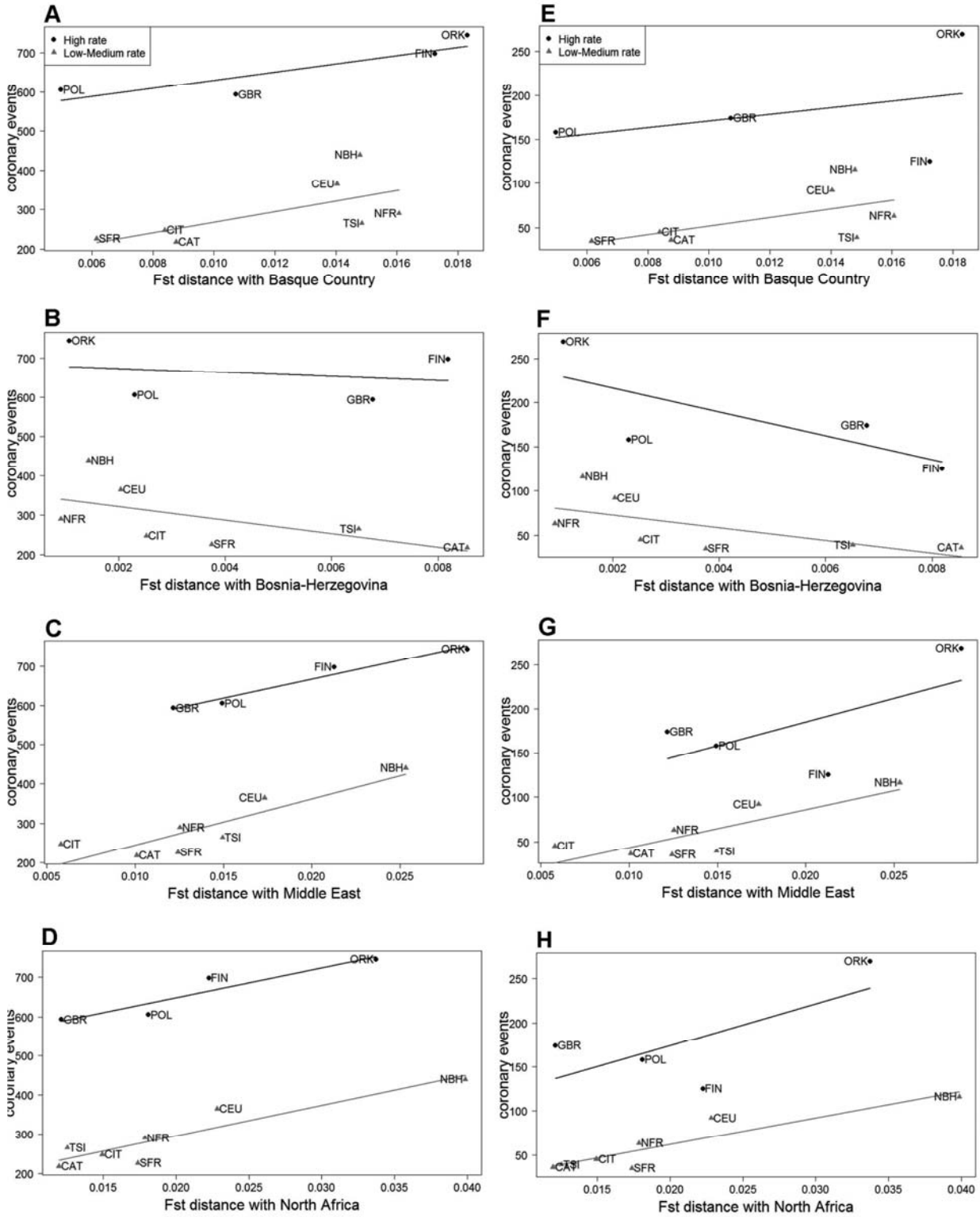


Figure S3- Correlation plots between coronary events rates and SNP's cluster proportion by incidence group and by gender (A, B and C men incidence vs cluster proportion; D, E and F women incidence vs cluster proportion).

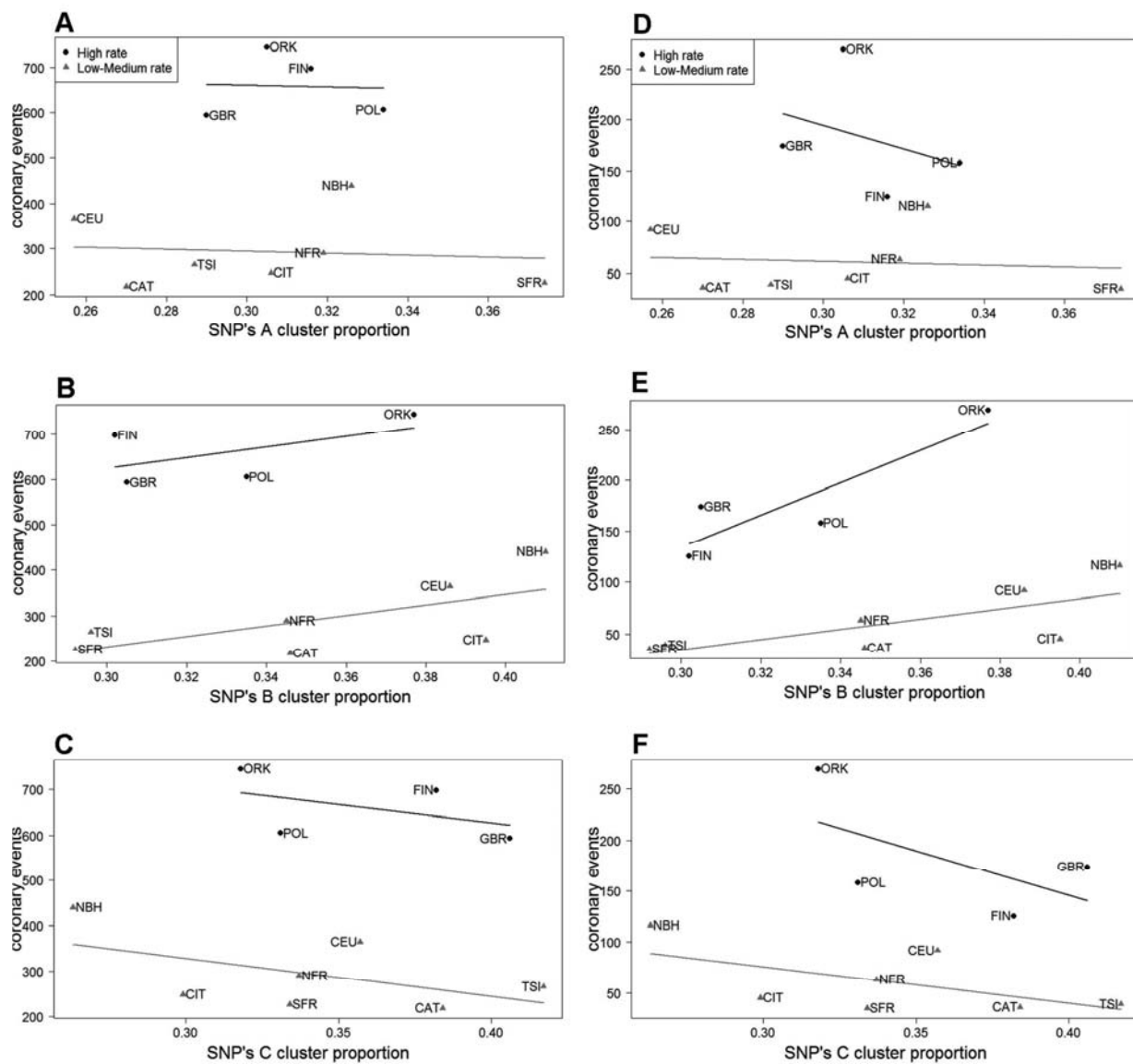


Figure S4- Correlation plots between coronary events rates and R_{ST} genetic distances with Basque Country, Bosnia-Herzegovina, Middle East, and North Africa by incidence group and by gender (A, B, C and D men incidence vs distances; E, F, G and H women incidence vs distances).

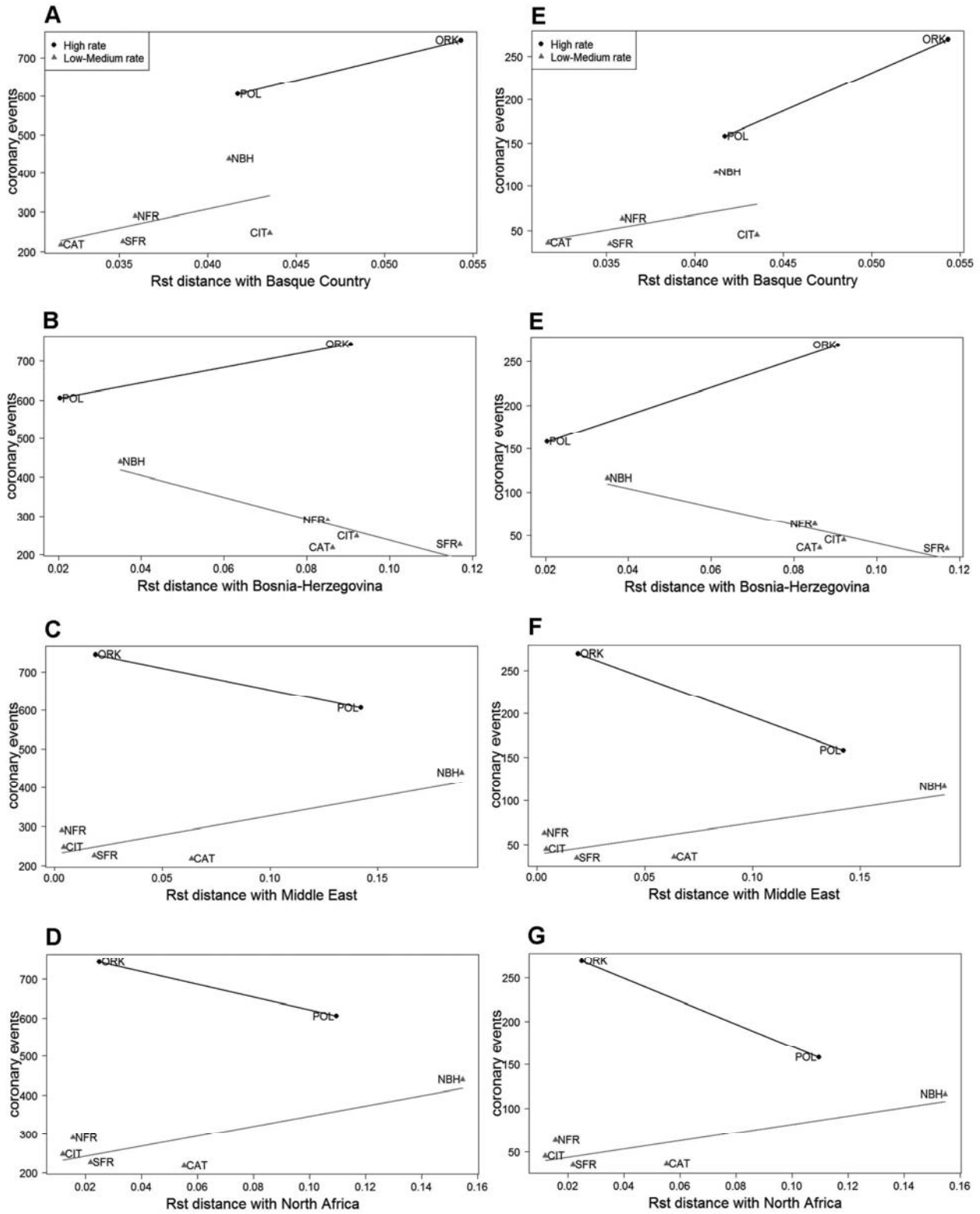
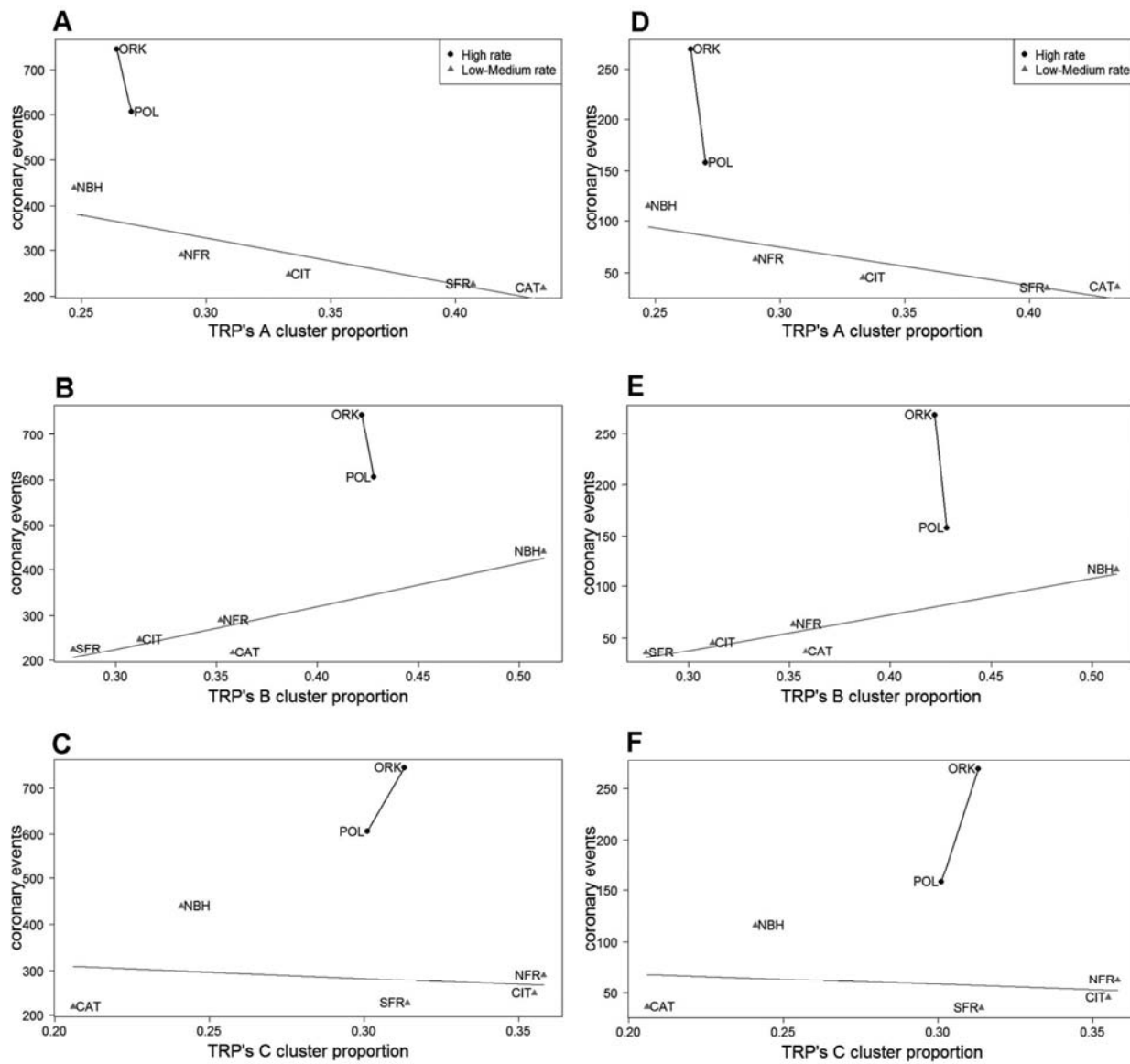


Figure S5- Correlation plots between coronary events rates and TRP's cluster proportion by incidence group and by gender (A, B and C men incidence vs cluster proportion; D, E and F women incidence vs cluster proportion).



Supporting Table (ST) Tab 1- Geographic origin, population codification, sample size and geographic coordinates in decimal degrees for the population samples.

Geographic origin	Country	Population	Code	N	Latitude	Longitude
Europe	United Kingdom	Orcadians	ORK	48	59.00	-3.00
	Poland	Warsaw	POL	44	52.25	21.00
	France	North France - Lille	NFR	48	50.60	3.05
		South France - Toulouse	SFR	48	43.65	1.50
		North Corsica	NCO	33	42.60	9.50
		Central Corsica	CCO	45	41.90	8.60
	Spain	Basque Country -Guipuzcoa	BQC	41	42.80	-2.60
		Catalonia - Girona	CAT	43	41.95	2.75
		East Castilla-Leon - Burgos/Valladolid	NSP	39	42.10	-3.73
		West Andalucia - Seville/Cadiz	SSP	46	37.38	-5.97
		Balearic island - Menorca	MEN	49	40.00	3.80
	Italy	Central Italy - Lazio	CIT	37	43.15	13.40
		South Italy - Naples	SIT	39	40.60	15.78
		North Sardinia	NSR	43	40.33	9.40
		South Sardinia	SSR	41	39.20	9.12
		Sicily	SIC	32	38.20	12.50
	Bosnia-Herzegovina	North Bosnia-Herzegovina - Krajina	NBH	40	44.80	17.20
		South Bosnia-Herzegovina - Adriatic coast	SBH	37	43.40	17.80
	Greece	North Greece - Tesaloniki	NGR	48	40.60	22.98
		South Greece - Athens	SGR	43	37.50	22.45
Crete		CRT	44	35.25	25.00	
Turkey	General population	TRK	48	39.95	32.90	
Middle East	Jordania	General Jordanian	GJD	48	31.65	36.50
		Bedouins of Jordania	BJD	42	30.20	35.75
North Africa	Morocco	High Atlas - Asni	ASM	48	31.15	-8.00
		High Atlas - Khenifra	KHM	43	32.90	-5.60
		Low Atlas - Bouhria	BOM	35	34.60	-3.69
	Algeria	M'zab	MZB	50	28.00	3.00
	Tunisia	North-Central Tunisia	NTZ	47	36.80	10.16
		South Tunisia	STZ	49	33.88	10.03
Egypt	Siwa	SIW	23	29.15	25.50	
South-Sahara	Cotê d'Ivori	Ahizi	AHZ	26	6.12	-3.82
	Nigeria	Bamileke	BAM	25	4.25	9.25
Amerindian	Bolivia	Aymara	AYM	43	-16.63	-68.47
		Quechua	QCH	36	-19.45	-66.20
European 1000 Genomes Project	United Kingdom	GBR	GBR	89	52.22	-0.97
	Central Europe	CEU	CEU	85	52.38	8.14
	Finland	FIN	FIN	93	62.02	24.07
	Italy	TSI	TSI	98	43.20	11.67

ST Tab 2 – Single nucleotide variants chromosome location and genotyping details. Chromosome positions from the NCBI Build 37. GEN: Genotyped.

Genetic variants chromosome information						GoldenGate Genotyping Assay	
ID reference	Chr	Position	Alleles	Coding Region	Gene Region	Genotyped %	Status
rs6951150	7	150681914	C/T	NOS3	5UTR	0.00	Not GEN
rs11771443	7	150687687	C/T	NOS3	5UTR	90.78	GEN
rs1800781	7	150692444	A/G	NOS3	intron	87.55	GEN
rs3793342	7	150695195	A/G	NOS3	intron	90.56	GEN
rs1799983	7	150696111	T/G	NOS3	coding	83.33	GEN
rs3918182	7	150702071	A/G	NOS3	intron	0.00	Not GEN
rs3918188	7	150702781	A/C	NOS3	intron	90.78	GEN
rs753482	7	150706383	T/G	NOS3	intron	89.32	GEN
rs7830	7	150709571	T/G	ATG9B	3UTR	84.93	GEN
rs2373929	7	150714812	T/C	ATG9B	intron	90.34	GEN
rs13307588	7	150720902	A/G	ATG9B	intron	90.99	GEN
rs34459931	7	150721614	-/ACA	ATG9B	5UTR	0.00	Not GEN
rs3763486	7	150723265	A/G	ATG9B	5UTR	0.00	Not GEN
rs219240	7	150978543	T/C	Non-coding		91.03	GEN
rs219243	7	150982510	T/C	Non-coding		90.85	GEN
rs219245	7	150988172	A/G	Non-coding		91.18	GEN
rs10230272	7	150992963	T/C	Non-coding		0.00	Not GEN
rs219269	7	150995229	C/G	Non-coding		90.81	GEN
rs12538729	7	151000447	T/C	Non-coding		91.18	Not polymorphic
rs6964641	7	151006529	A/G	Non-coding		90.81	GEN
rs973531	7	151007330	T/C	Non-coding		89.94	Not polymorphic
rs378140	7	151009517	T/C	Non-coding		90.99	GEN
rs17134612	7	151013785	T/C	Non-coding		0.00	Not GEN
rs367656	7	151014303	A/G	Non-coding		89.58	GEN
rs6977945	7	151019711	A/G	Non-coding		90.99	GEN
rs17705456	7	151024986	T/G	Non-coding		90.81	GEN
rs2792448	7	151027954	A/G	Non-coding		90.70	GEN
rs2487161	7	151033628	A/G	Non-coding		91.18	GEN
rs7972894	12	117011124	A/G	Non-coding		86.96	GEN
rs2393073	12	117015367	T/C	Non-coding		91.03	GEN
rs11834175	12	117021136	T/C	Non-coding		91.10	GEN
rs11068094	12	117026005	A/G	Non-coding		0.00	Not GEN
rs7971771	12	117026246	T/C	Non-coding		0.00	Not GEN
rs12425221	12	117031799	T/G	Non-coding		89.29	Not polymorphic
rs10850680	12	117032036	G/C	Non-coding		90.99	GEN
rs11068102	12	117038152	T/C	Non-coding		91.14	GEN
rs7958610	12	117043340	T/C	Non-coding		90.67	GEN

rs4766797	12	117050587	T/C	Non-coding		91.03	GEN
rs7963154	12	117054530	T/C	Non-coding		90.99	GEN
rs7307369	12	117056637	A/G	Non-coding		0.00	Not GEN
rs10850684	12	117062831	T/C	Non-coding		91.10	GEN
rs12312305	12	117065910	A/G	Non-coding		0.00	Not GEN
rs12301907	12	117070979	T/A	Non-coding		90.81	GEN
rs7307249	12	117077849	A/G	Non-coding		90.92	GEN
rs903820	12	117084285	T/G	Non-coding		91.07	GEN
rs17500613	12	117090452	A/G	Non-coding		90.09	GEN
rs10850694	12	117093237	A/G	Non-coding		88.63	GEN
rs7298460	12	117099811	A/G	Non-coding		89.43	GEN
rs1009981	12	117103935	T/C	Non-coding		91.18	Not polymorphic
rs17500655	12	117112127	G/C	Non-coding		83.19	GEN
rs2393083	12	117118028	A/G	Non-coding		0.00	Not GEN
rs11068140	12	117119675	A/G	Non-coding		90.20	Not polymorphic
rs10507272	12	117123376	A/G	Non-coding		89.91	GEN
rs7974008	12	117127540	A/G	Non-coding		90.96	GEN
rs17428858	12	117133400	A/G	Non-coding		90.52	GEN
rs10774907	12	117647403	A/G	NOS1	3UTR	90.99	GEN
rs2682826	12	117652838	T/C	NOS1	3UTR	90.92	GEN
rs816362	12	117657345	A/G	NOS1	intron	91.21	Not polymorphic
rs816363	12	117660467	G/C	NOS1	intron	91.14	GEN
rs2271986	12	117660873	T/C	NOS1	intron	88.45	GEN
rs904658	12	117665544	A/C	NOS1	intron	85.73	GEN
rs3741475	12	117669914	A/G	NOS1	coding	90.78	GEN
rs13377860	12	117672539	C/G	NOS1	coding	0.00	Not GEN
rs7959232	12	117676698	A/G	NOS1	intron	89.87	GEN
rs9658474	12	117681509	G/C	NOS1	intron	90.99	GEN
rs1047735	12	117685270	T/C	NOS1	coding	81.01	GEN
rs845820	12	117688699	T/C	NOS1	intron	90.89	GEN
rs11068428	12	117693817	A/G	NOS1	intron	91.10	GEN
rs1093330	12	117696534	A/G	NOS1	intron	90.89	GEN
rs2293054	12	117701714	T/C	NOS1	coding	90.56	GEN
rs3741480	12	117705774	T/C	NOS1	intron	91.07	GEN
rs6490121	12	117708195	A/G	NOS1	intron	90.96	GEN
rs1607817	12	117712186	A/C	NOS1	intron	90.63	GEN
rs2293052	12	117715620	T/C	NOS1	intron	91.07	GEN
rs4766842	12	117720130	G/C	NOS1	intron	91.10	GEN
rs9658354	12	117724225	A/T	NOS1	intron	89.94	GEN
rs7309163	12	117729274	T/C	NOS1	intron	90.96	GEN
rs539291	12	117735141	A/G	NOS1	intron	90.56	GEN
rs12811676	12	117741229	A/G	NOS1	intron	91.03	GEN

rs7295972	12	117747368	A/G	NOS1	intron	0.00	Not GEN
rs530393	12	117751258	T/C	NOS1	intron	89.58	GEN
rs2077171	12	117756502	A/G	NOS1	intron	90.74	GEN
rs1483757	12	117761540	A/G	NOS1	intron	91.03	GEN
rs9658281	12	117767578	T/C	NOS1	intron	90.81	GEN
rs570234	12	117770982	A/C	NOS1	intron	90.20	GEN
rs545654	12	117777049	T/C	NOS1	intron	90.81	GEN
rs482555	12	117781391	T/C	NOS1	intron	90.99	GEN
rs693534	12	117784718	A/G	NOS1	intron	91.10	GEN
rs3782219	12	117788240	T/C	NOS1	intron	89.36	GEN
rs4767533	12	117793530	T/C	NOS1	intron	0.00	Not GEN
rs3782221	12	117795881	A/G	NOS1	intron	90.74	GEN
rs9658255	12	117799756	G/C	NOS1	5UTR	90.81	GEN
rs9658253	12	117800626	T/C	NOS1	5UTR	90.89	GEN
rs4767535	12	117804875	T/C	NOS1	5UTR	90.96	GEN
rs816292	12	117811409	T/C	NOS1	5UTR	90.16	GEN
rs492623	12	117813981	T/C	NOS1	5UTR	90.89	GEN
rs1552229	12	117822187	A/G	NOS1	5UTR	88.93	GEN
rs2091	12	117824885	A/C	NOS1	5UTR	89.54	Not polymorphic
rs178672	17	25682796	G/C	Non-coding		91.18	GEN
rs322202	17	25688330	T/C	Non-coding		90.70	GEN
rs322207	17	25691145	T/G	Non-coding		87.76	GEN
rs1695670	17	25698029	T/G	Non-coding		91.07	GEN
rs1676587	17	25701710	A/C	Non-coding		91.03	GEN
rs322193	17	25702438	T/G	Non-coding		90.92	GEN
rs322187	17	25707580	A/G	Non-coding		90.92	GEN
rs4794888	17	25712965	T/C	Non-coding		90.92	GEN
rs2086033	17	25714436	A/G	Non-coding		90.67	GEN
rs9908619	17	25715677	T/C	Non-coding		88.96	GEN
rs8069120	17	25720245	T/C	Non-coding		91.03	GEN
rs4327097	17	25726587	T/C	Non-coding		0.00	Not GEN
rs7215048	17	25732356	T/C	Non-coding		91.18	GEN
rs11652073	17	25739737	T/C	Non-coding		88.45	GEN
rs2012824	17	25744276	A/G	Non-coding		89.91	GEN
rs1963517	17	25750795	T/C	Non-coding		90.78	GEN
rs7220478	17	25756158	T/C	Non-coding		0.00	Not GEN
rs1382046	17	25762715	A/G	Non-coding		87.11	GEN
rs7212692	17	25767504	T/C	Non-coding		0.00	Not GEN
rs8068829	17	25768794	T/G	Non-coding		0.00	Not GEN
rs7216024	17	25773157	T/G	Non-coding		0.00	Not GEN
rs6505276	17	25778498	A/G	Non-coding		91.14	GEN
rs883127	17	25785742	T/C	Non-coding		90.45	GEN

rs6505292	17	25790721	T/G	Non-coding		90.74	GEN
rs1889024	17	26060246	A/G	NOS2A	3UTR	85.08	GEN
rs953527	17	26068742	A/G	NOS2A	3UTR	90.63	GEN
rs4796017	17	26074991	A/G	NOS2A	3UTR	87.76	GEN
rs8081248	17	26081964	A/G	NOS2A	3UTR	89.76	GEN
rs28944201	17	26087190	T/C	NOS2A	coding	90.99	Not polymorphic
rs2255929	17	26087967	T/A	NOS2A	intron	0.00	Not GEN
rs2297514	17	26093315	T/C	NOS2A	intron	0.00	Not GEN
rs2297518	17	26096597	A/G	NOS2A	coding	90.92	GEN
rs2248814	17	26100321	A/G	NOS2A	intron	89.91	GEN
rs4795067	17	26106675	A/G	NOS2A	intron	91.18	GEN
rs3729508	17	26109030	T/C	NOS2A	intron	87.00	GEN
rs6505469	17	26111886	A/T	NOS2A	intron	89.91	GEN
rs8072199	17	26116848	T/C	NOS2A	intron	91.14	GEN
rs16949	17	26124699	T/C	NOS2A	intron	90.23	GEN
rs2779248	17	26127832	A/G	NOS2A	5UTR	0.00	Not GEN
rs2301368	17	26130118	A/G	NOS2A	5UTR	90.70	GEN
rs2779251	17	26131326	A/G	NOS2A	5UTR	85.66	GEN
rs2531862	17	26136991	A/G	NOS2A	5UTR	91.18	GEN
rs10853181	17	26139618	A/G	NOS2A	5UTR	90.85	GEN
rs2531869	17	26144284	A/G	NOS2A	5UTR	90.74	GEN
rs2779259	17	26148284	G/C	NOS2A	5UTR	90.63	GEN
rs16966603	17	26152445	A/T	NOS2A	5UTR	91.10	GEN

ST Tab 3 – Tandem repeat variants chromosome location and repeat motive, and PCR and genotyping details. Chromosome positions from the NCBI Build 37.

ID reference	Chr	Position	Repeat Motive	Coding Region	Gene Region
ss825679095	7	150694394	(GAAGTCTAGACCTGCTGCAGGGGTGAG)n	NOS3	intron
ss825679096	7	150701955	(TGTT)n	NOS3	intron
ss825679092	7	150992977	(CAAAA)n	Non-coding	
ss825679093	7	151007096	(ATTT)	Non-coding	
ss825679094	7	151013614	(AAAAT)n	Non-coding	
ss825679082	12	117056845	(GAAAA)n	Non-coding	
ss825679083	12	117119473	(ATTT)n	Non-coding	
ss825679084	12	117140196	(ATT)n	Non-coding	
ss825679079	12	117825088	(TTCT)n	NOS1	5UTR
ss825679080	12	117681334	(GGAT)n	NOS1	5UTR
ss825679081	12	117672712	(ATA)n	NOS1	intron
ss825679085	17	25682526	(GGAA)n	Non-coding	
ss825679087	17	26137166	(TGGA)n	NOS2A	5UTR
ss825679088	17	26130147	(CCTTT)n	NOS2A	5UTR
ss825679089	17	26128573	(TAAA)n	NOS2A	5UTR
ss825679090	17	26106315	(AAT)n	NOS2A	intron
ss825679091	17	26062652	(AAAAC)n	NOS2A	3UTR

Forward Primer	Reverse Primer	Forward primer labeling	Annealing Temp.	Genotyping rate (%)
CTTAGAGACTGGAAAGGTAGGG	GCTTGCTTCTCTTAGTGCTG	FAM	62	79.55
GAGTTAGTCACTGGGATGGG	GAGAAACATTACCCGTGGCT	HEX	61	80.55
GCTACAAAGAAATACTTGCCAG	CTTGTGAAGACATGCCAGGT	HEX	62	86.54
AAGTGCCAACTGCATATCCC	AGAGTCTACAGTTTCTCACTGG	HEX	62	65.09
AAAGATGCTTAAGGCCGGAC	TCTTACAGTCATCATCCCAGG	FAM	62	71.08
AAACACATCCTCAGTCTCACTC	ATCTAGGCACCTTCATTTCCC	HEX	61	73.76
GCTAGGCAATACCTTTGATGTG	TATATTGGCCCAACACAGCA	FAM	62	85.44
CTGATGATTTCACTCCCTAGCT	CTGCCTGCTGATGAGAAGAG	HEX	61	71.77
GATAGTCATGGTGTGGCAG	TACCAGGAACCTGCAGTCCC	FAM	61	82.65
GGAAGAGTAGATGAGTGGATGG	GTACACAAGGCTTTCAGTGAC	HEX	61	74.29
GTGAGGAGTTCACTGGTGAC	GCAGAGGAATCATAGGACAAGG	HEX	61	79.71
GGCACAAAGTAAGAACCCAG	GACAGGCCTAAATCCTGGAG	HEX	61	68.82
GTCCCAATCAGTCCACATCC	AAATCAGAGCTCAGAACTCAGG	HEX	62	75.18
CAACTGCATTTCGTCTTGTCAC	CATGCCCAGCTACTGTAGTC	FAM	61	75.87
GTGGCTAACCTTGTAATCCT	ACTAACACCTCATCCTTTAGCA	FAM	61	83.12
CAGACAAGGTAGTTCACACCTG	TGCCATCTTGCTCAGCTAAT	FAM	61	87.12
GCTGTCCCAGTTATTCAGGAG	ATGCTCACTTCGTGTCTCTG	FAM	62	84.81

ST Tab 4 - Present study/MONICA population pairs with coronary event rate separated by gender. CER: MONICA mean coronary event rates per 100,000 people over 5 years.

NOS study populations		MONICA populations		Men	Women
Population	Code	Population	Code	CER	CER
UK - Orcadians	ORK	UK - Glasgow	GLA	744	269
UK - GBR	GBR	UK - Belfast	BEL	593	174
Central Europe - CEU	CEU	Germany - Bremen	BRE	365	92
Poland - Warsaw	POL	Poland - Warsaw	WAR	605	158
Finland - FIN	FIN	Finland - North Karelia	NKA	697	125
North France – Lille	NFR	France - Lille	LIL	290	63
South France – Toulouse	SFR	France - Toulouse	TOU	226	35
North-West Spain – Girona	CAT	Spain - Catalonia	CAT	218	36
Italy - TSI	TSI	Italy - Area Brianza	BRI	265	39
Central Italy – Lazio	CIT	Italy - Friuli	FRI	248	45
North Bosnia and Herzegovina – Krajina	NBH	Serbia - Novi Sad	NOS	440	116

ST Tab 5. Minor allele frequencies (MAF) and heterozygosities (mean and standard deviation (SD)) per marker and per population.

Marker information		Populations													
ID reference	MAF Allele	ORK	GBR	CEU	POL	FIN	NFR	SFR	NCO	CCO	BQC	CAT	NSP	SSP	MEN
rs11771443	T	0.125	0.157	0.135	0.159	0.161	0.094	0.146	0.076	0.089	0.073	0.093	0.218	0.152	0.224
rs1800781	A	0.198	0.163	0.147	0.170	0.129	0.135	0.198	0.182	0.089	0.061	0.081	0.141	0.130	0.163
rs3793342	A	0.229	0.157	0.147	0.170	0.129	0.135	0.198	0.182	0.089	0.061	0.081	0.141	0.130	0.163
rs1799983	T	0.302	0.326	0.359	0.239	0.220	0.313	0.396	0.379	0.433	0.293	0.395	0.346	0.359	0.367
rs3918188	A	0.365	0.331	0.341	0.330	0.296	0.396	0.271	0.333	0.356	0.476	0.419	0.410	0.293	0.235
rs753482	G	0.208	0.230	0.218	0.159	0.124	0.281	0.260	0.273	0.356	0.232	0.314	0.256	0.283	0.255
rs7830	T	0.417	0.365	0.382	0.455	0.446	0.271	0.354	0.333	0.189	0.341	0.209	0.282	0.348	0.255
rs2373929	T	0.490	0.455	0.482	0.455	0.538	0.323	0.438	0.394	0.289	0.293	0.267	0.346	0.424	0.337
rs13307588	A	0.052	0.062	0.065	0.080	0.048	0.094	0.094	0.121	0.144	0.085	0.140	0.064	0.065	0.041
rs219240	C	0.271	0.292	0.259	0.318	0.306	0.354	0.385	0.348	0.322	0.268	0.267	0.205	0.304	0.327
rs219243	G	0.260	0.292	0.259	0.250	0.306	0.313	0.313	0.318	0.289	0.220	0.256	0.167	0.261	0.286
rs219245	C	0.271	0.292	0.253	0.307	0.306	0.354	0.365	0.348	0.322	0.268	0.267	0.192	0.326	0.327
rs219269	G	0.271	0.303	0.259	0.330	0.301	0.354	0.344	0.348	0.300	0.268	0.256	0.192	0.326	0.316
rs6964641	G	0.344	0.264	0.300	0.227	0.285	0.281	0.188	0.258	0.278	0.195	0.174	0.167	0.337	0.306
rs378140	T	0.354	0.337	0.347	0.409	0.355	0.469	0.396	0.394	0.322	0.268	0.267	0.256	0.380	0.316
rs367656	T	0.479	0.539	0.447	0.489	0.462	0.406	0.479	0.470	0.489	0.585	0.558	0.667	0.435	0.469
rs6977945	T	0.125	0.140	0.165	0.182	0.156	0.229	0.156	0.182	0.167	0.159	0.140	0.154	0.152	0.153
rs17705456	A	0.146	0.112	0.129	0.102	0.102	0.115	0.063	0.121	0.156	0.134	0.140	0.077	0.120	0.153
rs2792448	T	0.375	0.315	0.365	0.386	0.387	0.427	0.333	0.303	0.389	0.451	0.384	0.346	0.413	0.367
rs2487161	C	0.135	0.169	0.141	0.136	0.145	0.167	0.188	0.152	0.167	0.195	0.198	0.154	0.174	0.153
rs7972894	A	0.490	0.444	0.459	0.409	0.500	0.500	0.406	0.409	0.378	0.183	0.384	0.308	0.391	0.378
rs2393073	T	0.219	0.191	0.159	0.148	0.145	0.177	0.177	0.182	0.189	0.305	0.128	0.205	0.141	0.204
rs11834175	A	0.260	0.236	0.224	0.250	0.161	0.240	0.260	0.197	0.256	0.439	0.209	0.269	0.228	0.265
rs10850680	G	0.510	0.461	0.494	0.318	0.570	0.458	0.427	0.394	0.411	0.195	0.384	0.359	0.359	0.429
rs11068102	C	0.292	0.281	0.276	0.307	0.312	0.281	0.344	0.273	0.333	0.537	0.407	0.308	0.380	0.357
rs7958610	C	0.094	0.101	0.106	0.136	0.086	0.167	0.073	0.045	0.022	0.024	0.070	0.103	0.109	0.082
rs4766797	A	0.594	0.522	0.588	0.477	0.565	0.500	0.427	0.470	0.533	0.378	0.419	0.462	0.446	0.500
rs7963154	A	0.490	0.427	0.453	0.466	0.446	0.500	0.427	0.364	0.356	0.305	0.291	0.423	0.446	0.449
rs10850684	T	0.500	0.500	0.424	0.466	0.500	0.479	0.542	0.424	0.456	0.634	0.640	0.551	0.576	0.510
rs12301907	T	0.448	0.348	0.412	0.330	0.403	0.385	0.302	0.333	0.300	0.171	0.221	0.244	0.293	0.337
rs7307249	G	0.458	0.388	0.424	0.364	0.403	0.385	0.396	0.318	0.278	0.159	0.302	0.295	0.293	0.367
rs903820	G	0.625	0.590	0.612	0.580	0.634	0.531	0.542	0.485	0.467	0.622	0.512	0.577	0.511	0.541
rs17500613	C	0.198	0.163	0.182	0.159	0.113	0.146	0.229	0.136	0.178	0.110	0.151	0.231	0.152	0.235
rs10850694	T	0.438	0.494	0.488	0.523	0.597	0.552	0.552	0.485	0.500	0.537	0.488	0.474	0.511	0.469
rs7298460	T	0.417	0.247	0.318	0.250	0.280	0.323	0.302	0.409	0.378	0.329	0.279	0.346	0.315	0.439
rs17500655	C	0.438	0.601	0.606	0.602	0.634	0.531	0.552	0.470	0.356	0.659	0.523	0.513	0.565	0.510

TSI	CIT	SIT	NSR	SSR	SIC	NBH	SBH	NGR	SGR	CRT	TRK	GJD	BJD	ASM	KHM	BOM	MZB
0.148	0.108	0.167	0.163	0.146	0.250	0.100	0.149	0.156	0.151	0.205	0.250	0.260	0.167	0.146	0.105	0.086	0.050
0.148	0.149	0.167	0.105	0.110	0.141	0.225	0.162	0.146	0.186	0.216	0.063	0.104	0.107	0.146	0.174	0.057	0.100
0.148	0.149	0.167	0.116	0.122	0.141	0.225	0.162	0.135	0.186	0.205	0.146	0.146	0.107	0.156	0.198	0.057	0.140
0.388	0.338	0.372	0.523	0.451	0.219	0.238	0.297	0.281	0.407	0.250	0.208	0.250	0.179	0.260	0.326	0.357	0.350
0.296	0.338	0.295	0.314	0.280	0.344	0.375	0.378	0.396	0.244	0.352	0.323	0.406	0.464	0.313	0.198	0.357	0.240
0.327	0.297	0.308	0.395	0.268	0.156	0.138	0.203	0.167	0.349	0.216	0.146	0.240	0.250	0.240	0.372	0.400	0.460
0.245	0.324	0.295	0.186	0.354	0.375	0.425	0.338	0.313	0.349	0.364	0.260	0.271	0.226	0.354	0.407	0.229	0.320
0.398	0.311	0.346	0.360	0.451	0.344	0.413	0.338	0.406	0.326	0.409	0.385	0.271	0.321	0.458	0.674	0.371	0.450
0.077	0.108	0.103	0.105	0.098	0.016	0.063	0.068	0.031	0.058	0.045	0.042	0.042	0.024	0.021	0.012	0.043	0.000
0.301	0.270	0.333	0.419	0.317	0.234	0.263	0.284	0.281	0.233	0.284	0.365	0.313	0.333	0.344	0.174	0.186	0.230
0.301	0.270	0.308	0.395	0.280	0.219	0.250	0.203	0.250	0.221	0.250	0.344	0.333	0.369	0.344	0.198	0.186	0.330
0.301	0.284	0.333	0.407	0.329	0.234	0.263	0.270	0.271	0.233	0.284	0.375	0.323	0.333	0.365	0.209	0.186	0.290
0.306	0.284	0.321	0.407	0.317	0.250	0.263	0.270	0.302	0.256	0.295	0.406	0.281	0.286	0.323	0.221	0.286	0.220
0.235	0.351	0.282	0.174	0.220	0.188	0.300	0.351	0.260	0.256	0.250	0.302	0.302	0.298	0.313	0.326	0.329	0.230
0.362	0.297	0.372	0.407	0.366	0.375	0.325	0.378	0.354	0.314	0.352	0.365	0.438	0.512	0.365	0.349	0.371	0.260
0.505	0.432	0.513	0.407	0.415	0.547	0.438	0.446	0.469	0.523	0.523	0.417	0.438	0.429	0.500	0.593	0.571	0.610
0.179	0.149	0.141	0.291	0.220	0.203	0.188	0.108	0.188	0.116	0.136	0.198	0.146	0.167	0.135	0.105	0.100	0.100
0.092	0.189	0.077	0.140	0.159	0.047	0.213	0.135	0.167	0.128	0.114	0.135	0.063	0.012	0.115	0.035	0.043	0.020
0.439	0.311	0.410	0.453	0.354	0.375	0.350	0.338	0.292	0.349	0.386	0.229	0.417	0.214	0.396	0.419	0.371	0.440
0.209	0.135	0.128	0.186	0.183	0.188	0.150	0.176	0.219	0.151	0.193	0.250	0.240	0.190	0.188	0.291	0.214	0.220
0.444	0.311	0.385	0.407	0.439	0.500	0.463	0.419	0.406	0.547	0.307	0.323	0.292	0.262	0.354	0.279	0.386	0.220
0.128	0.216	0.128	0.151	0.110	0.125	0.125	0.162	0.115	0.116	0.239	0.167	0.188	0.119	0.156	0.209	0.129	0.240
0.189	0.230	0.179	0.198	0.171	0.156	0.188	0.243	0.198	0.128	0.284	0.229	0.313	0.250	0.281	0.314	0.286	0.350
0.444	0.324	0.397	0.372	0.366	0.406	0.413	0.365	0.375	0.477	0.318	0.323	0.313	0.310	0.365	0.291	0.400	0.350
0.260	0.311	0.321	0.291	0.207	0.281	0.313	0.284	0.385	0.256	0.330	0.396	0.406	0.286	0.292	0.326	0.229	0.340
0.092	0.108	0.115	0.128	0.159	0.125	0.113	0.149	0.094	0.035	0.125	0.042	0.167	0.083	0.094	0.151	0.200	0.240
0.510	0.419	0.410	0.465	0.488	0.500	0.525	0.581	0.427	0.570	0.409	0.438	0.302	0.488	0.448	0.360	0.457	0.350
0.439	0.297	0.436	0.465	0.463	0.484	0.475	0.527	0.406	0.465	0.330	0.365	0.375	0.429	0.385	0.360	0.414	0.400
0.418	0.568	0.449	0.442	0.427	0.422	0.438	0.459	0.510	0.384	0.557	0.500	0.594	0.429	0.396	0.384	0.429	0.480
0.383	0.230	0.308	0.314	0.329	0.344	0.400	0.392	0.323	0.453	0.205	0.260	0.240	0.274	0.344	0.221	0.314	0.260
0.367	0.189	0.256	0.337	0.341	0.344	0.375	0.338	0.260	0.349	0.273	0.250	0.271	0.345	0.354	0.233	0.300	0.220
0.439	0.405	0.423	0.570	0.573	0.469	0.550	0.554	0.344	0.581	0.477	0.417	0.448	0.548	0.490	0.419	0.486	0.500
0.179	0.230	0.256	0.198	0.207	0.156	0.263	0.243	0.146	0.140	0.170	0.167	0.208	0.274	0.125	0.186	0.143	0.050
0.546	0.554	0.487	0.547	0.610	0.609	0.475	0.459	0.594	0.581	0.568	0.594	0.479	0.500	0.583	0.442	0.557	0.590
0.337	0.311	0.385	0.337	0.244	0.281	0.275	0.338	0.271	0.302	0.284	0.208	0.333	0.310	0.281	0.233	0.271	0.350
0.505	0.541	0.333	0.442	0.610	0.516	0.525	0.527	0.573	0.523	0.580	0.667	0.479	0.381	0.469	0.477	0.414	0.490

rs10507272	G	0.563	0.382	0.388	0.386	0.344	0.458	0.417	0.500	0.578	0.366	0.500	0.474	0.402	0.480
rs7974008	G	0.510	0.270	0.335	0.284	0.306	0.396	0.292	0.394	0.456	0.317	0.360	0.385	0.326	0.429
rs17428858	C	0.531	0.315	0.371	0.307	0.339	0.396	0.385	0.409	0.467	0.354	0.465	0.410	0.370	0.500
rs10774907	A	0.313	0.331	0.306	0.318	0.333	0.240	0.219	0.258	0.244	0.220	0.209	0.167	0.250	0.286
rs2682826	T	0.219	0.309	0.306	0.227	0.306	0.250	0.271	0.318	0.467	0.220	0.360	0.269	0.283	0.235
rs816363	G	0.302	0.466	0.465	0.364	0.392	0.396	0.365	0.409	0.556	0.317	0.442	0.397	0.435	0.357
rs2271986	T	0.104	0.118	0.153	0.125	0.081	0.115	0.052	0.061	0.011	0.098	0.023	0.115	0.109	0.071
rs904658	C	0.250	0.393	0.359	0.307	0.258	0.281	0.292	0.318	0.489	0.293	0.360	0.321	0.304	0.265
rs3741475	A	0.135	0.230	0.206	0.182	0.172	0.156	0.219	0.258	0.444	0.195	0.267	0.205	0.185	0.153
rs7959232	G	0.260	0.404	0.359	0.330	0.274	0.302	0.281	0.318	0.489	0.305	0.372	0.295	0.304	0.276
rs9658474	G	0.125	0.174	0.171	0.136	0.086	0.146	0.083	0.106	0.056	0.134	0.128	0.141	0.141	0.122
rs1047735	T	0.292	0.360	0.335	0.307	0.269	0.271	0.260	0.303	0.400	0.317	0.314	0.295	0.283	0.235
rs845820	C	0.344	0.478	0.465	0.364	0.409	0.406	0.375	0.379	0.556	0.329	0.453	0.385	0.424	0.388
rs11068428	A	0.271	0.360	0.341	0.307	0.269	0.271	0.260	0.303	0.400	0.305	0.302	0.295	0.272	0.235
rs1093330	C	0.344	0.478	0.465	0.364	0.409	0.396	0.385	0.379	0.556	0.329	0.453	0.385	0.424	0.388
rs2293054	T	0.240	0.326	0.312	0.239	0.317	0.271	0.250	0.182	0.356	0.195	0.244	0.295	0.293	0.265
rs3741480	C	0.427	0.449	0.488	0.386	0.430	0.521	0.344	0.303	0.556	0.427	0.500	0.449	0.467	0.459
rs6490121	G	0.260	0.354	0.341	0.295	0.344	0.271	0.281	0.258	0.389	0.244	0.326	0.410	0.326	0.286
rs1607817	A	0.354	0.354	0.335	0.409	0.371	0.333	0.448	0.409	0.322	0.488	0.372	0.321	0.370	0.408
rs2293052	T	0.354	0.354	0.335	0.409	0.371	0.333	0.448	0.409	0.322	0.488	0.360	0.321	0.370	0.408
rs4766842	C	0.500	0.438	0.482	0.409	0.414	0.500	0.375	0.409	0.356	0.402	0.384	0.436	0.380	0.378
rs9658354	T	0.500	0.421	0.488	0.591	0.414	0.500	0.625	0.591	0.644	0.598	0.628	0.577	0.620	0.622
rs7309163	T	0.396	0.360	0.341	0.432	0.371	0.323	0.448	0.394	0.311	0.488	0.360	0.321	0.380	0.429
rs539291	A	0.385	0.382	0.353	0.443	0.376	0.344	0.458	0.409	0.322	0.476	0.360	0.346	0.391	0.459
rs12811676	A	0.198	0.101	0.153	0.114	0.108	0.240	0.063	0.106	0.156	0.195	0.140	0.090	0.152	0.112
rs530393	T	0.385	0.382	0.353	0.432	0.371	0.354	0.469	0.409	0.322	0.476	0.360	0.346	0.391	0.459
rs2077171	A	0.313	0.331	0.335	0.273	0.323	0.292	0.292	0.303	0.200	0.244	0.256	0.333	0.250	0.235
rs1483757	G	0.542	0.433	0.494	0.386	0.430	0.510	0.385	0.409	0.378	0.378	0.453	0.462	0.391	0.347
rs9658281	T	0.292	0.315	0.353	0.261	0.301	0.240	0.250	0.258	0.178	0.220	0.291	0.359	0.250	0.194
rs570234	C	0.354	0.388	0.341	0.455	0.366	0.385	0.469	0.424	0.278	0.524	0.326	0.372	0.391	0.418
rs545654	C	0.500	0.478	0.465	0.466	0.565	0.510	0.469	0.515	0.611	0.329	0.547	0.462	0.467	0.449
rs482555	C	0.229	0.180	0.182	0.227	0.269	0.250	0.229	0.167	0.411	0.171	0.279	0.231	0.196	0.184
rs693534	A	0.354	0.399	0.359	0.409	0.328	0.354	0.427	0.455	0.267	0.573	0.349	0.333	0.435	0.500
rs3782219	T	0.167	0.163	0.188	0.227	0.285	0.250	0.229	0.167	0.411	0.171	0.279	0.218	0.196	0.173
rs3782221	A	0.198	0.169	0.218	0.250	0.285	0.281	0.250	0.182	0.433	0.171	0.302	0.231	0.207	0.194
rs9658255	C	0.365	0.416	0.365	0.409	0.328	0.365	0.427	0.439	0.300	0.561	0.349	0.359	0.435	0.490
rs9658253	T	0.208	0.225	0.235	0.193	0.188	0.177	0.167	0.106	0.156	0.195	0.163	0.218	0.130	0.173
rs4767535	C	0.552	0.466	0.535	0.466	0.409	0.438	0.417	0.424	0.467	0.341	0.512	0.449	0.391	0.347
rs816292	T	0.313	0.202	0.229	0.284	0.237	0.344	0.281	0.379	0.333	0.354	0.267	0.256	0.391	0.439
rs492623	T	0.188	0.298	0.253	0.205	0.317	0.208	0.292	0.152	0.211	0.293	0.198	0.269	0.250	0.224
rs1552229	A	0.313	0.326	0.341	0.409	0.263	0.396	0.354	0.485	0.389	0.427	0.419	0.385	0.500	0.480
rs178672	C	0.385	0.309	0.353	0.307	0.317	0.302	0.302	0.333	0.411	0.427	0.384	0.359	0.348	0.388

0.490	0.419	0.615	0.523	0.402	0.484	0.488	0.473	0.417	0.488	0.398	0.333	0.458	0.512	0.458	0.465	0.486	0.470
0.378	0.297	0.487	0.419	0.390	0.313	0.325	0.405	0.313	0.372	0.330	0.271	0.375	0.405	0.406	0.349	0.371	0.440
0.398	0.338	0.500	0.407	0.390	0.344	0.325	0.446	0.354	0.384	0.330	0.281	0.427	0.417	0.448	0.419	0.486	0.520
0.209	0.135	0.231	0.198	0.244	0.125	0.238	0.230	0.188	0.221	0.125	0.135	0.198	0.202	0.333	0.314	0.257	0.350
0.316	0.392	0.372	0.465	0.378	0.375	0.175	0.270	0.302	0.337	0.341	0.313	0.396	0.274	0.396	0.372	0.286	0.240
0.434	0.459	0.513	0.558	0.451	0.438	0.288	0.365	0.406	0.395	0.443	0.406	0.500	0.393	0.583	0.523	0.457	0.390
0.082	0.054	0.090	0.023	0.024	0.031	0.088	0.068	0.115	0.035	0.011	0.063	0.073	0.036	0.031	0.035	0.057	0.070
0.372	0.392	0.449	0.500	0.415	0.328	0.213	0.324	0.365	0.291	0.420	0.229	0.479	0.393	0.563	0.453	0.457	0.320
0.255	0.297	0.321	0.430	0.329	0.281	0.125	0.216	0.250	0.244	0.352	0.281	0.396	0.321	0.375	0.314	0.286	0.220
0.378	0.392	0.462	0.500	0.415	0.328	0.225	0.324	0.354	0.314	0.432	0.365	0.490	0.417	0.573	0.442	0.486	0.390
0.128	0.108	0.141	0.070	0.098	0.078	0.100	0.108	0.125	0.058	0.091	0.115	0.104	0.083	0.219	0.174	0.200	0.180
0.347	0.351	0.397	0.442	0.329	0.328	0.213	0.297	0.333	0.291	0.375	0.365	0.469	0.321	0.344	0.395	0.414	0.300
0.449	0.459	0.487	0.547	0.451	0.453	0.263	0.378	0.427	0.419	0.477	0.396	0.531	0.440	0.667	0.570	0.557	0.510
0.347	0.351	0.397	0.442	0.329	0.328	0.213	0.297	0.333	0.291	0.375	0.333	0.469	0.333	0.333	0.395	0.414	0.320
0.444	0.459	0.500	0.547	0.439	0.453	0.263	0.378	0.406	0.430	0.477	0.396	0.521	0.440	0.656	0.570	0.557	0.510
0.311	0.284	0.397	0.302	0.207	0.266	0.175	0.243	0.229	0.233	0.227	0.260	0.313	0.274	0.323	0.326	0.129	0.270
0.459	0.473	0.577	0.500	0.402	0.359	0.313	0.405	0.375	0.349	0.455	0.417	0.490	0.381	0.573	0.558	0.414	0.470
0.372	0.297	0.462	0.372	0.232	0.313	0.250	0.284	0.271	0.279	0.341	0.271	0.427	0.298	0.365	0.349	0.171	0.290
0.327	0.392	0.321	0.314	0.341	0.375	0.425	0.338	0.406	0.395	0.341	0.354	0.344	0.476	0.406	0.337	0.443	0.330
0.321	0.405	0.321	0.314	0.341	0.375	0.425	0.338	0.385	0.395	0.341	0.344	0.344	0.440	0.333	0.291	0.386	0.310
0.423	0.365	0.346	0.419	0.488	0.375	0.438	0.473	0.438	0.337	0.409	0.479	0.365	0.250	0.240	0.291	0.443	0.290
0.408	0.635	0.654	0.581	0.537	0.625	0.575	0.527	0.583	0.663	0.614	0.510	0.635	0.750	0.760	0.709	0.557	0.700
0.327	0.405	0.321	0.302	0.317	0.375	0.425	0.311	0.406	0.419	0.330	0.333	0.323	0.524	0.448	0.372	0.429	0.410
0.357	0.392	0.321	0.326	0.354	0.375	0.425	0.324	0.385	0.419	0.341	0.323	0.313	0.500	0.365	0.349	0.429	0.430
0.077	0.122	0.115	0.105	0.159	0.078	0.125	0.108	0.135	0.093	0.091	0.188	0.073	0.095	0.156	0.140	0.257	0.040
0.357	0.405	0.321	0.326	0.354	0.359	0.425	0.324	0.396	0.419	0.341	0.323	0.313	0.512	0.354	0.349	0.414	0.380
0.332	0.243	0.256	0.314	0.366	0.344	0.325	0.392	0.313	0.256	0.330	0.292	0.302	0.167	0.104	0.140	0.171	0.200
0.408	0.378	0.372	0.465	0.537	0.422	0.450	0.500	0.417	0.337	0.386	0.448	0.385	0.262	0.260	0.279	0.457	0.270
0.296	0.257	0.231	0.256	0.268	0.313	0.300	0.378	0.333	0.244	0.330	0.260	0.323	0.179	0.094	0.128	0.086	0.180
0.332	0.432	0.295	0.384	0.402	0.406	0.388	0.365	0.354	0.430	0.330	0.354	0.333	0.464	0.313	0.302	0.371	0.340
0.536	0.486	0.551	0.547	0.573	0.469	0.538	0.554	0.458	0.535	0.500	0.573	0.448	0.369	0.490	0.453	0.486	0.470
0.240	0.257	0.321	0.302	0.256	0.172	0.213	0.257	0.167	0.233	0.193	0.271	0.188	0.119	0.292	0.279	0.257	0.130
0.337	0.405	0.333	0.360	0.329	0.375	0.388	0.324	0.427	0.465	0.420	0.333	0.417	0.476	0.344	0.477	0.343	0.360
0.240	0.257	0.321	0.302	0.256	0.141	0.213	0.230	0.188	0.221	0.205	0.229	0.188	0.131	0.302	0.291	0.243	0.090
0.276	0.257	0.372	0.326	0.280	0.125	0.225	0.230	0.219	0.233	0.250	0.292	0.188	0.179	0.323	0.302	0.286	0.100
0.347	0.405	0.333	0.360	0.329	0.406	0.388	0.324	0.427	0.465	0.432	0.354	0.427	0.488	0.344	0.477	0.414	0.410
0.219	0.149	0.154	0.128	0.171	0.125	0.175	0.203	0.167	0.116	0.136	0.167	0.167	0.107	0.167	0.081	0.100	0.200
0.444	0.419	0.526	0.477	0.500	0.438	0.438	0.486	0.479	0.302	0.409	0.458	0.396	0.405	0.521	0.419	0.557	0.480
0.245	0.405	0.256	0.291	0.354	0.219	0.288	0.257	0.240	0.395	0.170	0.198	0.229	0.143	0.313	0.314	0.343	0.290
0.260	0.162	0.295	0.174	0.171	0.281	0.263	0.216	0.302	0.233	0.432	0.323	0.333	0.440	0.208	0.302	0.171	0.190
0.291	0.527	0.321	0.395	0.524	0.313	0.325	0.311	0.313	0.477	0.307	0.240	0.281	0.167	0.406	0.360	0.457	0.330
0.393	0.486	0.397	0.395	0.354	0.391	0.363	0.419	0.292	0.453	0.352	0.375	0.479	0.429	0.521	0.442	0.429	0.460

rs322202	A	0.396	0.320	0.400	0.352	0.290	0.375	0.313	0.212	0.256	0.341	0.349	0.359	0.359	0.286
rs322207	T	0.385	0.320	0.394	0.341	0.290	0.354	0.323	0.227	0.256	0.317	0.337	0.372	0.359	0.286
rs1695670	G	0.396	0.331	0.406	0.341	0.296	0.375	0.333	0.242	0.256	0.305	0.360	0.385	0.359	0.296
rs1676587	G	0.396	0.331	0.406	0.341	0.296	0.375	0.333	0.242	0.256	0.305	0.372	0.385	0.359	0.296
rs322193	G	0.396	0.331	0.406	0.341	0.296	0.375	0.333	0.242	0.256	0.305	0.360	0.385	0.359	0.296
rs322187	A	0.396	0.326	0.400	0.341	0.296	0.375	0.333	0.242	0.244	0.305	0.349	0.385	0.359	0.286
rs4794888	G	0.448	0.393	0.482	0.398	0.398	0.427	0.396	0.394	0.322	0.451	0.453	0.474	0.413	0.388
rs2086033	C	0.396	0.326	0.400	0.341	0.296	0.375	0.333	0.242	0.244	0.305	0.349	0.385	0.359	0.286
rs9908619	A	0.396	0.326	0.400	0.341	0.296	0.375	0.333	0.242	0.244	0.305	0.349	0.385	0.359	0.286
rs8069120	A	0.052	0.067	0.082	0.057	0.102	0.052	0.063	0.152	0.078	0.146	0.105	0.090	0.054	0.102
rs7215048	T	0.469	0.399	0.488	0.398	0.392	0.427	0.396	0.394	0.322	0.451	0.453	0.474	0.424	0.388
rs11652073	G	0.479	0.404	0.494	0.398	0.387	0.427	0.396	0.394	0.333	0.439	0.465	0.487	0.424	0.388
rs2012824	A	0.479	0.404	0.494	0.409	0.387	0.427	0.396	0.394	0.333	0.439	0.465	0.487	0.424	0.388
rs1963517	T	0.406	0.326	0.406	0.364	0.323	0.354	0.333	0.227	0.256	0.329	0.349	0.372	0.337	0.306
rs1382046	A	0.469	0.416	0.488	0.477	0.452	0.417	0.365	0.439	0.311	0.463	0.430	0.526	0.457	0.378
rs6505276	A	0.542	0.545	0.488	0.477	0.409	0.531	0.469	0.485	0.578	0.500	0.512	0.423	0.457	0.439
rs883127	T	0.354	0.292	0.341	0.330	0.398	0.344	0.313	0.258	0.244	0.341	0.291	0.410	0.380	0.398
rs6505292	T	0.469	0.421	0.447	0.443	0.516	0.500	0.417	0.348	0.400	0.427	0.349	0.474	0.478	0.500
rs1889024	G	0.292	0.483	0.394	0.511	0.511	0.469	0.417	0.545	0.467	0.366	0.465	0.474	0.435	0.469
rs953527	A	0.292	0.455	0.388	0.489	0.489	0.469	0.406	0.530	0.422	0.390	0.442	0.474	0.424	0.459
rs4796017	G	0.292	0.449	0.382	0.489	0.489	0.458	0.406	0.530	0.422	0.390	0.453	0.474	0.424	0.449
rs8081248	A	0.542	0.399	0.482	0.432	0.409	0.333	0.406	0.348	0.378	0.451	0.384	0.436	0.457	0.459
rs2297518	A	0.083	0.247	0.182	0.295	0.269	0.177	0.188	0.258	0.256	0.232	0.186	0.269	0.217	0.224
rs2248814	A	0.479	0.376	0.459	0.398	0.392	0.302	0.417	0.348	0.344	0.476	0.349	0.436	0.435	0.367
rs4795067	G	0.271	0.393	0.347	0.432	0.376	0.427	0.313	0.394	0.389	0.329	0.337	0.359	0.315	0.337
rs3729508	T	0.490	0.371	0.435	0.432	0.382	0.344	0.417	0.333	0.333	0.500	0.384	0.449	0.435	0.429
rs6505469	T	0.521	0.466	0.476	0.489	0.441	0.396	0.490	0.394	0.444	0.561	0.500	0.526	0.500	0.490
rs8072199	T	0.479	0.427	0.500	0.602	0.484	0.375	0.396	0.470	0.289	0.549	0.267	0.474	0.391	0.439
rs16949	C	0.188	0.270	0.218	0.318	0.285	0.260	0.219	0.333	0.256	0.232	0.221	0.256	0.217	0.255
rs2301368	G	0.510	0.399	0.447	0.409	0.414	0.313	0.427	0.318	0.367	0.512	0.395	0.436	0.391	0.459
rs2779251	A	0.146	0.169	0.182	0.216	0.194	0.219	0.156	0.258	0.133	0.171	0.093	0.205	0.141	0.173
rs2531862	A	0.344	0.360	0.347	0.386	0.414	0.396	0.448	0.394	0.467	0.341	0.349	0.385	0.370	0.429
rs10853181	G	0.344	0.360	0.347	0.386	0.414	0.396	0.448	0.394	0.467	0.341	0.360	0.385	0.370	0.429
rs2531869	A	0.229	0.213	0.206	0.159	0.183	0.188	0.240	0.212	0.367	0.159	0.174	0.179	0.174	0.255
rs2779259	C	0.354	0.365	0.335	0.386	0.409	0.396	0.448	0.394	0.467	0.341	0.372	0.385	0.370	0.418
rs16966603	A	0.125	0.146	0.135	0.227	0.226	0.208	0.208	0.182	0.100	0.183	0.186	0.205	0.174	0.163
Heterozygosities per population	Mean	0.418	0.417	0.421	0.422	0.410	0.424	0.418	0.413	0.411	0.405	0.412	0.419	0.420	0.417
	SD	0.100	0.099	0.099	0.092	0.104	0.090	0.100	0.100	0.107	0.103	0.104	0.097	0.097	0.097

0.301	0.432	0.333	0.279	0.244	0.359	0.438	0.459	0.375	0.337	0.352	0.385	0.417	0.417	0.177	0.244	0.171	0.260
0.291	0.432	0.321	0.291	0.244	0.375	0.450	0.473	0.375	0.337	0.341	0.375	0.417	0.417	0.177	0.244	0.214	0.270
0.301	0.419	0.333	0.326	0.280	0.438	0.450	0.459	0.375	0.337	0.341	0.385	0.438	0.440	0.188	0.291	0.257	0.270
0.301	0.419	0.333	0.326	0.280	0.438	0.450	0.459	0.375	0.337	0.352	0.385	0.438	0.452	0.208	0.326	0.271	0.270
0.301	0.419	0.333	0.326	0.280	0.438	0.450	0.459	0.375	0.337	0.341	0.385	0.438	0.440	0.188	0.291	0.257	0.270
0.301	0.419	0.333	0.326	0.280	0.422	0.450	0.459	0.375	0.337	0.341	0.385	0.438	0.429	0.167	0.279	0.229	0.280
0.362	0.446	0.385	0.407	0.366	0.484	0.513	0.486	0.417	0.407	0.432	0.438	0.521	0.500	0.260	0.407	0.343	0.410
0.301	0.419	0.333	0.326	0.280	0.422	0.450	0.459	0.375	0.337	0.341	0.385	0.438	0.429	0.167	0.279	0.229	0.270
0.301	0.419	0.333	0.326	0.280	0.422	0.450	0.459	0.375	0.337	0.341	0.385	0.438	0.429	0.167	0.279	0.229	0.270
0.066	0.027	0.051	0.081	0.085	0.063	0.063	0.027	0.042	0.070	0.091	0.052	0.083	0.071	0.094	0.128	0.114	0.140
0.378	0.446	0.385	0.395	0.366	0.500	0.513	0.486	0.396	0.407	0.432	0.438	0.510	0.488	0.271	0.419	0.343	0.410
0.378	0.446	0.372	0.384	0.378	0.484	0.513	0.473	0.365	0.430	0.420	0.438	0.500	0.476	0.281	0.419	0.343	0.410
0.378	0.446	0.385	0.372	0.378	0.484	0.513	0.473	0.365	0.430	0.420	0.427	0.500	0.476	0.281	0.384	0.343	0.410
0.306	0.432	0.308	0.256	0.329	0.375	0.450	0.446	0.396	0.326	0.295	0.323	0.385	0.298	0.188	0.291	0.229	0.280
0.388	0.473	0.410	0.430	0.378	0.422	0.513	0.473	0.438	0.430	0.375	0.385	0.448	0.464	0.333	0.337	0.300	0.270
0.505	0.392	0.487	0.523	0.561	0.469	0.413	0.405	0.469	0.512	0.466	0.542	0.458	0.321	0.604	0.558	0.471	0.560
0.332	0.500	0.346	0.233	0.256	0.391	0.425	0.473	0.406	0.337	0.330	0.344	0.344	0.405	0.188	0.314	0.357	0.270
0.459	0.541	0.449	0.326	0.317	0.469	0.550	0.649	0.500	0.442	0.455	0.458	0.438	0.536	0.229	0.372	0.386	0.350
0.439	0.473	0.526	0.535	0.561	0.531	0.450	0.405	0.427	0.570	0.580	0.448	0.583	0.536	0.656	0.605	0.629	0.700
0.423	0.459	0.526	0.535	0.561	0.531	0.463	0.392	0.427	0.570	0.545	0.458	0.573	0.512	0.635	0.547	0.543	0.700
0.429	0.432	0.500	0.535	0.561	0.531	0.463	0.378	0.417	0.547	0.534	0.458	0.573	0.512	0.635	0.535	0.543	0.700
0.439	0.324	0.321	0.314	0.305	0.359	0.500	0.419	0.385	0.244	0.239	0.323	0.240	0.262	0.260	0.326	0.386	0.200
0.209	0.243	0.231	0.279	0.341	0.141	0.100	0.149	0.240	0.256	0.261	0.219	0.302	0.345	0.354	0.267	0.314	0.240
0.393	0.324	0.269	0.279	0.329	0.344	0.488	0.392	0.344	0.221	0.239	0.323	0.219	0.202	0.271	0.302	0.257	0.190
0.321	0.284	0.385	0.407	0.402	0.266	0.188	0.243	0.406	0.372	0.318	0.333	0.375	0.393	0.396	0.349	0.457	0.210
0.444	0.378	0.385	0.372	0.402	0.516	0.563	0.446	0.344	0.326	0.341	0.406	0.271	0.202	0.344	0.314	0.314	0.330
0.495	0.432	0.462	0.442	0.476	0.563	0.600	0.459	0.385	0.419	0.352	0.448	0.385	0.405	0.406	0.349	0.414	0.410
0.434	0.419	0.372	0.477	0.500	0.406	0.538	0.527	0.427	0.442	0.466	0.354	0.240	0.202	0.333	0.279	0.286	0.220
0.245	0.243	0.295	0.337	0.317	0.203	0.100	0.216	0.271	0.302	0.273	0.229	0.292	0.310	0.292	0.244	0.329	0.180
0.464	0.351	0.385	0.279	0.390	0.438	0.550	0.459	0.365	0.337	0.375	0.427	0.271	0.179	0.385	0.407	0.386	0.300
0.194	0.176	0.205	0.314	0.232	0.156	0.063	0.203	0.229	0.221	0.261	0.156	0.167	0.119	0.219	0.151	0.214	0.070
0.418	0.378	0.474	0.488	0.500	0.375	0.375	0.500	0.354	0.395	0.455	0.438	0.521	0.488	0.333	0.372	0.371	0.360
0.418	0.378	0.474	0.488	0.512	0.375	0.375	0.500	0.354	0.395	0.466	0.438	0.521	0.488	0.333	0.372	0.371	0.370
0.224	0.243	0.333	0.151	0.244	0.266	0.213	0.365	0.229	0.279	0.239	0.271	0.375	0.321	0.219	0.291	0.186	0.290
0.423	0.378	0.474	0.500	0.488	0.375	0.375	0.500	0.354	0.395	0.455	0.438	0.521	0.488	0.323	0.360	0.371	0.370
0.199	0.122	0.128	0.337	0.244	0.109	0.163	0.135	0.125	0.128	0.227	0.177	0.156	0.179	0.094	0.058	0.157	0.090
0.416	0.421	0.425	0.425	0.420	0.412	0.412	0.424	0.418	0.412	0.418	0.415	0.424	0.406	0.400	0.409	0.406	0.391
0.100	0.101	0.094	0.103	0.094	0.115	0.109	0.099	0.096	0.106	0.096	0.099	0.099	0.119	0.102	0.103	0.111	0.114

Marker information									Heterozygosity per marker	
ID reference	MAF Allele	NTZ	STZ	SIW	AHZ	BAM	AYM	QCH	Mean	SD
rs11771443	T	0.149	0.143	0.174	0.077	0.100	0.581	0.625	0.253	0.086
rs1800781	A	0.085	0.143	.	0.000	0.000	0.012	0.014	0.240	0.090
rs3793342	A	0.085	0.153	0.043	0.077	0.120	0.012	0.028	0.250	0.074
rs1799983	T	0.255	0.224	.	0.077	0.060	0.023	0.042	0.437	0.121
rs3918188	A	0.319	0.367	0.239	0.462	0.360	0.081	0.042	0.445	0.078
rs753482	G	0.266	0.235	0.239	0.269	0.280	0.012	0.014	0.384	0.109
rs7830	T	0.287	0.296	.	0.173	0.160	0.791	0.764	0.439	0.082
rs2373929	T	0.479	0.357	0.370	0.077	0.160	0.791	0.764	0.485	0.065
rs13307588	A	0.053	0.020	0.000	0.000	0.000	0.012	0.000	0.119	0.067
rs219240	C	0.287	0.265	0.196	0.154	0.040	0.616	0.653	0.415	0.071
rs219243	G	0.330	0.286	0.261	0.423	0.400	0.616	0.653	0.405	0.055
rs219245	C	0.298	0.296	0.239	0.231	0.200	0.605	0.639	0.418	0.048
rs219269	G	0.287	0.286	0.217	0.154	0.100	0.605	0.639	0.417	0.057
rs6964641	G	0.362	0.367	0.326	0.423	0.540	0.500	0.417	0.398	0.064
rs378140	T	0.394	0.449	0.457	0.385	0.300	0.488	0.375	0.462	0.037
rs367656	T	0.457	0.398	0.370	0.404	0.320	0.209	0.292	0.500	0.029
rs6977945	T	0.170	0.133	0.152	0.115	0.060	0.279	0.264	0.269	0.065
rs17705456	A	0.085	0.092	0.130	0.000	0.000	0.000	0.000	0.196	0.087
rs2792448	T	0.383	0.408	0.435	0.135	0.300	0.012	0.000	0.467	0.102
rs2487161	C	0.266	0.276	0.348	0.519	0.680	0.977	0.903	0.303	0.081
rs7972894	A	0.234	0.296	0.152	0.000	0.020	0.105	0.222	0.479	0.109
rs2393073	T	0.245	0.163	0.152	0.115	0.220	0.012	0.014	0.282	0.077
rs11834175	A	0.372	0.296	0.370	0.615	0.680	0.860	0.736	0.369	0.066
rs10850680	G	0.245	0.296	0.174	0.135	0.220	0.140	0.222	0.481	0.064
rs11068102	C	0.426	0.378	0.217	0.327	0.540	0.047	0.042	0.436	0.080
rs7958610	C	0.117	0.173	0.326	0.519	0.300	0.791	0.722	0.194	0.100
rs4766797	A	0.340	0.306	0.304	0.077	0.180	0.174	0.236	0.502	0.066
rs7963154	A	0.223	0.347	0.348	0.423	0.340	0.942	0.958	0.486	0.082
rs10850684	T	0.585	0.582	0.522	0.712	0.640	0.035	0.028	0.500	0.087
rs12301907	T	0.181	0.214	0.087	0.058	0.100	0.942	0.931	0.436	0.100
rs7307249	G	0.266	0.224	0.109	0.019	0.140	0.221	0.236	0.441	0.085
rs903820	G	0.447	0.408	0.435	0.058	0.260	0.233	0.319	0.502	0.060
rs17500613	C	0.202	0.194	0.130	0.115	0.040	0.000	0.000	0.292	0.093
rs10850694	T	0.574	0.439	0.283	0.212	0.180	0.163	0.292	0.499	0.049
rs7298460	T	0.351	0.357	0.174	0.077	0.140	0.477	0.431	0.430	0.069
rs17500655	C	0.436	0.480	0.000	0.231	0.020	0.163	0.292	0.501	0.102
rs10507272	G	0.543	0.469	0.674	0.769	0.880	0.442	0.361	0.496	0.045
rs7974008	G	0.468	0.439	0.652	0.712	0.800	0.442	0.361	0.465	0.037
rs17428858	C	0.532	0.469	0.717	0.846	0.940	0.779	0.597	0.482	0.070

rs10774907	A	0.362	0.276	0.739	1.000	1.000	0.721	0.625	0.379	0.100
rs2682826	T	0.457	0.378	0.304	0.154	0.140	0.116	0.125	0.435	0.072
rs816363	G	0.553	0.480	0.522	0.712	0.720	0.791	0.736	0.492	0.035
rs2271986	T	0.032	0.031	0.065	0.019	0.020	0.663	0.611	0.134	0.092
rs904658	C	0.489	0.429	.	0.615	0.620	0.791	0.736	0.464	0.079
rs3741475	A	0.394	0.337	0.348	0.154	0.300	0.116	0.125	0.388	0.083
rs7959232	G	0.553	0.449	0.565	0.769	0.780	0.791	0.736	0.473	0.049
rs9658474	G	0.149	0.133	0.217	0.673	0.580	0.674	0.625	0.221	0.094
rs1047735	T	0.415	0.327	.	0.154	0.200	0.791	0.736	0.444	0.079
rs845820	C	0.628	0.490	0.630	0.904	0.840	0.802	0.736	0.499	0.062
rs11068428	A	0.415	0.327	0.391	0.154	0.220	0.791	0.736	0.443	0.052
rs1093330	C	0.628	0.490	0.630	0.942	0.880	0.814	0.736	0.499	0.073
rs2293054	T	0.351	0.337	0.283	0.173	0.080	0.791	0.736	0.402	0.064
rs3741480	C	0.649	0.449	0.630	0.731	0.740	0.802	0.736	0.496	0.038
rs6490121	G	0.426	0.357	0.283	0.481	0.340	0.756	0.722	0.437	0.043
rs1607817	A	0.266	0.459	0.304	0.442	0.460	0.221	0.250	0.466	0.036
rs2293052	T	0.223	0.378	0.283	0.000	0.020	0.221	0.250	0.460	0.092
rs4766842	C	0.319	0.194	0.283	0.269	0.240	0.663	0.653	0.481	0.043
rs9658354	T	0.681	0.806	0.717	0.692	0.740	0.337	0.347	0.495	0.041
rs7309163	T	0.319	0.520	0.500	0.519	0.600	0.221	0.250	0.471	0.036
rs539291	A	0.277	0.459	0.478	0.558	0.400	0.221	0.236	0.471	0.036
rs12811676	A	0.191	0.020	0.087	0.077	0.080	0.244	0.264	0.217	0.080
rs530393	T	0.266	0.439	0.478	0.538	0.440	0.221	0.236	0.470	0.035
rs2077171	A	0.128	0.173	0.109	0.096	0.120	0.442	0.361	0.401	0.082
rs1483757	G	0.340	0.235	0.261	0.250	0.320	0.651	0.611	0.484	0.039
rs9658281	T	0.128	0.153	0.087	0.192	0.100	0.477	0.431	0.385	0.085
rs570234	C	0.298	0.388	0.152	0.096	0.040	0.233	0.222	0.468	0.079
rs545654	C	0.543	0.408	0.370	0.519	0.560	0.349	0.333	0.501	0.014
rs482555	C	0.266	0.214	0.043	0.000	0.040	0.081	0.042	0.354	0.104
rs693534	A	0.383	0.469	0.348	0.192	0.220	0.442	0.500	0.475	0.039
rs3782219	T	0.234	0.194	0.000	0.000	0.040	0.081	0.042	0.347	0.109
rs3782221	A	0.277	0.224	0.152	0.154	0.300	0.081	0.056	0.370	0.084
rs9658255	C	0.394	0.469	0.391	0.365	0.180	0.419	0.500	0.499	0.035
rs9658253	T	0.117	0.102	0.130	0.000	0.020	0.488	0.431	0.281	0.090
rs4767535	C	0.511	0.398	0.543	0.385	0.480	0.547	0.472	0.495	0.017
rs816292	T	0.223	0.327	0.065	0.000	0.000	0.058	0.083	0.405	0.118
rs492623	T	0.202	0.306	0.304	0.327	0.400	0.581	0.708	0.384	0.070
rs1552229	A	0.330	0.408	0.065	0.000	0.040	0.058	0.069	0.463	0.122
rs178672	C	0.489	0.357	0.457	0.212	0.340	0.430	0.597	0.475	0.033
rs322202	A	0.234	0.214	0.348	0.135	0.300	0.419	0.625	0.441	0.059
rs322207	T	0.223	0.204	0.348	0.154	0.300	0.419	0.625	0.440	0.055
rs1695670	G	0.277	0.224	0.457	0.192	0.340	0.419	0.625	0.448	0.049

rs1676587	G	0.277	0.255	0.457	0.231	0.420	0.419	0.625	0.450	0.045
rs322193	G	0.277	0.224	0.457	0.192	0.340	0.419	0.625	0.448	0.049
rs322187	A	0.266	0.204	0.413	0.173	0.300	0.419	0.625	0.446	0.053
rs4794888	G	0.330	0.347	0.522	0.365	0.600	0.605	0.764	0.486	0.029
rs2086033	C	0.266	0.204	0.413	0.173	0.300	0.419	0.625	0.446	0.053
rs9908619	A	0.266	0.204	0.413	0.173	0.300	0.419	0.625	0.446	0.053
rs8069120	A	0.064	0.143	0.109	0.192	0.300	0.186	0.139	0.149	0.074
rs7215048	T	0.351	0.347	0.522	0.365	0.600	0.605	0.764	0.486	0.028
rs11652073	G	0.383	0.357	0.500	0.385	0.600	0.605	0.764	0.486	0.027
rs2012824	A	0.383	0.367	0.500	0.385	0.600	0.605	0.764	0.486	0.026
rs1963517	T	0.298	0.194	0.391	0.212	0.140	0.384	0.597	0.441	0.056
rs1382046	A	0.383	0.316	0.543	0.308	0.280	0.558	0.736	0.487	0.030
rs6505276	A	0.468	0.571	0.326	0.481	0.360	0.372	0.181	0.501	0.031
rs883127	T	0.372	0.173	0.500	0.135	0.200	0.372	0.611	0.447	0.058
rs6505292	T	0.511	0.296	0.543	0.231	0.340	0.349	0.611	0.494	0.034
rs1889024	G	0.585	0.592	0.630	0.923	0.900	0.628	0.472	0.505	0.067
rs953527	A	0.521	0.531	0.478	0.692	0.680	0.570	0.444	0.502	0.025
rs4796017	G	0.521	0.520	0.478	0.692	0.680	0.570	0.444	0.502	0.025
rs8081248	A	0.287	0.235	0.152	0.077	0.100	0.372	0.514	0.469	0.079
rs2297518	A	0.170	0.235	0.174	0.058	0.080	0.128	0.056	0.359	0.086
rs2248814	A	0.266	0.184	0.109	0.077	0.100	0.360	0.500	0.454	0.085
rs4795067	G	0.309	0.327	0.283	0.135	0.160	0.395	0.458	0.454	0.057
rs3729508	T	0.319	0.255	0.109	0.212	0.120	0.081	0.069	0.476	0.090
rs6505469	T	0.362	0.337	0.152	0.173	0.100	0.151	0.111	0.497	0.079
rs8072199	T	0.309	0.214	0.196	0.135	0.140	0.081	0.056	0.489	0.094
rs16949	C	0.213	0.245	0.217	0.154	0.140	0.093	0.056	0.379	0.075
rs2301368	G	0.362	0.276	0.196	0.462	0.260	0.093	0.069	0.479	0.077
rs2779251	A	0.191	0.173	0.152	0.058	0.100	0.012	0.014	0.297	0.089
rs2531862	A	0.415	0.398	0.674	0.654	0.640	0.128	0.083	0.483	0.062
rs10853181	G	0.415	0.398	0.674	0.654	0.640	0.128	0.083	0.483	0.062
rs2531869	A	0.245	0.276	0.543	0.481	0.380	0.047	0.028	0.363	0.090
rs2779259	C	0.415	0.398	0.652	0.635	0.640	0.128	0.083	0.483	0.062
rs16966603	A	0.138	0.102	0.087	0.038	0.020	0.081	0.056	0.276	0.083
Heterozygosity	Mean	0.408	0.398	0.361	0.297	0.313	0.329	0.333		
per population	SD	0.100	0.105	0.155	0.162	0.160	0.161	0.165		

ST Tab 6. Tandem repeats frequencies and number of alleles for tested population samples, and heterozygosities (mean and standard deviation (SD)) per marker and per population. N rep: number of repeats. N al: number of alleles.

Marker information		Population												
ID reference	N rep	ORK	POL	NFR	SFR	NCO	CCO	BQC	CAT	NSP	SSP	MEN	CIT	SIT
ss825679095	4	0.208	0.208	0.177	0.181	0.197	0.111	0.077	0.048	0.152	0.129	0.149	0.176	0.149
	5	0.792	0.792	0.823	0.819	0.803	0.889	0.923	0.952	0.848	0.871	0.851	0.824	0.851
	N al	2	2	2	2	2	2	2	2	2	2	2	2	2
ss825679096	3
	4
	5	0.350	0.329	0.427	0.274	0.348	0.364	0.500	0.446	0.411	0.300	0.293	0.355	0.333
	6	0.650	0.671	0.573	0.714	0.652	0.636	0.500	0.554	0.589	0.683	0.707	0.645	0.667
	7	.	.	.	0.012	0.017	.	.	.
	8
	9
	N al	2	2	2	3	2	2	2	2	2	2	3	2	2
ss825679092	3	0.012
	4	.	0.013
	5	0.723	0.679	0.646	0.619	0.652	0.700	0.500	0.726	0.779	0.682	0.717	0.722	0.662
	6	0.015	0.011	0.033	.	.
	7
	8	0.245	0.282	0.281	0.286	0.288	0.211	0.431	0.226	0.132	0.284	0.217	0.208	0.297
	9	0.021	0.013	0.052	0.095	0.045	0.033	0.069	0.012	0.074	0.023	0.033	0.056	0.027
	10	0.011	0.013	0.021	.	0.015	0.056	.	0.024	.	.	.	0.014	0.014
	13
	N al	4	5	4	3	4	4	3	5	4	4	4	4	4
ss825679093	2
	5
	7
	8	0.319	0.145	0.250	0.128	0.233	0.189	0.250	.	.	.	0.202	0.281	0.205
	9	0.011	.	0.017	0.026
	10	0.096	0.161	0.100	0.103	0.133	0.044	0.025	.	.	.	0.107	0.063	0.128
	11	0.021	0.016	0.067	0.051	0.050	0.011	0.025	.	.	.	0.071	0.094	0.115
	12	0.394	0.452	0.350	0.538	0.400	0.533	0.575	.	.	0.500	0.452	0.391	0.410
	13	0.117	0.210	0.183	0.115	0.167	0.211	0.125	.	.	0.500	0.167	0.172	0.141
	14	0.043	0.016	0.033	0.038	0.017	0.011
	15
	N al	7	6	7	7	6	6	5	0	0	2	5	5	5

NSR	SSR	SIC	NBH	SBH	NGR	SGR	CRT	TRK	GJD	BJD	ASM	KHM	BOM	MZB
0.091	0.096	0.146	0.160	0.150	0.163	0.167	0.071	0.167	0.147	0.125	0.172	0.130	0.100	0.100
0.909	0.904	0.854	0.840	0.850	0.838	0.833	0.929	0.833	0.853	0.875	0.828	0.870	0.900	0.900
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
.	0.021	0.030	.
.	0.042	0.045	.
0.310	0.306	0.476	0.375	0.375	0.405	0.230	0.363	0.400	0.459	0.526	0.368	0.250	0.348	0.324
0.690	0.694	0.500	0.625	0.607	0.514	0.716	0.638	0.586	0.500	0.474	0.632	0.667	0.530	0.588
.	.	0.024	.	0.018	0.081	0.054	.	0.014	0.041	.	.	0.021	0.030	0.088
.	0.015	.
.
2	2	3	2	3	3	3	2	3	3	2	2	5	6	3
.
.	0.027
0.598	0.654	0.833	0.730	0.714	0.730	0.750	0.716	0.590	0.725	0.744	0.611	0.803	0.721	0.803
.	0.013
.	.	.	.	0.029	0.015	.	.
0.305	0.269	0.148	0.203	0.243	0.243	0.224	0.189	0.333	0.175	0.115	0.347	0.121	0.221	0.158
0.037	0.038	0.019	0.068	0.014	0.027	0.026	0.054	0.064	0.088	0.141	0.014	0.045	0.059	0.039
0.061	0.038	0.014	0.013	.	.	0.028	0.015	.	.
.
4	4	3	3	4	3	3	5	4	4	3	4	5	3	3
.	0.025
.	0.050	0.014	.
.	0.025
0.226	0.220	0.050	0.381	0.297	0.243	0.167	0.171	0.300	0.094	0.060	0.160	0.053	0.129	0.083
.	0.024	.	0.021
.	0.024	0.200	0.048	0.078	0.071	0.051	0.073	0.150	0.167	0.167	0.074	0.118	0.071	0.069
0.095	0.024	0.050	0.071	0.063	0.071	0.154	0.085	0.075	0.146	0.143	0.053	0.118	0.086	0.028
0.488	0.451	0.450	0.214	0.359	0.429	0.487	0.366	0.175	0.323	0.381	0.426	0.474	0.414	0.458
0.190	0.256	0.250	0.262	0.156	0.157	0.128	0.268	0.150	0.208	0.250	0.255	0.224	0.286	0.361
.	0.024	.	0.024	0.047	0.029	0.013	0.012	0.050	0.031	.	0.032	0.013	.	.
.	0.010
4	6	5	6	6	6	6	7	9	8	5	6	6	6	5

ss825679094	6	0.013	.	.
	7	0.018	.
	8
	9	.	0.016	0.015	0.034	0.050	0.057	.	.	.	0.013	0.038	0.018	0.028
	10	0.057	.	0.059	0.069	0.050	0.045	.	.	.	0.050	0.075	0.107	0.069
	11	0.091	0.016	0.206	0.172	0.233	0.205	.	.	0.118	0.088	0.163	0.179	0.167
	12	0.386	.	0.221	0.103	0.117	0.261	.	.	0.162	0.175	0.250	0.232	0.222
	13	0.068	.	0.088	0.293	0.100	0.125	.	.	0.162	0.138	0.063	0.161	0.069
	14	0.045	0.032	0.088	.	0.017	0.023	.	.	0.162	0.125	0.050	0.036	0.069
	15	0.250	0.323	0.118	0.172	0.333	0.091	.	.	0.088	0.138	0.163	0.071	0.139
	16	0.034	0.194	0.088	0.052	0.033	0.091	.	.	0.147	0.163	0.088	0.107	0.083
	17	0.068	0.339	0.074	0.052	0.033	0.068	.	.	0.059	0.050	0.075	0.036	0.097
	18	.	0.016	0.044	0.052	0.033	0.023	.	.	0.059	0.063	0.025	.	0.014
	19	.	0.065	0.044	.	.	.	0.028
	20	0.011	0.018	.
	21
	22
	25
	26	0.018	.
	27	0.014
28	
	N al	8	8	10	9	10	10	0	0	9	10	11	12	12
ss825679082	2
	3	0.013	.	.	.	0.029	.	0.014
	4	0.490	0.413	0.500	0.380	0.333	0.333	0.592	0.279	.	1.000	0.400	0.319	0.432
	5	0.417	0.400	0.457	0.554	0.561	0.556	0.197	0.640	.	.	0.557	0.556	0.500
	6	0.094	0.088	0.043	0.065	0.106	0.111	0.066	0.081	.	.	0.014	0.125	0.054
	7	.	0.100	0.132
	8
	9
	10
	11
	12
		N al	3	4	3	3	3	3	5	3	0	1	4	3
ss825679083	5	0.010	0.038	0.037	.	.	.
	6	.	0.038	0.029	0.012	0.011	.	.
	7	0.010	0.013	0.012	.	.	.
	8	0.396	0.577	0.579	0.595	0.468	0.444	0.625	0.409	0.529	0.488	0.533	0.565	0.400
	9	0.469	0.282	0.368	0.369	0.500	0.389	0.321	0.364	0.343	0.402	0.435	0.387	0.443
	10	0.115	0.051	0.053	0.036	0.032	0.167	0.054	0.227	0.086	0.049	0.022	0.048	0.157
	11	0.014
	12
	N al	5	6	3	3	3	3	3	3	5	6	4	3	3

.	0.013	0.028
.
.
.	0.039	0.056	.	.	0.045	0.013	0.068	.	0.113	0.042	0.125	0.054	0.086	0.231
0.045	0.026	0.056	0.027	0.029	0.113	0.111	0.038	.	0.034	0.077
0.303	0.276	0.444	.	0.074	0.136	0.132	0.149	0.214	0.175	0.181	0.213	0.250	0.103	0.250
0.227	0.171	0.167	.	0.185	0.227	0.250	0.243	0.214	0.175	0.111	0.250	0.107	0.138	0.038
0.167	0.158	.	.	0.093	0.182	0.158	0.108	0.086	0.013	0.097	0.100	0.107	0.086	0.038
0.015	0.013	0.056	.	0.056	0.045	0.039	0.068	0.071	0.050	0.097	0.025	0.107	0.069	0.038
0.136	0.224	.	0.207	0.167	0.125	0.197	0.257	0.129	0.100	0.153	0.075	0.018	0.069	0.173
0.061	0.053	0.111	0.293	0.185	0.057	0.079	0.027	0.100	0.125	0.014	0.038	0.089	0.086	0.019
.	0.013	0.111	0.414	0.222	0.114	0.092	0.054	0.057	0.038	0.139	.	0.179	0.241	0.077
0.015	.	.	0.017	0.019	0.034	0.013	.	0.043	0.063	.	0.063	.	.	0.058
.	0.013	.	0.069	.	0.011	.	.	0.043	.	.	0.013	0.018	.	.
.	0.013	.	0.014	0.025	.	.	0.018	.	.
.	0.013	0.028	0.038	0.018	0.034	.
0.030	0.023	0.013	0.036	0.017	.
.
.	0.013
.	0.034	.
.	0.013	.	.	.
9	11	7	5	8	11	11	9	11	12	11	13	12	12	10
0.024	0.075
0.012	0.013	.	0.025	0.029	.	.	.	0.013	.	0.013	.	0.015	.	.
0.333	0.313	0.500	0.388	0.500	0.366	0.434	0.292	0.275	0.310	0.397	0.314	0.273	0.400	0.314
0.524	0.475	0.431	0.375	0.338	0.585	0.474	0.569	0.638	0.607	0.526	0.593	0.636	0.471	0.629
0.107	0.125	0.069	0.075	0.088	0.037	0.092	0.125	0.075	0.036	0.064	0.070	0.045	0.114	0.057
.	.	.	0.138	0.044	0.012	.	.	.	0.048	.	.	.	0.014	.
.	0.014
.	0.030	.	.
.	0.023	.	.	.
.
5	5	3	5	5	4	3	4	4	4	4	4	5	4	3
.	0.013	.	.	.	0.011
.	.	.	0.026	0.030
.	.	.	0.038	0.015	0.038	.	.	0.013	.	0.030
0.451	0.600	0.423	0.513	0.515	0.513	0.538	0.557	0.638	0.534	0.439	0.547	0.500	0.500	0.513
0.439	0.238	0.423	0.385	0.348	0.333	0.397	0.398	0.275	0.352	0.394	0.349	0.419	0.353	0.342
0.110	0.150	0.154	0.038	0.091	0.103	0.051	0.045	0.075	0.080	0.136	0.105	0.081	0.132	0.145
.	0.013	0.013	.	.	0.011	.	.	.	0.015	.
.	0.011
3	4	3	5	5	5	4	3	4	6	4	3	3	4	3

ss825679084	7	0.132
	8	0.016	0.012	.	.	.
	9	0.188	0.068	0.104	0.200	0.094	0.078	.	.	0.041	0.085	.	0.037	0.039
	10	0.104	0.041	0.135	0.100	0.094	0.122	.	.	0.054	0.110	.	0.056	0.105
	11	0.041	0.049	.	.	.
	12	0.016	0.011	.	.	0.014	.	.	.	0.013
	13	0.052	0.014	0.083	.	0.016	0.033	.	.	0.108	0.085	.	0.037	0.092
	14	0.229	0.392	0.292	0.300	0.328	0.300	0.132	.	0.230	0.207	.	0.519	0.395
	15	0.333	0.338	0.208	0.100	0.219	0.233	.	.	0.311	0.354	.	0.222	0.158
	16	0.073	0.122	0.135	0.200	0.156	0.111	.	.	0.176	0.085	.	0.093	0.171
	17	0.021	0.027	0.042	.	0.031	0.100	0.289	.	0.027	0.012	.	0.037	0.026
	18	.	.	.	0.100	0.031	0.011	0.447
	N al	7	7	7	6	10	8	4	0	9	9	0	7	8
ss825679079	8
	9
	11
	12
	13	0.011	0.045	0.021	0.026	.	.	.	0.012	.	.	0.051	0.016	.
	14	0.011	.	0.010	0.026	.	0.022	0.018	0.024	.	0.041	0.013	.	0.014
	15	0.054	0.076	0.063	0.038	0.031	0.056	0.036	0.024	.	0.027	0.090	0.047	0.042
	16	0.185	0.167	0.146	0.141	0.125	0.100	0.125	0.146	0.190	0.135	0.154	0.125	0.153
	17	0.141	0.197	0.208	0.192	0.203	0.089	0.214	0.207	0.241	0.122	0.205	0.266	0.222
	18	0.163	0.152	0.281	0.179	0.234	0.233	0.286	0.195	0.172	0.203	0.244	0.234	0.222
	19	0.196	0.167	0.135	0.282	0.250	0.389	0.232	0.220	0.190	0.297	0.115	0.156	0.222
	20	0.087	0.106	0.083	0.051	0.094	0.044	0.054	0.085	0.103	0.108	0.064	0.063	0.069
	21	0.087	0.076	0.031	0.038	0.047	0.033	.	0.037	0.052	0.041	0.064	0.047	0.028
	22	0.054	.	0.010	0.013	0.016	0.033	0.018	0.024	0.052	0.027	.	0.047	0.014
	23	.	.	.	0.013	.	.	0.018	0.012
	24	0.011	.	0.010	0.012	0.014
25	.	0.015	
N al	11	9	11	11	8	9	9	12	7	9	9	9	10	
ss825679080	8	0.022	0.045	0.068	0.152
	9	.	.	.	0.012	0.050	.	0.273	0.318
	10	.	0.013	0.083	0.073	0.050	0.091	0.068	0.015	0.097	0.068	0.049	0.063	0.014
	11	0.957	0.700	0.875	0.841	0.900	0.773	0.455	0.515	0.806	0.865	0.915	0.938	0.932
	12	0.022	0.225	0.031	0.061	.	0.091	0.136	.	0.081	0.041	0.037	.	0.054
	13	.	0.063	0.010	0.016	0.027	.	.	.
	14	.	.	.	0.012
N al	3	4	4	5	3	4	5	4	4	4	3	2	3	

.
.
.	0.013	0.058	0.016	0.047	0.076	0.149	.	0.042	0.096	0.083	0.043	0.039	0.106	0.200
0.107	0.063	0.038	0.081	0.063	0.185	0.095	.	0.125	0.074	0.071	0.053	0.118	0.045	0.167
.	.	0.019	.	.	0.011	0.014	.	0.042	.	0.012	0.011	0.013	.	.
.	0.013	0.021	.	0.024	0.064	0.026	0.030	0.067
0.119	0.088	0.154	0.016	0.094	0.076	0.068	.	0.021	0.074	0.083	.	.	0.030	0.100
0.464	0.488	0.308	0.355	0.344	0.337	0.284	.	0.417	0.468	0.345	0.404	0.316	0.424	0.233
0.202	0.200	0.250	0.355	0.375	0.174	0.257	.	0.250	0.191	0.190	0.298	0.263	0.182	0.100
0.095	0.138	0.135	0.129	0.063	0.087	0.108	.	0.063	0.032	0.119	0.096	0.171	0.167	0.100
0.012	.	0.038	0.048	0.016	0.054	0.027	.	0.021	0.043	0.071	0.032	0.053	0.015	0.033
.	0.021
6	7	8	7	7	8	8	1	9	8	9	8	8	8	8
.	0.030
.	0.014
.
0.013	0.054	0.068	0.012	.	0.014
0.013	0.018	.	0.017	.	.	0.025	0.027	0.012	.	0.014
0.013	0.054	.	.	0.029	0.011	0.025	0.041	0.048	0.013	0.014	0.023	0.015	0.048	0.091
0.050	0.018	0.045	0.067	0.015	0.057	0.050	0.135	0.143	0.092	0.014	0.080	0.059	0.065	.
0.050	0.107	0.159	0.117	0.132	0.045	0.125	0.162	0.119	0.039	0.176	0.068	0.074	0.129	0.091
0.275	0.143	0.250	0.200	0.250	0.295	0.200	0.189	0.202	0.158	0.149	0.159	0.132	0.081	0.106
0.200	0.268	0.159	0.250	0.206	0.239	0.363	0.230	0.214	0.224	0.257	0.307	0.235	0.226	0.318
0.163	0.143	0.114	0.250	0.279	0.182	0.150	0.122	0.179	0.303	0.216	0.216	0.324	0.339	0.182
0.163	0.125	0.091	0.067	0.044	0.102	0.050	0.041	0.060	0.118	0.081	0.045	0.103	0.097	0.106
0.050	0.036	0.068	0.033	0.029	0.045	0.013	0.027	0.012	0.039	0.027	0.057	0.059	.	0.045
0.013	.	0.023	.	.	0.023	0.014	0.023	.	0.016	0.015
.	.	0.023	0.014	0.011	.	.	0.015
.	0.036	.	.	0.015	.	.	0.014	.	.	.	0.011	.	.	.
.	0.014	.	0.013
11	11	10	8	9	9	9	11	10	9	13	11	8	7	10
.
.	.	0.020
0.053	0.048	0.140	.	0.067	0.100	0.064	0.048	0.037	0.114	0.100	0.143	0.143	0.096	0.118
0.908	0.919	0.800	0.895	0.733	0.878	0.885	0.887	0.926	0.829	0.900	0.786	0.814	0.788	0.816
0.039	0.032	0.040	0.066	0.117	0.022	0.051	0.065	0.037	0.057	.	0.048	0.043	0.058	0.053
.	.	.	0.039	0.083	0.024	.	0.058	0.013
.
3	3	4	3	4	3	3	3	3	3	2	4	3	4	4

ss825679081	5	0.154
	6	0.077
	7
	8	0.245	0.348	0.268	0.341	0.328	0.489	0.333	.	0.361	0.297	0.271	0.368	0.474
	9	0.755	0.652	0.732	0.659	0.656	0.511	0.667	0.731	0.639	0.703	0.729	0.632	0.526
	10	0.016	.	.	0.038
	11
	N al	2	2	2	2	3	2	2	4	2	2	2	2	2
ss825679085	14	.	0.016
	15	0.010	.	.
	16	.	0.032	.	0.025	0.030	0.011	0.010	0.016	0.026
	17	0.304	0.194	0.190	0.288	0.227	0.400	0.143	.	.	.	0.260	0.371	0.368
	18	0.283	0.339	0.345	0.263	0.409	0.289	0.762	.	.	.	0.333	0.226	0.224
	19	0.261	0.355	0.381	0.275	0.227	0.222	0.095	.	.	.	0.302	0.306	0.276
	20	0.141	0.048	0.048	0.150	0.091	0.067	0.073	0.065	0.066
	21	0.011	0.016	0.036	.	0.015	0.011	0.010	0.016	0.039
	22
	23
	N al	5	7	5	5	6	6	3	0	0	0	7	6	6
ss825679087	7
	8	0.011	0.017	.	0.033	.	0.013
	9	0.181	0.227	.	0.271	0.194	0.311	0.188	0.128	0.103	0.274	0.267	0.265	0.250
	10	0.787	0.750	.	0.686	0.790	0.678	0.788	0.849	0.776	0.694	0.678	0.735	0.697
	11	0.021	0.023	.	0.029	0.016	0.011	0.025	0.023	0.086	0.032	0.022	.	0.039
	12
	13	.	.	.	0.014	0.017
	N al	4	3	0	4	3	3	3	3	5	3	4	2	4
ss825679088	6
	7	.	0.023
	8	0.022	.	.	0.017	0.016	0.022	0.044	0.036	0.021	0.015	0.011	0.015	0.026
	9	0.078	0.068	.	0.083	0.031	0.056	0.015	0.024	0.042	0.029	0.109	0.015	0.026
	10	0.133	0.136	.	0.267	0.094	0.144	0.147	0.155	0.104	0.176	0.120	0.152	0.105
	11	0.222	0.227	.	0.250	0.188	0.189	0.235	0.238	0.292	0.250	0.163	0.152	0.250
	12	0.333	0.227	.	0.150	0.266	0.222	0.221	0.286	0.292	0.324	0.326	0.364	0.276
	13	0.078	0.091	.	0.200	0.172	0.200	0.250	0.131	0.167	0.118	0.109	0.152	0.158
	14	0.089	0.205	.	0.017	0.156	0.089	0.059	0.060	0.063	0.044	0.098	0.061	0.105
	15	0.011	0.023	.	0.017	0.063	0.056	0.029	0.060	0.021	0.029	0.011	0.061	0.039
	16	0.022	.	.	.	0.016	0.022	.	0.012	.	0.015	0.043	0.015	0.013
17	0.011	0.011	0.015	.	
N al	10	8	0	8	9	9	8	9	8	9	10	10	9	

.
.
.	0.010	.	.	.
0.500	0.419	0.341	0.200	0.258	0.409	0.321	0.434	0.387	0.510	0.417	0.594	0.514	0.486	0.379
0.500	0.581	0.659	0.800	0.726	0.580	0.667	0.566	0.613	0.479	0.583	0.396	0.486	0.514	0.621
.	0.011	0.013	.	.	0.010
.	.	.	.	0.016
2	2	2	2	3	3	3	2	2	3	2	3	2	2	2
.	0.014	.	.	0.011	.	.	.
.	.	0.022	.	.	0.013
.	.	.	0.032	.	0.026	0.027	0.028	0.014	0.042	0.024	0.032	.	0.014	0.056
0.343	0.188	0.283	0.258	0.250	0.171	0.392	0.250	0.311	0.375	0.310	0.447	0.329	0.400	0.375
0.343	0.422	0.217	0.355	0.483	0.342	0.203	0.347	0.365	0.281	0.321	0.106	0.229	0.143	0.111
0.286	0.313	0.348	0.274	0.167	0.329	0.257	0.306	0.162	0.135	0.179	0.223	0.271	0.157	0.208
0.029	0.078	0.043	0.065	0.083	0.105	0.081	0.069	0.095	0.115	0.131	0.160	0.129	0.229	0.167
.	.	0.065	0.016	0.017	0.013	0.041	.	0.041	0.052	0.024	.	0.043	0.029	0.014
.	.	0.022	0.012	0.021	.	0.014	0.056
.	0.014	0.014
4	4	7	6	5	7	6	5	7	6	7	7	5	8	8
.	0.037	0.011
0.013	0.012	.	.	0.016	.	.	0.024	0.054	0.036
0.200	0.232	0.333	0.100	0.306	0.429	.	0.262	0.304	0.349	0.295	0.111	0.171	0.030	0.161
0.788	0.695	0.667	0.900	0.645	0.571	.	0.714	0.598	0.640	0.705	0.867	0.817	0.970	0.804
.	0.024	.	.	0.032	.	.	.	0.022	0.012	.	0.022	0.012	.	.
.	0.011
.
3	5	2	2	4	2	0	3	6	3	2	3	3	2	3
.	0.015	.
.	0.014	.	.	0.012
0.014	0.012	0.035	.	0.122	0.085	.	0.063
0.028	0.069	.	.	0.018	0.057	0.025	0.083	0.047	0.023	0.026	0.111	0.085	0.045	0.094
0.083	0.139	.	0.059	0.036	0.071	0.125	0.071	0.128	0.093	0.026	0.067	0.110	0.076	0.141
0.236	0.208	.	0.088	0.214	0.143	0.163	0.155	0.233	0.221	0.276	0.044	0.171	0.136	0.094
0.181	0.278	.	0.750	0.411	0.300	0.363	0.333	0.279	0.279	0.342	0.333	0.183	0.500	0.406
0.292	0.208	.	0.044	0.196	0.186	0.175	0.190	0.140	0.116	0.118	0.156	0.195	0.121	0.156
0.069	0.056	.	0.029	0.107	0.100	0.088	0.083	0.058	0.116	0.158	0.078	0.110	0.061	0.016
0.056	0.042	.	0.015	0.018	0.114	0.025	0.060	0.070	0.116	0.026	0.056	0.049	0.015	0.031
0.042	.	.	0.015	.	0.014	0.013	0.012	0.012	.	0.026	0.033	0.012	0.030	.
.	0.025	0.012	0.012
9	7	0	7	7	9	9	9	11	8	8	9	9	9	8

ss825679089	4	0.870	0.872	0.867	0.875	0.891	0.822	0.768	0.890	0.891	0.875	0.920	0.786	0.855
	5	0.130	0.128	0.133	0.125	0.109	0.178	0.232	0.110	0.109	0.125	0.080	0.214	0.145
	N al	2	2	2	2	2	2	2	2	2	2	2	2	2
ss825679090	5
	7	0.015
	8	0.170	0.317	0.261	0.239	0.303	0.273	0.327	0.214	0.241	0.181	0.219	0.250	0.289
	9	0.074	0.085	0.141	0.057	0.045	0.102	0.115	0.114	0.069	0.042	0.073	0.042	0.092
	10	0.638	0.488	0.533	0.625	0.500	0.455	0.481	0.600	0.603	0.625	0.542	0.597	0.566
	11	0.096	0.073	0.043	0.068	0.076	0.114	0.077	0.071	0.069	0.111	0.156	0.083	0.053
	12	0.021	0.024	0.022	0.011	0.030	0.045	.	.	0.017	0.042	0.010	0.014	.
	13	.	0.012	.	.	0.030	0.011	0.014	.
	14
	15
	N al	5	6	5	5	7	6	4	4	5	5	5	6	4
ss825679091	6	.	0.029	0.017
	7	0.309	0.603	0.500	0.453	0.532	0.422	0.638	0.886	0.455	0.456	0.500	0.387	0.514
	8	0.011	0.015	.	.	.	0.011	.	.	0.015
	9	0.511	0.265	0.261	0.349	0.355	0.356	0.207	0.071	0.394	0.397	0.357	0.355	0.306
	10	0.170	0.088	0.239	0.186	0.113	0.211	0.138	0.043	0.136	0.147	0.143	0.242	0.181
	11	.	.	.	0.012	0.016	.
	N al	4	5	3	4	3	4	4	3	4	3	3	4	3
Heterozygosities per population	Mean	0.485	0.544	0.521	0.524	0.511	0.546	0.502	0.460	0.499	0.492	0.481	0.512	0.513
	SD	0.219	0.177	0.206	0.196	0.211	0.190	0.172	0.226	0.215	0.205	0.220	0.199	0.213

0.845	0.842	0.850	0.865	0.871	0.814	0.833	0.842	0.909	0.707	0.750	0.872	0.878	0.843	0.875
0.155	0.158	0.150	0.135	0.129	0.186	0.167	0.158	0.091	0.293	0.250	0.128	0.122	0.157	0.125
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
.
.	0.013	.	0.016	.	.	0.011	.	.	0.011
0.314	0.305	0.138	0.300	0.236	0.255	0.276	0.218	0.156	0.302	0.321	0.330	0.269	0.329	0.233
0.105	0.073	0.069	0.113	0.014	0.106	0.079	0.038	0.047	0.083	0.071	0.032	0.090	0.029	0.056
0.500	0.500	0.672	0.413	0.681	0.489	0.487	0.615	0.391	0.563	0.464	0.543	0.487	0.543	0.556
0.058	0.110	0.103	0.100	0.042	0.138	0.145	0.090	0.391	0.052	0.131	0.064	0.128	0.057	0.133
0.012	0.012	0.017	0.063	.	0.011	.	0.038	.	.	.	0.021	0.026	.	0.011
0.012	.	.	0.013	0.014	0.012	.	.	0.043	.
.	.	.	.	0.014
.
6	5	5	6	6	5	5	5	5	4	5	6	5	5	6
.	.	.	0.038	0.014
0.536	0.561	0.143	0.425	0.457	0.390	0.577	0.534	0.513	0.593	0.529	0.620	0.600	0.544	0.606
.	0.012	.	0.075	0.071	.	.	.	0.013	0.011
0.310	0.244	0.643	0.375	0.314	0.378	0.205	0.250	0.250	0.233	0.229	0.109	0.200	0.250	0.223
0.131	0.146	0.214	0.088	0.143	0.232	0.218	0.216	0.213	0.163	0.229	0.261	0.183	0.176	0.149
0.024	0.037	.	.	0.014	.	.	.	0.013	0.012	.	0.011	0.017	0.029	0.011
4	5	3	5	5	3	3	3	5	4	4	4	4	4	5
0.500	0.508	0.506	0.476	0.537	0.548	0.524	0.490	0.522	0.555	0.533	0.511	0.511	0.480	0.505
0.206	0.205	0.225	0.215	0.178	0.195	0.237	0.244	0.219	0.183	0.207	0.202	0.217	0.223	0.216

Marker information									Heterozygosity	
ID reference	N	NTZ	STZ	SIW	AHZ	BAM	AYM	QCH	Mean	SD
ss825679095	4	0.114	0.157	.	0.260	0.460	.	0.059	0.236	0.083
	5	0.886	0.843	.	0.740	0.540	1.000	0.941		
	N al	2	2	0	2	2	1	2		
ss825679096	3	.	0.013	0.026	0.019	0.040	.	.	0.462	0.095
	4	0.018		
	5	0.268	0.421	0.316	0.538	0.540	0.088	0.059		
	6	0.714	0.539	0.553	0.365	0.280	0.913	0.912		
	7	.	.	0.105	0.038	0.080	.	0.029		
	8	.	0.026		
	9	.	.	.	0.038	0.060	.	.		
	N al	3	4	4	5	5	2	3		
ss825679092	3	0.431	0.075
	4		
	5	0.734	0.697	0.714	0.808	0.881	0.395	0.353		
	6	0.024	.	.		
	7		
	8	0.234	0.263	0.143	0.135	.	0.605	0.618		
	9	0.031	0.039	0.119	0.019	0.048	.	0.029		
	10	.	.	0.024	0.019	0.048	.	.		
	13	.	.	.	0.019	.	.	.		
	N al	3	3	4	5	4	2	3		
ss825679093	2	.	.	0.026	0.700	0.075
	5		
	7	.	.	0.026		
	8	0.122	0.156	0.184	.	.	0.186	0.074		
	9	0.012		
	10	0.098	0.135	0.053	0.038	0.021	.	0.015		
	11	0.110	0.146	0.211	0.192	0.167	.	0.059		
	12	0.439	0.323	0.316	0.442	0.313	0.628	0.647		
	13	0.207	0.219	0.158	0.269	0.458	0.163	0.206		
	14	0.012	0.021	0.026	0.038	0.042	0.023	.		
	15	.	.	.	0.019	.	.	.		
	N	7	6	8	6	5	4	5		

ss825679094	6	.	0.012	0.837	0.048
	7		
	8	.	.	.	0.024	.	.	.		
	9	0.043	0.061	0.088	0.214	0.184	.	.		
	10	0.029	0.049	0.088	0.190	0.211	.	.		
	11	0.257	0.280	0.029	0.071	0.026	0.195	0.219		
	12	0.171	0.122	0.382	0.071	0.105	0.439	0.391		
	13	0.057	0.122	0.147	0.095	0.026	0.024	.		
	14	0.086	0.024	0.029	0.071	0.211	0.110	0.094		
	15	0.114	0.098	0.029	0.095	0.105	0.024	0.047		
	16	0.043	0.037	0.088	0.048	0.053	0.037	.		
	17	0.086	0.049	0.088	.	.	0.073	0.078		
	18	0.086	0.037	0.029	0.095	0.079	0.085	0.125		
	19	.	0.012	.	0.024	.	0.012	0.031		
	20	.	0.037	0.016		
	21		
	22	.	0.024		
	25	0.014		
	26		
	27	0.014	0.012		
28	.	0.024			
	N al	12	16	10	11	9	9	8		
ss825679082	2	0.508	0.124
	3		
	4	0.200	0.398	0.206	0.404	0.286	0.942	0.956		
	5	0.650	0.545	0.588	0.558	0.667	0.058	0.044		
	6	0.150	0.057	0.176		
	7		
	8	.	.	0.029	0.038	.	.	.		
	9		
	10		
	11	0.024	.	.		
	12	0.024	.	.		
		N al	3	3	4	3	4	2		
ss825679083	5	0.582	0.051
	6	0.016		
	7	.	.	.	0.060	.	.	.		
	8	0.529	0.478	0.333	0.220	0.119	0.536	0.641		
	9	0.357	0.344	0.524	0.480	0.429	0.464	0.344		
	10	0.114	0.144	0.119	0.220	0.381	.	.		
	11	.	0.033	0.024	0.020	0.071	.	.		
	12		
	N al	3	4	4	5	4	2	3		

ss825679084	7	0.733	0.109
	8		
	9	0.125	.	0.167	0.120	0.200	0.036	0.152		
	10	0.071	0.045		
	11	.	.	.	0.020	0.040	.	.		
	12	0.375	.	0.042	0.020	.	.	.		
	13	0.125	.	.	0.080	0.120	0.012	.		
	14	0.125	.	0.208	0.100	0.060	0.798	0.758		
	15	.	.	0.375	0.360	0.380	0.060	.		
	16	0.250	.	0.208	0.220	0.100	0.024	0.045		
	17	.	.	.	0.080	0.100	.	.		
	18		
	N al	5	0	5	8	7	6	4		
ss825679079	8	0.040	.	.	0.830	0.025
	9		
	11	0.013	.	.	.	0.020	.	.		
	12	.	.	0.043	.	.	0.023	.		
	13	0.026	0.011	0.022	0.100	.	.	.		
	14	0.039	0.033	0.109	0.160	0.080	0.047	0.042		
	15	0.079	0.022	0.022	0.180	0.120	0.244	0.264		
	16	0.132	0.156	0.283	0.020	0.080	0.186	0.167		
	17	0.171	0.133	0.174	0.160	0.220	0.279	0.319		
	18	0.250	0.256	0.217	0.100	0.060	0.093	0.125		
	19	0.158	0.167	0.109	0.140	0.220	0.105	0.069		
	20	0.079	0.156	.	0.100	0.100	0.012	.		
	21	0.053	0.033	0.022	0.020	0.040	0.012	0.014		
	22	.	0.022	.	0.020	.	.	.		
	23		
	24	.	0.011	.	.	0.020	.	.		
	25		
N al	10	11	9	10	11	9	7			
ss825679080	8	0.278	0.162
	9	.	.	0.111	0.095	0.023	.	.		
	10	0.083	.	0.167	0.310	0.409	0.186	0.221		
	11	0.833	.	0.667	0.500	0.455	0.791	0.750		
	12	0.083	.	0.056	0.095	0.091	0.023	0.029		
	13	0.023	.	.		
	14		
N al	3	0	4	4	5	3	3			

ss825679081	5	0.450	0.059
	6		
	7	0.011	.	0.026		
	8	0.522	0.459	0.526	0.788	0.795	0.791	0.724		
	9	0.467	0.541	0.447	0.212	0.205	0.209	0.276		
	10		
	11		
	N al	3	2	3	2	2	2	2		
ss825679085	14	0.015	0.709	0.076
	15	0.011		
	16	0.011	0.020	.	0.077	0.048	.	0.015		
	17	0.456	0.378	0.476	0.269	0.286	0.419	0.439		
	18	0.256	0.224	0.310	0.269	0.071	0.233	0.197		
	19	0.133	0.224	0.119	0.192	0.333	0.012	0.091		
	20	0.100	0.112	0.071	0.115	0.095	0.314	0.242		
	21	0.022	0.031	0.024	0.058	0.095	0.023	.		
	22	0.011	0.010	.	0.019	0.071	.	.		
	23		
	N al	8	7	5	7	7	5	6		
ss825679087	7	0.371	0.099
	8	0.012	.		
	9	0.091	0.191	0.348	0.250	0.200	0.262	0.364		
	10	0.716	0.766	0.652	0.708	0.775	0.726	0.636		
	11	0.193	0.043	.	0.042	0.025	.	.		
	12		
	13		
	N al	3	3	2	3	3	3	2		
ss825679088	6	0.794	0.078
	7	0.045	.	.		
	8	0.059	0.060	0.026	0.100	0.159	.	.		
	9	0.044	0.024	0.105	0.260	0.068	0.036	.		
	10	0.088	0.119	0.132	0.160	0.136	0.048	0.106		
	11	.	0.083	0.105	0.060	0.045	0.083	0.030		
	12	0.515	0.405	0.211	0.080	0.159	0.190	0.364		
	13	0.118	0.083	0.158	0.120	0.182	0.083	0.076		
	14	0.103	0.119	0.158	0.100	0.136	0.202	0.061		
	15	0.059	0.095	0.079	0.120	0.045	0.238	0.242		
	16	0.015	0.012	.	.	0.023	0.119	0.091		
	17	.	.	0.026	.	.	.	0.030		
	N al	8	9	9	8	10	8	8		

ss825679089	4	0.859	0.819	0.733	0.940	0.896	0.931	1.000	0.240	0.071
	5	0.141	0.181	0.267	0.060	0.104	0.069	.		
	N al	2	2	2	2	2	2	1		
ss825679090	5	0.015	0.624	0.061
	7	.	0.010		
	8	0.221	0.229	0.278	0.200	0.208	0.103	0.059		
	9	0.103	0.073	0.139	0.060	0.063	0.279	0.368		
	10	0.529	0.542	0.389	0.360	0.313	0.574	0.500		
	11	0.103	0.135	0.167	0.320	0.354	0.029	0.059		
	12	0.015	0.010	0.028	.	0.042	.	.		
	13	.	.	.	0.060	0.021	0.015	.		
	14		
	15	0.029		
	N al	6	6	5	5	6	5	5		
ss825679091	6	0.581	0.083
	7	0.607	0.532	0.477	0.708	0.727	0.453	0.409		
	8	.	.	.	0.042	.	0.116	0.045		
	9	0.190	0.170	0.114	0.063	0.091	0.372	0.485		
	10	0.190	0.287	0.341	0.146	0.159	0.047	0.061		
	11	0.012	0.011	0.068	0.042	0.023	0.012	.		
	N al	4	4	4	5	4	5	4		
Heterozygosities per population	Mean	0.511	0.560	0.621	0.541	0.536	0.418	0.410		
	SD	0.199	0.192	0.206	0.220	0.222	0.247	0.243		

ST Tab 7- Averaged F_{ST}/R_{ST} genetic distances per population. Mean genetic distance averaged across distances to all other populations. SD: Standard deviation.

Population	F_{ST} distances		R_{ST} distances	
	Mean	SD	Mean	SD
ORK	0.016	0.0002	0.036	0.0017
GBR	0.006	0.0000		
CEU	0.010	0.0001		
POL	0.008	0.0001	0.122	0.0026
FIN	0.011	0.0001		
NFR	0.007	0.0001	0.027	0.0020
SFR	0.007	0.0001	0.036	0.0023
NCO	0.007	0.0000	0.021	0.0017
CCO	0.015	0.0001	0.027	0.0022
BQC	0.018	0.0001	0.043	0.0004
CAT	0.006	0.0000	0.075	0.0009
NSP	0.008	0.0001	0.037	0.0010
SSP	0.006	0.0000	0.030	0.0013
MEN	0.008	0.0000	0.031	0.0023
TSI	0.006	0.0001		
CIT	0.007	0.0000	0.028	0.0021
SIT	0.007	0.0000	0.019	0.0012
NSR	0.012	0.0001	0.032	0.0023
SSR	0.009	0.0001	0.031	0.0025
SIC	0.007	0.0001	0.042	0.0009
NBH	0.018	0.0002	0.159	0.0043
SBH	0.011	0.0001	0.043	0.0007
NGR	0.007	0.0000	0.027	0.0012
SGR	0.008	0.0001	0.026	0.0010
CRT	0.006	0.0000	0.034	0.0025
TRK	0.009	0.0001	0.039	0.0015
GJD	0.012	0.0000	0.026	0.0017
BJD	0.017	0.0000	0.024	0.0018
ASM	0.021	0.0001	0.025	0.0015
KHM	0.012	0.0001	0.031	0.0006
BOM	0.011	0.0001	0.026	0.0004
MZB	0.017	0.0000	0.050	0.0031
NTZ	0.015	0.0001	0.024	0.0011
STZ	0.016	0.0001	0.015	0.0009
SIW	0.030	0.0001	0.064	0.0022

ST Tab 8- Evanno results. K: estimated ancestral clusters. Reps: Iterations. L(K): Ln(probability of data). sd: standard deviation. L'(K): Rate of change of the likelihood distribution. |L''(K)|: Absolute value of the 2nd order rate of change of the likelihood distribution (mean). Delta K: $\text{mean}(|L''(K)|)/\text{sd}(L(K))$

	K	Reps	Mean L(K)	sd L(K)	L'(K)	L''(K)	Delta K
SNP data	2	5	-427327.04	329206.57	-131834.14	290866.00	0.88
	3	5	-268295.18	7.41	159031.86	154843.96	20890.17
	4	5	-264107.28	373.28	4187.90	802.40	2.15
	5	5	-259116.98	18.13	4990.30	2737.56	150.96
	6	5	-256864.24	72.44	2252.74	48.10	0.66
	7	5	-254563.40	214.59	2300.84	1220.08	5.69
TRP data	2	5	-54980.24	97.72	286.38	277.16	2.84
	3	5	-54416.70	18.37	563.54	355.34	19.34
	4	5	-54208.50	27.35	208.20	73.10	2.67
	5	5	-54073.40	26.34	135.10	74.60	2.83
	6	5	-54012.90	79.82	60.50	110.72	1.39
	7	5	-53841.68	80.36	171.22	47.58	0.59

ST Tab 9- Population membership coefficients of the estimated clusters for SNPs and TRPs.

Population	SNP's coefficients			TRP's coefficients		
	cluster A	cluster B	cluster C	cluster A	cluster B	cluster C
ORK	0.305	0.377	0.318	0.264	0.422	0.313
GBR	0.29	0.305	0.406			
CEU	0.257	0.386	0.357			
POL	0.334	0.335	0.331	0.27	0.428	0.301
FIN	0.316	0.302	0.382			
NFR	0.319	0.345	0.337	0.29	0.352	0.358
SFR	0.374	0.292	0.334	0.407	0.279	0.314
NCO	0.386	0.244	0.37	0.425	0.266	0.309
CCO	0.259	0.243	0.498	0.355	0.288	0.357
BQC	0.434	0.29	0.277	0.525	0.227	0.248
CAT	0.27	0.346	0.384	0.435	0.358	0.206
NSP	0.28	0.372	0.348	0.297	0.409	0.294
SSP	0.326	0.347	0.326	0.28	0.393	0.328
MEN	0.39	0.281	0.329	0.339	0.344	0.317
TSI	0.287	0.296	0.417			
CIT	0.306	0.395	0.299	0.333	0.312	0.355
SIT	0.277	0.317	0.406	0.317	0.352	0.331
NSR	0.262	0.298	0.44	0.42	0.276	0.304
SSR	0.286	0.269	0.445	0.431	0.267	0.302
SIC	0.298	0.388	0.314	0.291	0.336	0.373
NBH	0.326	0.41	0.263	0.247	0.512	0.241
SBH	0.259	0.438	0.303	0.249	0.431	0.32
NGR	0.285	0.338	0.377	0.338	0.325	0.337
SGR	0.355	0.318	0.326	0.297	0.362	0.341
CRT	0.267	0.316	0.416	0.318	0.339	0.344
TRK	0.283	0.362	0.355	0.301	0.344	0.354
GJD	0.241	0.383	0.377	0.306	0.299	0.395
BJD	0.358	0.365	0.277	0.341	0.303	0.356
ASM	0.31	0.188	0.502	0.418	0.278	0.304
KHM	0.318	0.263	0.42	0.354	0.324	0.322
BOM	0.35	0.232	0.418	0.337	0.405	0.258
MZB	0.336	0.275	0.389	0.396	0.312	0.292
NTZ	0.239	0.277	0.484	0.37	0.317	0.313
STZ	0.395	0.203	0.402	0.369	0.278	0.353
SIW	0.251	0.409	0.34	0.305	0.287	0.409

ST Tab 10- Detection of selection results for SNP markers: Observed (Obs) Fst with selection significance (highlighed)('P value<0.1; *: P value<0.05; **:P value<0.01.), and SPA scores for continental European (EU) and North African (NA) allele gradients (highlighed SPA scores were above 95th percentile).

ID reference	Chr	Coding Region	Gene Region	Obs Fst	EU SPA score	NA SPA score
rs11771443	7	NOS3	5UTR	0.0095	1.36	0.90
rs1800781	7	NOS3	intron	0.0040'	1.09	1.25
rs3793342	7	NOS3	intron	0.0026'	0.20	0.55
rs1799983	7	NOS3	coding	0.0198	1.57	0.71
rs3918188	7	NOS3	intron	0.0077	0.47	1.18
rs753482	7	NOS3	intron	0.0260	1.36	0.76
rs7830	7	ATG9B	3UTR	0.0175	1.29	0.71
rs2373929	7	ATG9B	intron	0.0251	1.57	1.61
rs13307588	7	ATG9B	intron	0.0138	0.45	0.10
rs219240	7	Non-coding		0.0031'	0.35	0.58
rs219243	7	Non-coding		0.0019*	0.61	0.65
rs219245	7	Non-coding		0.0005**	0.40	0.45
rs219269	7	Non-coding		0.0007	0.26	0.09
rs6964641	7	Non-coding		0.0061	0.31	0.14
rs378140	7	Non-coding		0.0043'	0.71	0.94
rs367656	7	Non-coding		0.0055	0.77	0.92
rs6977945	7	Non-coding		0.0004**	0.16	0.30
rs17705456	7	Non-coding		0.0123	0.28	0.20
rs2792448	7	Non-coding		0.0031'	0.65	0.62
rs2487161	7	Non-coding		0.0070	0.72	0.10
rs7972894	12	Non-coding		0.0343'	1.31	0.30
rs2393073	12	Non-coding		0.0029'	0.44	0.22
rs11834175	12	Non-coding		0.0158	0.47	0.16
rs10850680	12	Non-coding		0.0286	1.60	0.32
rs11068102	12	Non-coding		0.0074	0.62	0.38
rs7958610	12	Non-coding		0.0226	0.15	0.25
rs4766797	12	Non-coding		0.0239	1.29	0.21
rs7963154	12	Non-coding		0.0089	0.48	0.53
rs10850684	12	Non-coding		0.0077	0.44	0.80
rs12301907	12	Non-coding		0.0247	1.00	0.34
rs7307249	12	Non-coding		0.0212	1.19	0.13
rs903820	12	Non-coding		0.0186	1.48	0.25
rs17500613	12	Non-coding		0.0065	0.68	0.78
rs10850694	12	Non-coding		0.0034'	0.73	0.42
rs7298460	12	Non-coding		0.0031'	0.49	0.36
rs17500655	12	Non-coding		0.0356'	1.08	0.24
rs10507272	12	Non-coding		0.0116	0.83	0.45

rs7974008	12	Non-coding		0.0132	0.41	0.06
rs17428858	12	Non-coding		0.0180	0.61	0.27
rs10774907	12	NOS1	3UTR	0.0394*	1.22	0.90
rs2682826	12	NOS1	3UTR	0.0122	0.62	0.33
rs816363	12	NOS1	intron	0.0113	0.39	0.54
rs2271986	12	NOS1	intron	0.0168	0.54	0.15
rs904658	12	NOS1	intron	0.0261	0.84	0.33
rs3741475	12	NOS1	coding	0.0319'	1.47	0.32
rs7959232	12	NOS1	intron	0.0289	0.98	0.59
rs9658474	12	NOS1	intron	0.0074	0.64	0.79
rs1047735	12	NOS1	coding	0.0048	0.69	0.58
rs845820	12	NOS1	intron	0.0244	0.37	0.83
rs11068428	12	NOS1	intron	0.0053	0.71	0.54
rs1093330	12	NOS1	intron	0.0239	0.36	0.85
rs2293054	12	NOS1	coding	0.0039'	0.13	0.12
rs3741480	12	NOS1	intron	0.0147	0.50	0.70
rs6490121	12	NOS1	intron	0.0052	0.25	0.38
rs1607817	12	NOS1	intron	<0.0001**	0.42	0.22
rs2293052	12	NOS1	intron	0.0007*	0.29	0.52
rs4766842	12	NOS1	intron	0.0206	1.01	0.09
rs9658354	12	NOS1	intron	0.0497**	1.82	0.08
rs7309163	12	NOS1	intron	0.0056	0.35	0.15
rs539291	12	NOS1	intron	0.0007*	0.06	0.28
rs12811676	12	NOS1	intron	0.0133	0.35	0.56
rs530393	12	NOS1	intron	0.0007*	0.17	0.42
rs2077171	12	NOS1	intron	0.0259	0.54	0.78
rs1483757	12	NOS1	intron	0.0220	1.17	0.19
rs9658281	12	NOS1	intron	0.0302	0.37	1.10
rs570234	12	NOS1	intron	0.0050	0.09	0.55
rs545654	12	NOS1	intron	0.0044'	0.48	0.51
rs482555	12	NOS1	intron	0.0106	0.13	0.80
rs693534	12	NOS1	intron	0.0057	0.78	0.49
rs3782219	12	NOS1	intron	0.0149	0.21	0.75
rs3782221	12	NOS1	intron	0.0129	0.23	0.75
rs9658255	12	NOS1	5UTR	0.0432*	1.45	0.38
rs9658253	12	NOS1	5UTR	0.0043'	0.61	0.27
rs4767535	12	NOS1	5UTR	0.0044'	0.77	0.70
rs816292	12	NOS1	5UTR	0.0173	0.77	0.95
rs492623	12	NOS1	5UTR	0.0143	1.05	1.20
rs1552229	12	NOS1	5UTR	0.0256	1.42	1.24
rs178672	17	Non-coding		0.0090	1.16	0.14
rs322202	17	Non-coding		0.0181	0.55	1.61

rs322207	17	Non-coding		0.0167	0.58	1.55
rs1695670	17	Non-coding		0.0130	0.70	1.50
rs1676587	17	Non-coding		0.0105	0.74	1.39
rs322193	17	Non-coding		0.0130	0.70	1.50
rs322187	17	Non-coding		0.0152	0.63	1.58
rs4794888	17	Non-coding		0.0052	0.45	1.36
rs2086033	17	Non-coding		0.0155	0.63	1.59
rs9908619	17	Non-coding		0.0155	0.63	1.59
rs8069120	17	Non-coding		0.0036'	0.06	0.31
rs7215048	17	Non-coding		0.0038'	0.32	1.20
rs11652073	17	Non-coding		0.0022*	0.27	1.05
rs2012824	17	Non-coding		0.0022*	0.27	1.12
rs1963517	17	Non-coding		0.0097	0.17	0.81
rs1382046	17	Non-coding		0.0083	0.11	1.02
rs6505276	17	Non-coding		0.0065	0.82	1.24
rs883127	17	Non-coding		0.0125	0.50	0.77
rs6505292	17	Non-coding		0.0172	0.62	1.23
rs1889024	17	NOS2A	3UTR	0.0281	1.19	0.60
rs953527	17	NOS2A	3UTR	0.0187	1.23	0.46
rs4796017	17	NOS2A	3UTR	0.0186	1.21	0.44
rs8081248	17	NOS2A	3UTR	0.0319'	1.73	0.35
rs2297518	17	NOS2A	coding	0.0112	1.03	0.29
rs2248814	17	NOS2A	intron	0.0345'	1.75	0.43
rs4795067	17	NOS2A	intron	0.0051	0.38	0.15
rs3729508	17	NOS2A	intron	0.0222	1.21	0.68
rs6505469	17	NOS2A	intron	0.0128	0.83	0.16
rs8072199	17	NOS2A	intron	0.0491*	1.55	0.54
rs16949	17	NOS2A	intron	0.0012*	0.58	0.26
rs2301368	17	NOS2A	5UTR	0.0191	1.27	1.16
rs2779251	17	NOS2A	5UTR	0.0056	0.11	0.40
rs2531862	17	NOS2A	5UTR	0.0060	1.12	1.13
rs10853181	17	NOS2A	5UTR	0.0062	1.14	1.12
rs2531869	17	NOS2A	5UTR	0.0180	1.34	0.84
rs2779259	17	NOS2A	5UTR	0.0056	1.10	1.17
rs16966603	17	NOS2A	5UTR	0.0134	0.14	0.55

ST Tab 11- Detection of selection results for TRP markers: Observed (Obs) Fst with selection significance (highlighed); *: P value<0.05; **:P value<0.01.

ID reference	Chr	Repeat Motive	Coding Region	Gene Region	Obs FST
ss825679095	7	27bp VNTR	NOS3	Intron 5	0.0022*
ss825679096	7	(TGTT)n	NOS3	Intron 14	0.0089
ss825679092	7	(CAAAA)n	Non-coding		0.0106
ss825679093	7	(ATTT)	Non-coding		0.0199
ss825679094	7	(AAAAT)n	Non-coding		0.07833**
ss825679082	12	(GAAAA)n	Non-coding		0.0155
ss825679083	12	(ATTT)n	Non-coding		0.0234
ss825679084	12	(ATT)n	Non-coding		0.0098
ss825679079	12	(TTCT)n	NOS1	Upstream	0.0096
ss825679080	12	(GGAT)n	NOS1	Intro18	0.0594**
ss825679081	12	(ATA)n	NOS1	Intro20	0.0426*
ss825679085	17	(GGAA)n	Non-coding		0.0014*
ss825679087	17	(TGGA)n	NOS2A	Upstream	0.0282
ss825679088	17	(CCTTT)n	NOS2A	Upstream	0.0121
ss825679089	17	(TAAA)n	NOS2A	Upstream	0.0099
ss825679090	17	(AAT)n	NOS2A	Intro9	0.0080
ss825679091	17	(AAAAC)n	NOS2A	Downstream	0.0206

