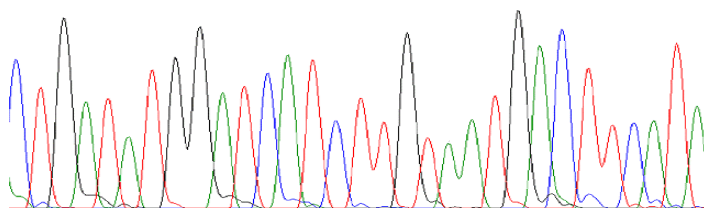


IDENTIFICACIÓ DELS FACTORS GENÈTICS QUE DETERMINEN LA VARIABILITAT DELS NIVELLS DE FVII A LA POBLACIÓ ESPANYOLA

Resultats del Projecte GAIT



MARIA SABATER LLEAL
2006

Tesi Doctoral
Departament de Genètica
Universitat de Barcelona

DISCUSSIÓ

Els estudis realitzats que formen part d'aquesta tesi han investigat la variabilitat genètica en el gen *F7* en població espanyola i l'efecte d'aquesta variabilitat sobre els nivells plasmàtics de la proteïna que codifica, el factor VII de la coagulació (FVII), ja que els nivells elevats de FVII en plasma han estat identificats en diversos estudis com a un potencial factor de risc per malalties cardiovasculars (Eriksson-Berg, et al., 2005; Green, et al., 1994; Meade, et al., 1986; Meade, et al., 1993).

La trombosi venosa i arterial, amb les seves manifestacions clíniques més greus (infart de miocardi, infart cerebral, embòlia pulmonar) són la major causa de mortalitat en les societats occidentals (Endler and Mannhalter, 2003). Concretament, la trombosi venosa afecta anualment 1/1000 individus als països industrialitzats (Bates and Ginsberg, 2004; Rosendaal, 1999).

La tendència genètica a desenvolupar trombosis, la trombofília hereditària, es considera una malaltia multifactorial i complexa, en la que interactuen múltiples gens entre sí i amb els factors ambientals (Bertina, 2001; Rosendaal, 1999). Aquestes interaccions generen en cada individu un gradient de predisposició (susceptibilitat) a la trombosi (Bovill, et al., 1999; Rosendaal, 1999).

Tot i que s'estima que el 61% de la predisposició a la trombosi en famílies espanyoles és atribuïble a factors genètics (Souto, et al., 2000b) només s'identifiquen factors genètics de risc trombòtic coneguts en el 60% de les famílies amb trombofília a Espanya (Mateo, et al., 1997). La resta és encara deguda a factors genètics desconeguts.

En els darrers anys, l'intent de localitzar *loci* en el genoma que determinin risc de patir malalties s'ha convertit en un dels principals reptes de la recerca biomèdica actual. D'aquests, els *loci* que confereixen susceptibilitat a malalties complexes i multifactorials són els més difícils d'identificar, ja que l'estat final de la malaltia depèn de la interacció de múltiples *loci* amb determinants ambientals. En aquesta direcció, un dels objectius més ambiciosos del Projecte Genoma és conèixer tota la variabilitat genètica que determina la susceptibilitat de patir malalties i d'aquesta manera poder arribar a la “medicina personalitzada”, on cada individu és medicat d'una manera individualitzada segons el seu perfil genètic. Per arribar a aquest nivell de coneixement del perfil individual (diferent risc de determinades malalties, diferent tolerància a determinats medicaments) cal analitzar no només els gens que determinen aquest perfil, sinó també arribar a conèixer l'estructura al·lèlica dels gens en qüestió. Òbviament, no tots els al·lels d'un gen determinat seran rellevants per al risc de la malaltia. Arribar a saber quins ho són és un repte important de la biomedicina actual.

Per identificar factors de risc genètic en malalties complexes com la trombofília, s'utilitzen dos tipus diferents d'estratègies; els estudis d'associació i els estudis basats en l'anàlisi de lligament genètic.

Ambdues estratègies tenen avantatges i inconvenients i actualment hi ha un intens debat sobre quina és la millor estratègia.

Encara que hi ha molta bibliografia sobre estudis que han utilitzat anàlisis d'associació entre individus no relacionats per localitzar factors genètics responsables de risc de patir determinades malalties complexes, aquests han donat molt sovint resultats contradictoris (Ioannidis, 2001). Aquest mètode es limita a fer inferències sobre l'associació entre un polimorfisme d'un gen candidat i el fenotip, i depèn del fet que es tracti just del polimorfisme funcional o d'un en desequilibri de lligament amb ell. Amb l'avenç dels mètodes de genotipació a gran escala, ha augmentat molt el nombre d'anàlisis d'associació entre fenotips complexos i polimorfismes triats de forma aleatòria al llarg d'un gen candidat, la qual cosa ha generat un nombre molt important d'errors deguts a l'estratificació de la població, l'organització "caòtica" dels blocs de desequilibri de lligament, o simplement a artefactes estadístics (Weiss, 1993). A més, l'absència d'associació entre un determinat SNP en un gen i la malaltia no significa que el gen no estigui implicat en la malaltia, sinó que el marcador en particular no ho està. Per tant, cal prendre amb certa precaució els resultats obtinguts d'estudis d'associació d'aquest tipus, especialment en l'àmbit de les malalties cardiovasculars (Lane and Grant, 2000).

L'anàlisi de lligament, en canvi, ha demostrat ser una estratègia eficient per la identificació de *loci* de susceptibilitat. Especialment interessants són els estudis de lligament genètic amb fenotips intermediaris que correlacionen amb la malaltia estudiada. Examinant aquesta variació contínua dels fenotips intermediaris, molt més propers a l'acció del gen, obtenim senyals genètics menys atenuats que ens garanteixen molt més poder per localitzar i identificar gens relacionats amb la susceptibilitat de determinades malalties complexes. A més, presenten un clar avantatge respecte de la malaltia final i és que tots els individus de la mostra contribueixen a la informació genètica i no només els afectes.

Un bon exemple de la validesa d'aquest disseny i de la metodologia emprada en aquest treball el constitueix el Projecte GAIT (*Genetic Analysis of Idiopathic Thrombophilia*). El Projecte GAIT es va dissenyar l'any 1995 a partir d'una mostra d'individus relacionats en famílies extenses espanyoles mitjançant el qual s'han pogut determinar les heretabilitats de més de 60 fenotips relacionats amb l'hemostàsia (Souto, et al., 2000b). També s'ha pogut establir que existeix una correlació genètica entre molts d'aquests fenotips i el risc de patir trombosi (Souto, et al., 2000a), probablement perquè estiguin regulats per gens comuns, i el que és molt més rellevant, mitjançant un model matemàtic s'ha pogut estimar l'heretabilitat del risc de patir trombosi en un 61% (Souto, et al., 2000b).

Un dels resultats més rellevants del Projecte GAIT indica que existeix una correlació genètica significativa (0,35) entre els nivells plasmàtics de FVII i la malaltia trombòtica (Souto, et al., 2000a)

demostrant que alguns dels factors genètics que influeixen en la variació d'aquesta proteïna estan influïnt alhora el risc de patir trombosi. Aquest resultat posa de manifest la importància de l'estudi de les variants genètiques que determinen aquest fenotip. L'estudi dels nivells plasmàtics de FVII representa doncs un clar exemple de fenotip intermediari per a l'anàlisi de la trombofília hereditària. El primer pas previ a qualsevol anàlisi genètica és el càlcul de l'heretabilitat del FVII, dada indispensable en l'estudi de la variabilitat genètica de qualsevol proteïna. Si obtenim que una proporció important de la variabilitat de la proteïna és atribuïble a l'efecte dels gens (com en el cas del FVII, que és d'un 53%), té sentit abordar la recerca dels *loci* responsables d'aquesta variabilitat.

Si bé és cert que s'han publicat una gran quantitat d'estudis sobre les variants genètiques en el gen *F7* i dels efectes d'aquestes variants sobre els nivells de la proteïna en plasma (Bernardi, et al., 1996; Marchetti, et al., 1993; Peyvandi, et al., 2000; Pollak, et al., 1996; van 't Hooft, et al., 1999), no existeix cap treball fins ara on s'hagi establert la variabilitat total de la seqüència completa del gen del *F7*. Això ha portat a una generació de resultats contradictoris en els quals els polimorfismes analitzats, triats sovint sense obeir a cap base fisiològica, eren analitzats de forma individual sense tenir en compte l'arquitectura del gen complet, i per tant passant per alt el desequilibri de lligament entre diferents polimorfismes. És important esmentar que cada vegada hi ha més evidències que polimorfismes localitzats en zones no codificants del gen poden tenir una importància cabdal en la determinació fenotípica. Amb l'anàlisi d'un subgrup de variants al·lèliques del gen, sovint triades d'una manera aleatòria, és doncs bastant possible que s'ometin les variants al·lèliques causants del fenotip d'interès, especialment si, com en la majoria de casos, només es consideren les variants al·lèliques localitzades en la regió codificant.

Un dels principals avenços en la identificació de gens causals de malalties o de determinants genètics de fenotips intermediaris per la malaltia en estudi és l'ús de l'Anàlisi Global del Genoma, una tecnologia que ha convertit la identificació de gens humans de susceptibilitat a malalties complexes en una realitat. La primera localització d'un QTL humà utilitzant l'anàlisi global del genoma va ser el descobriment d'un QTL relacionat amb l'obesitat (Comuzzie, et al., 1997) i des de llavors s'han localitzat molts més QTL relacionats amb diferents fenotips complexos com ara la diabetis o la tendència a l'alcoholisme mitjançant aquest mètode (Duggirala, et al., 1999; Williams, et al., 1999). En aquesta tesi es presenten resultats del primer Anàlisi Global del Genoma realitzat en el nostre país, i el primer a nivell mundial en malaltia tromboembòlica. Concretament, els nivells de FVII de la coagulació presenten un senyal únic amb una probabilitat de lligament altament significativa (*Lod score* 3.18) al cromosoma 13, exactament on mapa el gen estructural del FVII, i que ofereix una sòlida base pel cos d'aquesta tesi, l'anàlisi de la variabilitat genètica en el gen estructural del FVII.

Així doncs, el primer resultat destacable d'aquesta anàlisi genètica ha estat la publicació de la seqüència completa del gen *F7* (GeneBank AY212252). Tot i que ja havia estat descrita el 1987 (O'Hara, et al., 1987) la seqüència publicada per aquest grup derivava de dos clons parcialment solapats provinents d'una genoteca. La seqüència obtinguda en el nostre treball presenta nombroses diferències respecte de la publicació original degut al fet que la nostra seqüència prové de les dades de la seqüenciació completa del DNA genòmic en 40 individus no relacionats.

En segon lloc, en aquesta tesi s'ha aconseguit identificar les posicions polimòrfiques del gen *F7*, per primera vegada incloent també l'anàlisi de totes les zones intròniques. L'objectiu de l'estudi de la variabilitat en les seqüències de DNA i la recerca de noves posicions polimòrfiques és el d'aprofundir en el coneixement de la susceptibilitat individual de patir malalties complexes comunes, com en aquest cas la malaltia tromboembòlica. L'estudi de les diferències genètiques entre individus té una aplicació molt clara en epidemiologia genètica: el coneixement de com les diferències genètiques expliquen la diferent susceptibilitat a malalties complexes.

En el present treball, la majoria de les variants identificades són canvis de nucleòtid (SNPs), tot i que també s'han detectat dues repeticions en tàndem de nombre variable (VNTRs), dues insercions i dues delecions.

La proporció de variants no sinònimes en la població estudiada és notablement baixa. De fet, de les 49 variants al·lèliques trobades en la mostra, només dues provoquen canvis no sinònims, la qual cosa suggereix que la variabilitat en els nivells plasmàtics de la proteïna està determinada per zones reguladores localitzades al promotor i als introns del gen, més que per canvis en la seqüència proteica. Aquesta observació és important a l'hora de dissenyar estratègies de localització de variants funcionals. Tot i que la majoria d'estudis només han considerat les zones codificants i les fronteres *d'splicing*, aquests resultats evidencien que la seqüenciació completa del gen és l'única estratègia que garanteix que totes les variants al·lèliques són considerades i que el polimorfisme estudiat és realment el que causa l'efecte fenotípic i no qualsevol altre polimorfisme no considerat però en desequilibri de lligament amb el primer. De fet, amb els resultats obtinguts en aquest treball s'observa que el desequilibri de lligament presenta un patró bastant aleatori, i que no necessàriament es correlaciona amb la distància genètica, fet que ja s'ha observat en altres estudis (Gabriel, et al., 2002; Patil, et al., 2001). Per exemple, s'observen algunes variants (-2989 C/A, 7147G/C) que tot i estar situades a unes 10 Kb de distància, presenten un desequilibri de lligament pràcticament total. Sense genotipar tots els polimorfismes, no és possible arribar a discernir quina de les variants és la causant de l'efecte fenotípic. Aquests resultats són rellevants a l'hora d'interpretar estudis d'associació realitzats a partir

d'un o dos marcadors seleccionats per gen. Com hem esmentat anteriorment, un resultat positiu detectat en aquests tipus d'estratègies pot estar causat per qualsevol altra variant no analitzada.

Un cop s'ha localitzat el gen i s'ha analitzat tota la seva variabilitat, el pas següent és identificar quina o quines de les variants al·lèliques localitzades són les causants de la variació fenotípica.

La nova metodologia estadística utilitzada en aquest treball ha permès estimar quines de les 49 variants al·lèliques (o quin grup de lligament) té una probabilitat major d'explicar la variabilitat fenotípica. L'avantatge d'aquesta metodologia és que considera *tota* la variabilitat del gen, evitant així ometre l'efecte de cap variant al·lèlica. Incorpora models multivariants amb totes les combinacions possibles de polimorfismes, i selecciona aquell que explica millor la variabilitat fenotípica (Blangero, et al., 2003). Com a resultat, s'han pogut estimar per primera vegada els polimorfismes del gen *F7* que expliquen la variabilitat del FVII en plasma (taula 18). De fet, en una nova anàlisi de lligament on s'exclouen aquests polimorfismes del model, el senyal de lligament cau pràcticament a zero, la qual cosa indica que, efectivament, s'han aconseguit identificar totes les variants funcionals que determinen la variabilitat del fenotip estudiat.

Variant al·lèlica	Probabilitat de ser funcional
9830 C/T	>0.999 (0.00004)
-402G/A, 707G/A, -670A/C, -630A/G, 700T/C, 712C/G	>0.999 (0.00016)
10782C/T (A354V)	0.941 (0.0080)
7147G/C, -122T/C, 835T/C, -2989C/A, 73G/A, -323ins0/10, -401G/T, -122T/C	0.931 (0.0096)
6441G/T	0.819 (0.0473)
5506T/A	0.811 (0.0531)
IVS7, 9762T/C	0.655 (0.1160)

Taula 17: Variants al·lèliques funcionals segons l'estimació estadística realitzada mitjançant el *Bayesian Quantitative Trait Nucleotide Analysis*. A la columna de la dreta, probabilitat de cada variant de determinar la variabilitat dels nivells de FVII.

Segons els nostres resultats, doncs, hi ha dos polimorfismes aïllats (que no es presenten en desequilibri de lligament amb cap altre) que clarament afecten els nivells de FVII en plasma. El primer (9830C/T), amb una probabilitat superior al 99% de ser funcional, és sorprenentment un polimorfisme intrònic descrit per primera vegada en aquest treball. El segon, amb una probabilitat del

94%, és un polimorfisme que provoca un canvi d'aminoàcid A→V a la posició 354 i que ha estat descrit en diversos casos de deficiència de FVII (Bernardi, et al., 1994; Toso, et al., 2002).

A més, també hi ha dos grups de polimorfismes que, degut al fort desequilibri de lligament entre ells, no és estadísticament possible separar-los. En cada un d'aquests grups, un o més polimorfismes són els responsables de l'efecte fenotípic, i calen experiments *in vitro* per tractar-los individualment i demostrar quin o quins dels polimorfismes dins de cada grup és el realment responsable d'aquest efecte.

El primer grup conté polimorfismes tant del promotor com del primer intró i té una probabilitat superior al 99% d'influir en els nivells de FVII. S'ha descrit que l'intró 1 de les proteïnes dependents de vitamina K conté també zones reguladores, per la qual cosa té sentit pensar que qualsevol d'aquests polimorfismes podria estar alterant la unió d'alguna proteïna reguladora de l'expressió gènica (Shamsher, et al., 2000).

El segon grup conté polimorfismes del promotor i de l'intró 1, i també un polimorfisme de l'intró 5.

D'aquestes variants identificades, ja se'n coneixien algunes dades. Dels polimorfismes localitzats a les posicions -402 i -401 del promotor s'ha descrit que influeixen la unió de dos factors de transcripció encara desconeguts, la qual cosa modifica els nivells de transcripció, en el primer cas augmentant-los i en el segon cas reduint-los (van 't Hooft, et al., 1999). La inserció del decanucleòtid a la posició -323, també en el promotor, és probablement el polimorfisme més estudiat en el gen *F7* i diversos estudis d'expressió indiquen que redueix l'activitat del promotor en un 33% (Bernardi, et al., 1996; Pollak, et al., 1996). A més dels estudis *in vitro*, nombrosos estudis d'associació han analitzat l'efecte de determinats polimorfismes del promotor en els nivells plasmàtics de FVII. Tot i així, i com s'ha esmentat anteriorment, la majoria d'estudis han donat fins ara resultats bastant contradictoris: mentre que alguns estudis demostren l'efecte d'un polimorfisme, altres estudis no aconsegueixen demostrar variacions en els nivells d'expressió. Aquest fenomen s'explica pel fort desequilibri de lligament que existeix entre aquests polimorfismes, que dificulta molt la identificació de l'efecte de cada polimorfisme de forma individual.

Per evitar aquest problema, altres autors han optat per estudiar l'efecte dels grups de lligament majoritaris en la població, tant en estudis poblacionals com en estudis d'expressió *in vitro*. Segons aquests estudis, apart de l'haplotip majoritari, que presenta tots els al·lels "normals", s'ha estudiat l'haplotip que conté les variants -670C/-630G/-402A, associat a un increment significatiu dels nivells de FVII (Carew, et al., 2003) i el que conté les variants -323ins10 /-401T /-122C que redueix significativament l'activitat del promotor (Dell'Acqua, et al., 1997; van 't Hooft, et al., 1999). Aquestes observacions corroboren els resultats obtinguts en l'estudi dels haplotips del present treball, tot i que no aprofundeixen en el coneixement de l'efecte individual de cada polimorfisme. Qualsevol

dels polimorfismes de cada grup de lligament (o una suma dels efectes de tots ells) podria ser el responsable de la variació en els nivells de transcripció.

D'altra banda, fora de la zona del promotor, també existeixen nombrosos estudis sobre el VNTR a la posició 9707 de l'intró 7 (IVS7), que ha estat associat amb els nivells d'expressió de l'mRNA (Pinotti, et al., 2000) així com amb els nivells de FVII antigènic i funcional (Di Castelnuovo, et al., 1998; Iacoviello, et al., 1998) tot i que la probabilitat de ser funcional no és, segons el nostre treball, significativa.

Pel que fa a les variants que provoquen un canvi d'aminoàcid, el canvi A354V (Hunault, et al., 1997) [A294V segons la nomenclatura original (O'Hara, et al., 1987)] al nucleòtid 10782, ha estat identificat en diferents estudis en famílies amb deficiència lleu de FVII (Bernardi, et al., 1994; Toso, et al., 2002; Wulff and Herrmann, 2000).

Curiosament, el canvi aminoacídic R413Q [R353Q segons la nomenclatura d'O'Hara (1987)] situat al nucleòtid 10959, ha estat associat amb variacions dels nivells antigènics i funcionals de FVII en múltiples estudis (Eriksson-Berg, et al., 2005; Feng, et al., 2000; Girelli, et al., 2000; Humphries, et al., 1996) tot i que segons els resultats obtinguts en aquest treball, no sembla que tingui un efecte sobre els nivells fenotípics de FVII. Aquest polimorfisme, però, presenta un desequilibri de lligament molt fort amb els polimorfismes del promotor -323ins10 /-401T /-122C, per la qual cosa l'efecte sobre els nivells plasmàtics de FVII podria ser degut a qualsevol altre d'aquests polimorfismes. De fet, en un estudi *in silico* de genòmica comparativa (Figura 17) s'aprecia que la posició 413 està molt poc conservada entre espècies, suggerint que el canvi podria ser irrellevant per la funcionalitat de la proteïna.

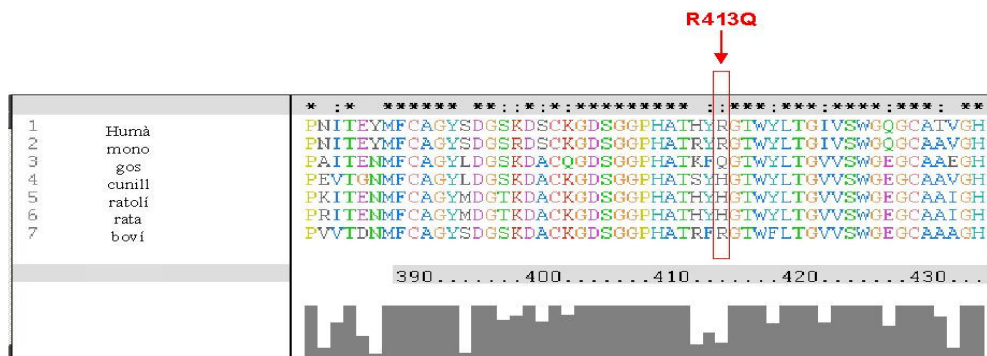


Figura 17: Anàlisi comparativa entre diferents mamífers en la posició aminoacídica 413 del FVII.

Un altre resultat interessant d'aquest treball és l'estudi de l'organització haplotípica de la variabilitat en el gen *F7* en els individus de la població espanyola estudiada, i l'anàlisi dels possibles processos evolutius que l'han pogut generar i que han fet que es mantingués en la nostra població.

Una de les dades rellevants és el valor de l'estadístic *D* de Tajima obtingut. Aquest estadístic bàsicament analitza el grau de polimorfisme de la població, i calcula si és o no significativament diferent del que seria esperat en condicions de neutralitat, és a dir, en absència de selecció. El valor de la *D* de Tajima de la nostra mostra (1,13, $p=0,035$) rebutja la hipòtesi nul·la de neutralitat i indica que hi ha un excés de posicions polimòrfiques en la població, o el que és el mateix, un excés d'al·lels amb freqüències intermèdies. Aquest resultat generalment s'interpreta com a un efecte de selecció equilibradora, molt probablement deguda a un avantatge dels heterozigots.

A partir de les dades genotípiques de tots els individus de la mostra, i degut al fet que es tracta de famílies extenses, es poden analitzar les fases dels polimorfismes i s'han pogut resoldre 212 dels 245 cromosomes independents de la mostra. Les possibles relacions evolutives entre els haplotips indica que hi ha tres grups d'haplotips (Hp) clarament diferenciats: Hp1 amb el 53,8% dels haplotips, Hp2 amb el 28,8% i Hp3 amb el 12,5% (figura 18). A més, la data d'expansió en la població d'aquests grups d'haplotips es va poder estimar, i van ser de 40.000 ± 17.000 anys per l'Hp1, 36.000 ± 12.000 per l'Hp2, i 20.000 ± 10.000 anys per l'Hp3. Segons aquestes dades, el temps d'expansió dels Hp1 i 2

podria coincidir amb la colonització d'Europa pels humans moderns, encara que caldria una mostra amb més representació geogràfica per discutir millor aquesta hipòtesi.

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Figura 18: Haplotips obtinguts en la mostra GAIT. Els polimorfismes que defineixen els Haplogrups 2 i 3 estan marcats en color, i coincideixen amb els polimorfismes estimats com a funcionals en les anàlisis estadístiques anteriorment esmentades (taula 17).

Sota una hipòtesi de selecció equilibradora, els nostres resultats indiquen que els tres haplogrups han estat mantinguts alhora en la nostra població. Per explicar aquest resultat, vam estudiar la relació de cada haplogrup amb els nivells plasmàtics de FVII funcional. Els resultats obtinguts són un bon exemple per explicar com un fenotip implicat en malaltia complexa pot estar modulant les

freqüències haplotípiques mitjançant la selecció natural; es va trobar que l'Hp2 estava format principalment per individus amb els nivells alts de FVII, i l'Hp3 format principalment per individus amb els nivells baixos de FVII. Aquest fenomen no només demostra que els haplogrups tenen una clara contribució sobre els nivells de FVII sinó que, a més, corroboren els resultats obtinguts en l'anàlisi estadística dels polimorfismes del gen. Així, l'Hp2 conté el grup d'al·lels funcionals amb una probabilitat superior al 99% de contenir una variant que influeixi el fenotip elevant-ne els nivells (en la figura 18 marcats en verd), i l'Hp3 conté el grup d'al·lels funcionals amb una probabilitat superior al 93% de disminuir els nivells de FVII en plasma (en la figura 18 marcats en groc).

Intentar explicar com l'efecte del fenotip ha pogut modular les freqüències haplotípiques en la població no és fàcil, però sembla raonable hipotetitzar que els haplotips que confereixen nivells alts de FVII s'han mantingut en la població gràcies al seu efecte protector sobre un excés d'hemorràgies degudes al part, la menstruació o les ferides, com s'ha demostrat per altres polimorfismes protrombòtics (FVL, PT20210) (Corral, et al., 2001; Lindqvist, et al., 1998; Lindqvist, et al., 2001). Per altra banda, el fenomen evolutiu que ha mantingut els haplotips que confereixen nivells baixos de FVII és més difícil d'explicar, tot i que és plausible pensar que, d'alguna manera, la protecció envers una situació protrombòtica també ha pogut modular la història evolutiva del gen.

D'altra banda, fruit de la seqüenciació del gen *F7* en quatre primats (ximpanzés, bonobos, gorilles i orangutans) es van poder establir els estats ancestrals dels polimorfismes detectats en humans. La incorporació de l'haplotip ancestral en l'arbre filogenètic el situa pràcticament al punt de separació dels haplogrups Hp1, Hp2 i Hp3.

La comparació d'aquestes seqüències també va permetre comparar el nombre de canvis no sinònims (K_a) *versus* canvis sinònims (K_s) per a cada parella d'espècies. Un excés de canvis no sinònims (canvis d'aminoàcids) seria indicatiu d'una pressió de selecció positiva que afavoriria un canvi en la proteïna. Per contra, un excés de canvis sinònims indicaria una pressió de selecció purificadora, on els canvis d'aminoàcid són seleccionats en contra, i per tant la proteïna es manté pràcticament inalterada al llarg de l'evolució. En el nostre cas, amb un K_a/K_s de 0.2, s'indica que el FVII ha estat sotmès a una pressió purificadora que ha mantingut la cadena aminoacídica del FVII en el llinatge que porta als humans, fet que no és gens sorprenent tenint en compte la rellevància de la funció d'aquesta proteïna en l'inici de la cascada de la coagulació. De fet, s'han descrit nombroses mutacions que alteren l'estructura o la funció de la proteïna i que resulten en tendències hemorràgiques (McVey, et al., 2001). A més, l'absència completa de FVII és incompatible amb la vida, fet que s'ha observat tant en mutacions humanes (McVey, et al., 2001) com en experiments en ratolins (Rosen, et al., 1997).

Així doncs, aquest treball suggereix un clar efecte de la selecció natural sobre el procés evolutiu que ha portat a la variabilitat actual observada en el gen *F7*. Aquesta selecció actuarà mitjançant la diferent eficàcia biològica dels individus amb diferents nivells de FVII en sang.

Aquest estudi demostra, a més, la importància d'establir els haplotips d'una determinada regió genòmica d'interès a l'hora de realitzar estudis d'associació. Malgrat el creixent interès en els últims anys de crear un mapa dels patrons de desequilibri de lligament en el genoma (HapMap), és imprescindible la seqüenciació de la totalitat dels gens candidats analitzats per establir els haplotips i refinar la informació dels grans blocs de lligament definits en l'HapMap. La utilització dels haplotips en lloc de polimorfismes individuals a l'hora de fer estudis d'associació evita molts dels errors produïts en aquests tipus d'estudis. D'altra banda, només així, i mitjançant l'anàlisi evolutiva de la relació entre els diferents haplogrups, es pot crear una hipòtesi de la relativa importància que han tingut les diferents variants al·lèliques sobre el fenotip, com ha pogut ser demostrat en altres treballs similars. Per exemple, mitjançant l'anàlisi dels haplotips en el gen de l'angiotensinogen (Fejerman, et al., 2004) s'ha pogut establir la relació entre cada haplogrup i els nivells de la proteïna en plasma, i s'ha pogut concloure que els haplogrups associats a nivells baixos de la proteïna han estat afavorits per la selecció per causar una disminució en la pressió arterial.

En conseqüència, tant mitjançant la genètica de poblacions i l'estudi dels haplotips com mitjançant models estadístics per explicar la funcionalitat dels polimorfismes, arribem a les mateixes conclusions: hi ha dos grups d'haplotips comuns a la població que determinen un augment o una disminució dels nivells plasmàtics de FVII. A més, existeixen també variants rares que no formen part d'un haplotip concret que també poden tenir un efecte molt notable sobre els nivells de FVII. Cal destacar que realitzant només una anàlisi dels haplotips és possible detectar les variables comunes responsables de la variabilitat en els nivells de FVII, però que aquest tipus d'anàlisi no permet detectar l'efecte de les variants rares. En el nostre cas, les variants al·lèliques 9830C/T i 10782C/T no haurien estat localitzades. De tota manera, segons la teoria més acceptada actualment de *Common Disease/Common variant* (CD/CV), les malalties comunes són degudes a al·lels ancestrals, comuns en la població, que interaccionen per causar la malaltia (Reich and Lander, 2001). Assumint aquesta teoria, queda clar que és molt important analitzar l'estructura al·lèlica de la població i la història evolutiva del gen d'interès per detectar els haplotips ancestrals que determinen susceptibilitat per la malaltia en la població estudiada.

A més, malgrat que aquestes anàlisis han permès localitzar els haplotips responsables de la variabilitat fenotípica en la nostra població, l'estructura al·lèlica és específica de cada població fruit d'una història evolutiva diferent. Calen doncs estudis funcionals d'expressió *in vitro* per confirmar la

funcionalitat de les variants trobades i per poder discernir entre els polimorfismes compresos dins d'un grup de desequilibri de lligament quin és el realment responsable.

Com a part d'aquesta tesi, els estudis funcionals d'expressió *in vitro* demostren l'efecte individual de cada un dels polimorfismes del promotor sobre els nivells de FVII. D'acord amb els nostres resultats, les variants al·lèliques –323ins10 i –122C, redueixen significativament els nivells d'expressió *in vitro* de forma individual, i són responsables dels baixos nivells d'expressió de FVII que presenten els individus amb l'haplotip que les conté. Aquestes variants poden ser utilitzades com a marcadors genètics de nivells reduïts de FVII en qualsevol població, i podrien ser marcadors útils de factors genètics protectors de risc cardiovascular.

D'altra banda, la variant al·lèlica –2989A, descrita en aquest estudi per primera vegada, causa un augment significatiu dels nivells d'expressió *in vitro* quan s'estudia el seu efecte de forma individual. No obstant, aquesta variant al·lèlica presenta un desequilibri de lligament total amb les variants –323ins10 i –122C en la nostra població. Com s'ha esmentat anteriorment, aquest haplotip redueix significativament els nivells d'expressió del gen i emmascaren totalment l'efecte de la variant –2989A. Aquest resultat indica una vegada més que s'ha de tenir molta cautela a l'hora de dissenyar estudis d'associació entre un fenotip i variants al·lèliques triades sovint sense saber-ne l'efecte molecular i que és important conèixer *tota* la variabilitat genètica del gen estudiat per poder testar l'efecte de cada polimorfisme. En aquest cas, si s'hagués utilitzat el polimorfisme –2989C/A en un estudi d'associació amb els nivells de FVII en població espanyola, hauria resultat associat a una reducció d'aquests nivells, quan en realitat el seu efecte és totalment contrari. A més, també cal tenir en compte la importància d'analitzar l'efecte de les combinacions de polimorfismes. Així com el resultat de la combinació de les variants –2989A i –122C indica una acció dominant de la primera variant, i resulta en nivells d'expressió clarament més alts que la construcció WT, la combinació de la mateixa variant –2989A amb la variant –323ins10 indica una clara dominància de la segona, i per tant resulta en nivells pràcticament iguals als que presenta la inserció individualment.

De la mateixa manera, la variant al·lèlica –401T, també provoca un augment en els nivells d'expressió que queda emmascarat per l'efecte global de l'haplotip en el què apareix.

Per últim, la variant al·lèlica –402A també provoca un augment significatiu de l'expressió *in vitro*. Aquest polimorfisme es troba en desequilibri de lligament amb les variants –670C i –630G i formen l'haplotip relacionat amb un augment dels nivells d'expressió. D'acord amb els nostres resultats, aquest augment seria degut a l'efecte de les variants –402A i –630G, ambdues relacionades amb un augment de l'expressió *in vitro*. A més, aquest haplotip ha estat relacionat amb un augment

significatiu de risc cardiovascular, el qual suggereix que aquests polimorfismes podrien ser usats com a marcadors de risc cardiovascular.

No obstant, cal recordar que tot i que s'ha demostrat que aquestes variants del promotor tenen un efecte sobre els nivells finals de proteïna, això no vol dir que siguin els únics determinants genètics responsables de la seva variabilitat. Caldrien estudis funcionals en els altres polimorfismes de cada haplotip, així com en els polimorfismes individuals detectats en l'anàlisi estadística amb alta probabilitat de ser també responsables d'aquesta variabilitat en els nivells de FVII. Finalment, caldria també estudiar l'efecte de les interaccions entre aquestes variants al·lèliques.

Per cloure l'estudi de la variabilitat genètica en el gen *F7* és interessant estudiar l'associació d'aquesta variabilitat amb la variabilitat dels nivells de FVII en altres mostres, i estudiar també la seva implicació en la malaltia tromboembòlica. Per això, es presenta un estudi d'associació cas-control en una mostra de pacients amb trombofília on s'analitza l'efecte de dos SNPs, cadascun d'ells representant d'un dels haplotips descrits anteriorment com a responsables de la variació dels nivells de la proteïna en la mostra GAIT (*tag* SNP). Cal destacar que els resultats obtinguts en aquesta darrera anàlisi confirmen els obtinguts anteriorment, tant amb l'anàlisi estadística com amb l'anàlisi dels haplotips i l'estudi funcional pel que fa a la relació entre els polimorfismes del *F7* i els nivells plasmàtics de proteïna. Així doncs, es demostra una vegada més la validesa de l'estratègia i es certifica que, una vegada localitzats els polimorfismes funcionals i els blocs de lligament, els resultats són reproduïbles en diferents estudis.

No obstant això, l'anàlisi d'associació entre els polimorfismes del *F7* i el risc trombòtic no dona significativa, ni tampoc és significativa l'anàlisi entre els nivells de FVII i l'esmentat risc, fet que coincideix amb resultats obtinguts per altres autors (Koster, et al., 1994, Corral, et al., 1998). Probablement, aquest resultat negatiu sigui degut a una manca de poder estadístic ocasionat per una mida mostral massa petita. La malaltia tromboembòlica és un fenotip complex, amb la qual cosa és raonable pensar que l'efecte del FVII tingui una aportació massa petita sobre el fenotip final (la malaltia tromboembòlica) com per ser detectada amb la mida mostral utilitzada en el nostre estudi. De fet, s'ha suggerit ja per altres factors genètics amb una contribució modesta en el fenotip final, que caldrien milers d'individus en estudis cas-control per poder-ne obtenir dades significatives (Lohmueller, et al., 2003). Per obtenir aquestes mides mostrals, probablement caldran col·laboracions entre varis grups d'investigació per acabar de clarificar l'efecte del FVII en el risc trombòtic.

Finalment, en el darrer treball d'aquesta tesi es presenten dues famílies amb deficiència hereditària de FVII que, tot i presentar el mateix genotip per les mutacions que els causen la deficiència, presenten una notable heterogeneïtat fenotípica. Els resultats de la seqüenciació del gen del FVII en aquestes famílies suggereix que, tot i que les mutacions rares trobades els causen una deficiència pràcticament total de FVII, existeixen altres polimorfismes en aquests individus que modulen els nivells d'aquesta proteïna igual que els modulen en els individus de la població normal. Aquests polimorfismes són probablement part dels que controlen el delicat equilibri de l'hemostàsia, i que determinen també un major o menor risc de malaltia tromboembòlica.

A més, els resultats també suggereixen que, al contrari del que està descrit fins al moment, la deficiència de FVII no és una malaltia mendeliana sinó una malaltia complexa on els factors ambientals i el perfil genètic de cada individu determina la diferent expressivitat clínica de la malaltia, amb una major o menor tendència a l'hemorràgia tot i compartir la mateixa mutació greu que la causa.

Concloent, la suma dels efectes de tots els polimorfismes d'un individu donaran uns nivells finals de FVII que, al seu torn, i en combinació amb tots els altres factors de risc cardiovascular, situaran cada individu en un punt concret dins l'escala de risc cardiovascular. L'absència total de FVII, d'altra banda, no està causada per la combinació dels efectes de polimorfismes comuns en la població sinó que és deguda a mutacions rares que no permeten la síntesi de proteïna funcional. En aquests casos, es parla de deficiència de FVII i el resultat solen ser greus problemes d'hemorràgia. Tot i així, l'expressivitat variable de la malaltia quedarà influïda pel *background* genètic de cada individu, format per les diferents variants funcionals, cada una amb un efecte additiu sobre el fenotip.

Les variants funcionals trobades tenen un especial interès epidemiològic, ja que són excel·lents marcadors per quantificar el risc de la malaltia estudiada en la població general. Arribar a conèixer tots els factors genètics que contribueixen al desenvolupament dels fenòmens trombòtics ajudarà al disseny d'estratègies de tractament i prevenció basades en el perfil genètic de cada individu.

Finalment, cal esmentar que un dels resultats més rellevants d'aquest treball és la demostració que l'estratègia utilitzada permet disseccionar amb èxit un QTL i localitzar, d'entre totes les variants, aquelles que realment tenen un efecte funcional sobre un fenotip complex com és la malaltia tromboembòlica.

De la mateixa manera, altres estudis recents han verificat l'eficàcia d'aquesta estratègia. Un bon exemple dels passos a seguir per la completa identificació de les variants al·lèliques responsables d'un determinat fenotip intermediari queda exposat en un treball recent sobre la variabilitat genètica de la selenoproteïna S en relació amb la malaltia inflamatòria crònica (Curran, et al., 2005). En primer lloc,

coneguda una important heretabilitat del fenotip d'interès, es realitza una anàlisi global del genoma per localitzar el gen responsable. Donada la presència d'un gen candidat en la zona de lligament, es procedeix a la identificació de *totes* les variants al·lèliques en aquest gen en un subgrup de la mostra i, a continuació, es genotipen les variants localitzades al total de la mostra. Mitjançant anàlisis estadístiques (BQTN) se seleccionen les variants amb més probabilitat de tenir un efecte funcional sobre el fenotip. Per comprovar el seu efecte, es realitzen estudis funcionals *in vitro* i, finalment, es realitzen estudis d'associació amb les variants funcionals en diferents poblacions per analitzar la importància de la mutació en relació amb la malaltia d'interès en cada població.

Actualment hi ha molt de debat respecte a les estratègies més adequades a l'hora de cercar *loci* que afectin a fenotips complexes, especialment si són factors de risc de malalties comunes en la població com les malalties cardiovasculars. La gran millora en les tècniques de genotipació, com s'ha esmentat anteriorment, ha permès la realització d'anàlisis globals del genoma d'associació, que, a diferència dels estudis de lligament, utilitzen poblacions i no individus emparentats, i comparen els genotips d'un gran nombre de marcadors polimòrfics en individus que manifesten una malaltia respecte d'un grup de controls. Aquest tipus d'estudi és més sensible que els estudis de lligament genètic a l'hora de detectar *loci* amb un efecte baix sobre el fenotip i és capaç d'acotar una regió genòmica més petita. No obstant això, cal definir una bona estratègia per l'elecció dels marcadors per analitzar. El patró aleatori del desequilibri de lligament afegeix molta dificultat a l'hora de seleccionar aquests marcadors. Actualment, la metodologia d'elecció dels marcadors no assegura la localització de les variants funcionals (o en desequilibri de lligament amb elles), sense les quals els resultats obtinguts en l'estudi d'associació s'han de prendre amb certa cautela.

Amb aquesta tesi, queda palès que l'estratègia emprada per a l'anàlisi de la variabilitat dels nivells de FVII pot ser utilitzada com a model per analitzar QTLs en qualsevol altre fenotip complex d'interès.