Hospital de la Santa Creu i Sant Pau

Facultat de Medicina

Departament de Medicina

Universitat Autònoma de Barcelona

BASES GENÉTICAS DE LOS COMPONENTES DE LA HEMOSTASIA Y DEL RIESGO DE ENFERMEDAD TROMBOEMBÓLICA

Tesis presentada por

Juan Carlos Souto Andrés

Para optar al grado de Doctor en Medicina y Cirugía

Barcelona, Mayo de 2001

JORDI FONTCUBERTA BOJ, Jefe de la Unitat d'Hemostàsia i Trombosi del Hospital de la Santa Creu i Sant Pau de Barcelona

CERTIFICA

Que la tesis "BASES GENETICAS DE LOS COMPONENTES DE LA HEMOSTASIA Y DEL RIESGO DE ENFERMEDAD TROMBOEMBOLICA" presentada por Juan Carlos Souto Andrés para acceder al grado de Doctor en Medicina y Cirugía, ha sido realizada bajo mi dirección y se halla en condiciones de ser leída

Barcelona, 20 de febrero de 2001

A mi madre, Montserrat

A mis hermanos, Ramón, Jordi y Mª José

A mis sobrinos, Andrea y Victor

Al profesor William Stone

AGRADECIMIENTOS

Al Dr. Jordi Fontcuberta, por la confianza depositada en mí desde hace más de 10 años y por haber compartido su enorme experiencia clínica y científica en tantos proyectos. Responsable último del proyecto GAIT y Director de esta Tesis, su apoyo institucional como responsable de la Unitat d'Hemostàsia i Trombosis del Hospital de la Santa Creu i Sant Pau ha sido crucial para que hayamos llegado donde hoy nos encontramos.

Al profesor William Stone, por abrirnos a todos nosotros las puertas de la Genética. Por su amistad y su magisterio sin condiciones, que desde el primer momento me obligaron a poner lo mejor de mí en esta investigación. De todas las deudas de gratitud que aquí pongo por escrito, la suya es la única impagable.

A John Blangero, por la enorme generosidad con que ha compartido conmigo y el resto de nuestro equipo su ingente sabiduría estadística. Los resultados más importantes de la Tesis son, en primer lugar, mérito de la impresionante metodología matemática e informática que John Blangero y otros pocos estadísticos eminentes están poniendo al servicio de toda la comunidad científica mundial.

A Laura Almasy, por haber realizado personalmente los cálculos que conducen a nuestros resultados. Por sus enseñanzas sobre estas técnicas realmente complicadas y por su imprescindible colaboración en la escritura de los artículos publicados.

Al Dr. Miquel Rutllant, Cap del Departament d'Hematologia de l'Hospital de la Santa Creu i Sant Pau. Por aceptar la tutoría de la Tesis y por haber favorecido y estimulado el desarrollo del proyecto GAIT en nuestro Centro.

Al Dr. José Manuel Soria y a Alfonso Buil, entre otros muchos motivos, por su entusiasmo y por compartir conmigo la aventura intelectual del descubrimiento de la Genética de Rasgos Complejos. La ayuda de ambos en la revisión del manuscrito me ha sido también muy valiosa.

A la Dra. Montserrat Borrell por haber dirigido la realización de todos los fenotipos relacionados con la Hemostasia y al Dr. José Mateo por su participación en el diseño clínico y el reclutamiento de pacientes.

A los técnicos de laboratorio de la Unitat d'Hemostàsia i Trombosi, del Servei de Bioquimica y del Banc de Sang del Hospital de la Santa Creu i Sant Pau por el esfuerzo que supuso la determinación de los fenotipos estudiados en la Tesis, tarea que sobrepusieron al duro trabajo cotidiano propio de un gran Hospital: Teresa Urrutia, Rosa Felices, Cristina Vallvè, Isabel Tirado, Imma Coll, Elisabeth Martínez, Dolors Llobet, Mercè Garí, Joaquín Murillo, Laia Bayén, Rosa Mª Arcelús, Pilar Santo Domingo, Marina Arilla y Neus Boto.

A las administrativas de la Unitat d'Hemostàsia i Trombosi, Mª Jesús Gallego y Beatriz Carreras por la contribución métodica y ordenada en la atención a las familias y almacenamiento de datos.

A Laura Domingo, Maite Royo, Asunción Petitbó y de nuevo a Rosa Felices, compañeras en los laboratorios del Hospital de Sant Pau por la colaboración en el reclutamiento de las familias control.

Al Dr. Eduardo Muñiz-Díaz y al Dr. Pedro Madoz, del Banc de Sang, por poner a nuestra disposición personal y métodos de su laboratorio para la realización del fenotipo eritrocitario extenso y del genotipo ABO.

Al Dr. Francisco Blanco y al Dr. Jordi Ordoñez, del Servei de Bioquimica por sus sugerencias y por las determinaciones en su laboratorio para el estudio de homocisteína, vitaminas y lípidos.

A todos los componentes del Servei d'Extraccions del Hospital de Sant Pau, por la eficacia con que se realizó la obtención de muestras sanguíneas de más de 400 personas analizadas, a pesar de la interferencia que supuso el estudio en su actividad habitual durante 2 años.

A cualquier otra persona que se sienta en mayor o menor grado unida a este proyecto, y que por una omisión involuntaria no figure en esta relación.

Y para terminar, mi agradecimiento más profundo a las 22 familias que generosamente han participado en este proyecto. Como representantes de la comunidad a la que todo esfuerzo científico debería servir, ellas son a la vez el origen y el fin último de nuestro trabajo. Por razones deontológicas este recuerdo debe ser anónimo.

ÍNDICE

Algo, que ciertamente no se nombra con la palabra azar, rige estas cosas

J. L. Borges, *Poema de los dones*.

GLOSARIO DE TERMINOS Y ABREVIATURAS

Alelo: cada una de las posibles formas alternativas de un gen concreto o de una secuencia de ADN en un lugar cromosómico específico (*locus*). En cada *locus* autosómico un individuo posee dos alelos, uno heredado del padre y el otro de la madre.

AVC: accidente vascular cerebral.

Correlación: concepto estadístico que indica la asociación entre dos variables cuantitativas. El parámetro clásico para estimar el grado de correlación lineal es el coeficiente de correlación de Pearson (ρ).

Desequilibrio de ligamiento (δ): asociación no aleatoria, observada en una población, de alelos en dos o más *loci* ligados. El desequilibrio de ligamiento disminuye cuando aumenta la distancia (y por tanto la recombinación) entre los *loci*.

Deriva genética: es la fluctuación azarosa de las frecuencias de los alelos o los genes que se observa por los errores de muestreo. Ocurre en cualquier población pero sus efectos, por ejemplo provocando desequilibrio de ligamiento, pueden ser muy evidentes en poblaciones pequeñas.

Epistasis: efecto fenotípico resultado de la interacción de dos o más genes distintos.

Fenotipo: característica física de un organismo, observable y mensurable, producida por su constitución genética (genotipo) en combinación con el ambiente.

Fenotipo complejo: cualquier fenotipo cuya expresión es influida por múltiples genes y por uno o más factores ambientales.

Frecuencia alélica: porcentaje de un alelo específico del total de los alelos posibles observado en un *locus* concreto y en una población determinada.

Gen: segmento del ADN que codifica un polipéptido funcional o un producto de ARN. La secuencia de ADN que constituye un gen incluye intrones, exones y regiones regulatorias.

Genotipo: constitución genética de un individuo. También se aplica al tipo de alelos que presenta un individuo en un *locus* concreto.

Haplotipo: serie de alelos en varios *loci* ligados en un mismo cromosoma.

Heredabilidad (h^2): proporción de la variancia fenotípica total que se atribuye al efecto de los genes.

HRG: de *histidine-rich glycoprotein*, glicoproteína rica en histidina.

IBD: del inglés *identity by descent* (identidad por descendencia); Dos alelos en un *locus* de un individuo o en dos personas emparentadas son idénticos por descendencia cuando han sido heredados de un familiar antecesor común. El número de alelos IBD compartidos por dos individuos emparentados (0,1 o 2) es uno de los parámetros básicos que usa el Análisis de Ligamiento en los métodos de componentes de la variancia.

Ligamiento: tendencia de los genes u otras secuencias de ADN en *loci* específicos a ser heredados juntos como consecuencia de su proximidad física en un cromosoma.

LOD: logaritmo de la *odds ratio* que se usa como medida de la certeza estadística para la hipótesis de que dos *loci* están ligados. La *odds ratio* es el cociente entre dos verosimilitudes o probabilidades: probabilidad de que un conjunto específico de datos refleje el ligamiento entre los dos *loci, versus* la probabilidad de que esos mismos datos reflejen la ausencia de ligamiento.

Locus: una posición en la secuencia del ADN, definida con relación a las otras. En contextos diferentes puede significar un sitio polimórfico específico o una gran región de la secuencia del ADN en dónde se puede localizar un gen. En latín, el plural es **loci**.

Microsatélite: un segmento corto de ADN que contiene repeticiones de 2 a 5 pares de bases y que presenta polimorfismo en el número de repeticiones. Estos *loci* también se llaman STR (de *short tandem repeats*) y se usan como marcadores anónimos para localizar posiciones en los mapas cromosómicos y en el Análisis de Ligamiento.

Monogénico: un rasgo fenotípico es monogénico cuando es influido primariamente o enteramente por sólo un *locus* genético.

Odds ratio (OR): cociente o razón entre 2 *odds*. A su vez la *odds* es una razón en la que el numerador es la probabilidad (p) de que ocurra un suceso y el denominador es la probabilidad de que tal suceso no ocurra (1-p). La OR es una medida clásica en Epidemiología de la asociación entre un factor de riesgo y una enfermedad *(Martínez-González y col 1999)*.

PAI-1: del inglés *plasminogen activator inhibitor*, inhibidor del activador del plasminógeno tipo 1.

PCR: del inglés *polymerase chain reaction*, reacción en cadena de la polimerasa.

Pleiotropía: fenómeno por el cual un único gen influye en varios fenotipos distintos.

Poligénico: un fenotipo es poligénico si se ve influido por múltiples genes de efectos individuales relativamente pequeños, de modo que la influencia de cada *locus* individual es muy difícil o imposible de detectar por sí sola.

Polimorfismo: la existencia de dos o más variantes en un *locus* determinado. El polimorfismo es **neutro** si las distintas variantes no causan diferencias en ningún fenotipo. Por el contrario, el polimorfismo es **funcional** si a cada variante le corresponden valores distintos de un fenotipo.

QTL: del inglés *quantitative trait locus*; cualquier *locus* que influye sobre la variabilidad de un fenotipo complejo.

RPCa: resistencia a la proteína C activada.

TFPI: de *tissue factor pathway inhibitor*, inhibidor de la via del factor tisular.

t-PA: de *tissue plasminogen activator*, activador tisular del plasminógeno.

TTPA: tiempo de tromboplastina parcial activado.

TVP: trombosis venosa profunda.

Variancia (σ^2): dispersión de los valores de una variable alrededor de la media. Cuando esta variable es un fenotipo complejo que se mide en una escala cuantitativa continua, se observa la variancia fenotípica (σ^2_p). La variabilidad entre los individuos que se refleja en la variancia fenotípica tiene dos orígenes; uno es genético y se debe a las diferencias genéticas entre los sujetos que causan el **componente genético (σ^2_g)** de la variancia total fenotípica; el otro es ambiental, fruto de las diferentes circunstancias ambientales de cada individuo y se puede cuantificar en la **variancia ambiental (σ^2_e)**.

1. INTRODUCCIÓN

1.1 DEFINICIÓN DE TROMBOFILIA HEREDITARIA

Por trombofilia se designa a una especial tendencia del individuo a la trombosis. Los defectos genéticos o adquiridos subyacentes no necesariamente causan una afectación clínica continua. Su papel en la patogenia de la trombosis se debería a la disminución de la capacidad fisiológica para enfrentarse con fluctuaciones normales producto de las interacciones con el ambiente (Tabla 1). Se suele aplicar el término trombofilia sólo a un subgrupo de los pacientes con trombosis: aquellos con una gran expresividad clínica. Las características que se presentan con mayor frecuencia en este subgrupo selecto se citan en la Tabla 2. Existen unos pocos pacientes con trombofilia fulminante, que en ausencia de tratamiento pueden sufrir trombosis repetitivas continuamente. Pero en la mayoría de pacientes la trombosis sucede en episodios separados a menudo por periodos asintomáticos prolongados. Esta discontinuidad sugiere la necesidad de factores desencadenantes para cada episodio, tal vez estímulos directos, deterioro temporal de resistencia intrínseca, o alguna combinación de estos factores. Asimismo este característico modo de presentación, como sucesos separados en el tiempo, subraya la importancia de la interacción con el ambiente.

La definición de trombofilia hereditaria ha seguido un refinamiento continuo hasta llegar a la propuesta actualmente por un comité conjunto de expertos de la OMS y la ISTH (International Society on Thrombosis and Hemostasis): ***trombofilia hereditaria*** *es una tendencia genéticamente determinada al tromboembolismo venoso. Tanto anomalías dominantes, en algunos casos, como combinaciones de defectos más leves, en otros, pueden ser aparentes clínicamente con edad temprana de inicio, recidivas frecuentes o historia familiar de trombosis. Los rasgos leves pueden ser sólo revelados mediante investigación de laboratorio. Todavía no se conocen ni se han comprendido todas la influencias genéticas ni sus interacciones (Lane y col 1996 a).*

Tabla 1

Factores adquiridos que pueden interactuar con la base genética de la trombofilia

Edad avanzada

Inmovilización

Cirugía mayor

Cirugía ortopédica

Neurocirugía

Embarazo

Puerperio

Uso de hormonas estrogénicas

Cáncer

Síndrome antifosfolípido

Tabla 2

Características clínicas de la trombofilia. No necesariamente deben presentarse todas.

Edad joven de la 1ª trombosis (< 45 años)

Trombosis de repetición

Historia familiar positiva

Localizaciones inusuales de la trombosis

Severidad desproporcionada con un estímulo reconocido

1.2 EVOLUCIÓN DEL CONCEPTO DE TROMBOFILIA COMO ENFERMEDAD COMPLEJA

Como punto de partida de la relación entre los genes y la trombosis, o de la interpretación de la trombofilia como una enfermedad hereditaria debemos situarnos en 1965. En ese año se describió la primera familia con un déficit de antitrombina en la que se demostró la cosegregación entre un defecto congénito y la tendencia trombótica *(Egeberg 1965)*. Desde entonces el modelo explicativo de la base genética subyacente a la trombosis ha evolucionado conceptualmente en tres fases progresivas.

A partir de los primeros años ochenta se descubrieron nuevas alteraciones hereditarias, relacionadas con proteínas de la hemostasia: déficit de proteína C *(Griffin y col 1981)*, déficit de proteína S *(Comp y Esmon 1984, Schwarz y col 1984)* o disfibrinogenemias *(Havertake y Samama 1995)*. También se comprobó que cada uno de estos déficits era un factor de riesgo independiente asociado con trombofilia familiar siguiendo modelos hereditarios mendelianos típicos. En este contexto, la complejidad de la enfermedad trombótica posee un sentido muy concreto: el de la **heterogeneidad genética**, es decir, el fenómeno producido cuando mutaciones en uno cualquiera de entre varios genes pueden provocar fenotipos idénticos (la trombosis en nuestro caso). Esta situación aparece típicamente si los genes son necesarios para una ruta bioquímica o fisiológica común, como sucede con los procesos de la Hemostasia.

La idea de la trombofilia familiar como una enfermedad monogénica (con un modelo de herencia mendeliana simple) quedó en entredicho a través del estudio sistemático de la deficiencia de proteína C. En estos estudios, tres observaciones contradicen la hipótesis de un modelo de herencia dominante para la trombosis: (1) Los pacientes homocigotos para una deficiencia de proteína C, sufren habitualmente una clínica trombótica muy severa. Sin embargo, muy pocos de sus familiares portadores heterocigotos de la deficiencia presentan episodios trombóticos; (2) La prevalencia de heterocigotos para el déficit de proteína C en la población sana es de 0,2 a 0,5%. Ni estos individuos ni sus familiares también portadores heterocigotos experimentan eventos trombóticos (portadores de deficiencia de proteína C asintomáticos); (3) En familias trombofílicas (sintomáticas) portadoras de una deficiencia de proteína C, alrededor del 45% de los familiares heterocigotos son sintomáticos (han sufrido algún episodio

trombótico); es decir, el 55% restante de los portadores del déficit son asintomáticos. Además, alrededor del 8% de los miembros no portadores de la deficiencia de proteína C también han sufrido episodios trombóticos. Estas observaciones indican una reducida penetrancia y una alta frecuencia de fenocopias (individuos que expresan la enfermedad sin ser portadores del rasgo analizado, en este caso la deficiencia de proteína C). Algo similar ocurre en familias trombofílicas portadoras de una deficiencia de proteína S, en las que el 45% de los familiares afectos del déficit permanecen asintomáticos.

Como explicación a estas observaciones se empezó a especular que puedan ser necesarios varios defectos genéticos, actuando en combinación, para ocasionar la trombosis *(Miletich y col 1993)*. Desde esta perspectiva, la trombofilia seguiría el modelo conocido como **herencia poligénica**: el rasgo clínico patológico requiere la presencia simultánea de mutaciones en múltiples genes. El descubrimiento en 1993 del fenotipo de laboratorio conocido como resistencia a la proteína C activada (RPCa) *(Dahlbäck y col 1993)* y en 1994 de su principal base genética, la mutación G1691A en el factor V de la coagulación, también denominada factor V Leiden *(Bertina y col 1994)*, vinieron a corroborar esta hipótesis en un gran número de familias, previamente diagnosticadas con alguno de los otros déficits descritos. Posteriormente, la descripción de la mutación G20210A en el gen de la protrombina *(Poort y col 1996)* ha reafirmado la hipótesis poligénica: los individuos portadores de más de un defecto genético tienen más riesgo de trombosis que los individuos portadores de uno solo de los defectos.

Sin embargo, a pesar de los importantes avances en las bases moleculares de la enfermedad tromboembólica sucedidos en estas últimas décadas, los modelos genéticos anteriores todavía son insatisfactorios para explicar un gran porcentaje de casos de trombofilia. La mayoría de rasgos fenotípicos de relevancia clínica no siguen modelos de herencia monogénica mendeliana simple. Este es el caso, por ejemplo, de la susceptibilidad a cardiopatía isquémica, hipertensión arterial, diabetes, cáncer o infecciones. Son los denominados rasgos o **fenotipos complejos** y con este término nos referimos a cualquier fenotipo que no exhibe una herencia mendeliana clásica regulada por un gen único. En general, la complejidad aparece cuando se rompe la correspondencia sencilla entre genotipo y fenotipo, bien porque el mismo genotipo puede resultar en diferentes fenotipos (por efectos del azar, ambiente o interacciones con otros genes) o bien porque diferentes genotipos pueden tener como resultado el mismo fenotipo *(Lander y Schork 1994)*. Una de

las ventajas de contemplar a la trombofilia con este prisma genético mucho más general es la inclusión de los factores ambientales, que tanta influencia parecen tener en el desarrollo de los episodios de trombosis. El modelo teórico actual para entender la fisiopatología de la enfermedad trombótica venosa acepta que se trata de una enfermedad multicausal, en la que se implican factores genéticos que interaccionan con factores ambientales *(Rosendaal 1997, Rosendaal 1999)*. Es decir, existe un acuerdo general sobre su naturaleza de enfermedad compleja. Esta visión coincide con la definición de trombofilia hereditaria realizada por la OMS y la ISTH y mencionada anteriormente.

1.3 LA TROMBOFILIA COMO RASGO FENOTÍPICO COMPLEJO PRECISA NUEVAS TÉCNICAS DE INVESTIGACIÓN

Pese al avance experimentado en el conocimiento de las causas de trombofilia y en el diagnóstico y tratamiento de los pacientes, tan solo conocemos claramente unos pocos factores de riesgo. A partir de la experiencia clínica de casos familiares de trombofilia inexplicada (después de descartar todas las causas biológicas conocidas) parece evidente que un número indeterminado de factores genéticos involucrados está por identificar. La Hemostasia plasmática es un sistema enzimático en el que intervienen gran cantidad de proteínas con función activadora o inhibidora tanto en la Coagulación *(Mann 1999)* como en la Fibrinolisis *(Collen 1999)*. En principio, cualquiera de ellas podría estar implicada como factor de riesgo para la trombosis. Sin embargo, a día de hoy, tan sólo se ha demostrado claramente la responsabilidad de unas pocas anomalías: disminución de inhibidores (antitrombina, proteína C, proteína S), proteínas disfuncionales (disfibrinogenemia), mutación G1691A en el gen del factor V, mutación G20210A en el gen del factor II y aumento de factores procoagulantes como el factor VIII. Otras anomalías plasmáticas que no involucran directamente a componentes de la Hemostasia, como la presencia de anticuerpos antifosfolípido o la hiperhomocisteinemia moderada, también parecen ser factor de riesgo de trombosis *(Bertina 1999)*, si bien el mecanismo fisiopatológico por el que provocan la enfermedad aún no ha sido explicado. En cuanto al resto de componentes de la Hemostasia, algunos se han relacionado con casos esporádicos familiares de trombosis (déficits de plasminógeno, trombomodulina, cofactor II de la heparina o TFPI) pero en grandes estudios epidemiológicos su relevancia es prácticamente nula *(Lane y col 1996 a, Bertina 1999)*.

Los mencionados déficits asociados con un mayor riesgo de trombosis (por ejemplo, déficit de antitrombina) siempre se han entendido, quizás por una simplificación operativa, como una variable cualitativa, dicotómica, a partir de un valor umbral (por ejemplo el percentil 5 de la distribución en la población general). Dicho valor determina la ausencia / presencia del déficit. Sin embargo es posible que los niveles de las proteínas se comporten como un factor de riesgo continuo (al igual que sucede con el colesterol o la tensión arterial) y la estratificación del riesgo tenga mucho más de 2 niveles. En otras palabras, que no sea lo mismo presentar, por ejemplo, una antitrombina del 85% que una

del 95%, valores que hoy consideramos normales (por encima del umbral patológico). Se hace pues necesario abordar el estudio de los componentes de la Hemostasia desde su naturaleza de variables continuas. Ello nos conducirá a conocer cuales son las causas de su variabilidad y si alguna de estas causas también influye en el riesgo de enfermedad trombótica. Conviene recordar en este punto que los expertos se inclinan a defender con insistencia la co-heredabilidad de varios factores de riesgo trombofílico *leve* o *moderado*, que actuando en combinación serían los responsables de una mayor expresión clínica *(Lane y col 1996 b)*. Una vez más nos topamos con la implicación de múltiples factores.

El concepto de enfermedad poligénica y multifactorial obliga a utilizar nuevos métodos de investigación que se acomoden a la mayor complejidad subyacente en estas entidades. Como ya se ha dicho, las enfermedades multifactoriales se deben a la interacción de múltiples genes entre sí y con el ambiente. Estas interacciones provocan un *gradiente* de susceptibilidad genética a la enfermedad (trombosis, en nuestro caso) y explican las diferencias clínicas observadas en individuos diferentes. La identificación y caracterización de los genes implicados requiere el estudio de grandes colecciones de datos familiares, marcadores genéticos altamente informativos que se extiendan por todo el genoma y, de modo primordial, estrategias estadísticas específicamente diseñadas para trabajar con enfermedades complejas *(Lander y Schork 1994, Weeks y Lathrop 1995)*.

Para relacionar un factor genético con una enfermedad existen dos tipos principales de diseño: los estudios de ligamiento genético y los estudios de asociación.

Los estudios de ligamiento exploran si una enfermedad y un alelo tienen transmisión común dentro de una familia, mientras que los estudios de asociación buscan esta relación en la población *(Lander y Schork 1994)*. La diferencia fundamental entre ambos diseños radica en el tipo de muestra que se debe reclutar: los de ligamiento necesitan individuos emparentados, mientras que los de asociación se realizan con cohortes de casos y controles, en las que se reclutan, por definición, individuos no emparentados.

1.3.1 Estudios de asociación

Pretenden demostrar si un marcador genético concreto (un alelo) se asocia a una enfermedad. En caso de asociación la frecuencia del alelo será distinta en pacientes que en controles sanos. Cuando aparece un resultado positivo verdadero (es decir, si la población estudiada es homogénea genéticamente y no hay sesgos ni en los pacientes ni en los controles) se puede concluir que o bien el propio marcador es el *locus* responsable de la enfermedad o se encuentra muy próximo, en desequilibrio de ligamiento (ver punto 2.6) con el *locus* responsable. Si bien para los estudios de asociación positivos es imposible distinguir entre una situación u otra, el marcador involucrado servirá para detectar individuos con riesgo de padecer la enfermedad.

Es importante destacar que los estudios de asociación presentan grandes limitaciones *(Gambaro y col 2000)*, entre las que figuran:

- Una gran facilidad para obtener falsos positivos. La heterogeneidad genética de las poblaciones, sobre todo en áreas multirraciales o en países receptores de migraciones históricas desde puntos muy diversos, hace casi imposible obtener cohortes de pacientes realmente comparables (desde el punto de vista genético) con cohortes de controles sanos por muchos esfuerzos que se realicen durante el reclutamiento para asegurar la comparabilidad de edad, sexo u otros factores ambientales como dieta o nivel socioeconómico. Si por motivos azarosos, en el grupo control se reclutan individuos de un estrato genético distinto al del grupo de casos, las frecuencia alélicas del polimorfismo investigado pueden ser distintas y, automáticamente, el estudio resultará en una *odds ratio* significativa, aunque el polimorfismo no tenga ninguna relación fisiopatológica con la enfermedad.

- Debido al punto anterior, la comparabilidad entre diferentes estudios de un mismo polimorfismo candidato realizados en zonas geográficas distintas es muy difícil. Por este motivo los estudios subsiguientes de metanálisis, tan populares y ampliamente aceptados en la medicina actual, pueden generar resultados completamente falsos.

- La elección del polimorfismo suele ser absolutamente arbitraria en la mayoría de casos. Es decir casi nunca se tienen evidencias previas de una relación biológica entre el gen implicado y la enfermedad o al menos un fenotipo intermediario con la enfermedad.

En caso de existir esta relación o influencia se define al polimorfismo como *funcional*. Si el polimorfismo no influye sobre el fenotipo en cuestión, se denomina polimorfismo *neutro*. En general, la relación entre un polimorfismo funcional y un fenotipo sólo se puede establecer con estudios de ligamiento, mucho más robustos estadísticamente que los de asociación (ver 1.3.2. y los ejemplos presentados en 6.4.1 y 6.5). Es decir, antes de realizar estudios de asociación entre un polimorfismo genético y una enfermedad es muy conveniente haber realizado estudios de ligamiento que demuestren inequívocamente la relación causal entre el polimorfismo y la enfermedad.

- Un estudio de asociación perfecto, sin sesgos, con un resultado positivo verdadero, nunca puede establecer una relación causal entre el polimorfismo genético y una enfermedad compleja (porque es incapaz de excluir el desequilibrio de ligamiento con un polimorfismo desconocido cercano, realmente causante de la enfermedad). En el mejor de los casos sólo servirá para generar nuevas hipótesis de trabajo. Por el contrario, un resultado negativo tiende a interpretarse como que el gen candidato no está relacionado con la enfermedad, y en realidad sólo puede descartarse el polimorfismo investigado. Otros polimorfismos desconocidos o no explorados del mismo gen podrían estar asociados con la enfermedad.

1.3.2. Estudios de ligamiento

La mejor (y muchas veces única) manera de establecer que un factor genético determina una enfermedad es el análisis de ligamiento genético. Los métodos de ligamiento se basan en un hecho biológico simple: los genes están concatenados en los cromosomas y sus posiciones son constantes. Dos genes estan "ligados" cuando se encuentran muy cerca el uno del otro, prácticamente nunca se observa recombinación entre ellos en las meiosis y, por tanto, los alelos concretos contenidos en ellos se transmiten juntos a la descendencia. Precisamente es la observación de esta transmisión familiar la que permite relacionar una enfermedad con un punto concreto del genoma.

El objetivo básico del análisis de ligamiento es localizar uno o más genes que influyen sobre un rasgo concreto en una región cromosómica específica. Esto se consigue examinando la co-segregación (transmisión conjunta a la descendencia) del rasgo de

interés con marcadores genéticos. Los parientes que sean fenotípicamente parecidos compartirán alelos comunes en los marcadores que rodean el gen influyente sobre el fenotipo, mientras que otros parientes que presenten un fenotipo distinto (o valores muy discrepantes de un fenotipo cuantitativo) portarán alelos distintos. Así pues, la información mínima que se necesita para el análisis de ligamiento proviene de un grupo de familias con individuos fenotipados, cuyas relaciones de parentesco entre sí son conocidas, y los genotipos de estos mismos individuos en uno o varios marcadores genéticos.

A diferencia de los estudios de asociación, los de ligamiento:

- No se limitan sólo a genes candidatos, ya conocidos. Permiten la localización de genes desconocidos, mediante el análisis de marcadores genéticos situados a lo largo de todo el genoma (microsatélites). Pese a que estos marcadores no codifican ninguna proteína, son muy útiles para generar mapas de los cromosomas.

- No se ven afectados por los diferentes estratos genéticos ocultos en la población, puesto que siempre se realizan en individuos emparentados. Por ese motivo, no son susceptibles de errores estadísticos tipo I (falsos positivos).

- Permiten cuantificar la importancia relativa de un factor genético en la variación del riesgo de enfermedad en la población.

- Permiten distinguir al desequilibrio de ligamiento como causa de una asociación estadística entre un marcador genético y una enfermedad.

Por todos estos motivos, los análisis de ligamiento son mucho más robustos y potentes que los de asociación *(Weiss 1993)*.

Sin embargo, en la práctica resulta mucho más fácil reclutar individuos no emparentados que grupos familiares, especialmente si las familias deben ser de gran tamaño y complejidad. Esto explicaría en parte que, hasta el momento, en el campo de la trombofilia prácticamente sólo se han realizado estudios epidemiológicos de *asociación* y apenas existen ejemplos de estudios de ligamiento *(Scott y col 2001)*. Si bien, algunos factores de riesgo genético para trombosis venosa se han visto confirmados en múltiples estudios de asociación y podemos aceptarlos como muy probable causa de enfermedad (caso de la mutación factor V Leiden, de la mutación G20210A en el gen de la protrombina o de los déficits plasmáticos de antitrombina y de proteína C) otras muchas variantes genéticas en genes candidatos de la Hemostasia no han hecho más que generar

una enorme confusión. La causa se encuentra en la abundancia de estudios de asociación publicados en los últimos años, con resultados contradictorios. Es sobre todo en el campo de la trombosis arterial dónde se ha producido una mayor incertidumbre por la inconsistencia de resultados *(Lane y Grant 2000)*. Esta situación es un ejemplo muy claro de las limitaciones y de los riesgos metodológicos inherentes a los estudios de asociación. En definitiva, parece llegado el momento de cambiar el tipo habitual de diseño utilizado en la investigación epidemiológica de la trombofilia. Especialmente, cuando el objetivo sea la búsqueda de genes desconocidos, objetivo imposible de alcanzar mediante estudios de casos y controles (asociación).

1.4 EL PROYECTO GAIT

La investigación clínica de las causas biológicas de trombosis venosa, especialmente en pacientes jóvenes o en casos de trombofilia familiar, es una práctica obligada en la Medicina moderna. Desgraciadamente, no siempre se consigue identificar una anomalía biológica de entre las que componen la actual batería de estudio: disfibrinogenemia, anticoagulante lúpico, anticuerpos antifosfolípido, déficits de antitrombina / proteína C / proteína S, mutaciones factor V Leiden y G20210A en el gen del factor II, resistencia a la proteína C activada (RPCa) o hiperhomocisteinemia. En nuestro medio, aproximadamente un 50% de los casos investigados revela una anomalía biológica subyacente, a la que se le pueda atribuir parte de la responsabilidad en la trombosis *(Mateo y col 1997, Souto y col 1998, González y col 1998)*. En el 50% restante de los casos, la trombosis permanece inexplicada desde el punto de vista biológico. Para intentar avanzar en el conocimiento de las causas de la trombosis, principalmente en las de los casos idiopáticos, y como continuación del Estudio Múlticéntrico Español de Trombosis (EMET) *(Mateo y col 1997, Mateo y col 1998)*, se está ejecutando desde hace varios años el proyecto GAIT (*Genetic Analysis of Idiophatic Thrombophilia*).

Este proyecto se inició en 1995 y ha sido diseñado, dirigido, financiado (en parte) y realizado por la Unitat d'Hemostàsia i Trombosi del Hospital de la Santa Creu i Sant Pau de Barcelona. Su objetivo principal es la identificación de los genes que determinan la variabilidad de los fenotipos relacionados con la Hemostasia, y muy especialmente la localización de aquellos que influyen sobre el riesgo de enfermedad tromboembólica. El análisis de los datos clínicos y biológicos se ha realizado en colaboración con el Department of Genetics de la Southwest Foundation for Biomedical Research de San Antonio, Texas, utilizando una tecnología estadística e informática de última generación que permite, esencialmente, explotar al máximo la información genética contenida en grandes familias.

Este proyecto es pionero en el uso de esta metodología en el ámbito de la trombofilia y supone un cambio de estrategia radical en la investigación de las bases genéticas de la Hemostasia y de la enfermedad tromboembólica. Una de las diferencias fundamentales respecto a estudios previos es el tipo de muestra: el proyecto GAIT analiza individuos emparentados pertenecientes a familias de gran tamaño. Tal y como ya se ha

apuntado, hasta el presente la práctica totalidad de estudios sobre las causas genéticas de la trombosis se basan en el diseño epidemiológico clásico de "casos / controles" a partir de sujetos no emparentados.

El estudio de los fenotipos incluidos en el proyecto GAIT tiene una importancia capital para proseguir la investigación de las causas de la trombosis. En caso de demostrarse una gran influencia de los genes en los componentes de la Hemostasia, o en la propia enfermedad tromboembólica, se justifica de manera natural la necesidad de proseguir la búsqueda. Por el contrario si no se detectara una influencia suficientemente grande de los factores genéticos sería un absurdo metodológico y económico obstinarse en la búsqueda de genes inexistentes o con un peso clínico irrelevante.

Anticipando información cabe decir que los resultados obtenidos demuestran una gran influencia de los genes en la Hemostasia y la trombosis. Como consecuencia inmediata, en el momento de redactar esta Tesis, el proyecto GAIT ya se encuentra en una fase ulterior de análisis sistemático de las relaciones (ligamiento) entre los fenotipos estudiados en el proyecto y los genes estructurales de la Hemostasia u otros genes aún desconocidos a lo largo de todo el genoma humano (*Full Genome Scan*). Los resultados de esta fase del proyecto, en vías de publicación, no se contemplan en la presente Tesis.

Antes de entrar en la exposición de los métodos y en el análisis de los resultados conviene introducir brevemente algunos conceptos básicos de Epidemiología Genética, que han sido utilizados en nuestro trabajo.

2. CONCEPTOS BASICOS DE EPIDEMIOLOGIA GENETICA

2.1 RASGOS (FENOTIPOS) COMPLEJOS

Un fenotipo complejo es cualquier característica mensurable en un organismo que está determinada simultáneamente por múltiples genes, por factores ambientales y por las posibles interacciones entre ellos (gen/gen y gen/ambiente). Esta multiplicidad de factores determinantes se traduce en fenotipos de tipo continuo por lo que se suele aceptar que "rasgo complejo" es sinónimo de "rasgo cuantitativo continuo" *(Weiss 1993)*. La práctica totalidad de proteínas plasmáticas y de pruebas de laboratorio relacionadas con la Hemostasia se miden en escalas cuantitativas continuas. En oposición a los rasgos complejos o continuos existen los rasgos mendelianos, de tipo monogénico, que se observan como variables cualitativas discretas. Un ejemplo clásico sería la hemofilia A, causada por anomalías en un sólo gen que determinan la presencia o la ausencia de la enfermedad.

El análisis matemático de variables biológicas cuantitativas es un poderoso instrumento para obtener información acerca de los factores que determinan su variabilidad (genéticos y/o ambientales) cuando estas variables son medidas en individuos emparentados. Por otro lado, las variables cuantitativas continuas permiten un estudio mucho más preciso y pormenorizado de las posibles relaciones con la enfermedad porque:

1º Muchas enfermedades se pueden definir según su medición en una escala cuantitativa subyacente (p.e. presión sanguínea o niveles de glucosa)

2º Las variables cuantitativas proporcionan mayor precisión para definir el fenotipo de los individuos afectados y su variación puede estudiarse en los pacientes sanos, con lo que se aumenta intensamente el poder estadístico de los análisis de ligamiento.

3º A veces, el rasgo cuantitativo puede servir como un fenotipo intermediario, influido por un número menor de factores ambientales o genéticos que la enfermedad misma. En tal caso será mucho más manejable para identificar *loci* o zonas del genoma relacionadas con la enfermedad. Existen ejemplos recientes del hallazgo de regiones genéticas implicadas en enfermedades complejas, a través de fenotipos intermediarios como la leptina para la obesidad *(Comuzzie y col 1997)* o el péptido amiloide β42 para la enfermedad de Alzheimer *(Bertram y col 2000, Ertekin-Taner y col 2000.)* La variable cuantitativa (imaginemos, por ejemplo, el

valor en plasma de la RPCa) puede estar más intensamente correlacionada con polimorfismos de un gen candidato (pongamos por caso el del factor VIII) que el rasgo cualitativo dicotómico (p.e. trombosis si o no).

Así entendido, el valor de un fenotipo complejo en un individuo concreto i ($\mathbf{y_i}$) se puede modelizar como una función lineal simple:

$$y_i = \mu + \Sigma\beta_j + g_i + e_i$$

Donde μ es la media del rasgo en la población, β_j es el coeficiente de regresión para la covariable j (por ejemplo, edad o sexo), $\mathbf{g_i}$ es el efecto conjunto de los genes, y $\mathbf{e_i}$ es el efecto de factores ambientales que influyen en el individuo (por ejemplo, la dieta, el clima o incluso el error de medida del fenotipo).

En la actualidad existen métodos para localizar dentro del genoma humano a los genes (simbolizados por $\mathbf{g_i}$ en la ecuación) que influyen sobre un rasgo complejo. Estos métodos se basan en el análisis mediante técnicas de genética molecular de secuencias de ADN y en el análisis mediante estadística genética de las relaciones entre las variantes de ADN (polimorfismos) y los niveles del fenotipo complejo *(Rogers y col 1999)*.

Los lugares dentro del genoma dónde se encuentran genes que contribuyen a la variabilidad de los fenotipos complejos se conocen como QTL (del inglés *quantitative trait loci*), término introducido en 1975 *(Geldermann 1975)*. Cada uno de los QTL puede ser responsable de una pequeña proporción de la variabilidad observada en un fenotipo complejo.

2.2 HEREDABILIDAD DE UN RASGO COMPLEJO

Cuando medimos un fenotipo cuantitativo continuo (por ejemplo, la estatura en cm, los niveles en plasma de factor von Willebrand en ui/L o el tiempo de protrombina en segundos) en diferentes individuos, obtenemos resultados distintos para cada sujeto. En otras palabras, observamos una variabilidad. Cuanto mayor es el grado de parentesco entre dos individuos, mayor cantidad de genes (alelos) idénticos son compartidos por ambos. Si los valores de un fenotipo complejo tienden a ser más próximos (estadísticamente) en los sujetos emparentados que en los no emparentados, y más parecidos cuanto mayor es el grado de parentesco (más entre abuelo-nieto que entre primos segundos, por ejemplo), se puede establecer que los genes tienen una influencia en la variabilidad observada de dicho fenotipo. Para cuantificar esa influencia se usa un parámetro llamado "heredabilidad" y que es el porcentaje de la variabilidad de un fenotipo que se atribuye en exclusiva al efecto de los genes *(Falconer y Mackay 1996).*

Existen diversos modos matemáticos de estimar la heredabilidad. Todos ellos parten de una premisa inexcusable: los fenotipos deben medirse en individuos emparentados. Las estimaciones realizadas por el proyecto GAIT incluyen 46 fenotipos y se han calculado utilizando todos los tipos posibles de parentesco que aparecen en las familias grandes incluidas. El método matemático se conoce como "Análisis de componentes de la variancia" *(Almasy y Blangero 1998).* La variancia de un fenotipo cuantitativo describe la dispersión de los valores que toma el fenotipo en los distintos individuos alrededor de la media en la población.

Muy sucintamente, la variancia total observada en un fenotipo (σ^2_p) es el resultado de la suma de las variancias debidas a los factores determinantes genéticos (σ^2_g) y ambientales (σ^2_e)

$$\sigma^2_p = \sigma^2_g + \sigma^2_e$$

Se define la heredabilidad (h^2) como la proporción de la variancia total debida a la variancia genética:

$$h^2 = \sigma^2_g / \sigma^2_p$$

La h^2 toma valores entre 0 y 1. El valor 0 significa que el fenotipo no está determinado genéticamente y que toda su variabilidad es producto de efectos ambientales. Por el contrario, el valor 1 implica que la totalidad de la variancia se debe al efecto de los genes. Este valor extremo es puramente teórico y en la práctica nunca se observa, aunque sólo sea porque el error de medida de cualquier variable siempre introduce un efecto ambiental en la variancia del fenotipo.

Previamente a la estimación de la h^2 para un fenotipo, el modelo matématico controla el efecto de las covariables ambientales conocidas para cada sujeto (en nuestro caso, se midieron la edad, el sexo, el consumo de tabaco y el uso de anticonceptivos orales, en mujeres). La h^2 suele expresarse en % y el valor complementario hasta llegar al 100% de la variabilidad corresponde a los factores ambientales que influyen en el fenotipo. Dentro de los factores ambientales, se pueden delimitar los que se comparten cuando los sujetos viven en un mismo domicilio (*household effect*) como por ejemplo los debidos a una dieta común. Para ello es necesario registrar la composición de los distintos domicilios en cada una de las familias.

Una vez se demuestra que un fenotipo es heredable ($h^2 > 0$), o lo que es lo mismo, cierta proporción de su variancia se debe a los efectos de las diferencias genéticas entre los individuos, tiene sentido que nos propongamos determinar las localizaciones cromosómicas de los genes responsables de estos efectos. En teoría, cualquier *locus* que influye en la variabilidad puede clasificarse como un QTL, pero no siempre es posible localizarlo. El poder actual de localización mediante análisis de ligamiento (ver punto 2.5) depende de varios factores, siendo el principal de ellos la intensidad del efecto del QTL en el fenotipo. Cuanta más proporción de la h^2 total del fenotipo es causada por un QTL concreto, más fácil es localizar ese QTL. Hoy día es razonable esperar que, con los tamaños de muestra y el diseño adecuados, aquellos genes individuales que suponen más del 10-15% de la variancia en un fenotipo complejo pueden ser localizados en regiones cromosómicas específicas *(Rogers y col 1999).*

2.3 RIESGO O SUSCEPTIBILIDAD DE ENFERMEDAD COMO UNA FUNCIÓN CONTÍNUA

En la práctica clínica y en la epidemiología clásica, el estado de enfermedad se registra como una variable discreta dicotómica (es decir, un sujeto está afecto o sano). Sin embargo, en la moderna Epidemiología Genética se asume que existe un rasgo cuantitativo contínuo, inobservable, llamado "susceptibilidad" o "riesgo" (*liability,* en inglés) que determina el estado de afección: si el valor de susceptibilidad de un sujeto excede un valor umbral específico, la enfermedad se observa, mientras que si este valor no alcanza al umbral, el individuo aparece como sano *(Falconer y Mackay 1996)*. Dicho de otro modo, los valores de la susceptibilidad subyacente en los individuos no afectos de la enfermedad son inferiores al valor umbral. Este umbral depende de la prevalencia de la enfermedad, de manera que el área bajo la curva situada por encima del umbral debe ser igual a la prevalencia en la población (Figura 1).

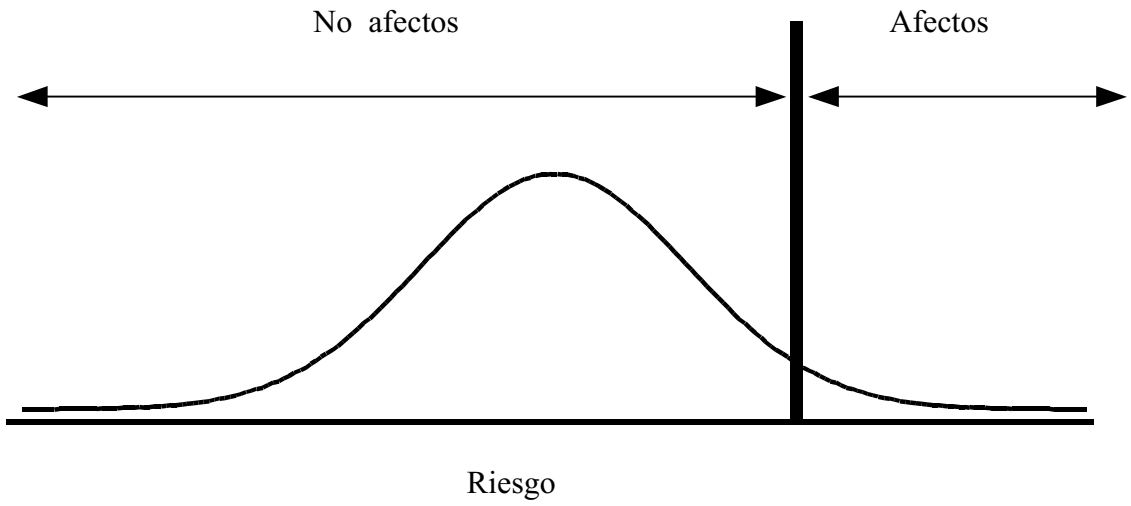
El valor que toma la función "susceptibilidad" en cada individuo depende en primer lugar de su estado de afección (los sanos tienen un valor inferior al del umbral, los enfermos un valor superior). En segundo lugar, depende del grado de parentesco con familiares enfermos y del número de parientes enfermos. Intuitivamente se entiende que un sujeto sano, pero con un padre y un hermano enfermos, tiene un riesgo mayor de enfermar que otro sujeto sano con tan solo un bisabuelo enfermo. El valor de la susceptibilidad del primero estará más cerca del umbral que el valor del segundo. Se puede distinguir así la diferente predisposición a la enfermedad de los individuos sanos.

El valor umbral puede depender de la presencia de otros factores de riesgo ambientales. Así, por ejemplo, si la edad influye en el riesgo, observaremos que en la gente de edad avanzada el umbral se desplaza más a la izquierda de la distribución, reflejando una mayor prevalencia de la enfermedad (mayor área bajo la curva situada por encima del umbral). Del mismo modo si el riesgo de trombosis fuera mayor en mujeres que en hombres, el valor umbral en la población femenina, se situaría más a la izquierda que en la población masculina.

Puesto que la mayoría de fenómenos biológicos aparecen como procesos continuos, resulta útil realizar inferencias sobre la escala cuantitativa continua subyacente. Los modelos matemáticos del tipo descrito son mucho más congruentes con los modelos

actuales de acción genética. Para esta modelización del efecto de los genes, se asume que la *susceptibilidad* sigue una distribución normal (Figura 1). La construcción de esta función se realiza mediante métodos matemáticos cercanos conceptualmente a las técnicas estadísticas ampliamente conocidas de regresión logística. Recientemente, los métodos basados en los componentes de la variancia, han sido ampliados para permitir el estudio de rasgos dicotómicos mediante modelos umbral *(Duggirala y col 1997).*

Figura 1. Modelo umbral de susceptibilidad a la trombosis



Susceptibilidad o riesgo *(liability)*:

El riesgo individual (en abcisas) sigue una distribución normal en la población. La mayor parte de los individuos no sufren la enfermedad y se situan en los valores centrales de la distribución. Cuando el riesgo de un individuo supera el umbral (*threshold*) representado en la figura por la linea vertical, el sujeto presenta un evento trombótico

2.4 CORRELACIONES ENTRE FENOTIPOS COMPLEJOS. DESCOMPOSICIÓN EN CORRELACIÓN GENÉTICA Y CORRELACIÓN AMBIENTAL

La estadística clásica ofrece la posibilidad de estudiar la correlación entre dos variables medidas en los mismos individuos. El parámetro más usado es el coeficiente de correlación de Pearson (ρ_p), que mide el grado de asociación lineal y que se puede interpretar como una medida de la variación conjunta de las 2 variables debido a causas subyacentes comunes. El coeficiente de Pearson tiene valores posibles entre -1 (máxima correlación negativa) y $+1$ (máxima correlación positiva). Cuando el valor es 0 o no distinto estadísticamente de 0 se suele aceptar que las variables no están correlacionadas. Esta asociación entre dos fenotipos, que podemos observar directamente, se debe a la correlación de los valores de los fenotipos o *correlación fenotípica.*

Cuando las variables se miden en individuos emparentados (familias) y se utiliza el método de componentes de la variancia es posible separar (y cuantificar) la correlación debida a factores genéticos de la debida a factores ambientales *(Comuzzie y col 1996).* Existe una relación matemática entre estos parámetros de modo que la correlación fenotípica (coeficiente de Pearson, ρ_p) se deriva de las 2 correlaciones subyacentes, *genética* (ρ_g) y *ambiental* (ρ_e) y de las heredabilidades de los 2 fenotipos (h^2_1 y h^2_2):

$$\rho_p = \sqrt{(h^2_1\, h^2_2)}\, \rho_g + \sqrt{(1- h^2_1)}\, \sqrt{(1- h^2_2)}\, \rho_e$$

Una correlación genética significativa entre 2 fenotipos debe interpretarse como la demostración estadística de la existencia de genes que influyen simultáneamente en los 2 fenotipos. Este fenómeno genético (un gen que afecta a la vez a dos o más fenotipos) se conoce como *pleiotropía*. El grado de correlación provocado por la pleiotropía (valor numérico entre -1 y $+1$) expresa hasta qué punto los mismos genes influyen en los fenotipos. Por su parte, una correlación ambiental significativa implica la existencia de factores ambientales con efectos sobre ambos fenotipos *(Falconer y Mackay 1996)*.

Por el contrario, la ausencia de correlación fenotípica (coeficiente de Pearson no distinto de 0) en ocasiones puede resultar engañosa, y no siempre significa que los 2 fenotipos no están correlacionados. A veces existe una correlación genética significativa y

una correlación ambiental también significativa, pero de signo opuesto a la correlación genética. En esta situación, las 2 fuentes de correlación tienden a anular la correlación fenotípica, según se desprende de la fórmula matemática. Paradójicamente, estos dos fenotipos intensamente correlacionados (por factores comunes ambientales y genéticos) aparecerán como completamente no correlacionados si se usan los métodos de la estadística convencional.

2.5 LIGAMIENTO ENTRE *LOCUS* GENÉTICO Y FENOTIPO COMPLEJO

En sentido estricto, el ligamiento genético es un fenómeno que sólo puede producirse entre dos *loci*, cuando estos se encuentran muy cercanos en un mismo cromosoma y por lo tanto los alelos que contienen se transmiten juntos con mayor frecuencia que si estuvieran muy separados o en distintos cromosomas. En palabras más técnicas, la probabilidad de recombinación entre ellos es estadísticamente inferior a 0.5 *(Ott 1999)*. Esto significa que los alelos se transmiten juntos a la descendencia con una frecuencia mayor que el 50% de las veces porque más del 50% de los gametos formados durante el proceso biológico de meiosis contienen los mismos alelos (50% es la frecuencia de transmisión conjunta de 2 alelos o genes escogidos al azar entre todo el genoma). Según la intensidad del ligamiento, esta transmisión conjunta aumenta hasta llegar a ser del 100% de las meiosis si los dos *loci* están completamente ligados.

¿Cómo se mide el ligamiento genético? Para saber si dos *loci* están ligados existen diferentes tests estadísticos. El parámetro clásico es la escala de LOD o logaritmo de la *odds ratio* entre 2 probabilidades alternativas:

$$\textbf{LOD} = \log_{10} \frac{\text{Probabilidad de que ambos } \textit{loci} \text{ estén ligados}}{\text{Probabilidad de no-ligamiento}}$$

Por ejemplo, si el LOD es 3, la posibilidad de ligamiento es 1000 veces mayor que la contraria. Cuanto mayor sea el LOD, mayor seguridad estadística o certeza podemos tener en el ligamiento. Esto se comprende observando la equivalencia entre el valor del LOD y el valor ”p” usado para excluir falsos positivos en la inferencia estadística convencional:

LOD	valor p
0.60	0.05
1.17	0.01
1.90	0.0015
3.0	0.0001
5.0	0.000001

¿ Por qué se habla de ligamiento entre un *locus* y un fenotipo, si el ligamiento es un fenómeno que ocurre necesariamente entre dos *loci* ?

Imaginemos dos *loci* (*locus* A y *locus* B) que presentan ligamiento genético. Si uno de ellos (*locus* A) ejerce un efecto sobre un fenotipo determinado, se define como un *locus* funcional. En caso de que el fenotipo sea de tipo complejo (cuantitativo y continuo) este *locus* A se denomina en terminología moderna QTL, como ya se ha mencionado anteriormente. En esta situación, por extensión se dice que el *locus* B, ligado con el QTL también esta "ligado" al fenotipo complejo que el QTL regula. Por razones prácticas y mediante una licencia del lenguaje se acepta el ligamiento entre un fenotipo y un *locus* funcional y también cualquier otro *locus* que no influye sobre el fenotipo pero que está ligado con el *locus* funcional.

Este concepto es importante porque las técnicas más avanzadas de búsqueda de nuevos genes están basadas en la utilización de marcadores genéticos anónimos llamados microsatélites, no funcionales, altamente polimórficos y que en caso de presentar ligamiento con fenotipos complejos, permiten detectar la presencia cercana del verdadero gen funcional o QTL. Cuando estos marcadores se genotipan sistemáticamente a lo largo de todo el genoma, a una distancia aceptablemente pequeña entre cada dos marcadores consecutivos, se obtiene un análisis global del genoma (en inglés, *full genome scan analysis*). Realizado este mismo análisis en sujetos emparentados (familias) a los que también se les ha medido un fenotipo concreto, es posible mediante técnicas muy sofisticadas de Estadística Genética establecer si alguno de los marcadores está ligado con el fenotipo en cuestión, y por tanto nos está señalando un gen cercano que actúa como QTL *(Almasy y Blangero 1998)*.

2.6 DESEQUILIBRIO DE LIGAMIENTO

Existe un concepto cercano al ligamiento conocido como "desequilibrio de ligamiento". Se trata de un fenómeno observado en Genética de Poblaciones consistente en la asociación no aleatoria, dentro de una misma población, de alelos en 2 o más sitios ligados *(Hartl y Clark 1997)*. La frecuencia con que se observan juntos los alelos en desequilibrio es mayor de la esperada si se heredaran de manera aleatoria (en este caso los alelos estarían en "equilibrio" y la frecuencia sería igual al producto de las frecuencias de los dos o más alelos). En otras palabras, si tenemos el caso más sencillo de 2 *loci* con 2 alelos cada uno, se observa una asociación entre un alelo específico del primer *locus* con otro alelo específico del segundo *locus*. Es decir, no son independientes. Aunque esto suele suceder cuando los *loci* están ligados, también puede observarse entre *locus* muy separados o en distintos cromosomas. Veamos un ejemplo numérico muy simple:

Sea el *locus A* con 2 alelos distintos: A, a

Y con las frecuencias alélicas f(A) = 0.6 y f(a) = 0.4

Sea el *locus B* con 2 alelos distintos: B, b

Y con las frecuencias alélicas f(B) = 0.3 y f(b) = 0.7

Si no hay asociación, la probabiliad de observar el alelo A y el alelo B en un individuo es simplemente el producto de las frecuencias alélicas o probabilidades de cada loci: f (AB) = f(A) x f(B) = 0.6 x 0.3 = 0.18. Lo mismo sucederá con los otros 3 haplotipos posibles, Ab, aB y ab. Así f(Ab) = 0.42; f(aB) = 0.12 y f(ab) = 0.28

Pero si observamos una diferencia entre estas probabilidades esperadas y las frecuencias obtenidas para los haplotipos en una población concreta, por ejemplo

f (AB) = 0.26

f (Ab) = 0.34

f (aB) = 0.04

f (ab) = 0.36

Existe un desequilibrio. El alelo *A* se asocia con el *B* y por tanto el alelo *a* lo hace con el *b* más a menudo de lo que sucedería si fuesen independientes. La discrepancia se representa por $\delta = 0.08$. El valor δ es el desequilibrio de ligamiento.

$f(AB) = f(A) \times f(B) + \delta = (0.6 \times 0.3) + 0.08$

En cambio, cuando $\delta = 0$ los alelos están en equilibrio y $f(AB) = f(A) \times f(B)$

El desequilibrio de ligamiento se puede producir por diferentes motivos: mezcla en el pasado reciente de la población estudiada de otras dos poblaciones precursoras pero dispares genéticamente; deriva genética; selección natural; y por una mutación relativamente reciente. Esta es la causa más fácil de entender y la que suele provocar desequilibrios más intensos y duraderos, porque en realidad se debe a que los *loci* se hallan ligados. La nueva mutación origina un nuevo polimorfismo sobre un trasfondo genético previo. Con el paso del tiempo, la deriva genética azarosa y/o la selección natural puede aumentar la frecuencia del alelo nuevo en la población. Cuando este nuevo alelo se transmite a la siguiente generación, lleva consigo a todos los alelos de los *loci* cercanos (haplotipo). Por eso se observa una asociación (no-independencia) entre todos ellos. Sin embargo, también con el paso del tiempo el desequilibrio de ligamiento tiende a desaparecer de la población, a través de mecanismos de recombinación. El segmento cromosómico próximo al alelo puede ser ocasionalmente intercambiado con segmentos homólogos de otros cromosomas, que pueden contener alelos diferentes en algunos de los loci. De esta manera el nuevo alelo aparece en distintos haplotipos. A mayor distancia, mayor tasa de recombinación y mayor velocidad en la desaparición del desequilibrio (en otras palabras, se observará durante un número menor de generaciones en esa población). Por el contrario, cuanto más estrecho sea el ligamiento entre los *loci*, más duradero será el desequilibrio.

Finalmente, conviene recordar lo que se ha apuntado en la Introducción al respecto de los estudios de asociación a partir de cohortes de casos y controles (ver 1.3.1). El desequilibrio de ligamiento es responsable de muchas de las asociaciones demostradas entre polimorfismos genéticos y enfermedades. Estos polimorfismos no tienen por qué ser funcionales, es decir no están implicados en la patogenia de la enfermedad. Sin embargo, se hallan en desequilibrio de ligamiento con los polimorfismos realmente funcionales, casi siempre desconocidos, situados en su mismo gen o en otro *locus* cercano, y cuyos alelos presentan desequilibrio con los alelos del polimorfismo conocido.

3. OBJETIVOS DE LA TESIS

El trabajo presentado en la Tesis constituye parte de la primera fase de un proyecto global encaminado a la búsqueda y análisis de los genes que determinan la enfermedad tromboembólica. Este proyecto, denominado GAIT (del inglés *Genetic Analysis of Idiopathic Thrombophilia*) comprende tres grandes fases:

1ª. Es la fase que ha originado esta Tesis y además una gran cantidad de información y resultados que todavía no han sido publicados. En esta parte del proyecto GAIT se ha obtenido una muestra adecuada de familias y se han determinado un gran número de fenotipos plasmáticos relacionados con la Hemostasia. También se han estudiado los genes ya conocidos que codifican las proteínas de la Hemostasia y algún otro gen potencialmente implicado en la trombofilia. Son los llamados genes candidatos.

2ª. Realización de un Análisis Global de Genoma y posterior estudio del Ligamiento Genético entre los fenotipos y cualquier zona del genoma, en busca de todos los *locus* posibles que contienen genes influyentes sobre la Hemostasia y la enfermedad tromboembólica (QTLs). Esta fase ya ha sido concluida y los resultados se encuentran en periodo de publicación.

3ª. Exploración de todos los *locus* detectados en la fase 2 con objeto de identificar a los genes responsables de las señales de ligamiento y a los polimorfismos intragénicos que determinan las diferencias entre los individuos en la población general. Esta fase todavía no se ha iniciado. Sin duda, será necesaria la incorporación de nuevos equipos de investigación para llevarla a término por la magnitud de la empresa. Su duración es imprevisible, pero en cualquier caso ocupará varios años más.

Este preámbulo a los objetivos de la Tesis pretende explicar que las tres fases son necesariamente consecutivas, siendo la primera imprescindible para la secuencia lógica del proyecto. En ella se ha realizado el análisis de los fenotipos plasmáticos de la Hemostasia y de un fenotipo especial, no observable empíricamente, pero deducible mediante modelos matemáticos: el riesgo individual de enfermedad tromboembólica.

OBJETIVOS ESPECIFICOS

1. Estudio de la heredabilidad de los fenotipos de la Hemostasia.
2. Estudio de la heredabilidad del riesgo de trombosis.
3. Investigación de la influencia de algunas variables ambientales sobre los fenotipos de la Hemostasia: edad, sexo, anticonceptivos orales, tabaco y dieta.
4. Análisis de las correlaciones entre los fenotipos plasmáticos, para determinar la existencia de correlaciones genéticas que suponen la presencia de genes con efectos pleiotrópicos sobre los citados fenotipos.
5. Análisis de las correlaciones entre el fenotipo "riesgo de trombosis" y el resto de fenotipos plasmáticos. En caso de detectar correlaciones genéticas, los genes pleiotrópicos responsables tendrán necesariamente interés clínico y los fenotipos involucrados podrían ser de gran utilidad diagnóstica.
6. Demostrar la utilidad y potencia del método utilizado basado en el Análisis de ligamiento genético de caracteres cuantitativos. Como ejemplo, se ha aplicado en el estudio de dos genes candidatos:
   - 6.1 Estudio de la relación entre el gen de la protrombina (factor II), más concretamente del polimorfismo G/A en la posición 20210, los niveles en plasma de protrombina y el riesgo de trombosis.
   - 6.2 Estudio de la relación entre el grupo sanguíneo ABO y algunos componentes de la Hemostasia a los que parece estar asociado (factor VIII y factor von Willebrand). En caso de demostrar esta asociación, se pretende aclarar su naturaleza: ¿se debe a ligamiento entre el *locus* ABO y los fenotipos o bien a desequilibrio de ligamiento entre el *locus* ABO y otro *locus* próximo aún no identificado?
7. En función de los resultados, es decir de la demostración de factores genéticos subyacentes a todos los fenotipos o a parte de ellos, y de la intensidad de sus efectos, el objetivo culminante de la Tesis es responder a la pregunta:
   ¿Tiene sentido proseguir la búsqueda de los hipotéticos genes responsables?

4. MÉTODOS

4.1 MUESTRA ESTUDIADA

Criterios de inclusión y reclutamiento

El proyecto GAIT está basado en el estudio de familias. La muestra se compone de 21 familias elegidas por su tamaño (para maximizar el poder estadístico para detectar efectos genéticos). Todas tienen como mínimo 10 individuos repartidos en 3 o más generaciones. Doce familias se seleccionaron a través de un *propositus* con trombofilia inexplicada, según la definición de trombofilia presentada en la Introducción (punto 1.1). La trombofilia se consideró inexplicada o idiopática porque ninguna de las causas conocidas de trombosis en 1995 fue demostrada en el estudio biológico del *propositus*: se descartaron las deficiencias de antitrombina, proteína C, proteína S, plasminógeno, cofactor II de la heparina y la presencia de anticoagulante lúpico, anticuerpos antifosfolípido, disfibrinogenemia o RPCa. Estos factores de riesgo tampoco estaban presentes en ninguno de los familiares con trombosis. Las 9 familias restantes se seleccionaron con los mismos criterios de tamaño, de entre la población general y de forma aleatoria, es decir sin atender a ningún criterio clínico de trombosis.

El reclutamiento de los miembros de las familias se realizó fundamentalmente en Barcelona, aunque se completó con individuos emparentados y residentes en Lleida, Tarragona, Albacete, Córdoba, Málaga y Cádiz.

La Tabla 3 refleja la distribución por familias de los 398 individuos que se analizaron (182 varones y 216 mujeres). De ellos, 101 son fundadores (individuos cuyos padres no están incluidos en los análisis) y 297 son no fundadores.

La Tabla 4 refleja la gran cantidad de parejas de parientes que se obtienen gracias al estudio de familias grandes. En este número total de 2744 parejas distintas descansa, en gran parte, la potencia estadística del estudio.

En la Figura 2 se representan los árboles de las familias (ver páginas 47-50).

Todos los sujetos fueron entrevistados por un médico para obtener información sobre sus antecedentes de trombosis venosa o arterial, edad de los eventos, otras enfermedades (como diabetes, obesidad o dislipemia), consumo de tabaco y en las mujeres uso de anticonceptivos orales e historia reproductiva. Se registró la composición de los domicilios (para valorar el efecto de influencias ambientales compartidas, como la dieta)

Se obtuvo el consentimiento informado de todos los participantes.

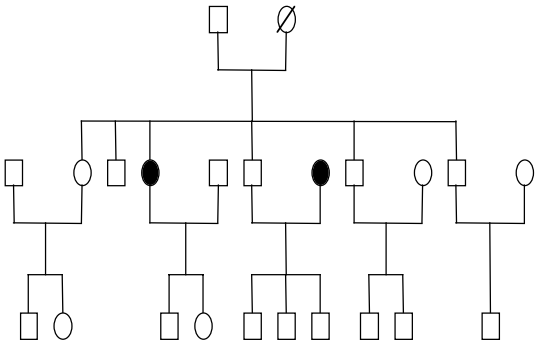
Tabla 3. Distribución por familias de los individuos examinados

Número de familia	Total individuos	Varones examinados	Mujeres examinadas	Número de domicilios	Tipo de familia
1	23	15	7	9	Control
2	30	14	13	14	Control
3	23	10	12	5	Control
4	16	6	9	5	Control
5	15	3	11	4	Control
6	28	10	14	13	Control
7	18	12	6	9	Control
8	24	7	15	7	Control
9	20	9	11	4	Control
10	29	10	12	7	Trombofilia
11	43	20	19	8	Trombofilia
12	24	5	8	10	Trombofilia
13	11	6	4	7	Trombofilia
14	10	6	4	7	Trombofilia
15	47	10	23	4	Trombofilia
16	27	9	10	9	Trombofilia
17	17	5	9	4	Trombofilia
18	15	8	5	9	Trombofilia
19	17	7	7	6	Trombofilia
20	19	7	11	8	Trombofilia
21	10	3	6	4	Trombofilia
Total	466	182	216	153	9 C / 12 Tr

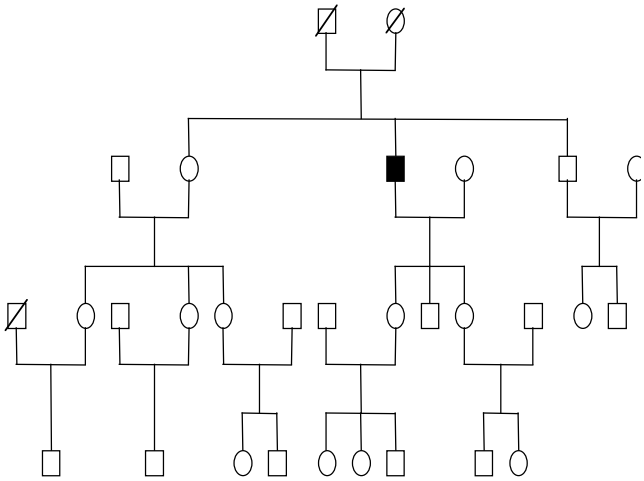
Tabla 4. Tipos de parentesco, grado de relación y número de parejas analizadas

Tipo de parentesco	Grado	Nº de parejas
Individuos (398)	0	–
Gemelos monocigotos	0	1
Padre-hijo	1	470
Hermanos	1	340
Abuelo-nieto	2	225
Tío-sobrino	2	693
Medio hermanos	2	13
Bisabuelo-bisnieto	3	13
Tío abuelo-nieto	3	137
Primos hermanos	3	547
Tío bisabuelo-bisnieto	4	9
Tío segundo-sobrino	4	233
Primos segundos	5	63

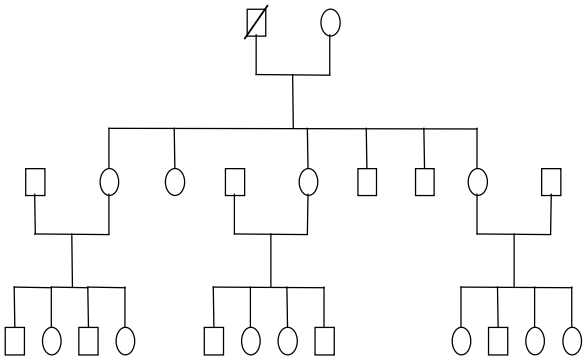
Figura 2. Arboles de las familias incluidas



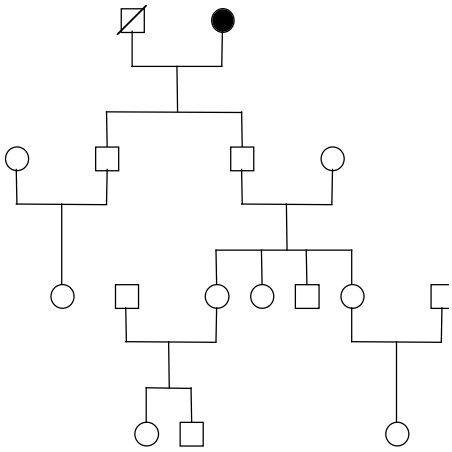
Familia 1



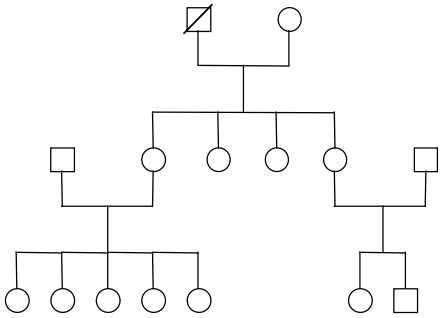
Familia 2



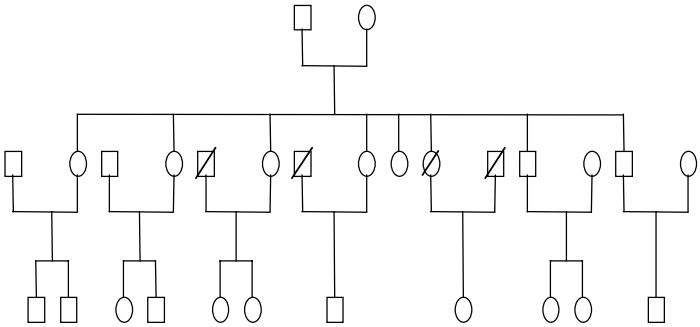
Familia 3



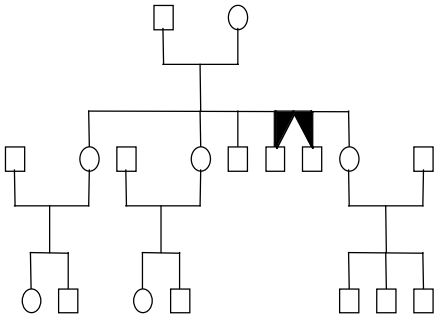
Familia 4



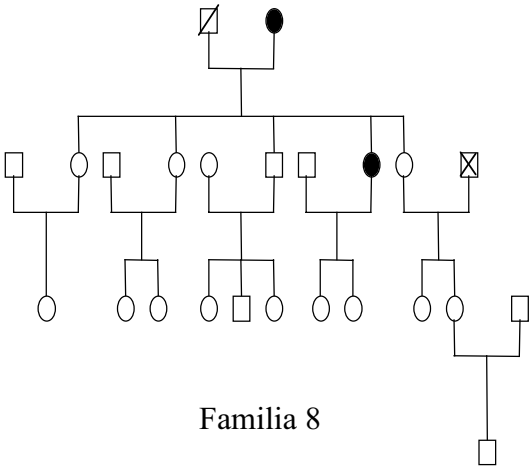
Familia 5



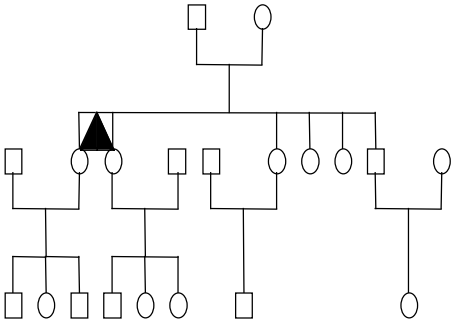
Familia 6



Familia 7

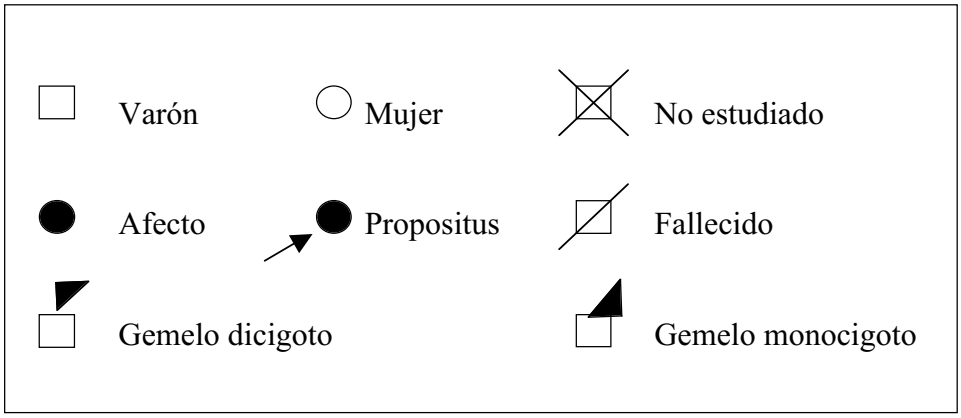


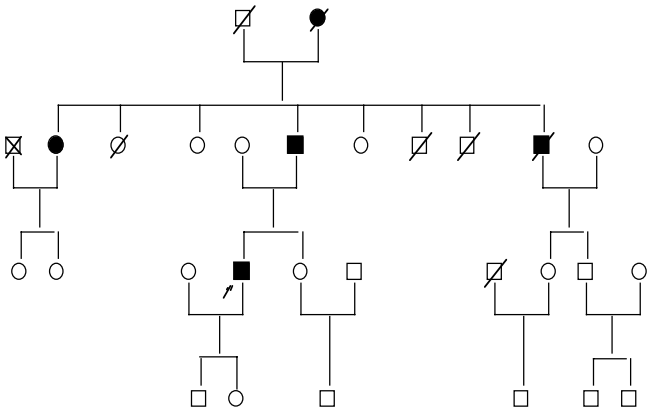
Familia 8



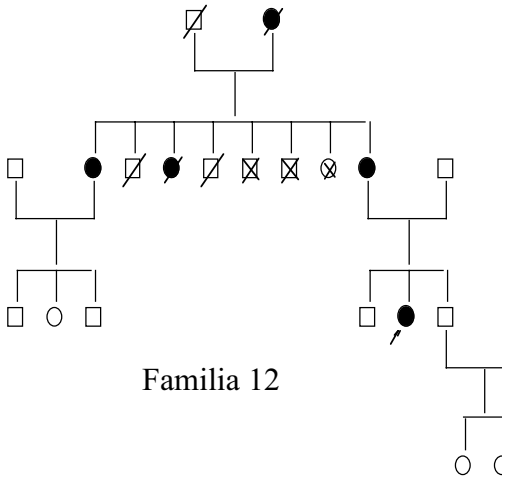
Familia 9

Leyenda

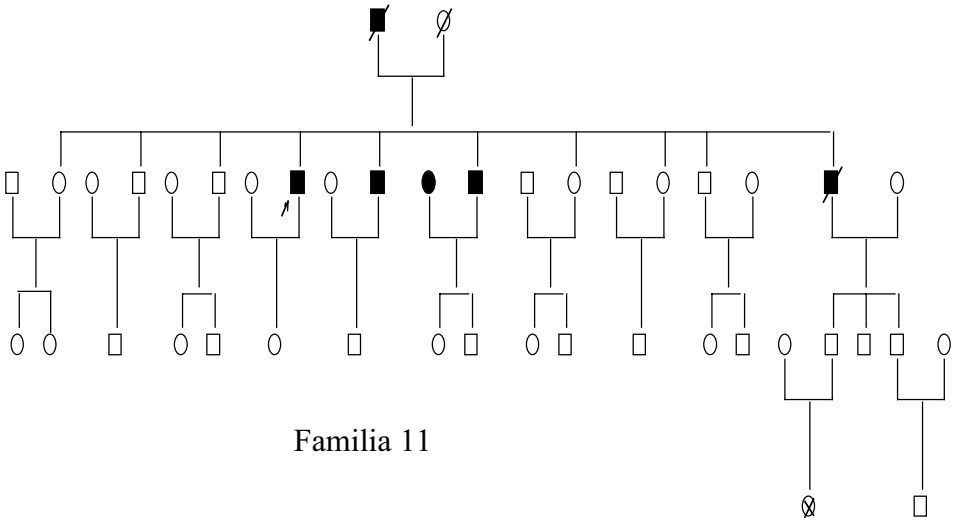




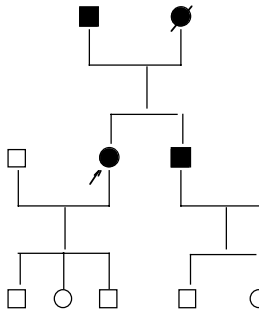
Familia 10



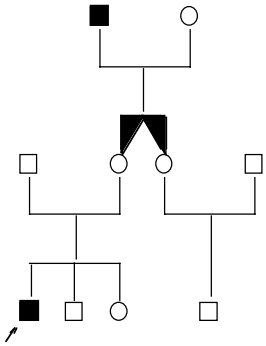
Familia 12



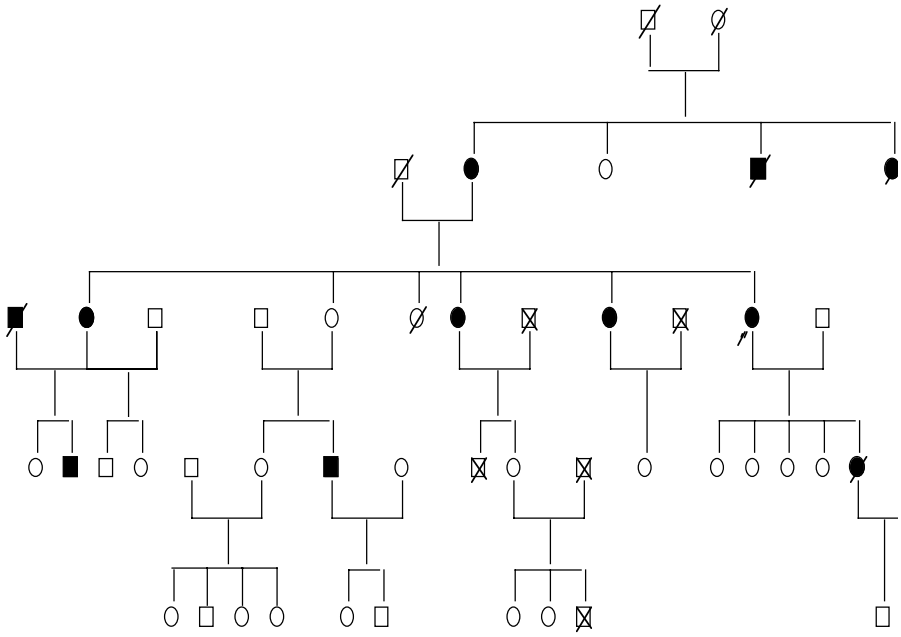
Familia 11



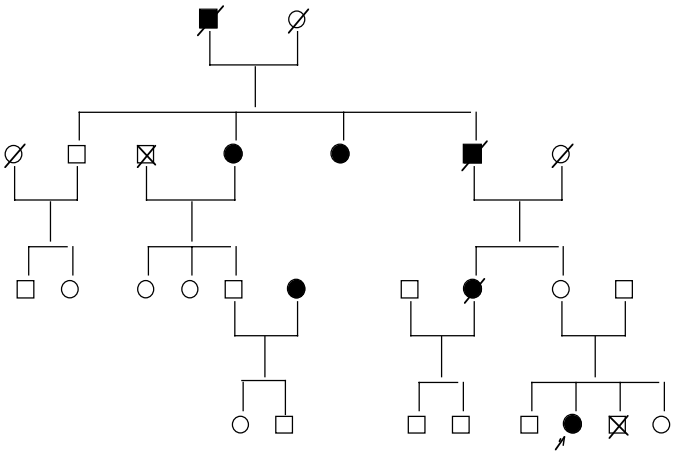
Familia 13



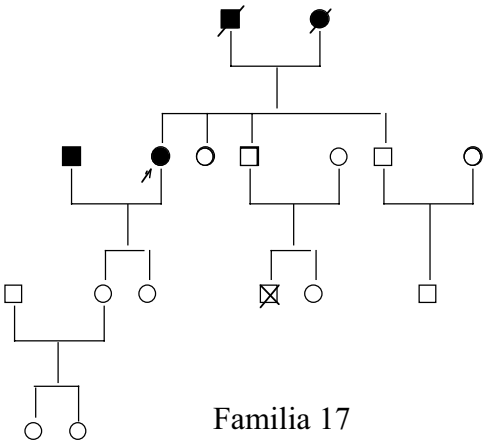
Familia 14



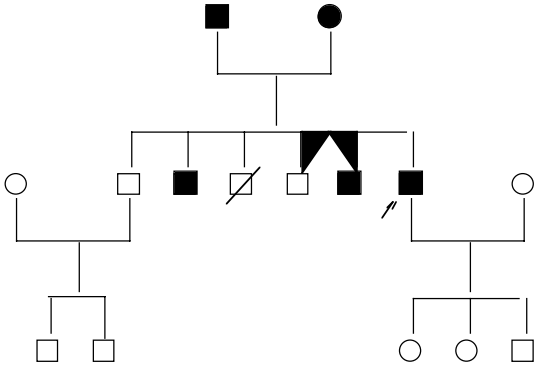
Familia 15



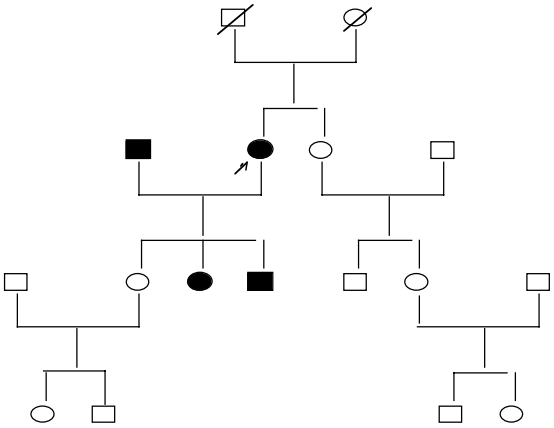
Familia 16



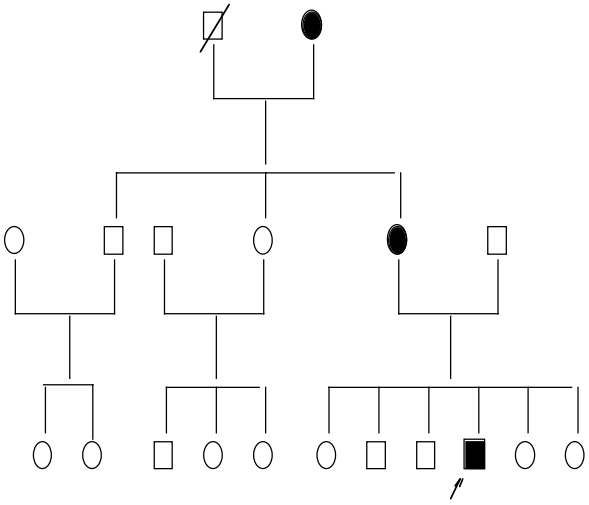
Familia 17



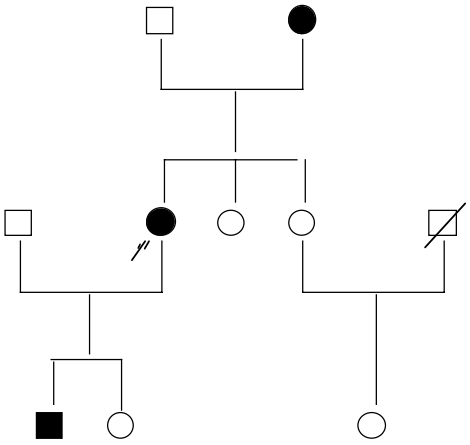
Familia 18



Familia 19



Familia 20



Familia 21

4.2 DETERMINACIONES DE LABORATORIO

Plasmáticas

Los fenotipos de la Hemostasia incluidos en el estudio se determinaron, en los 398 individuos reclutados, mediante técnicas convencionales de laboratorio:

a) Tiempos de coagulación: tiempo de tromboplastina parcial activado (TTPA), tiempo de protrombina, tiempo de protrombina en presencia (R1) o ausencia (R2) de trombomodulina.
b) Determinaciones funcionales mediante pruebas coagulométricas: fibrinógeno, factores II, V, VII, VIII, IX, X, XI, XII, proteína S funcional y resistencia a la proteína C activada (RPCa).
c) Determinaciones funcionales mediante substratos cromogénicos: antitrombina, proteína C, cofactor II de la heparina, plasminógeno, inhibidor del activador del plasminógeno tipo 1 (PAI-1), inhibidor de la via del factor tisular (TFPI), α_2-antiplasmina, precalicreína.
d) Enzimainmunoensayo (ELISA): proteína S total, proteína S libre, activador tisular del plasminógeno (t-PA), Dímero D, factor tisular, factor von Willebrand, β_2-glicoproteína I, P-selectina soluble.
e) Electroinmunoensayo: glicoproteína rica en histidina (HRG), factor XIII subunidad A, factor XIII subunidad S.

La homocisteína se determinó, tras separación mediante HPLC, con un método fluorométrico. El folato sérico, el folato en sangre total y la vitamina B12 se midieron con un ensayo comercial.

Finalmente se analizaron colesterol total, triglicéridos, fracciones del colesterol (HDL, LDL, VLDL) y lipoproteína A. Estos componentes del metabolismo lipídico no se presentan en la Tesis.

Genéticas

Mediante el uso de técnicas de PCR, también habituales en los laboratorios de Genética Molecular, se han genotipado distintos marcadores genéticos en los 398 individuos incluidos. Los marcadores se pueden clasificar en tres grupos:

1. Polimorfismos descritos en genes candidatos y asociados previamente con el riesgo de trombosis
   - mutación G1691A en el gen del factor V (factor V Leiden)
   - mutación G20210A en el gen de la protrombina
   - polimorfismo C677T en el gen de la metilen-tetrahidro folato reductasa
   - genotipo ABO

2. Otros marcadores intragénicos o muy próximos (ligados) a un total de 29 genes candidatos que codifican las proteínas plasmáticas de la Hemostasia.

3. Análisis Global del Genoma, mediante el genotipado de unos 400 marcadores del tipo microsatélite, altamente polimórficos y por lo tanto muy informativos para el Ligamiento. Con estos marcadores se ha establecido un mapa del genoma con señales aproximadamente cada 10 centimorgans (1cM supone alrededor de 10^6 pares de bases)

De toda esta enorme cantidad de información genética, en la presente Tesis sólo se presentan resultados relacionados con la mutación G20210A en el gen de la protrombina y con el genotipo ABO.

4.3 MÉTODOS ESTADÍSTICOS GENÉTICOS

A diferencia de las enfermedades monogénicas (como la fibrosis quística, neurofibromatosis o distrofia miotónica), cuyos genes responsables han sido localizados con éxito, por medio de métodos clásicos de ligamiento *(Ott 1999)*, la mayoría de enfermedades comunes envuelven múltiples componentes genéticos y ambientales y sus interacciones. El análisis estadístico genético de estas enfermedades requiere aproximaciones matemáticas diferentes para la localización y evaluación de la importancia relativa de los *locus* de rasgo cuantitativos (QTLs) implicados. En el proyecto GAIT se aplican modelos estadísticos genéticos basados en los *componentes de la variancia*. En contra de lo que pudiera parecer, estos modelos tienen una larga historia y han sido desarrollados a lo largo del siglo XX. En la actualidad podemos considerarlos como la herramienta más potente para la localización de genes relacionados con las enfermedades complejas más frecuentes. Seguidamente resumimos los hitos principales que mejor describen la evolución de dicha metodología.

Fue en 1918 cuando R.A. Fisher, en un artículo clásico, introdujo el término estadístico *variancia*. Mostró como el comportamiento mendeliano de múltiples genes actuando juntos explica perfectamente los rasgos cuantitativos. Con ello armonizó la herencia mendeliana con la teoría de la evolución y demostró que la variancia de un rasgo contínuo se puede descomponer en componentes genéticos aditivos y no-aditivos *(Fisher 1918)*.

En los años 30 se elaboró por primera vez un método de ligamiento entre un rasgo cuantitativo contínuo y un marcador genético evaluando las correlaciones en parejas de hermanos *(Penrose 1938)*.

En 1953, C.R. Henderson fue el primero en aplicar los métodos matemáticos de *Máxima Verosimilitud* en Genética para obtener estimaciones de parámetros estadísticos a partir de grandes genealogías (*pedigrees*) de animales *(Henderson 1953)*.

Hace ya 30 años, J.K. Haseman y R.C. Elston investigaron el ligamiento genético entre un rasgo cuantitativo y un marcador mediante análisis de regresión. Su aportación fundamental deriva del uso de los alelos idénticos por descendencia (en inglés *identical-by-descent* o IBD) compartidos por una pareja de parientes, como explicación de la correlación observada entre los respectivos valores del fenotipo *(Haseman y Elston 1972)*.

Poco después se determinaron las bases para la estimación de componentes de la variancia en grandes genealogías en humanos, también mediante métodos análiticos de Máxima Verosimilitud. Con ello culminó el desarrollo teórico de los métodos basados en la descomposición de la variancia de rasgos cuantitativos medidos en grandes familias *(Lange y col 1976)*. Pero la aplicación de esta teoría quedó limitada por el enorme costo computacional y de cálculo, inasequible para los ordenadores de la época.

En 1982, J.L. Hopper y J.D. Mathews refinaron el modelo añadiendo la medida de los efectos de un marcador genético específico dentro de la variancia genética total y la estimación de los efectos ambientales compartidos entre individuos de una misma familia. También describieron un método que permite corregir posibles sesgos debidos al reclutamiento de las familias a través de probandos con valores extremos del rasgo en estudio, por ejemplo sujetos afectos de una enfermedad *(Hopper y Mathews 1982)*.

Fue en la década de los 90 cuando el desarrollo espectacular de la computación mediante ordenadores permitió la aplicación de toda esta formidable herramienta al análisis de ligamiento de rasgos o fenotipos cuantitativos, en combinación con nuevos progresos teóricos.

En el año 1990, D.E Goldgar presentó un método para descomponer la variancia genética cuantitativa en efectos debidos a regiones cromosómicas específicas *(Goldgar 1990)*. Se basó en la estimación de la proporción de material genético compartido entre dos hermanos, proveniente de un antepasado común a ambos. Era, de nuevo el concepto de *identidad por descendencia* que Haseman y Elston habían usado en 1972 para estimar regresiones.

En 1994 se estableció el modelo, de nuevo a partir de los componentes de la variancia, en el que la variabilidad entre los distintos individuos de una misma familia se expresa a partir de efectos debidos a las covariables ambientales, efectos causados por un *locus* principal, efectos poligénicos debidos a múltiples *loci* menores a lo largo de todo el genoma y por último efectos residuales debidos a factores no genéticos *(Amos 1994)*.

Finalmente, L. Almasy y J. Blangero en 1998 extienden estos métodos para obtener análisis de ligamiento de múltiples puntos a la vez, en familias de cualquier tamaño y complejidad estructural *(Almasy y Blangero 1998)*. Con ello se alcanza la máxima potencia estadística. Este es el modelo general aplicado en el proyecto GAIT:

En resumen, la idea básica subyacente es atribuir la variancia observada en la población a una variedad de causas tanto genéticas como no-genéticas (ambientales). El modelo busca explicar las correlaciones de un fenotipo complejo observadas entre miembros de una misma familia, descomponiendo la variancia medida en ese fenotipo en distintos componentes: efecto causado por un QTL específico ligado a un marcador conocido (genotipado), efectos causados por otros QTL (en número indefinido) y no ligados a la región en estudio (aquella que marca el mencionado marcador), efectos causados por factores ambientales compartidos por distintos familiares (por ejemplo, la dieta) y otras fuentes de variabilidad ambiental específicas para cada individuo (por ejemplo, el error de medida del fenotipo). Mediante técnicas de Máxima Verosimilitud y con el concurso imprescindible de ordenadores muy potentes, se estima el valor de cada parámetro (heredabilidades, correlaciones, LODs) que mejor encaja o que explica con la mayor *verosimilitud* todos los datos biológicos (por ejemplo las determinaciones de laboratorio) recogidos en el trabajo de campo.

Todas las estimaciones que se efectúan en la tesis se han obtenido una vez controlado el efecto de las covariables ambientales recogidas en el estudio (edad, sexo, tabaquismo y uso de anticonceptivos orales) sobre los fenotipos plasmáticos medidos en el laboratorio o inferidos matemáticamente (riesgo de trombosis). Debido a que 12 de las familias se reclutaron a través de un *propositus* con trombofilia, las estimaciones presentadas también han sufrido una corrección estadística (corrección de reclutamiento) para evitar sesgos y obtener resultados aplicables a la población general *(Hopper y Mathews 1982, Boehnke y Lange 1984).*

Los resultados se han obtenido mediante el programa de ordenador denominado SOLAR (*Sequential Oligogenic Linkage Analysis Routines*) que puede obtenerse libremente a través de Internet en la dirección www.sfbr.org

Nota: en cada uno de los artículos 2 a 7 que componen la Tesis se detalla con mayor amplitud la metodología estadística aplicada.

5. COPIA DE LAS PUBLICACIONES

CONTENIDO DE LOS ARTÍCULOS

El artículo número 5.1 de los que se presentan sirve como ejemplo de la complejidad de la trombosis, y de cuánto nos queda todavía por conocer sobre la fisiopatogenia de la enfermedad. Se trata de un caso familiar de trombofilia, en que se identificó un factor genético de riesgo, la mutación G20210A en el gen de la protrombina descrita en 1996 *(Poort y col 1996).* Se sabe que los portadores heterocigotos de esta mutación presentan un riesgo de trombosis unas 3 veces mayor que la población normal. Paradójicamente, en la familia hay individuos homocigotos para esta mutación asintomáticos, mientras que los miembros con clínica de tromboembolismo venoso son heterocigotos. En teoría, los homocigotos para un defecto genético causante de enfermedad deben tener un riesgo notablemente superior que los heterocigotos. Esto es muy claro en las enfermedades hereditarias de tipo monogénico (mendeliano), pero tal y como demuestra este caso, la patogenia de la trombosis venosa es más compleja, y otros factores genéticos y ambientales propios de cada individuo, son los que finalmente condicionan quién y cuando presenta un evento trombótico.

Además, este caso familiar ha sido uno de los primeros publicados en la literatura internacional de portadores homocigotos de la mutación PT20210A sin clínica tromboembólica.

En el artículo número 5.2 se publica la heredabilidad y el efecto domiciliario de 27 fenotipos. En total, el proyecto GAIT ha estimado la heredabilidad de 40 fenotipos relacionados con la Hemostasia (los resultados de otros 12 fenotipos están en proceso de publicación, artículo 5.7).

En el artículo 5.3 se describe la construcción de la variable "riesgo de trombosis" a partir de las familias reclutadas en el proyecto GAIT y los individuos afectos de trombosis. A continuación se estima su heredabilidad.

En los artículos 5.3 y 5.7 se presentan los resultados del estudio de correlaciones entre el fenotipo "riesgo de trombosis" y los restantes fenotipos plasmáticos.

En el artículo 5.4 se analizan las correlaciones entre todos los fenotipos relacionados con la vitamina K.

En los artículos 5.5 y 5.6 de la Tesis, se utilizan los conceptos de “ligamiento” y de “desequilibrio de ligamiento” para explorar la influencia de dos conocidos marcadores genéticos en algunos fenotipos de la hemostasia. La mutación G20210A en el gen de la protrombina (factor II) resulta ligada a los niveles plasmáticos de factor II y al fenotipo “riesgo de trombosis”. Además se demuestra que es una mutación funcional por sí misma, es decir no está marcando a otro *locus* cercano. Un resultado similar se obtiene con el *locus* del grupo sanguíneo ABO, en el cromosoma 9, y los niveles plasmáticos de factor VIII y de factor von Willebrand.

5.1 Homocigotos para el alelo 20210 A en el gen de la protrombina en una familia trombofílica sin manifestaciones clínicas de tromboembolismo venoso

(*Haematologica 1999;84:627-632)*

Haematologica 1999; 84:627-632 original paper

Homozygotes for prothrombin gene 20210 A allele in a thrombophilic family without clinical manifestations of venous thromboembolism

JUAN CARLOS SOUTO, JOSÉ MATEO, JOSÉ MANUEL SORIA, DOLORS LLOBET, INMA COLL, MONTSERRAT BORRELL, JORDI FONTCUBERTA
Unitat d'Hemostàsia i Trombosi, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain

ABSTRACT

Background and Objective. **A new genetic risk factor for venous thromboembolism has recently been described which involves a G to A transition at position 20210 in the 3' untranslated region of the prothrombin gene. To date, only a few homozygotes for this mutation have been reported and in most of cases, they suffered from thrombotic disease. Here, we describe a pedigree including both heterozygous and homozygous subjects for prothrombin (PT) 20210 A.**

Design and Methods. **This family was recruited in 1996 as part of our GAIT (*Genetic Analysis of Idiopathic Thrombophilia*) project. To qualify for the GAIT study, a pedigree was required to have at least 10 living individuals in three or more generations (i.e. extended pedigree). The pedigrees were selected through probands with idiopathic thrombophilia. A complete set of plasma and DNA determinations related to hemostasis was performed on this family.**

Results. **The plasma studies yielded normal results in all of the individuals. The family members who had a history of thromboembolism were heterozygous carriers of the PT 20210 A variant. In addition, 4 relatives who were heterozygous, and two who were homozygous for this A allele, failed to show clinical manifestations. These two homozygotes were 51 and 19 years old.**

Interpretation and Conclusions. **This case exemplifies the complexity of thrombotic disease since individuals homozygous for a mutant gene do not exhibit symptoms while heterozygous individuals often do exhibit the disease. This case suggests that the new genetic risk factor for thrombosis (i.e. PT 20210 A) may not be as strong as most of the previously described genetic risk factors.**

Key words: thrombophilia, prothrombin gene, factor II, prothrombin gene 20210 A mutation

Correspondence: Juan Carlos Souto, M.D., Unitat d'Hemostàsia i Trombosi, Hospital de la Santa Creu i Sant Pau, Avda Sant Antoni Mª Claret 167, 08025 Barcelona, Spain.
Phone: international +34-93-2919193 – Fax: international +34-93-2919192 – E-mail: jsouto@hsp.santpau.es

Thrombophilia is a common disease clinically defined by early age of onset, repeated episodes of venous thromboembolism (VTE) and frequent co-existence in related individuals.[1] This last suggests that heredity plays a role in susceptibility to thrombophilia. Only a few inherited deficiencies are considered as independent risk factors for VTE. Among these are mutations in structural genes encoding for antithrombin, protein C, protein S, and fibrinogen.[1] Some individuals with VTE are heterozygous carriers of one of these mutant genes. In contrast, the rare homozygous individuals exhibit very severe thrombotic symptoms.[2] In 1993, activated protein C resistance was identified as a very frequent cause of inheritable thrombophilia determined in the great majority of cases by the factor V Leiden mutation.[3] In 1996, Poort *et al.* described a G to A transition at position 20210 in the 3' untranslated region (UT) of the prothrombin gene, which was also a new genetic risk factor for VTE. An important finding from this seminal work was the significant increase of plasma levels of prothrombin in the carriers of the A allele. Unfortunately, we do not know the pathogenic mechanisms associated with this genetic variant. The risk of VTE in heterozygous carriers of the 20210 A allele was estimated to be 2.8 times higher than in non-carriers.[4]

More than 40 epidemiological studies appeared during 1997 and 1998 that reported the prevalence of this variant in different countries or ethnic groups, ranging from 0% to 18% in patients with VTE or arterial disease and ranging from 0% to 8.1% in control individuals.[5-7] The biggest study published up to now estimated a prevalence in the normal population between 1.4% and 2.6%.[8] Recently, our study in Spain found one of the highest prevalences reported to date in healthy people: 6.5% (confidence interval 95% 3.5-10.8). Furthermore, the 20210 A variant appears to be the most prevalent genetic risk factor for thrombosis in our geographical area, accounting for the condition in 17.2% of the patients.[9]

It is important to note that the great majority of individuals described in these studies were heterozy-

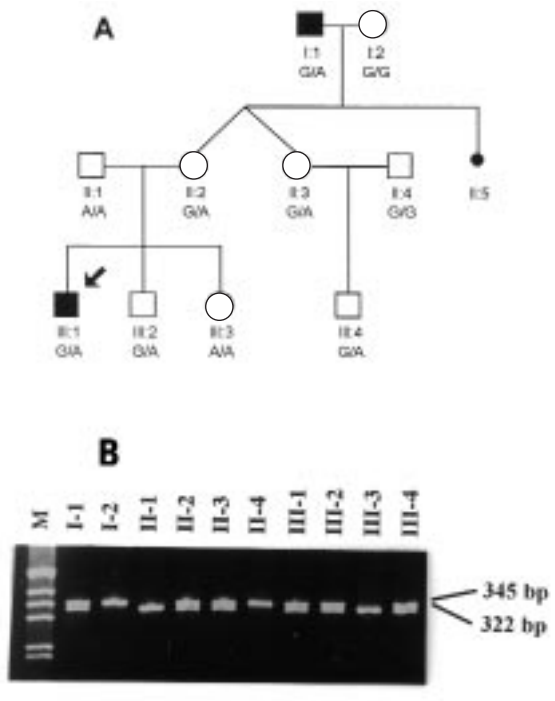


Figure 1. A: Pedigree of the reported family: The proband is indicated by an arrow. The 20210 variant genotype present in each member is also shown under his / her symbol. Filled-black symbols indicate thrombotic disease. Symbol II-5 refers to spontaneous abortion. B: Familial segregation of the 20210 variant. A new *HindIII* site is introduced in the amplified fragment when the 20210 A allele is present, yielding two fragments (322 bp and 23 bp in length) after digestion. The 20210 G allele lacks the restriction site and therefore generates only a 345 bp fragment by PCR-*HindIII* digestion. M is the Φ174 DNA / *HaeIII* marker. Individual numbers along the top refer to the same numbers as the pedigree.

gous. To our knowledge, only 34 cases of homozygous individuals for the 20210 A allele have been reported.[4,7,10-22] Of these, 9 were in 3 families.[16,18,22] The remaining 25 individuals are unrelated. A total of 17 cases had thrombosis, including 14 individuals with venous thrombosis. Nine cases remain asymptomatic; four of them belonging to the same pedigree.[18] There is no clinical information about the 8 cases mentioned by Zivelin *et al*.[17] Here, we present two new cases of homozygous individuals for this mutation. Remarkably, neither of these homozygotes has experienced thrombosis in spite of the fact that they belong to a family in which hereditary thrombophilia is clearly evident.

Design and Methods

Case Report

The family was recruited in 1996 as part of our GAIT (Genetic Analysis of Idiopathic Thrombophilia) project.[23] To qualify for the GAIT study, a pedigree was required to have at least 10 living individuals in three or more generations (i.e. extended pedigree). The pedigrees were selected through probands with idiopathic thrombophilia. The proband's thrombophilia was considered idiopathic because all known (during the recruitment period of 1995-1997) biological causes of thrombophilia were excluded (i.e., antithrombin deficiency, protein C and S deficiencies, activated protein C resistance and factor V Leiden, plasminogen deficiency, heparin cofactor II deficiency, dysfibrinogenemia, lupus anticoagulant and antiphospholipid antibodies).

The proband of this family (individual III-1, see Figure 1), is a 25 year-old male who suffered from spontaneous deep venous thromboses at the age of 19 in the vein cava and right iliac vein. There were diagnosed by means of venography and abdominal CT-scan. Initially, he received heparin treatment followed by a six-month treatment with acenocoumarol. As a sequel, a minor post-thrombotic syndrome remained in his right leg. At the age of 22, he developed a new episode of spontaneous left iliofemoral vein thrombosis. An objective diagnosis was made by ultrasonography. Since then, he has been under oral anticoagulant therapy as prophylaxis against the disease. His maternal grandfather (individual I-1, Figure 1) suffered from deep venous thrombosis in his right leg after a surgical repair of a groin hernia at the age of 62. He had no other predisposing conditions to thrombosis through his life.

None of the remaining pedigree members has had thromboembolic disease, although some of them have been exposed to some risk factors for thrombosis such as pregnancy and puerperium (individuals I-2, II-2 and II-3), surgical procedures (II-3) and oral contraceptives (II-2 and II-3). Specifically, II-1 and III-3 have not been exposed to acquired risk factors. The individual I-2 had a spontaneous abortion during the second trimester of her second pregnancy, presented as II-5 in the family tree.

After we had finished the required analyses of all of the members of this family for our GAIT project, we stored samples of plasma and DNA in the event that further investigations were warranted. When Poort *et al.* reported the discovery of the prothrombin gene 20210 A allele as a risk factor for thrombosis,[4] we retrospectively tested the probands in our GAIT families. We found that the proband of this family was a carrier of the A allele at position 20210 of the prothrombin gene. Consequently, we investigated all of his relatives.

Methods

Plasma study. APTT, PT, fibrinogen, coagulation factors IIc, Vc,VIIc,VIIIc, IXc, Xc, XIc and XIIc (coagulative methods); von Willebrand factor (antigen) ; total, free and functional protein S; APC resistance, antithrombin (functional), protein C (functional),

heparin cofactor II (functional), plasminogen (functional), t-PA (antigen), PAI-1 (functional), HRG (antigen), TFPI (functional), tissue factor (antigen) and homocysteine were determined by means of standard methods.

Factor V-Leiden detection. Factor V Leiden genotype was screened using the two primers described previously,[24] with minor modifications in the reaction conditions.

Detection of the prothrombin gene 20210 variant. The 3'-UT region of the prothrombin gene was obtained by PCR as previously described,[4] with minor modifications in the reaction conditions. The 345-bp fragment was digested with *Hind*III endonuclease (Life Technologies Inc. Gaithersburg, MD, USA) according to the recommendation from the supplier. Digestion products were analyzed by ethidium bromide UV-fluorescence after electrophoresis in 3% Nusieve GTG agarose gel (FMC Bioproducts, Rockland, ME, USA). To confirm the results, another 418-bp fragment, spanning position 19889 to 20307, from exon 14 and the 3'-UT region of the prothrombin gene was amplified and sequenced directly in an Applied Biosystem 310 DNA sequencer following the instructions from the supplier.

Results

Initially, all of the classical risk factors for thrombophilia tested normal in the family members. The 20210 prothrombin variant analysis using PCR-*Hind*III digestion (Figure 1) demonstrated that the proband (III-1) was heterozygous for the 20210 A allele as was his maternal grandfather (I-1). These were the only members of the family with a history of thromboembolic disease. We completed the analysis in the remaining family members and found 4 additional heterozygous (II-2, II-3, III-2 and III-4) and two homozygotes: the father (II-1) and the sister (III-3) of the proband. Figure 2 shows the genomic sequence of the propositus DNA (using the reverse primer), the DNA of his father and a control DNA known to be homozygous G/G for the 20210 variant (C/C in the antisense strand). Table 1 shows the clinical data, current ages, prothrombin levels and 20210 genotypes for all of the individuals in this study. The paternal grandparents of the proband (not tested) were not known to be related. The maternal grandparents were not likely to be related to the paternal ones, since the maternal branch came from Barcelona (Catalunya, in East Spain) and the paternal came from the region of Castilla-León, in Central Spain.

Discussion

To our knowledge, this is the second report of homozygous individuals for the prothrombin gene 20210 variant, belonging to the same thrombophilic pedigree, in whom there is no evidence of any thromboembolic events, despite the fact that one of the patients is 51 years old. A considerable number of

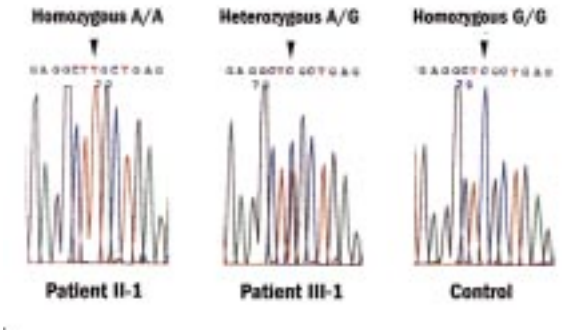


Figure 2. Partial sequences of genomic DNA of the 3'-UTR region of the prothrombin gene from the proband (III-1), his father (II-1) and a control. The antisense strand sequence is shown with the 20210 G→A variant (C to T in the antisense strand) indicated by an arrow.

Table 1. Clinical and laboratory features of the family.

Family member	*Current age (yr)*	*History of VTE*	*20210 genotype*	*Factor IIc levels**
I-1	78	Yes	G/A	149
I-2	77	No	G/G	152
II-1	51	No	A/A	168
II-2	49	No	G/A	156
II-3	49	No	G/A	133
II-4	51	No	G/G	147
III-1 (proband)	25	Yes	G/A	140°
III-2	21	No	G/A	142
III-3	19	No	A/A	183
III-4	17	No	G/A	145

**Values for Factor IIc plasma levels are given in %. Normal values in our laboratory are 70-125%. °The proband stopped oral anticoagulation 1 month before testing factor II levels.*

cases of homozygous individuals for the G20210A variant in the prothrombin gene, have been clinically reported. Table 2 summarizes the clinical information available from all of these reported PT 20210 AA individuals. In some of them, the associations with factor V Leiden or hyperhomocysteinemia make it difficult to interpret the role played by PT 20210 A allele in the thrombotic events, although a synergy could be suspected. Among the published series there are 17 cases of thromboembolic disease; at least 4 cases were spontaneous and in another 4 there were related triggering factors. At least 6 patients have had recurrent events.

Apart from these sporadic cases, there are no data about the specific risk associated with the homozygous state of the 20210 A allele. The fact that even homozygotes and heterozygotes may not show any symptoms makes the prognosis of the thrombotic risk extremely tenuous. More studies are needed to resolve this dilemma. But, from the observed clinical

Table 2. Reported cases of homozygous individuals PT 20210 AA.

Case	*Ref.*	*Sex*	*Current age*	*Thrombosis (age of first)*	*Location*	*Triggering factors*	*Recurrence*	*Associated risk factors*	*Factor IIc levels (%)*
1	4	F	#	Y (#)	#	#	#	F.VLeiden	#
2	7	M	Elderly	N	–	–	–	#	#
3	10	M	24	Y (24)	DVT/PE	MI	N	F.V Leiden	146
4	11	F	18	Y (18)	DVT	Pregnancy	N	N	136
5	12	M	>70	Y (66)	retina	#	Y	Hyper Hcy	#
6	13	M	26	Y (24)	stroke	#	Y	N	#
7	13	M	26	Y (26)	stroke	#	N	Foramen ovale	#
8	14	#	#	N	–	–	–	#	#
9–13	15	#	#	Y (#)	DVT	#	#	#	#
14	16	M	56	Y (40)	DVT	N	Y	N	154
15	16	F	52	Y (26)	STP	Pregnancy	Y	N	170
16–23	17	#	#	#	#	#	#	#	#
24	18	M	44	Y (#)	DVT/PE	N	N	N	132
25–28	18	F	33–74	N	–	–	–	N	113–129
29	19	M	65	Y (65)	DVT	Surgery	Y	N	142
30–31	20	#	#	N	–	–	–	#	96/137*
32	21	M	72	N	–	–	–	N	#
33	22	M	48	Y (40)	DVT/PE	N	Y	N	148*
34	22	F	>48	Y (30)	PE	N	N	N	205*
35	Ours	M	51	N	–	–	–	N	168
36	Ours	F	19	N	–	–	–	N	183

*F: female, M: male, #: not reported, –: not applicable, Y: yes, N: no; DVT: deep venous thrombosis, PE: pulmonary embolism, STP: superficial thrombophlebitis; MI: myocardial infarction, hyperHcy: moderate hyperhomocysteinemia; *antigen levels of factor II.*

data we can make some comparisons with other genetic thrombophilic defects.

Individuals homozygous for the 20210 A allele seem to be much less affected than individuals homozygous for protein C, protein S or antithrombin deficiency.[2] This can be reasonably concluded because none of the previously reported patients suffered from thrombosis in their childhood. Further, one of our cases was an asymptomatic homozygote even at the age of 51. In the report from Morange *et al.* the four asymptomatic individuals are a mother aged 74 and 3 sisters, all older than 33 years. Each of these women had several pregnancies without thrombotic complications.[18] Furthermore, the individual mentioned by Akar *et al.* is an asymptomatic grandfather.[7] The case reported by Alatri *et al.* is an asymptomatic man aged 72 who has had several risk situations for thrombosis during his life.[21] In this sense, the PT 20210 A allele would be more similar to factor V Leiden, since there are several cases of homozygous individuals for this mutation without thrombotic disease.[25, 26]

There are two family cases in which two homozygous siblings have suffered from recurrent venous thrombosis.[16,22] In addition, there are at least two other families including six homozygotes without thromboembolic disease (Morange *et al.* and the present study). Theoretically, it would be expected that individuals homozygous for a thrombophilia risk factor would have a higher probability of developing thrombotic disease than individuals who are heterozygous. In fact, our cases argue against this expectation, since thrombosis has appeared only in heterozygotes, and not in homozygotes. One explanation might be that thrombotic risk is in fact higher in homozygous than in heterozygous, but that there is an epistatic locus inhibiting the risk. Alternatively, there may be an unknown risk factor (genetic or not) associated with the heterozygote, leading perhaps to gene conversion and subsequent clinical manifestations. Moreover, we must emphasize that a complete set of hemostatic parameters was normal in all of the members of our family.

An interesting question arises concerning the higher prothrombin plasma levels in 20210 AA homozygotes than in the heterozygotes or normal relatives. As with any complex phenotype, plasma prothrombin levels are determined by the interaction of genetic and environmental factors. It is also likely that the prothrombin levels are controlled, in part, by multiple genes (mainly regulatory). For this reason, it is necessary to compare relatives (who share genetic and environmental backgrounds) because such family studies would avoid interfamilial heterogeneity.[27] As a general trend, in the four pedigrees mentioned here, the prothrombin levels are higher in homozygotes than in heterozygotes, and also higher in heterozygotes than in non-carriers. Nevertheless, in all of these homozygous individuals the plasma levels of prothrombin are far from those expected if the genet-

ic effect of this variant were additive. Another remarkable point is that some individuals with normal alleles have plasma levels above the upper limit of the normal range (Table 2, individuals I-2 and II-4). This must be due to the above-mentioned specific genetic and environmental factors. It has been demonstrated recently that prothrombin levels have a wide range of values both in carriers of PT 20210 A and in normal controls.[20] In relation to the prothrombotic state, it is perhaps more important to investigate the ability of the affected individuals to generate active thrombin, rather than their levels of circulating plasma prothrombin. Interestingly, the cases reported by Kyrle *et al.* had normal levels of prothrombin fragment F1+2, indicating the absence of ongoing hemostatic system activation but, simultaneously, they showed a clear increase in their endogenous thrombin potential.[16] Because physiopathological mechanisms responsible for thrombosis underlying this variant are unknown, further investigations, both epidemiological and biochemical are needed to answer the intriguing questions arising from this new thrombosis-related genetic abnormality, among which, why some homozygotes are asymptomatic.

Contributions and Acknowledgments

JCS and JM were responsible for the recruitment of the family, data analysis and writing the manuscript. JMS was responsible for the genetic analysis, wrote part of the manuscript and supplied the figures. DL and IC developed and carried out the molecular biology assays. MB was in charge of all the plasma studies and analyzed their results. JF was responsible for the conception of the study and its interpretation. We thank Elisabeth del Río, from the Servei de Genètica, Hospital de la Santa Creu i Sant Pau, Barcelona, for technical assistance with DNA sequencing and Professor William H. Stone, from the Department of Biology, Trinity University, San Antonio, TX, USA, for critically reviewing the manuscript.

Funding

Supported by grants DGICYT Sab 94-0170 and FIS 97/2032. JM.S is also supported by grant 1997 RED 00003, from Generalitat de Catalunya.

Disclosures

Conflict of interest: none

Redundant publications: no substantial overlapping with previous papers.

Manuscript processing

Manuscript received December 11, 1998; accepted March 15, 1999.

References

1. Lane DA, Mannucci PM, Bauer KA, et al. Inherited thrombophilia: Part 1. Thromb Haemost 1996; 76: 651-62.
2. Rosendaal FR. Risk factors for venous thrombosis: prevalence, risk, and interaction. Semin Hematol 1997; 34:171-87.
3. Zöller B, Svensson PJ, He X, Dahlbäck B. Identification of the same factor V gene mutation in 47 out of 50 thrombosis-prone families with inherited resistance to activated protein C. J Clin Invest 1994; 94:2521-4.
4. Poort SR, Rosendaal FR, Reitsma PH, Bertina RM. A common genetic variation in the 3'-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. Blood 1996; 88:3698-703.
5. Miyata T, Kawasaki T, Fujimura H, Uchida K, Tsushima M, Kato H. The prothrombin gene G20210A mutation is not found among Japanese patients with deep vein thrombosis and healthy individuals. Blood Coagul Fibrinol 1998; 9:451-2.
6. Martinelli I, Sacchi E, Landi G, Taioli E, Duca F, Mannucci PM. High risk of cerebral-vein thrombosis in carriers of a prothrombin-gene mutation and in users of oral contraceptives. N Engl J Med 1998; 338:1793-7.
7. Akar N, Misirlioglu, Akar E, Avcu F, Yalçin A, Sözüöz A. Prothrombin gene 20210 G-A mutation in the Turkish population. Am J Hematol 1998; 58:249-50.
8. Rosendaal FR, Doggen CJM, Zivelin A, et al. Geographic distribution of the 20210 G to A prothrombin variant. Thromb Haemost 1998; 79:706-8.
9. Souto JC, Coll I, Llobet D, et al. The prothrombin 20210 A allele is the most prevalent genetic risk factor for venous thromboembolism in the Spanish population. Thromb Haemost 1998; 80:366-9.
10. Howard TE, Marusa M, Channell C, Duncan A. A patient homozygous for a mutation in the prothrombin gene 3'-untranslated region associated with massive thrombosis. Blood Coagul Fibrinol 1997; 8:316-9.
11. Scott CM, Hanley JP, Ludlam CA, Stirling D. Homozygosity for a factor II polymorphism associated with thrombosis during pregnancy [abstract]. Thromb Haemost 1997; 77(suppl):770.
12. Kappur RK, Mills LA, Spitzer SG, Hultin MB. A prothrombin gene mutation is significantly associated with venous thrombosis. Arterioscler Thromb Vasc Biol 1997; 17:2875-9.
13. De Stefano V, Chiusolo P, Paciaroni K, et al. Prothrombin G20210A mutant genotype is a risk factor for cerebrovascular ischemic disease in young patients. Blood 1998; 91:3562-5.
14. Bowen DJ, Bowley S, John M, Collins PW. Factor V Leiden (G1691A), the prothrombin 3'-untranslated region variant (G20210A) and thermolabile methylenetetrahydrofolate reductase (C677T): a single genetic test genotypes all three loci - Determination of frequencies in the South Wales population of the UK. Thromb Haemost 1998; 79:949-54.
15. Margaglione M, Brancaccio V, Giuliani N, et al. Increased risk for venous thrombosis in carriers of the prothrombin G→A^{20210} gene variant. Ann Intern Med 1998; 129:89-93.
16. Kyrle PA, Mannhalter C, Béguin S, et al. Clinical studies and thrombin generation in patients homozygous or heterozygous for the G20210A mutation in the prothrombin gene. Arterioscler Thromb Vasc Biol 1998; 18:1287-91.
17. Zivelin A, Rosenberg N, Faier S, et al. A single genetic origin for the common prothrombotic G20210A polymorphism in the prothrombin gene. Blood 1998; 92.1119-24.
18. Morange PE, Barthet MC, Henry M, et al. A three-generation family presenting five cases of homozygosity for the 20210 G to A prothrombin variant [letter]. Thromb Haemost 1998; 80:859-60.
19. González AJ, Medina JM, Fernández CR, Macias MD, Coto E. A patient homozygous for mutation 20210 A in the prothrombin gene with venous thrombosis and

transient ischemic attacks of thrombotic origin [letter]. Haematologica 1998; 83:1050-1.
20. Simioni P, Tormene D, Manfrin D, et al. Prothrombin antigen levels in symptomatic and asymptomatic carriers of the 20210 A prothrombin variant. Br J Haematol 1998; 103:1045-50.
21. Alatri A, Franchi F, Moia M. Homozygous G 20210 A prothrombin gene mutation without thromboembolic events: a case report [letter]. Thromb Haemost 1998; 80:1028-9.
22. Zawadzki Ch, Gaveriaux V, Trillot N, et al. Homozygous G20210A transition in the prothrombin gene associated with severe venous thrombotic disease: two cases in a French family [letter]. Thromb Haemost 1998; 80:1027-8.
23. Souto JC. Trombofilia: enfermedad poligénica. Rev Iberoamer Tromb Hemostasia 1997; 10(Suppl):115-22.
24. Koeleman BPC, Reitsma PH, Allaart CF, Bertina RM. Activated protein C resistance as an additional risk factor for thrombosis in protein C deficient families. Blood 1994; 84:1031-5.
25. Samama MM, Trosaert M, Horellou MH, Elalamy I, Conard J. Risk of thrombosis in patients homozygous for factor V Leiden. Blood 1995; 86:4700-2.
26. Greengard JS, Eichinger S, Griffin JH, Bauer KA. Brief report: variability of thrombosis among homozygous siblings with resistance to activated protein C due to an Arg-Gln mutation in the gene for factor V. N Engl J Med 1994; 331:1559-62.
27. Ott J. Analysis of human genetic linkage. Baltimore and London: The John Hopkins University Press. 1991.

5.2 Determinantes genéticos de los fenotipos de la Hemostasia en familias españolas

(*Circulation 2000;101:1546-1551*)

Genetic Determinants of Hemostasis Phenotypes in Spanish Families

Juan Carlos Souto, MD; Laura Almasy, PhD; Montserrat Borrell, PhD; Merce Garí, BSc; Elisabet Martínez, BSc; José Mateo, MD; William H. Stone, PhD; John Blangero, PhD; Jordi Fontcuberta, MD, PhD

Background—Recent studies have described genetic mutations that affect the risk of thrombosis as a result of abnormal levels of such hemostatic parameters as protein C, protein S, and the activated protein C resistance ratio. Although these mutations suggest that genes play a part in determining variability in some hemostasis-related phenotypes, the relative importance of genetic influences on these traits has not been evaluated.

Methods and Results—The relative contributions of genetic and environmental influences to a panel of hemostasis-related phenotypes were assessed in a sample of 397 individuals in 21 extended pedigrees. The effects of measured covariates (sex, age, smoking, and exogenous sex hormones), genes, and environmental variables shared by members of a household were quantified for 27 hemostasis-related measures. All of these phenotypes showed significant genetic contributions, with the majority of heritabilities ranging between 22% and 55% of the residual phenotypic variance after correction for covariate effects. Activated protein C resistance ratio, activated partial thromboplastin time, and Factor XII showed the strongest heritabilities, with 71.3%, 83.0%, and 67.3%, respectively, of the residual phenotypic variation attributable to genetic effects.

Conclusions—These results clearly demonstrate the importance of genetic factors in determining variation in hemostasis-related phenotypes that are components of the coagulation and fibrinolysis pathways and that have been implicated in risk for thrombosis. The presence of such strong genetic effects suggests that it will be possible to localize previously unknown genes that influence quantitative variation in these hemostasis-related phenotypes that may contribute to risk for thrombosis. **(*Circulation*. 2000;101:1546-1551.)**

Key Words: genetics ■ coagulation ■ fibrinolysis ■ epidemiology ■ thrombosis

The physiological and biochemical pathways involved in hemostasis are complex. However, recently, important advances have been made in characterizing the major phenotypic components of the coagulation and fibrinolysis pathways. Epidemiological studies have focused on correlations among hemostatic parameters and their relation to risk of diseases such as thrombosis and coronary artery disease.[1] Although there is great interest in assessing genetic components of phenotypic variability in hemostasis and its relation to thrombosis,[2] most current work has focused on evaluating the role of structural candidate genes through population-based association studies.[3,4] Such approaches invariably underestimate the importance and complexity of genetic factors because of their reliance on linkage disequilibrium.[5] Comparatively few studies[6–8] have attempted to quantify the nature and extent of genetic determinants of phenotypic variation in hemostatic parameters through the use of family-based sampling designs. Such knowledge is critical to inform future genome-wide linkage studies to localize novel regulatory loci involved in coagulation and fibrinolysis.

Given the continuous nature of most commonly assayed hemostasis-related phenotypes, it is likely that there will be a number of interacting genetic and environmental factors that jointly determine their variable expression. Powerful new analytical methods have been developed that ultimately will be used to localize and evaluate the relative effects of these quantitative trait loci (QTLs).[9,10] Before such costly analyses, it is necessary to determine which phenotypes can be pursued profitably through linkage studies. Therefore, the primary purpose of this investigation was to examine the roles of genetic and environmental factors in determining hemostasis-related phenotypes. We studied a sample of extended Spanish kindreds, half of which were ascertained through individuals with thrombophilia. This study is the first large-scale family study of the genetics of quantitative variation in these putative risk factors for thrombosis and ischemic heart disease.

Received November 24, 1998; revision received October 26, 1999; accepted October 27, 1999.

From the Unit of Thrombosis and Hemostasis, Department of Hematology, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain (J.C.S., M.B., M.G., E.M., J.M., J.F.); the Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Tex (L.A., J.B.); and the Department of Biology, Trinity University, San Antonio, Tex (W.H.S.).

Correspondence to Dr John Blangero, Department of Genetics, Southwest Foundation for Biomedical Research, PO Box 760549, San Antonio, TX 78245-0549. E-mail john@darwin.sfbr.org

Circulation **is available at http://www.circulationaha.org**

TABLE 1. Distribution of Examined Individuals by Pedigree

Pedigree No.	Total in Pedigree	Examined Male Subjects	Examined Female Subjects	No. of Households
1	23	15	7	9
2	30	14	13	14
3	23	10	12	5
4	16	6	9	5
5	15	3	11	4
6	28	10	14	13
7	18	12	6	9
8	24	7	15	7
9	20	9	11	4
10	29	10	12	7
11	43	20	19	8
12	21	5	7	10
13	11	6	5	7
14	10	6	4	7
15	46	10	23	4
16	28	9	10	9
17	17	5	9	4
18	15	8	5	9
19	17	7	7	6
20	19	7	11	8
21	10	2	6	4
Total	463	181	216	153

Methods

Enrollment of Family Members

Recruitment of family members was based in Barcelona and was performed as part of the GAIT project. The sample included 21 families selected primarily for pedigree size to maximize the power to detect genetic effects. To be included, a family had to have ≥10 living individuals in ≥3 generations. Twelve families were selected through a proband with idiopathic thrombophilia, which was defined as multiple thrombotic events (≥1 spontaneous), a single spontaneous episode of thrombosis with a first-degree relative also affected, or onset of thrombosis before age 45 years. Ten of the 12 probands had onset before age 45 years, 8 had multiple thromboses, and only 2 were ascertained because of a single episode of thrombosis with a relative also affected. The proband's thrombophilia was considered idiopathic because all known (during the recruitment period of 1995 to 1997) biological causes (eg, antithrombin deficiency, protein S and C deficiencies, activated protein C resistance, plasminogen deficiency, HCII deficiency, Factor V Leiden, dysfibrinogenemia, lupus anticoagulant, and antiphospholipid antibodies) of thrombophilia were excluded. These thrombophilic factors were also absent in all affected relatives. The remaining 9 families were selected without regard to phenotype.

Subjects were interviewed by a physician to determine their health/reproductive history, current medications, including use of oral contraceptives, and smoking history. They were questioned about episodes of venous or arterial thrombosis, the age at which these events occurred, and the presence of potentially correlated disorders such as diabetes and lipid disease. The residence of each subject was determined to assess the contribution of shared environmental influences (such as diet) common to members of a household. All procedures were reviewed by the Institutional Review Board of the Hospital de la Santa Creu i Sant Pau. Adult subjects gave informed consent for themselves and for their minor children.

A total of 397 individuals were examined, with a mean of 19 individuals and 7 households per family. Subjects ranged in age from <1 year to 88 years, with a mean of 37.7 and approximately equal numbers of male (46%) and female (54%) subjects. Table 1 lists the number of individuals examined by sex for each pedigree as well as the number of additional unexamined family members (most deceased) required to account for biological links among pedigree members. Of the individuals examined, 101 were founders (individuals whose parents are not in the pedigree) and 296 were nonfounders. The number of households per pedigree ranged from 4 to 14 and the number of examined individuals per household ranged from 1 to 7, with a mean of 2.6. Most pedigrees contained 3 generations, with 8 families having 4 generations and 1 having 5. The depth and complexity of these pedigrees is illustrated by the number of relative pairs contained therein (Table 2).

Blood Collection

Blood was obtained by venipuncture after a 12-hour fast. Samples for hemostatic tests were collected in 1:10 volumes of 0.129 mol/L sodium citrate. Platelet-poor plasma was obtained by centrifugation at 2000g for 20 minutes at room temperature (22±2°C). Assays for APTT, prothrombin time, and coagulation factors were performed on fresh plasma samples. The remaining plasma samples were stored at −80°C until use. Samples for homocysteine determination were collected in EDTA and kept on ice until plasma was harvested by centrifugation. DNA extraction and storage were performed according to standard protocols.[11]

Phenotype Assays

APTT and PT were measured in an automated coagulometer (ACL 3000; IL) with the use of bovine cephalin and silica for APTT (IL) and human thromboplastin for PT (Thromborel S; Behring). Fibrinogen, coagulation factors, funcPS, and APCR were assayed in the STA automated coagulometer (Boehringer Mannheim). Fibrinogen was measured by the von Clauss method[12] with thrombin from BioMerieux (Marcy-l'Etoile). FII, FV, FVII, FVIII, FIX, FX, FXI, and FXII were assayed with deficient plasma from Diagnostica Stago (Asnières). funcPS was determined with a kit from Diagnostica Stago. APCR was measured with the kit Coatest APC Resistance from Chromogenix. AT, protein C, HCII, plasminogen, and PAI-1 were measured in a biochemical analyzer (CPA Coulter, Coulter Corp) with the use of chromogenic methods from Chromogenix for AT, protein C, and plasminogen and from Diagnostica Stago for HCII and PAI-1.

tPS and fPS, TPA, and DD were assayed with the use of ELISA methods from Diagnostica Stago. TF was tested by an ELISA method from American Diagnostica. von Willebrand factor was measured by an ELISA method with antibodies from Dako. Levels of histidine-rich glycoprotein were measured by electroimmunoassay with antibodies from Diagnostica Stago. TFPI was measured by a functional method as described by Sandset et al.[13] Basal homocysteine was separated by HPLC and determined by a fluorometric method.[14]

To reduce measurement error, assays were performed in duplicate, and the average value was calculated for each person. Intra-assay and interassay coefficients of variation were generally estimated to be between 2% and 6%. However, the interassay coefficients of variation were somewhat higher for DD (16.7%), TFPI (9.7%), and TPA (9%).

Statistical Methods

The goal of these analyses was to determine the contributions of genes, measured environmental factors specific to an individual, and environmental factors shared in common by members of a household to variation in hemostasis-related phenotypes. The phenotypic covariance among relatives was used to estimate the additive genetic and shared environmental components of variance.

TABLE 2. Examined Relative Pairs

n	Relation	Degree of Relation
470	Parent-offspring	1
340	Siblings	1
1	Monozygotic twins	0
225	Grandparent-grandchild	2
693	Avuncular	2
13	Half-siblings	2
13	Great grandparent-grandchild	3
137	Grand avuncular	3
547	1st cousins	3
9	Great grand avuncular	4
233	1st cousins, once removed	4
63	2nd cousins	5

The level of a trait, y, for individual i (y_i) was modeled as a linear function as follows:

$$y_i = \mu + \Sigma \beta_j x_{ij} + g_i + h_i + e_i$$

where μ is the trait mean in male subjects, x_{ij} is the j-th covariate, and β_j is its regression coefficient. Covariates included female sex, sex-specific age and age squared, smoking, and for female subjects, current use of oral contraceptives. Age-related covariates were scaled such that the regression coefficients represent the effect associated with a 10-year deviation from the mean age. Discrete covariates (female sex, smoking, and oral contraceptive use) were scaled so that the regression coefficients represent the effect of presence of the covariate versus absence. The remaining variables in the above formula, g_i, h_i, and e_i, represent the random deviations from μ for individual i that are attributable to additive genetic, household, and residual error effects, respectively. The residual error component includes true random error, measurement error, and any nonadditive genetic components. The effects of g_i, h_i, and e_i are assumed to be uncorrelated with one another and normally distributed with mean zero and variances σ_g^2, σ_h^2, and σ_e^2. The likelihood of the phenotypes of the family members is assumed to follow a multivariate normal distribution with a phenotypic covariance matrix that is a function of kinship between individuals and the additive genetic, household, and environmental variances.

This approach can be viewed intuitively as decomposing the observed phenotypic correlations among different classes of relatives in terms of underlying genetic and shared environmental factors. Once the expected means and covariance matrix of each pedigree are defined, the likelihood of a pedigree is evaluated with the multivariate normal density function and cumulated over pedigrees. Although we assume multivariate normality, this assumption is robust, and consistent parameter estimates are obtained when the assumption is violated.[15]

Because 12 pedigrees were ascertained through a thrombophilic proband, we performed an ascertainment correction to obtain unbiased parameter estimates relevant to the general population. This was achieved by conditioning on the probands' phenotype. Two pedigrees were ascertained through a thrombophilic proband and an affected relative. Ascertainment correction with both individuals did not produce different results than correction on the focal proband alone. Although ascertainment was based on thrombophilia, our ascertainment correction was performed by conditioning on the hemostasis-related phenotype being analyzed. This conservative correction can lead to larger standard errors of parameter estimates but protects against type I error.

Maximum likelihood methods were used to simultaneously estimate mean and variance values as well as the effects of covariates, heredity, and household through the use of the computer package SOLAR.[10] The significance of covariate effects was assessed with a Wald test. The relative proportions of the residual variance in a trait explained by genetic and household determinants were calculated as the variance attributable to that component divided by the residual phenotypic variance after adjustment for covariates. The significance of genetic and household effects was assessed by comparing the likelihoods of models in which these parameters were estimated to models in which they were constrained to zero. Twice the difference in ln-likelihood between these models is asymptotically distributed as a 1/2:1/2 mixture of χ_1^2 and χ_0^2.[16]

Results

Regression coefficients for the environmental covariates are shown in Table 3. Sex and age effects were significant for most traits examined. For example, Factor (F)V, FVII, FVIII, FXI, von Willebrand factor, D-dimer (DD), tissue factor (TF), protein C, TF pathway inhibitor (TFPI), fibrinogen, total protein S (tPS), free protein S (fPS), homocysteine, and tissue plasminogen activator (TPA) showed dramatic increases with age, whereas antithrombin (AT), prothrombin time (PT), and activated partial thromboplastin time (APTT) showed substantial decreases. Similarly, several traits showed significant sex differences, with female subjects generally showing lower age-corrected phenotypic values than male subjects. This is true for the protein S traits, FV, FIX, FX, activated protein C resistance (APCR), homocysteine, TPA, and plasminogen activator inhibitor-1 (PAI-1), for which female subjects have substantially lower mean values than male subjects. In contrast, female subjects exhibited significantly higher levels of FVIII than did male subjects. Smoking significantly increased levels of heparin cofactor II (HCII) but decreased FV, FVII, FVIII, protein C, and PT. Oral contraceptive use significantly increased FII, FX, HCII, and plasminogen levels and decreased levels of FVIII, PAI-1, and fPS.

Table 4 presents the estimated components of variance for the hemostasis-related phenotypes. Components of variance are shown for the most parsimonious model (ie, the model that best fits the observed data and exhibits the minimum of complexity) for each phenotype, including only significant sources of variation. The remaining variance not accounted for in Table 4 is attributable to individual-specific random environmental influences and random error. All of the traits studied except DD had significant genetic components, with most ranging between 22% and 55% of the residual phenotypic variability. APTT, APCR, and FXII showed exceptionally large genetic influences, accounting for 83%, 71%, and 67% of residual variance, respectively. In contrast, DD showed no significant heritable component, with an estimated heritability of 10.9% ($P=0.07$).

The proportion of the residual phenotypic variability accounted for by shared household effects tended to be considerably smaller than that accounted for by genetic effects. Household components were significant for only 8 traits: tPS, functional protein S (funcPS), FV, FX, FXI, fibrinogen, PAI-1, and fPS. Household membership accounted for ≈10% to 16% of the residual phenotypic variability in most of these traits, with only fPS having household effects accounting for >20% of its residual phenotypic variance.

Likelihood-based tests of heterogeneity allowing for ascertainment correction revealed no differences between ran-

TABLE 3. Regression Coefficients for Statistically Significant Covariate Effects

	Mean (Males)	Female Sex	Age: Male (10-y Change)	Age: Female (10-y Change)	Age^2: Male (10-y Change)	Age^2: Female (10-y Change)	Smoking	Use of Oral Contraceptives
APCR	3.27	−0.218§		−0.043‡				
APTT	0.96		−0.015‖	−0.017‖	0.002*	0.002*		
PT	0.97		−0.012‖	−0.010‖	0.005‖	0.004§	−0.014†	
AT	111.12	−2.982*	−1.898‖	−1.314‡	−0.361*			
FII	126.97		1.805‡		−1.183‖	−0.711‡		9.226†
FV	125.65	−7.763†	4.857‖	3.190‖	−0.584*	−0.560*	−4.084*	
FVII	118.89		3.894‖	6.360‖	−1.016†	−1.021†	−5.203*	
FVIII	135.62	12.303*	8.496‖	7.486‖	3.305‖	1.590†	−9.729*	−31.881‡
FIX	120.87	−6.48†	5.592‖	5.676‖	−0.116‡			
FX	124.89	−10.276§	1.614†	2.179‡	−2.232‖	−1.011‡		12.249†
FXI	104.00		2.073‡	4.203§	−0.063*	−0.115§		
FXII	115.48							
HCII	104.34		3.058‖	2.414§	−1.109§		3.616*	10.263†
Homocysteine	8.08	−1.058‡	0.579‖	0.420‖				
Histadine-rich glycoprotein	103.33			2.849‡				
Protein C	120.46		5.087‖	4.316‖	−1.840‖	−1.234‖	−5.063†	
Plasminogen	119.87		1.558‡		−0.822‡	−0.508†		12.429‡
TFPI	91.60		4.913‖	2.866†				
Fibrinogen	2.97		0.412‖	0.136‖	0.099‖	0.028‡		
PAI-1	17.40	−3.807‡	0.801†	0.798‡	−0.388‡			−5.295†
fPS	108.01	−10.252‖	3.132‖	2.479‖				−9.169†
funcPS	114.80	−21.225‖				−1.837‖		
tPS	111.73	−21.454‖	2.019‡	2.133§	−1.360‖			
DD	175.81		30.41‖	18.99‖	18.99‖	0.926‖	0.926‖	
TF	119.11		5.025*	14.300‖		0.353†		
von Willebrand factor	89.54		4.966§	6.035‖	0.281‖	0.202§		
TPA	7.40	−2.292‖	1.033‖	0.801‖		0.009†		−1.121*

*$P<0.10$, †$P<0.05$, ‡$P<0.01$, §$P<0.001$, ‖$P<0.0001$.

domly ascertained families and families ascertained through thrombophilic probands. This result suggests that the ascertainment correction used was successful in recovering population-based estimates of both covariate effects and the relative variance components.

Discussion

Our results document the importance of genetic factors influencing hemostasis-related phenotypes in this population. For most of the traits, genes appear to be the largest identifiable determinant of quantitative variation. The use of extended pedigrees and household-sharing information yielded precise information on the determinants of correlations among family members. Shared environment had a substantial effect on a few phenotypes and was most apparent for fPS. These hemostasis-related phenotypes are similar to other cardiovascular risk factors such as lipoprotein phenotypes,[17] in which shared environmental effects also appear to be of minor importance.

We have limited the estimation of genetic components to that attributable to additive effects. If other nonadditive sources of genetic variance exist, such as dominance or epistasis, then our observed heritabilities will represent lower bounds. Therefore, our estimates are conservative.

Heritability can be diminished by measurement error. One way to increase the genetic signal-to-noise ratio is to eliminate measurement error. In general, the measures considered have modest measurement errors, with interassay and intraassay coefficients of variation from 2% to 6%. The measurement error for DD is larger (16.7%) and may have contributed to its low observed heritability. However, measurement error of this magnitude is likely to have only a small effect on heritability. If measurement error were eliminated for DD by multiple measures, the estimated heritability would increase only slightly from 0.109 to 0.129. However, the complete elimination of measurement error is not feasible in large studies.

In this study, we have statistically controlled for the effects of demographic and exogenous covariates such as smoking behavior. We have consciously avoided the use of biological covariates that may be influenced by genes. For example, a composite phenotype such as APCR is influenced by a number of intermediate traits such as protein C, protein S, and FVIII. If we were to correct our APCR phenotype for these

TABLE 4. Components of Variance From the Most Parsimonious Model±SE

Variable	Heritability	Household
APTT	0.830±0.067‖	
APCR	0.713±0.078‖	
FXII	0.673±0.085‖	
FVII	0.523±0.089‖	
Histadine-rich glycoprotein	0.522±0.093‖	
TFPI	0.516±0.086‖	
PT	0.504±0.085‖	
Protein C	0.501±0.086‖	
FII	0.492±0.088‖	
AT	0.486±0.086‖	
tPS	0.460±0.088‖	0.108±0.057†
funcPS	0.453±0.096‖	0.095±0.060†
FXI	0.452±0.104‖	0.162±0.076†
FV	0.442±0.094‖	0.133±0.067†
HCII	0.439±0.086‖	
FX	0.434±0.127‖	0.135±0.076†
FVIII	0.400±0.088‖	
FIX	0.387±0.086‖	
Fibrinogen	0.336±0.101‖	0.137±0.065†
von Willebrand Factor	0.318±0.108‖	
PAI-1	0.298±0.080‖	0.139±0.061‡
TPA	0.268±0.072‖	
Homocysteine	0.244±0.077‖	
Plasminogen	0.236±0.096‡	
fPS	0.223±0.106‡	0.212±0.065§
TF	0.167±0.079‡	
DD	0.109±0.091*	

*$P<0.10$, †$P<0.05$, ‡$P<0.01$, §$P<0.001$, ‖$P<0.0001$B

correlated phenotypes, the relative genetic and environmental components would be altered unpredictably. Such purely phenotypic correction cannot disentangle genetic correlates from environmental correlates.

We expect that the same genes influence multiple phenotypes. Such pleiotropy is widespread in highly coordinated physiological systems such as coagulation/hemostasis. Additionally, the hemostasis-related phenotypes are correlated with phenotypes from other physiological systems such as the lipid pathway. We do not correct for such covariation because of the potential to eliminate genetic signals that may be important for mapping QTLs that influence hemostasis. For example, it is conceivable that a locus could influence both a hemostasis-related phenotype and a lipid trait. Only joint analysis of both traits could unequivocally determine whether a specific QTL is responsible for some of this covariation. Identification of such a QTL would also allow determination of whether the locus acts through direct effects on each phenotype or indirectly by affecting one phenotype, which then influences the second phenotype. Such information cannot be gleaned from simple regression-based corrections of the hemostasis-related phenotype and can be lost by such phenotypically based correction. Systematic analysis of pleiotropy will require multivariate genetic analysis, and any correlation between traits, whether caused by genes or environment, can be exploited to increase the power of a subsequent linkage study.[9]

The utility of genetic studies of quantitative intermediate risk factors is manifold. Intermediate risk factors are more proximal to gene action and thus provide less attenuated genetic signals than when a discrete clinical end point such as disease is analyzed. Also, susceptibility to disease is primarily a quantitative process that reflects an unobservable continuous liability. Evidence for the continuous relation between several of the hemostasis-related risk factors considered in this study and risk of venous thrombosis has been widely reported. For example, APCR shows an inverse continuous relation with risk of thrombosis,[18] whereas fibrinogen,[19] FVIII,[20] FII,[21] and homocysteine levels[22] all exhibit continuous positive relations with risk of thrombosis. Liability to thrombosis is influenced not only by abnormalities in these systems but also by quantitative variation within the normal physiological range. Such candidate risk factors can be utilized jointly with disease status to search the genome for QTLs that pleiotropically affect both risk factor and disease.

A primary goal of modern genetic analysis is to partition the genetic variability in a phenotype into components attributable to specific QTLs. Such goals now can be attained with the use of powerful new methods of quantitative trait linkage analysis on human pedigree data such as that collected for the Genetic Analysis of Idiopathic Thrombophilia (GAIT) study. These new linkage approaches will provide estimates of chromosomal location and, equally important, unbiased estimates of the relative importance of specific QTLs for the general population. Such estimates will be essential for the decomposition of the risk of disease in the general population and therefore are relevant to public health. Ultimately, the joint analysis of both thrombosis and its quantitative risk factors will lead to the identification of the genes determining risk of thrombosis. Such information then may be used for predicting individual-specific risk early enough in life to consider prophylactic intervention.

Candidate gene studies can provide some information regarding the likelihood of finding novel QTLs by linkage analysis. For example, data provided by de Ronde and Bertina[23] on the FV Leiden mutation suggest that 34% of the phenotypic variance in APCR in the Netherlands may be attributable to this gene. However, even considering the lower-bound nature of association-derived locus-specific effects,[5] it is unlikely that such genes account for all or most of the variability in the quantitative risk factors considered. The FV Leiden mutation results have particular relevance for the current study. Since this mutation is much rarer in the Spanish population,[24] it could account for little (<5%) of the variation in APCR in our sample, yet our estimated heritability for the Spanish population is very high (71%). Although some of this genetic variance may be attributable to unknown mutations in the FV gene, it is likely that some of it is attributable to other unknown genes. If some of these novel genes exhibit comparable effects on the hemostasis-related phenotypes as those seen in the candidate gene studies, it should be relatively easy to localize them in linkage-based designs whose power

depends solely on the relative heritability attributable to the QTL. To this end, our results showing substantial total heritabilities for most of the measured hemostasis-related phenotypes provide excellent support for our plan to perform a genomic search to identify and assess the importance of these genes in the Spanish population.

Acknowledgments

This study was supported by grants DGICYT Sab 94/0170 from the Ministerio de Educacion y Ciencia and FIS 97/2032 from the Ministerio de Sanidad y Consumo, Spain, and National Institutes of Health grants MH59490 and GM18897. We are grateful to the doctors who assisted in recruitment of thrombophilic pedigrees: Dr Javier Rodríguez Martorell (Hospital Universitario Puerta del Mar, Cádiz), Dr Carmen Araguás (Hospital Arnau de Vilanova, Lleida), Dr Francisco Velasco (Hospital Reina Sofía, Córdoba), Dr Montserrat Maicas (Hospital General de Albacete), and Dr Dilia Brito (Hospital Carlos Haya, Málaga). We also acknowledge the assistance of Rosa Felices, Cristina Vallvé, Isabel Tirado, Dolors Llobet, Inma Coll, Joaquín Murillo, Teresa Urrutia, Rosa Ma Arcelús, Laia Bayén, Ma Jesús Gallego, Beatriz Carreras, Laura Vicente, and María Teresa Royo. Finally, we are deeply grateful to the families who participated in this study.

References

1. Scarabin PY. Haemostatic variables and arterial thrombotic disease: epidemiological evidence. In: Seghatchian MJ, Samama MM, Hecker SP, eds. *Hypercoagulable States*. Boca Raton, Fla: CRC Press; 1996: 209–216.
2. Lane DA, Mannucci PM, Bauer KA, Bertina RM, Bochkov NP, Boulyjenkov V, Chandy M, Dahlback B, Ginter EK, Miletich JP, Rosendaal FR, Seligsohn U. Inherited thrombophilia: part 1. *Thromb Haemost*. 1996;76:651–662.
3. Humphries SE, Panahloo A, Montgomery HE, Green F, Yudkin J. Gene-environment interaction in the determination of levels of haemostatic variables involved in thrombosis and fibrinolysis. *Thromb Haemost*. 1997;78:457–461.
4. Di Minno G, Grandone E, Margaglione M. Clinical relevance of polymorphic markers of arterial thrombosis. *Thromb Haemost*. 1997;78: 462–466.
5. Ruiz A, Barbadilla A. The contribution of quantitative trait loci and neutral marker loci to the genetic variances and covariances among quantitative traits in randomly mating populations. *Genetics*. 1995;139: 445–455.
6. Hong H, Pederson NL, Egberg N, de Faire U. Genetic effects for plasma factor VII levels independent of and in common with triglycerides. *Thromb Haemost*. 1999;81:382–386.
7. Hennis BC, Van Boheemen PA, Koeleman BP, Boomsma DI, Engesser L, Van Wees AG, Novakova I, Brommer EJ, Kluft C. A specific allele of the histidine-rich glycoprotein (HRG) locus is linked with elevated plasma levels of HRG in a Dutch family with thrombosis. *Br J Haematol*. 1995;89:845–852.
8. Pankow JS, Folsom AR, Province MA, Rao DC, Williams RR, Eckfeldt J, Sellers TA. Segregation analysis of plasminogen activator inhibitor-1 and fibrinogen levels in the NHLBI Family Heart Study. *Arterioscler Thromb Vasc Biol*. 1998;18:1559–1567.
9. Almasy LA, Dyer TD, Blangero J. Bivariate quantitative trait linkage analysis: pleiotropy versus coincident linkage. *Genet Epidemiol*. 1997; 14:953–958.
10. Almasy L, Blangero J. Multipoint quantitative trait linkage analysis in general pedigrees. *Am J Hum Genet*. 1998;62:1198–1211.
11. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*. 1988; 16:1215.
12. Clauss A. Gerinnungsphysiologische Schnellmethode zur Bestimmung des Fibrinogens. *Acta Haematol*. 1957;17:237–246.
13. Sandset PM, Larsen ML, Abilgaard U, Lindahl KA, Odergaard OR. Chromogenic substrate assay of extrinsic pathway inhibitor (EPI): levels of the normal population and relation to cholesterol. *Blood Coagul Fibrinolysis*. 1991;2:425–433.
14. Hyland K, Bottiglieri T. Measurement of total plasma and cerebrospinal fluid homocysteine by fluorescence following high-performance liquid chromatography and precolumn derivatization with o-phthaldialdehyde. *J Chromatogr*. 1992;79:55–62.
15. Beaty TH, Self SG, Liang KY, Connolly MA, Chase GA, Kwiterovich PO. Use of robust variance components models to analyse triglyceride data in families. *Ann Hum Genet*. 1985;49:315–328.
16. Self SA, Liang KY. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc*. 1987;82:605–610.
17. Mitchell BD, Kammerer CM, Blangero J, Mahaney MC, Rainwater DL, Dyke B, Hixson JE, Henkel RD, Sharp M, Comuzzie AG, Vandeberg JL, Stern MP, MacCluer JW. Genetic and environmental contributions to cardiovascular risk factors in Mexican Americans: the San Antonio Family Heart Study. *Circulation*. 1996;94:2159–2170.
18. Koster T, Rosendaal FR, de Ronde H, Briët E, Vandenbroucke JP, Bertina RM. Venous thrombosis due to poor anticoagulant response to activated protein C: Leiden thrombophilia study. *Lancet*. 1993;342:1503–1506.
19. Koster T, Rosendaal FR, Reitsma PH, van der Velden PA, Briet E, Vandenbroucke JP. Factor VII and fibrinogen levels as risk factors for venous thrombosis. *Thromb Haemost*. 1994;71:719–722.
20. Koster T, Blann AD, Briet E, Vandenbroucke JP, Rosendaal FR. Role of clotting factor VIII in effect of von Willebrand factor on occurrence of deep-vein thrombosis. *Lancet*. 1995;345:152–155.
21. Poort SR, Rosendaal FR, Reitsma PH, Bertina RM. A common genetic variation in the 3′-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. *Blood*. 1996;88:3698–3703.
22. Den Heijer M, Koster T, Blom HJ, Bos GMJ, Briet E, Reitsma PH, Vandenbroucke JP, Rosendaal FR. Hyperhomocysteinemia as a risk factor for deep-vein thrombosis. *N Engl J Med*. 1996;334:759–762.
23. De Ronde H, Bertina RM. Laboratory diagnosis of APC-resistance: a critical evaluation of the test and the development of diagnostic criteria. *Thromb Haemost*. 1994;72:880–886.
24. Garcia-Gala JM, Alvarez V, Pinto CR, Soto I, Urgelles MF, Menendez MJ, Carracedo C, Lopez-Larrea C, Coto E. Factor V Leiden (R506Q) and risk of venous thromboembolism: a case-control study based on the Spanish population. *Clin Genet*. 1997;52:206–210.

5.3 Susceptibilidad genética para la trombosis y su relación con factores de riesgo fisiológico: el estudio GAIT

(*Am J Hum Genet 2000;67:1452-1459*)

Am. J. Hum. Genet. 67:1452–1459, 2000

Genetic Susceptibility to Thrombosis and Its Relationship to Physiological Risk Factors: The GAIT Study

Juan Carlos Souto,[1] Laura Almasy,[3] Montserrat Borrell,[1] Francisco Blanco-Vaca,[2] José Mateo,[1] José Manuel Soria,[1] Inma Coll,[1] Rosa Felices,[1] William Stone,[3,4] Jordi Fontcuberta,[1] and John Blangero[3]

[1]Unitat de Trombosi i Hemostasia, Departament d'Hematologia, and [2]Servei de Bioquimica i Institut de Recerca, Hospital de la Santa Creu i Sant Pau, Barcelona; and [3]Department of Genetics, Southwest Foundation for Biomedical Research, and [4]Department of Biology, Trinity University, San Antonio

Although there are a number of well-characterized genetic defects that lead to increased risk of thrombosis, little information is available on the relative importance of genetic factors in thrombosis risk in the general population. We performed a family-based study of the genetics of thrombosis in the Spanish population to assess the heritability of thrombosis and to identify the joint actions of genes on thrombosis risk and related quantitative hemostasis phenotypes. We examined 398 individuals in 21 extended pedigrees. Twelve pedigrees were ascertained through a proband with idiopathic thrombosis, and the remaining pedigrees were randomly ascertained. The heritability of thrombosis liability and the genetic correlations between thrombosis and each of the quantitative risk factors were estimated by means of a novel variance component method that used a multivariate threshold model. More than 60% of the variation in susceptibility to common thrombosis is attributable to genetic factors. Several quantitative risk factors exhibited significant genetic correlations with thrombosis, indicating that some of the genes that influence quantitative variation in these physiological correlates also influence the risk of thrombosis. Traits that exhibited significant genetic correlations with thrombosis included levels of several coagulation factors (factors VII, VIII, IX, XI, XII, and von Willebrand), tissue plasminogen activator, homocysteine, and the activated protein C ratio. This is the first study that quantifies the genetic component of susceptibility to common thrombosis. The high heritability of thrombosis risk and the significant genetic correlations between thrombosis and related risk factors suggest that the exploitation of correlated quantitative phenotypes will aid the search for susceptibility genes.

Introduction

Thrombosis is a common cause of morbidity and mortality in industrialized nations. Both venous and arterial forms of thrombosis are of great public-health importance. Although there is little direct information on prevalence, retrospective and prospective data (Coon et al. 1973; Anderson et al. 1991; Nordstrom et al. 1992) suggest a minimum lifetime prevalence of 5%–10% for deep-vein thrombosis. After the inclusion of arterial thromboses, other venous thromboses, and undiagnosed thrombotic conditions, the true lifetime prevalence of thrombosis must be substantially >10%.

The canonical causes of thrombosis include both environmental and genetic factors (Rosendaal 1999). The high prevalence of thrombosis and its known environmental influences, such as smoking and oral contraceptive use, suggest that multiple genes of varying effects will be involved in determining susceptibility to thrombosis. Such complex oligogenic inheritance is also likely to involve gene-gene and gene-environment interactions (Hasstedt et al. 1998). Although there are a number of well-characterized genetic defects that lead to increased thrombotic risk (Lane et al. 1996), it is unlikely that these comparatively infrequent mutations constitute the primary genetic influences on risk of common late-onset thrombosis. In fact, very little information is available on the relative importance of genetic factors in thrombosis risk in the general population. Because of the paucity of family-based studies, there are no extant estimates of the heritability of thrombosis risk.

The physiological cascade that underlies the normal formation of thrombin and the pathological endpoint of thrombosis is complex, with many components involved in the coagulation and fibrinolytic pathways. The identification of quantitative risk factors for thrombosis has accelerated in recent years. Numerous hemostatic factors—including fibrinogen, factor VII, factor VIII, von Willebrand factor, and homocysteine—have been implicated as possible concomitants of both

Received August 14, 2000; accepted for publication October 9, 2000; electronically published October 19, 2000.

Address for correspondence and reprints: Dr. John Blangero, Department of Genetics, Southwest Foundation for Biomedical Research, P.O. Box 760549, San Antonio, TX 78245-0549 (express delivery: 7620 NW Loop 410, San Antonio, TX). E-mail: john@darwin.sfbr.org

0002-9297/2000/6706-0011$02.00

venous (Koster et al. 1994, 1995; MacCallum et al. 1995; den Heijer et al. 1996) and arterial thrombosis (Meade et al. 1986; Hamsten et al. 1987; Ernst and Resch 1993; Ridker et al. 1993; Folsom et al. 1997; Nygard et al. 1997). Regardless of their causal relationships with thrombosis, such correlated phenotypes can provide additional information about the genetic basis of thrombosis risk. Recent advances in statistical genetics allow the simultaneous examination of the genetic and environmental sources of correlations between such continuous physiological measures and discrete disease outcomes (Williams et al. 1999*b*) through the examination of data from large families. Such approaches, when coupled with modern molecular genetic technologies, will soon permit the localization and identification of the quantitative trait loci (QTLs) that underlie thrombosis risk. Prior to embarking on the potentially expensive search for the actual loci involved, it is prudent to evaluate the magnitude of genetic effects on thrombosis and to test for the pleiotropic effects of genes on both risk factors and disease.

As a first step toward the ultimate goal of the identification of novel genes involved in thrombosis susceptibility, we performed a family-based study of the genetics of thrombosis in the Spanish population. This study design has allowed us to quantify the heritability of thrombosis and to identify the joint actions of genes on thrombosis risk and a number of related quantitative phenotypes. Previous analyses of these 27 quantitative phenotypes have already demonstrated strong heritabilities for most of these traits (Souto et al. 2000). The majority of the heritabilities ranged between 0.22 and 0.55, with somewhat higher values seen for factor XII (0.67), activated protein C resistance ratio (0.71), and activated partial thromboplastin time (0.83), and somewhat lower values observed for D-dimer (0.11) and tissue factor (0.17).

Subjects and Methods

Study Population and Diagnosis

The Genetic Analysis of Idiopathic Thrombophilia (GAIT) Study is composed of 21 extended families, 12 of which were ascertained through a proband with thrombophilia and 9 of which were obtained randomly. Thrombophilia was defined as multiple thrombotic events (at least one of which was spontaneous), a single spontaneous episode of thrombosis with a first-degree relative also affected, or onset of thrombosis at age <45 years. Ten of the 12 thrombophilic probands had onset at age <45 years, 8 had multiple episodes of thrombosis, and 2 probands were ascertained on the basis of family history. Diagnoses of the thrombophilic probands were verified by objective methods. Thrombosis in these individuals was considered idiopathic because of exclusion of all biological causes of thrombosis, including antithrombin deficiency, Protein S and C deficiencies, activated protein C resistance, plasminogen deficiency, heparin cofactor II deficiency, Factor V Leiden, dysfibrogenemia, lupus anticoagulant, and antiphospholipid antibodies, known at the time of recruitment (1995–97).

A total of 398 individuals (with a mean of 19 individuals per family) were examined. Most pedigrees contained three generations, although eight families had four generations and one family had five. Subjects had a mean age at examination of 37.7 years, and there were approximately equal numbers of males and females. The composition of the families and the collection of lifestyle, medical, and family-history data have been described elsewhere (Souto et al. 2000). Reported history of thrombosis in family members was verified by examination of medical records, when available. Although some deceased family members had a history of thrombosis, only individuals interviewed and examined in person were included in the analyses. The primary residence of each subject was also determined, to assess the contribution of shared environmental influences (such as diet) common to members of a household. The study was performed according to the Declaration of Helsinki of 1975, and all adult patients provided informed consent for themselves and for their minor children.

Laboratory Measurements and Techniques

A total of 27 quantitative phenotypes were measured in the plasma of each individual. None of the participants was being treated with anticoagulant therapy at the time of blood drawing. Activated partial thromboplastin time (APTT), prothrombin time (PT), coagulation factors (FII, FV, FVII, FVIII, FIX, FX, FXI, and FXII), functional protein S, and the activated protein C-sensitivity ratio (APCR) were measured by automated coagulometry. Antithrombin, protein C, heparin cofactor II, plasminogen, and plasminogen activator inhibitor were measured by chromogenic methods. Fibrinogen was measured by the Clauss method (Clauss 1957). Total and free protein S, tissue plasminogen activator (t-PA), D-dimer (DD), tissue factor (TF), and von Willebrand factor (vWF) were assayed by use of commercially available ELISA kits. Histidine-rich glycoprotein (HRG) was measured by electroimmunoassay, tissue factor pathway inhibitor (TFPI) by a functional method (Sandset et al. 1991), and homocysteine by a fluorimetric method (Hyland and Bottiglieri 1992). ABO blood groups and Factor V Leiden genotypes were assessed by means of standard techniques. Details of phenotype assays are available in Souto et al. (2000).

Statistical Genetic Analysis

The heritability (the proportion of the total phenotypic variability attributable to genetic effects) of susceptibility to thrombosis was evaluated by means of a pedigree-based maximum-likelihood method that models affection status as a threshold process (Duggirala et al. 1997, 1999*a*; Williams et al. 1999*b*). Although disease status is usually operationalized as a discrete trait, with individuals scored as unaffected or affected, it is generally assumed that there is an unobservable continuous trait, termed "liability" or "susceptibility," that determines affection status. If an individual's liability score exceeds some specified threshold, disease results; if it is below the threshold, the individual is unaffected. The threshold is placed in an age- and sex-specific manner, to produce the appropriate population prevalence. A specific individual's liability is only known to be above or below the threshold, depending on the individual's affection status, and an integral over the appropriate region of the curve is used to estimate each person's liability value. Since such continuous processes determine most biological phenomena, it is useful to make inferences on the underlying continuous scale, which is more consistent with current models of gene action. Threshold models permit such inferences regarding the latent underlying quantitative scale to be made. To use a threshold model, some weak assumptions regarding the form of the underlying continuous process are necessary. For genetic modeling, we assume that the underlying liability distribution is normal, and we calculate the joint probability of observing the disease statuses of family members by using a multivariate normal distribution that allows for correlations among family members.

The analysis of heritability of thrombosis susceptibility was performed using the variance component method. The total phenotypic variance in thrombosis susceptibility was partitioned into three components: (1) an additive genetic variance, caused by the sum of the average effects of all the genes that influence thrombosis; (2) a shared environmental variance, caused by the effects of environmental factors that are common to households; and (3) a random environmental variance specific to each individual. The random environmental variance also absorbs nonadditive genetic effects, such as interactions between alleles within loci (dominance effects), interactions between alleles at different loci (epistatic effects), and effects caused by gene-environment interactions. Therefore, such models will generally underestimate the role of genetics in the determination of the trait.

With this approach, the relative components of variance can be estimated by use of maximum-likelihood estimation. Evaluation of the likelihood function for a pedigree involves high dimensional integration of a multivariate normal distribution. The limits of integration may be different for each individual, depending on affection status as well as on any covariates that are introduced as fixed effects in the model for the mean liability. In the current analyses, these covariates included age and sex.

To study the genetic relationships between thrombosis susceptibility and quantitative variation in hemostatic parameters, we used a new mixed discrete/continuous trait variance component analysis (Williams et al. 1999*b*). This analysis used a modified variance component method to accommodate a mixture of discrete and continuous data and allows the phenotypic correlations between these traits to be decomposed into factors caused by common genetic influences and common environmental influences on the two traits. Examination of the underlying determinants of phenotypic correlations provides information on the role of pleiotropic genetic effects.

All the extant epidemiological evidence for the relationship between thrombosis and hemostatic parameters is based on the evaluation of phenotypic correlations. However, the decomposition of phenotypic correlations into genetic and environmental components is potentially valuable, since hidden relationships between traits can be revealed (Comuzzie et al. 1996). For example, if trait $y_1 = g_1 + e_1$ and trait $y_2 = g_2 + e_2$, where g and e denote genetic and environmental effects, the observed correlations between the phenotypic traits are determined by the latent genetic and environmental correlations between the component variables. By studying both traits in extended families, we can estimate both the genetic (ρ_g) and the environmental (ρ_e) correlations between traits. The phenotypic correlation (ρ_p) is derived from these two constituent correlations and the heritabilities of the traits:

$$\rho_p = \sqrt{(h_1^2 h_2^2)}\rho_g + \sqrt{(1 - h_1^2)}\sqrt{(1 - h_2^2)}\rho_e .$$

We have incorporated the threshold model (Duggirala et al. 1997, 1999*a*) and the mixed discrete/continuous trait variance component method (Williams et al. 1999*b*) into our statistical genetic computer package, *SOLAR* (Almasy and Blangero 1998). All statistical genetic analyses were performed using *SOLAR,* with these modifications. Estimates of variance component parameters, including the heritabilities of thrombosis and the quantitative measures and all the phenotypic, genetic, and environmental correlations between thrombosis and the quantitative phenotypes, were obtained by use of maximum-likelihood estimation. All hypothesis tests were performed using likelihood-ratio test statistics (Kendall and Stuart 1972; Self and Liang 1987).

Because 12 of the 21 pedigrees were ascertained

through a thrombophilic proband, all analyses included an ascertainment correction, to allow unbiased estimation of parameters relevant to the general population. To achieve this, the likelihood for each family ascertained through a thrombophilic proband was conditioned on the phenotype of the proband (Hopper and Mathews 1982; Boehnke and Lange 1984). Since two families were ascertained, in part, because of the family history of the proband, analyses were repeated conditioning on both the original proband and the affected first-degree relative in these two families. However, the results of the analyses were unchanged.

Results

Characteristics of Affected Individuals

A total of 53 people with venous or arterial thrombosis were identified, 47 in the families ascertained through thrombophilic probands and 6 in the randomly ascertained families. The number of affected individuals per family ascertained through a thrombophilic proband was 2–8, with a mean of 3.9. The distribution of thrombotic subjects in these extended families included many instances of affected first-degree relatives (siblings or parents and children) but also grandparents, aunts or uncles, and first cousins. Eight of these families contained cases of both arterial and venous thrombosis. Two of the randomly ascertained families each had two individuals with thrombophlebitis. One of these was a parent-child pair, but the other consisted of two unrelated individuals (in-laws). One randomly ascertained family had a single individual with deep-vein thrombosis, and one had an individual with transient ischemic attacks.

There were slightly more affected females ($n = 31$, 58.5%) than males ($n = 22$, 41.5%), and the age at diagnosis of first thrombosis was 12–76 years, with a mean of 44.5 (table 1). When venous and arterial thrombosis were considered separately, 40 individuals, with an average age at first diagnosis of 39.7 years, had one or more diagnoses of venous thrombosis; 17 individuals, with an average age at first diagnosis of 61.0 years, had one or more arterial thromboses. The early observed age at diagnosis for venous thrombosis is partially a function of the ascertainment criteria. Deep-vein thrombosis was the most common condition ($n = 28$) and superficial thrombophlebitis (SFT) the second most common ($n = 14$). Fifteen (28%) of the 53 affected people had multiple thrombotic diagnoses, and five (9.4%) of these people had both venous and arterial events. Twelve individuals had deep-vein thrombosis and one to three other venous or arterial thromboses; one person had ischemic stroke and transient ischemic attacks; one person had SFT and pulmonary embolism; and one had SFT and other venous thrombosis.

Table 1

Number and Percent of Individuals in Each Diagnostic Category of Thrombosis and Age at Diagnosis

Diagnosis	No. (and %) of Individuals with Thrombosis	Mean Age at Diagnosis (years)
Venous thrombosis:		
Deep-vein thrombosis	28 (52.8)	40.3
Pulmonary embolism	9 (17.0)	45.6
SFT	14 (26.4)	41.2
Other venous thrombosis	3 (5.7)	58.0
Any venous thrombosis	40 (75.5)	39.7
Arterial thrombosis:		
Myocardial infarction	4 (7.5)	66.5
Angina pectoris	4 (7.5)	57.3
Ischemic stroke	6 (11.3)	61.0
Transient ischemic attack	5 (9.4)	55.4
Any arterial thrombosis	17 (32.1)	61.0
Any thrombosis	53 (100.0)	44.5

NOTE.—Some individuals are represented in multiple diagnostic categories.

Genetic Determinants of Liability to Thrombosis

The evidence for a strong genetic influence on risk of thrombosis was striking. Liability to thrombosis exhibited an additive genetic heritability of 0.61 ± 0.16 ($P = 9 \times 10^{-5}$), indicating that, after correction for the effects of age and sex, 61% of the variation in liability to thrombosis at the population level can be attributed to genetic factors. No shared environmental effects were found among members of a household for liability to thrombosis. Therefore, the above heritability estimates are unlikely to be inflated by nongenetic correlations among family members, and environmental factors shared by members of a household, such as diet, do not have major effects on thrombosis susceptibility. When the diagnoses considered are restricted to venous thrombosis, excluding arterial thrombotic events, the additive genetic heritability is not significantly different from that obtained with any thrombosis. Similarly, when venous and arterial thrombosis are analyzed jointly as two distinct traits, the phenotypic correlation between these two manifestations of thrombosis is .333 ($P = .0126$), and the genetic correlation is .55 ($P = .09$). Additionally, the genetic correlation is not significantly different from 1. Both the robustness of the heritability when combining across venous and arterial diagnoses and the fact that the genetic correlation is not significantly different from one strongly suggest that arterial and venous thromboses are highly genetically correlated and that our broad phenotypic characterization will be useful to increase the power to detect genetic effects.

Correlations between Thrombosis Liability and Quantitative Risk Factors

Table 2 shows the results of bivariate genetic analyses of thrombosis, with each of the quantitative physiological traits considered. Only the nine quantitative traits showing at least one significant correlation ($P < .05$) are presented. Of these, seven exhibit significant phenotypic correlations with thrombosis susceptibility, eight demonstrate significant genetic correlations with thrombosis, and only two exhibit significant environmental correlations. The largest phenotypic correlations ($|\rho_p| > 0.2$) are seen between FVIII, vWF, APCR, FXI, homocysteine, and thrombosis.

The genetic correlations provide strong evidence for significant pleiotropy underlying the covariation between several of the quantitative traits and thrombosis risk. Those quantitative measures exhibiting the largest genetic correlations ($|\rho_g| > 0.6$) with thrombosis include vWF, t-PA, FVIII, homocysteine, and APCR. The only traits to exhibit significant environmental correlations with thrombosis were APCR and FVII. Table 2 provides a good demonstration of how low-phenotypic correlations may misrepresent the true underlying relationships. Both FIX and FVII failed to show significant phenotypic correlations with thrombosis. However, both provide strong evidence for correlations between genetic effects (FIX) and environmental effects (FVII) with thrombosis. Similarly, the genetic and environmental correlations between APCR and thrombosis are of similar magnitudes but exhibit different directions. When such differences in sign appear, the phenotypic correlation is attenuated, although the underlying components suggest much stronger correlations. Relationships between APCR and thrombosis were unchanged when the presence of the Factor V Leiden mutation (there were nine heterozygotes in the sample) was statistically controlled. Similarly, the correlations between FVIII and thrombosis were unchanged when ABO blood type was incorporated into the model.

Table 2

Phenotypic, Genetic, and Environmental Correlations of Quantitative Risk Factors with Thrombosis

Phenotype[a]	ρ_p	P[b]	ρ_g	P	ρ_e	P[b]
APCR	−.230	.0003	−.650	1×10^{-6}	.669	.0006
FVII	.025	NS	−.354	.0564	.568	.0091
FVIII	.288	.0002	.689	.0005	−.126	NS
FIX	.151	.0787	.597	.0131	−.198	NS
FXI	.209	.0180	.564	.0245	.070	NS
FXII	.172	.0339	.351	.0500	−.145	NS
Homocysteine	.227	.0018	.652	.0015	−.028	NS
t-PA	.180	.0002	.752	.0070	−.099	NS
vWF	.261	.0010	.729	.0005	−.181	NS

[a] Only phenotypes with one or more correlations having $P < .05$ are shown.

[b] NS = nonsignificant ($P > .10$).

Discussion

This is the first study that formally documents the large genetic component for risk of thrombosis. By gathering and analyzing data on extended pedigrees that have been methodically ascertained to allow general population inferences, we have begun to fill a critical gap in the study designs used in thrombosis genetics. Researchers in hemostasis/thrombosis generally have not actively pursued family studies, except for the occasional serendipitous collection of unusual families with high densities of affected individuals. Therefore, most of our knowledge regarding the genetic factors involved in common thrombosis has been limited to association studies that use case-control designs to look at known polymorphic variations in candidate genes (Poort et al. 1996; Rosendaal 1997; Rosendaal et al. 1997; Iacoviello et al. 1998). Although such studies provide important indirect evidence for the presence of genetic effects, they have a number of weaknesses. These include their limitation to known candidate genes, their propensity for type I errors caused by hidden population stratification, the lack of direct evaluation of familial transmission, and their general inability to reliably estimate the relative importance of genetic factors in determining within-population variation in thrombosis risk. Family-based studies eliminate these problems, although their costs tend to be greater.

The high additive genetic heritability that we estimated suggests that whole-genome approaches to localizing and characterizing QTLs that underlie thrombosis susceptibility will be feasible. The magnitude of the additive genetic heritability is greater than or equal to that seen in other common complex diseases such as type II diabetes (Duggirala et al. 1999*a*), gallbladder disease (Duggirala et al. 1999*b*), alcoholism (Williams et al. 1999*a*), and obesity (Comuzzie et al. 1997), whose contributing QTLs are currently being pursued through genome scans.

This is also the first study that attempts to decompose the phenotypic correlations between quantitative physiological risk factors and thrombosis into genetic and environmental components. Evidence for strong genetic correlations between FVIII, vWF, APCR, FIX, FXI, homocysteine, t-PA, and thrombosis indicate that there are sets of genes that jointly influence both disease risk and quantitative physiological variation. The detection of genetic effects that act jointly on both quantitative risk factors and disease liability is critically important for subsequent genetic analyses. When evidence of pleiotropy is detected, the correlational structure between the quantitative phenotypes and risk of thrombosis can be exploited to improve the power of joint linkage anal-

yses to detect QTLs contributing to thrombotic risk (Almasy et al. 1997).

Most of our observed phenotypic correlations are consistent with known epidemiological results. For example, there is previous evidence for a positive relationship between both vWF and FVIII levels and risk of venous (Koster et al. 1995) and arterial thrombosis (Folsom et al. 1997). High plasma homocysteine levels have been associated with deep-vein thrombosis (den Heijer et al. 1996) and with arterial thrombosis (Nygard et al. 1997). The quantitative measure of APCR is correlated with risk of venous thrombosis, even when the Factor V Leiden polymorphism is taken into account (De Visser et al. 1999). Similarly, levels of FXII (Kohler et al. 1998) and t-PA (Ridker et al. 1993; Carter et al. 1998) have been correlated with arterial thrombosis. Evidence regarding the association of FVII levels with thrombosis has been equivocal (Doggen et al. 1998; Iacoviello et al. 1998). Very recently, results from the LETS study have implicated high plasma levels of factor IX (Vlieg et al. 2000) and factor XI (Meijers et al. 2000) as risk factors for venous thrombosis.

The unique aspects of our correlational analyses lie in the ability to disentangle genetic and environmental sources of correlation. This technique allows us, for the first time, to conclude that most of the phenotypic correlations between thrombosis susceptibility and the quantitative physiological measures are due to pleiotropic effects of genes. There is little evidence that environmental effects induce much of the observed phenotypic correlations. In the two cases where we did observe significant environmental correlations, they were opposite in sign to the genetic correlations. Other investigators have reported similar results from bivariate genetic analyses of a wide variety of traits (e.g., Comuzzie et al. 1996; Brooks 2000; Mahaney et al. 2000; Stern et al. 2000). One interpretation of the difference in sign is that the genetic and environmental sources of variation on these traits act through different physiological mechanisms.

In this study, we have chosen a broad definition of thrombosis that includes both venous and arterial forms. Our justification for this is both empirical and theoretical. Pooling two genetically heterogeneous traits would decrease the genetic signal-to-noise ratio of the composite trait. However, our heritability analyses provided no evidence for such a depression in genetic signal, indicating that there must exist substantial overlap in the genetic determinants of venous and arterial forms of thrombosis. Similarly, the bivariate analysis of venous and arterial thrombosis yielded a genetic correlation not significantly different from 1 and suggests that many of the same genes are involved in the pathogenesis of venous and arterial events. Additionally, there is epidemiological evidence that similar pathways are involved in venous and arterial thrombosis, as evidenced by the correlation between critical risk factors (such as homocysteine, vWF, and FVIII) and both venous and arterial thrombosis. Although unique local environmental factors can separately influence thrombogenesis in veins and arteries, the evidence suggests that much of the underlying process is driven by a common set of genes. Pooling of these two categories of thrombosis clearly improved the power of the present study. However, even if we disaggregate these components and analyze only venous thrombosis, our results are effectively unchanged (data not shown) except for predictable alterations in observed significance values resulting from the decreased overall prevalence of disease.

Finally, these results provide strong support for using genome scans to localize and evaluate the specific QTLs involved in thrombosis susceptibility. We hope to use the information on the genetic correlations between thrombosis and quantitative phenotypes obtained in this study to maximize our potential for mapping the responsible QTLs in a genome scan currently under way.

Acknowledgments

This research was supported by grants DGICYT Sab 94/0170 from the Ministerio de Educacion y Ciencia, Spain; FIS 97/2032 from the Ministerio de Sanidad y Consumo, Spain; RED97/3 from the Generalitat de Catalunya, Spain; and National Institutes of Health grant MH59490.

We are grateful to a number of doctors who assisted in the ascertainment and recruitment of thrombophilic pedigrees: Dr. Javier Rodriguez Martorell, from Hospital Universitario Puerta del Mar, in Cádiz; Dr. Carmen Araguás, from Hospital Arnau de Vilanova, in Lleida; Dr. Francisco Velasco, from Hospital Reina Sofia, in Córdoba; Dr. Montserrat Maicas, from the Hospital General de Albacete; and Dr. Dilia Brito, from Hospital Carlos Haya, in Málaga. We would also like to acknowledge the technical assistance of Teresa Urrutia, Joaquín Murillo, Mercè Garí, Elisabet Martínez, Isabel Tirado, Dolors Llobet, Cristina Vallvè, Laia Bayén, and Rosa M. Arcelús, in performing laboratory assays; and the work of M. Jesús Gallego, Beatriz Carreras, Laura Vicente, and Maria Teresa Royo, in the day-to-day operations of interview scheduling and data management. We are deeply grateful to the families who participated in this study.

References

Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 62: 1198–1211

Almasy L, Dyer TD, Blangero J (1997) Bivariate quantitative trait linkage analysis: pleiotropy versus coincident linkage. Genet Epidemiol 14:953–958

Anderson FA, Wheeler HB, Goldberg RJ, Hosmer DW, Patwardhan NA, Jovanovic B, Forcier A, Dalen JE (1991) A population-based perspective of the hospital incidence and

case-fatality rates of deep vein thrombosis and pulmonary embolism: the Worcester DVT study. Arch Intern Med 151: 933–938

Boehnke M, Lange K (1984) Ascertainment and goodness of fit of variance component models for pedigree data. Prog Clin Biol Res 147:173–192

Brooks R (2000) Negative genetic correlation between male sexual attractiveness and survival. Nature 406:67–70

Carter AM, Catto AJ, Grant PJ (1998) Determinants of tPA antigen and associations with coronary artery disease and acute cerebrovascular disease. Thromb Haemost 80:632–636

Clauss A (1957) Gerinnungsphysiologische schenellmethode zur bestimmung des fibrinogens. Acta Haematol 17:237–246

Comuzzie AG, Blangero J, Mahaney MC, Haffner SM, Mitchell BD, Stern MP, McCluer JW (1996) Genetic and environmental correlations among hormone levels and measures of body fat accumulation and topography. J Clin Endocrinol Metab 81:597–600

Comuzzie AG, Hixson JE, Almasy L, Mitchell BD, Mahaney MC, Dyer TD, Stern MP, MacCluer JW, Blangero J (1997) A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2. Nat Genet 15:273–275

Coon WW, Park WW, Keller JB (1973) Venous thromboembolism and other venous disease in the Tecumseh Community Health Study. Circulation 48:839–846

den Heijer M, Koster T, Blom HJ, Bos GM, Briet E, Reitsma PH, Vandenbroucke JP, Rosendaal FR (1996) Hyperhomocysteinemia as a risk factor for deep-vein thrombosis. N Engl J Med 334:759–762

De Visser MCH, Rosendaal FR, Bertina RM (1999) A reduced sensitivity for activated protein C in the absence of factor V Leiden increases the risk of venous thrombosis. Blood 93: 1271–1276

Doggen CJM, Cats VM, Bertina RM, Reitsma PH, Vandenbroucke JP, Rosendaal FR (1998) A genetic propensity to high factor VII is not associated with risk of myocardial infarction in men. Thromb Haemost 80:281–285

Duggirala R, Blangero J, Almasy L, Dyer TD, Williams KL, Leach RJ, O'Connell P, Stern MP (1999*a*) Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. Am J Hum Genet 64:1127–1140

Duggirala R, Mitchell BD, Blangero J, Stern MP (1999*b*) Genetic determinants of variation in gallbladder disease in the Mexican-American population. Genet Epidemiol 16:191–204

Duggirala R, Williams JT, Williams-Blangero S, Blangero J (1997) A variance component approach to dichotomous trait linkage analysis using a threshold model. Genet Epidemiol 14:987–992

Ernst E, Resch KL (1993) Fibrinogen as a cardiovascular risk factor: a meta-analysis and review of the literature. Ann Intern Med 118:956–963

Folsom AR, Wu KK, Rosamond WD, Sharrett AR, Chambless LE (1997) Prospective study of hemostatic factors and incidence of coronary heart disease: the atherosclerosis risk in communities (ARIC) study. Circulation 96:1102–1108

Hamsten A, DeFaire U, Walldius G, Dahlen G, Szamosi A, Landou C, Blomback M, Wiman B (1987) Plasminogen activator inhibitor in plasma: risk factor for recurrent myocardial infarction. Lancet 2:3–9

Hasstedt SJ, Bovill EG, Callas PW, Long GL (1998) An unknown genetic defect increases venous thrombosis risk, through interaction with protein C deficiency. Am J Hum Genet 63:569–576

Hopper JL, Mathews JD (1982) Extensions to multivariate normal models for pedigree analysis. Ann Hum Genet 46: 373–383

Hyland K, Bottiglieri T (1992) Measurement of total plasma and cerebrospinal fluid homocysteine by fluorescence following high-performance liquid chromatography and precolumn derivatization with o-phthaldialdehyde. J Chromatogr 579:55–62

Iacoviello L, Di Castelnuovo A, de Knijff P, D'Orazio A, Amore C, Arboretti R, Kluft C, Benedetta Donati M (1998) Polymorphisms in the coagulation factor VII gene and the risk of myocardial infarction. N Engl J Med 338:79–85

Kendall MG, Stuart A (1972) Advanced theory of statistics. Hafner, New York

Kohler HP, Carter AM, Strickland MH, Grant PJ (1998) Levels of activated FXII in survivors of myocardial infarction: association with circulating risk factors and extent of coronary artery disease. Thromb Haemost 79:14–18

Koster T, Blann AD, Briet E, Vandenbroucke JP, Rosendaal FR (1995) Role of clotting factor VIII in effect of von Willebrand factor on occurrence of deep-vein thrombosis. Lancet 345:152–155

Koster T, Rosendaal FR, Reitsma PH, van der Velden PA, Briet E, Vandenbroucke JP (1994) Factor VII and fibrinogen levels as risk factors for venous thrombosis. Thromb Haemost 71: 719–722

Lane DA, Mannucci PM, Bauer KA, Bertina RM, Boehkov NP, Boulyjenkov V, Chandy M, Dahlback B, Ginter EK, Miletich JP, Rosendaal FR, Seligsohn U (1996) Inherited thrombophilia: part 1. Thromb Haemost 76:651–662

MacCallum PK, Meade TW, Cooper JA, Stirling Y, Howarth DJ, Ruddock V, Miller GJ (1995) Clotting factor VIII and risk of deep-vein thrombosis. Lancet 345:804

Mahaney MC, Czerwinski SA, Adachi T, Wilcken DEL, Wang XL (2000) Plasma levels of extracellular superoxide dismutase in an Australian population: genetic contribution to normal variation and correlations with plasma nitric oxide and apolipoprotein A–I levels. Arterioscler Thromb Vasc Biol 20:683–688

Meade TW, Mellows S, Brozovic M, Miller GJ, Chakrabarti RR, North WR, Haines AP, Stirling Y, Imeson JD, Thompson SG (1986) Hemostatic function and ischemic heart disease: principal results of the Northwick Park study. Lancet 2:533–537

Meijers JCM, Tekelenburg WLH, Bouma BN, Bertina RM, Rosendaal FR (2000) High levels of coagulation factor XI as a risk factor for venous thrombosis. N Engl J Med 342: 696–701

Nordstrom M, Lindblad B, Bergqvist D, Kjellstrom T (1992) A prospective study of the incidence of deep-vein thrombosis

within a defined urban population. J Intern Med 232:155–160

Nygard O, Nordrehaug JE, Refsum H, Ueland PM, Farstad M, Vollset SE (1997) Plasma homocysteine levels and mortality in patients with coronary artery disease. N Engl J Med 337:230–236

Poort SR, Rosendaal FR, Reitsma PH, Bertina RM (1996) A common genetic variation in the 3′-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. Blood 88:3698–3703

Ridker PM, Vaughan DE, Stampfer MJ, Manson JE, Hennekens CH (1993) Endogenous tissue-type plasminogen activator and risk of myocardial infarction. Lancet 341:1165–1168

Rosendaal FR (1997) Risk factors for venous thrombosis: prevalence, risk, and interaction. Semin Hematol 34:171–187

——— (1999) Venous thrombosis: a multicausal disease. Lancet 353:1167–1173

Rosendaal FR, Siscovick DS, Schwartz SM, Psaty BM, Raghunathan TE, Vos HL (1997) A common prothrombin variant (20210 G to A) increases the risk of myocardial infarction in young women. Blood 90:1747–1750

Sandset PM, Larsen ML, Abilgaard U, Lindahl KA, Odergaard OR (1991) Chromogenic substrate assay of extrinsic pathway inhibitor (EPI): levels of the normal population and relation to cholesterol. Blood Coagul Fibrinolysis 2:425–433

Self SG, Liang K-Y (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 82:605–610

Souto JC, Almasy L, Borrell M, Gari M, Martinez E, Mateo J, Stone WH, Blangero J, Fontcuberta J (2000) Genetic determinants of hemostasis phenotypes in Spanish families. Circulation 101:1546–1551

Stern MP, Bartley M, Duggirala R, Bradshaw B (2000) Birth weight and the metabolic syndrome: thrifty phenotype or thrifty genotype? Diabetes Metab Res Rev 16:88–93

Vlieg AH, Linden IK, Bertina RM, Rosendaal FR (2000) High levels of factor IX increase the risk for venous thrombosis. Blood 95:3678–3682

Williams JT, Begleiter H, Porjesz B, Edenberg HJ, Foround T, Reich T, Goate A, Van Eerdewegh P, Almasy L, Blangero J (1999*a*) Joint multipoint linkage analysis of multivariate qualitative and quantitative traits. II. Alcoholism and event-related potentials. Am J Hum Genet 65:1148–1160

Williams JT, Van Eerdwegh P, Almasy L, Blangero J (1999*b*) Joint multipoint linkage analysis of multivariate qualitative and quantitative traits. I. Likelihood formulation and simulation results. Am J Hum Genet 65:1134–1147

5.4 Regulación genética de los niveles en plasma de las proteínas dependientes de la vitamina K involucradas en la Hemostasia: resultados del proyecto GAIT

(*Thromb Haemost 2001;85:88-92*)

Genetic Regulation of Plasma Levels of Vitamin K-dependent Proteins Involved in Hemostasis

Results from the GAIT Project

Juan Carlos Souto[1], Laura Almasy[2], John Blangero[2], William Stone[2, 3], Montse Borrell[1], Teresa Urrutia[1], José Mateo[1], Jordi Fontcuberta[1]

[1]Unitat d'Hemostàsia i Trombosi. Hospital de la Santa Creu i Sant Pau. Barcelona, Spain;
[2]Department of Genetics, Southwest Foundation for Biomedical Research. San Antonio, TX,
[3]Department of Biology, Trinity University, San Antonio, TX, USA

Key words

Vitamin K, thrombosis, genetic correlation, heritability, pleiotropy

Summary

Vitamin K-dependent proteins play a critical role in hemostasis. We have analysed the genetic and environmental correlations between measures of several vitamin K-dependent proteins in 21 Spanish extended families, including 397 individuals. Plasma functional levels of factors II, VII, IX, X, protein C and functional protein S were assayed in an automated coagulometer. Antigenic levels of total and free protein S were measured using an ELISA method. A maximum likelihood-based covariance decomposition analysis was used to assess the heritability of each trait and the genetic and environmental correlations between all possible pairs. All of the plasma levels had a significant genetic component (heritability) ranging from 22% to 52% of the phenotypic variance. Among the 28 possible pairs of genetic correlations, 18 were significant at a level of $p < 0.05$ and six exhibited a p-value between 0.05 and 0.10. Positive environmental correlation was observed for 25 of the pairs ($p < 0.05$). We conclude that genetic effects account for a large proportion of the observed phenotypic variation in vitamin K-dependent proteins. Some of the genes appear to pleiotropically influence all of these traits, since most pairs of phenotypes exhibit significant genetic correlation. However, since these phenotypes show a high degree of environmental correlation, it is also likely that the same environmental factors influence them co-jointly.

Introduction

Vitamin K-dependent proteins participating in the hemostasis pathways include factors II (prothrombin), VII, IX and X, protein C and protein S. Vitamin K is required for the complete synthesis of these blood-clotting proenzymes or natural anticoagulants. All of these proteins contain γ-carboxyglutamic acid. The γ-carboxyglutamic acid (Gla) domain found in the vitamin K-dependent proteins contains 10 to 12 residues of γ-carboxyglutamic acid. These unique amino acids are critical for calcium-ion binding and are necessary for the interaction of these proteins with cell membranes (1). Apart from the Gla domain, the vitamin K-dependent serine proteinases exhibit substantial sequence and structural homology. Factors VII, IX, X and protein C all contain two epidermal growth-factor-like domains and a catalytic domain. Prothrombin posseses two kringle domains instead of the EGF domains found in the other factors (1). With the exception of prothrombin, the vitamin K-dependent factors of coagulation are encoded by genes with virtually identical exon/intron distributions (2), suggesting that they have evolved relatively recently from a common ancestor by a process of gene duplication and divergence (3). The probable amino acid sequences for some of the functional domains of this early mammalian ancestor have been reconstructed by employing cDNA sequence data from a range of mammal species and by using established phylogenies (4, 5). Considering the evolutionary relatedness of the current structural genes encoding the vitamin K-dependent proteins, it is likely that both the genes and the proteins have common regulatory mechanisms.

Although there have been a substantial number of biochemical and molecular biological studies of these vitamin K-dependent proteins, comparatively little is known about the relative importance of genetic factors in the determination of observed functional levels of variability in human populations. To investigate the genetic basis of these important phenotypes and to assess the common regulation of the phenotypic expression of vitamin K-dependent proteins related with hemostasis, we have analysed the genetic and environmental correlations among these proteins using data from the Spanish family-based GAIT project (Genetic Analysis of Idiopathic Thrombophilia).

Methods

Study Population and Diagnosis

The GAIT Study is composed of 21 extended families, 12 of which were ascertained through a proband with thrombophilia and 9 of which were obtained randomly from the general population. Thrombophilia was defined in the probands as multiple thrombotic events (≥1 spontaneous), a single spontaneous episode of thrombosis with a first-degree relative also affected, or onset of thrombosis before 45 years. Thrombosis in these individuals was considered idiopathic because of exclusion of all of the biological causes of thrombosis known at the time of recruitment (1995-1997) including antithrombin deficiency, protein S and C deficiencies, activated protein C resistance, plasminogen deficiency, heparin cofactor II deficiency, Factor V Leiden, dysfibrinogenemia, lupus anticoagulant, and antiphospholipid antibodies.

Correspondence to: Dr. Juan Carlos Souto, Unitat d'Hemostàsia i Trombosi, Hospital de la Santa Creu i Sant Pau, Sant Antoni M. Claret 167, 08025, Barcelona, Spain – Tel.: 34-93-2919193; Fax: 34-93-2919192; E-mail: jsouto@hsp.santpau.es

A total of 397 individuals were examined with a mean of 19 individuals per family. Most pedigrees contained three generations, with 8 families having four generations, and one family having five generations. Subjects exhibited a median age-at-examination of 35.9 years (range 1-88) and consisted of approximately equal numbers of males and females. The composition of the families and the collection of lifestyle, medical and family history data is detailed in Souto et al. (6). Current oral contraceptive and cigarette use were recorded as part of the lifestyle and medical history data. The primary residence of each subject was also determined to assess the contribution of shared environmental influences (such as diet) common to members of a household. The study was performed according to the Declaration of Helsinki of 1975, and all adult patients provided informed consent for themselves and for their minor children, when applicable.

Laboratory Measurements and Techniques

A total of 43 quantitative physiological phenotypes related to hemostasis and to lipid metabolism were measured in the plasma of each individual. Details of most of these assays are available in Souto et al. (6).

None of the sampled individuals was being treated with anticoagulant therapy at the time of blood drawing. Blood was obtained from the antecubital vein, with the subject in a 12-hour fasted state. Samples for hemostatic tests were collected in 1/10 volume of 0.129 M sodium citrate. Platelet poor plasma was obtained by centrifugation at 2000 × g for 20 min at room temperature (22 ± 2° C). Coagulation factors were determined immediately on fresh plasma samples. The remaining plasma samples were stored at –80° C until use.

Coagulation factors II, VII, IX and X were assayed using deficient plasma from Diagnostica Stago (Asnières), in the STA automated coagulometer (Boehringer Mannheim, Mannheim). Functional protein S was determined with a kit from Diagnostica Stago in the STA automated coagulometer. Protein C was measured in a biochemical analyzer (CPA Coulter, Coulter Corporation, Miami FL) using chromogenic methods from Chromogenix (Mölndal). Total and free protein S were assayed using ELISA methods from Diagnostica Stago.

To reduce measurement error, assays were performed in duplicate and the average value of assays was calculated for each person. Intra- and inter-assay coefficients of variation were generally estimated between 2 and 6%.

Statistical Genetic Analysis

The analysis of heritabilities (i.e., h^2, the relative proportion of phenotypic variance that is attributable to the additive effects of genes) was performed using a variance component method (7-9). The total phenotypic variance in the traits was partitioned into three components including: 1) an additive genetic variance that is due to the sum of the average effects of all of the genes that influence the trait; 2) a shared environmental variance due to the shared effects of environmental factors that are common to members of a household, and 3) a random environmental variance that is specific to each individual. The random environmental variance also absorbs non-additive genetic effects such as interactions between alleles within loci (dominance effects), interactions between alleles at different loci (epistatic effects), and effects due to gene-environment interactions. Therefore, this model generally underestimates the total role of genetics in the determination of the trait.

In the variance component approach, the covariances in any phenotype among pedigree members that are due to additive genetic effects are modeled as a function of their expected genetic kinship relationships and the h^2 (9). Covariances among individuals that are due to shared environmental effects are modeled by considering a component of covariance that is only present among individuals living in the same household (c^2). The power of this general variance component method to disentangle genetic and environmental effects stems from the high information content of extended pedigrees where families cut across multiple households.

Variance component analysis also allows us to study the genetic relationships between quantitative variation in the parameters of interest. This method allows the phenotypic correlations between these traits to be decomposed into factors due to genetic, shared environmental (e.g., household) and individual-specific environmental influences. Examination of the underlying determinants of phenotypic correlations provides information on the role of pleiotropic genetic effects (i.e., one gene may have effects on several phenotypes).

The classical epidemiological evidence for the relationship between hemostatic parameters is based on the evaluation of phenotypic correlations. However, the partitioning of observed phenotypic correlations into genetic and environmental components is potentially valuable since hidden relationships between traits can be revealed (10). By studying both traits in extended families, we can estimate the genetic (ρ_g), the shared household (ρ_c), and the environmental (ρ_e) correlations between traits. The phenotypic correlation (ρ_p) can be derived from these three constituent correlations and the heritabilities and household effects of the traits from the relationship:

$$\rho_p = \sqrt{(h^2_1 h^2_2)}\,\rho_g + \sqrt{(c^2_1 c^2_2\,\rho_c)} + \sqrt{(1-h^2_1-c^2_1)}\,\sqrt{(1-h^2_2-c^2_2)}\,\rho_e$$

where h^2_1 and h^2_2 are the heritabilities for trait one and trait two and c^2_1 and c^2_2 are the shared environmental (household) effects on trait "one" and trait "two".

All statistical genetic analyses were performed using *SOLAR* software program (9). Estimates of variance component parameters including the heritabilities of the related vitamin K-dependent proteins and all of the phenotypic, genetic, and environmental correlations between these quantitative phenotypes were obtained using maximum likelihood estimation. All hypothesis tests were performed using likelihood ratio test statistics (11, 12). Covariate effects including smoking, oral contraceptives use, sex and age effects were simultaneously estimated in all analyses.

Because 12 of the 21 pedigrees were ascertained through a thrombophilic proband, all analyses included an ascertainment correction to allow unbiased estimation of parameters relevant to the general population. To achieve this, the likelihood for each family ascertained through a thrombophilic proband was conditioned on the phenotype of the proband (8, 13).

Results

Summary statistics (means and standard deviations) for the various vitamin K-dependent phenotypes are shown by age category and sex in Table 1. Additionally, the frequency of smoking and contraceptive use is provided. It should be noted that these estimates do not account for the non-independence among family members and therefore statistical tests directly based on them are invalid. They are provided here for comparative purposes only with formal tests of effects being limited to the likelihood based analyses, which explicitly deal with non-independence of phenotypes among relatives.

Regression coefficients for the environmental covariates, estimated simultaneously along with the effects of heredity and household showed that sex and age effects were significant for most of the traits examined. For example, factors IX, X, and all three protein S measures exhibited lower mean values in females. All of the traits showed some relationship with age. In general, there was a tendency to increased levels with age. Factor VII and protein C also showed lower levels in smokers while oral contraceptive use increased factors II and X and decreased free protein S levels. The regression coefficients and p-values for these covariate effects can be found in Souto et al. (6).

Table 2 shows the estimates of h^2 and household effects for each vitamin K-dependent phenotype. All the explored plasma levels had a significant h^2 ranging from 22% to 52% of the phenotypic variance, while only total Protein S, functional Protein S, factor X and free protein S exhibited a household (shared environment) component of the variance. The household effect was much weaker than the genetic effect for each trait, ranging from 9.5% to 21%.

Table 3 shows the estimates of the phenotypic correlations between all possible pairs of the phenotypes. All of them are highly significant and correspond to the classical Pearson coefficient used in convention-

Table 1 Distribution of the levels of vitamin K-dependent phenotypes and mean age of subjects, percentage of smokers and women using oral contraceptives within eight groups stratified by age and sex

	Ages 0-19		Ages 20-39		Ages 40-59		Ages 60+	
Variable	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Number	39	45	64	75	55	62	23	34
Protein C (%)	90.3	101.4	112.6	111.6	123.1	120.3	115.1	124.7
	(15.0)	(19.5)	(16.0)	(15.7)	(18.7)	(25.3)	(23.3)	(23.9)
Total Protein S (%)	92.5	86.8	111.4	83.9	109.0	95.4	103.6	93.1
	(16.3)	(15.5)	(20.0)	(15.9)	(22.7)	(21.0)	(20.6)	(16.1)
Free Protein S (%)	97.5	96.8	105.3	88.0	111.2	104.4	110.9	107.1
	(14.3)	(20.4)	(18.6)	(13.5)	(17.8)	(19.8)	(23.9)	(20.4)
Functional	95.1	86.3	110.3	86.4	112.1	90.6	96.5	83.6
Protein S (%)	(16.6)	(18.8)	(21.0)	(18.4)	(26.8)	(21.2)	(21.6)	(21.9)
Factor II (%)	109.4	118.7	124.2	125.9	127.7	129.5	117.0	124.4
	(14.5)	(17.1)	(16.6)	(19.7)	(16.0)	(17.2)	(19.2)	(18.8)
Factor VII (%)	102.2	101.5	110.4	114.0	120.7	130.4	119.0	137.3
	(19.5)	(21.9)	(24.8)	(28.5)	(23.5)	(31.3)	(32.3)	(31.1)
Factor IX (%)	98.2	99.9	116.8	109.4	127.0	116.8	126.2	132.6
	(19.3)	(19.7)	(23.2)	(23.5)	(24.4)	(23.8)	(22.9)	(24.3)
Factor X (%)	99.1	100.7	123.2	112.2	124.2	119.0	103.7	105.8
	(19.8)	(16.3)	(23.1)	(23.7)	(20.0)	(22.6)	(21.2)	(16.4)
Age (years)	12.3	14.1	28.2	29.0	49.4	49.0	72.5	71.7
	(5.3)	(5.0)	(5.4)	(6.0)	(5.4)	(4.8)	(6.2)	(6.4)
Smoking	10.3	20.0	54.7	50.7	56.4	29.0	60.9	0
Use of oral contraceptives		2.2		17.3		3.2		0

The vitamin K-dependent phenotypes are expressed as means with the standard deviations in parenthesis below. Age is represented as mean (SD). Smokers and oral contraceptive users are expressed as proportions.

Table 2 Estimated components of relative variance with standard errors (SE). They are listed in descending order of heritabilities. Extracted from Souto et al. (6)

Phenotype	Heritability	SE	Household	SE
Factor VII	0.523 §	0.089		
Protein C	0.501 §	0.086		
Factor II	0.492 §	0.088		
Total Protein S	0.460 §	0.088	0.108 *	0.057
Functional Protein S	0.453 §	0.096	0.095 *	0.060
Factor X	0.434 §	0.127	0.135 *	0.076
Factor IX	0.387 §	0.086		
Free Protein S	0.223 †	0.106	0.212 ‡	0.065

* p<0.05, † p<0.01, ‡ p<0.001, § p<0.0001

Table 3 Estimated phenotypic, genetic, environmental, and household correlations between all pairs of vitamin K-dependent phenotypes

	Correlations			
Phenotypes	Phenotypic	Genetic	Household	Environmental
FII – FVII	0.39[1]	0.23[5]	-	0.55[1]
FII – FIX	0.46[1]	0.53[3]	-	0.41[2]
FII – FX	0.54[1]	0.35[4]	-	0.82[1]
FII – Prot C	0.45[1]	0.31[4]	-	0.59[1]
FII – total PS	0.29[1]	0.46[3]	-	0.30[3]
FII – free PS	0.41[1]	0.86[1]	-	0.19[6]
FII – funct PS	0.23[1]	0.26[6]	-	0.33[3]
FVII – FIX	0.35[1]	0.08[6]	-	0.55[1]
FVII – FX	0.48[1]	0.14[5]	-	0.71[1]
FVII – Prot C	0.46[1]	0.42[3]	-	0.51[1]
FVII – total PS	0.20[2]	0.22[5]	-	0.24[4]
FVII – free PS	0.29[1]	0.57[4]	-	0.21[5]
FVII – funct PS	0.23[1]	0.36[4]	-	-0.02[6]
FIX – FX	0.43[1]	0.23[5]	-	0.53[1]
FIX – Prot C	0.42[1]	0.33[4]	-	0.48[1]
FIX – total PS	0.30[1]	0.35[4]	-	0.23[4]
FIX – free PS	0.29[1]	0.51[5]	-	0.30[3]
FIX – funct PS	0.19[1]	-0.13[6]	-	0.35[3]
FX – Prot C	0.41[1]	0.15[5]	-	0.61[1]
FX – total PS	0.45[1]	0.65[2]	-0.11[6]	0.37[3]
FX – free PS	0.45[1]	0.86[3]	-0.23[6]	0.30[4]
FX – funct PS	0.42[1]	0.65[3]	0.12[6]	0.28[4]
Prot C – total PS	0.31[1]	0.33[3]	-	0.33[3]
Prot C – free PS	0.35[1]	0.67[3]	-	0.25[4]
Prot C – funct PS	0.28[1]	0.12[6]	-	0.37[3]
Total PS – free PS	0.63[1]	0.80[3]	0.80[1]	0.55[1]
Total PS – funct PS	0.62[1]	0.65[3]	0.61[4]	0.58[1]
Free PS – funct PS	0.42[1]	0.76[3]	0.12[6]	0.29[4]

(1) p<0.0001, (2) p<0.001, (3) p<0.01, (4) p< 0.05, (5) p<0.10, (6) non significant

al epidemiological studies. Among the 28 possible pairs of genetic correlations included in Table 3, 18 were significant at a level of $p < 0.05$ and 6 exhibited a p-value between 0.05 and 0.10. Only 4 pairs of genetic correlations were clearly not significant and they involved functional protein S. High environmental correlations were observed for the majority of the pairs (Table 3). Twenty-five of the pairs showed statistical significance ($p < 0.05$). Household correlations are shown only for those traits with significant household components, namely the protein S measures and factor X. Only measures related to protein S showed significant household correlations, between total protein S and both free and functional protein S.

Discussion

As quantitative traits, the plasma levels of the vitamin K-dependent proteins are the result of the interaction of multiple genes and environmental factors (14). Heritability represents the summed effects of these as yet unknown genes. To date, very few data on the relative contribution of genes in the determination of plasma levels of vitamin K-dependent proteins are available. Only the heritabilities of factors VII and IX have been investigated previously; a h^2 of 0.57 was estimated for factor VII in a twin study (15). This is very similar to our own estimate of 0.52. Another twin study by Orstavik et al. estimated a h^2 of 0.20 for factor IX, although this was not significantly different from zero (16). Due to the relatively small sample size in the Orstavik et al. study, their estimate of 0.20 is also not significantly different from our estimate of 0.39. Regarding the remaining phenotypes, our h^2 estimates represent the first ones to be published (6). The relative homogeneity of our h^2 estimates for these vitamin K-dependent proteins (i.e. 0.40 to 0.50) may be an indirect reflection of their shared regulatory mechanisms. In any case, the most important conclusion to be drawn from these data is that the genetic influences on these phenotypes are substantial and therefore it is likely that we will be able to localize at least some of the underlying genes.

We found substantial phenotypic correlations between all pairs of vitamin K-dependent proteins. Previous studies have found positive correlations among some of these pairs by means of classical epidemiological approaches (17, 18). In a cross-sectional study of healthy North American individuals over age 65, Sakkinen et al. (18) found

positive correlations for protein C and total protein S with factors VII, IX and X. In a larger sample of people aged 25-74 years, randomly obtained from the Scottish population, Lowe et al. (17), reported significant correlations among functional protein S, protein C and factors VII and IX. They also found a significant correlation between factor VII and factor IX. Nevertheless, such classical epidemiological methods are unable to distinguish the source of these correlations. Purely phenotypic correlations do not reflect the magnitude of underlying genetic and environmental correlations, particularly when the genetic and environmental correlations between traits have opposite signs. Clearly, the sums of disparate components (such as genetic and environmental effects) lead to a loss of information relative to their separate evaluation. Significant phenotypic correlations among traits can arise from three conditions: 1) entirely from shared genetic effects, 2) entirely from shared random environmental effects, or 3) a combination of both effects.

The present study provides the first direct quantification of the genetic and environmental correlations between the vitamin K-dependent hemostasis factors. Both genetic and environmental factors appear to co-jointly affect these traits. The ultimate causal determinants of these correlations are not identifiable using this approach, but we can speculate on their probable identities.

With regard to the genetic basis of our observed correlations, some of the underlying genes pleiotropically influence all of these traits, since most of the phenotypes exhibit significant genetic correlation among them. This implies the existence of common genetic regulatory mechanisms for these plasma proteins. Probably, part of this common biological control is due to a number of shared post-translational modifications to yield the active forms of the proteins. This process may involve *γ-glutamyl carboxylase*, an integral membrane microsomal enzyme located in the rough endoplasmic reticulum, which carboxylates glutamate residues located in the Gla domain of vitamin K-dependent proteins. The carboxylation reaction is dependent on reduced vitamin K, which is converted to vitamin K epoxide during carboxylation, and must be regenerated by the *vitamin K epoxide reductase* for carboxylation to continue (19). We hypothesise that the genes encoding both enzymes (γ-glutamyl carboxylase and vitamin K epoxide reductase) and their regulatory genes could be responsible of part of the observed genetic correlations in our study. There are several clinical cases reported in the literature of hereditary combined deficiencies, involving vitamin K-dependent proteins, which support this hypothesis. These rare bleeding disorders can be theoretically related to functional defects in some of the previously mentioned enzymes (20). Recently, a mutation in the γ-glutamyl carboxylase gene has been determined to cause this syndrome (21). In addition to these two potential influences, other common genetic regulatory mechanisms can be expected at the transcriptional or post-transcriptional levels of the respective structural genes. For instance, the promoter regions of these genes share some general features and they are regulated by the combined action of liver-specific and ubiquitous transcription factors (22). This expectation arises from the evolutionary history (3) and the high degree of similarity among the vitamin K-dependent proteins (1, 2).

Although genetic determinants appear to dominate the phenotypic correlations between vitamin K-dependent proteins, the observed random environmental correlations were substantial also. However, shared environmental components (as measured by household effect) were relatively unimportant, attaining significance only for factor X and protein S measures. For them, the estimated magnitude of household effects was small to moderate, ranging from 9% to 21% of the total phenotypic variation. Household effects are attributable to unmeasured non-genetic factors, which are shared more closely by individuals living in the same households than by individuals living in different households. They may represent unmeasured dietary (e.g. vitamin K intake) or other lifestyle factors. Thus, the environmental correlations among the other phenotypes, different from factor X or protein S may be due to other factors independent of dietary influences.

Although these results do not have immediate clinical applications, they imply the existence of quantitative trait loci (QTLs) that regulate these plasma phenotypes, which can be important in an individual's thrombotic risk profile. This possibility is reasonable since additional results from our GAIT Project demonstrate that plasma levels of factor II (23), factor VII (24) and factor IX (24) are genetically correlated with the risk of thrombosis. And very recently, it has been found in the LETS study that high levels of factor IX may be a risk factor for venous thromboembolism (25). Thus, it seems likely that any gene co-jointly influencing (pleiotropy) the vitamin K-dependent phenotypes may be clinically relevant for diagnosis, treatment and prevention of thrombosis.

In summary, our study indicates that the plasma levels of this group of closely related proteins are co-jointly influenced by a set of common genes. It is likely that currently unidentified genes mainly compose this set. Similarly, a set of common unmeasured environmental factors also appears to influence plasma levels of vitamin K-dependent proteins. These environmental factors are not correlated within households and therefore do not appear to represent shared environmental effects. Considering the genetic, chemical and functional relatedness of all these proteins, our results are in agreement with what one could expect *a priori*. They demonstrate the power of statistical genetic methodologies for studying the genetic regulation of quantitative complex phenotypes. We anticipate that the pleiotropy that was revealed in our analyses can be exploited in genomic scans (searches for specific genes) to improve the power to localize and characterize QTLs influencing variation in the vitamin K-related proteins.

Acknowledgements

This study was supported by grants DGICYT Sab 94/0170 from the Ministerio de Educacion y Ciencia, FIS 97/2032 from the Ministerio de Sanidad y Consumo and RED 97/3 from the Generalitat de Catalunya, Spain. Statistical genetic analysis was supported by NIH grants MH59490 and GM18897. We are grateful to a number of doctors who assisted in the ascertainment and recruitment of thrombophilic pedigrees: Dr Javier Rodríguez Martorell from Hospital Universitario Puerta del Mar, in Cádiz, Dr Carmen Araguás from Hospital Arnau de Vilanova, in Lleida, Dr Francisco Velasco from Hospital Reina Sofía, in Córdoba, Dr Montserrat Maicas from the Hospital General de Albacete and Dr Dilia Brito from Hospital Carlos Haya, in Málaga. We would also like to acknowledge the excellent technical contributions of Cristina Vallvè, Dolors Llobet, Rosa Felices and Joaquín Murillo, Finally, we are deeply grateful to all of the families who have participated in this study.

References

1. Furie B, Furie BC. Molecular and cellular biology of blood coagulation. N Engl J Med 1992; 326: 800-6.
2. Tuddenham EGD, Cooper DN. The molecular genetics of haemostasis and its inherited disorders. Oxford University Press. Oxford 1994.
3. Patthy L. Evolutionary assembly of blood coagulation proteins. Semin Thromb Hemost 1990; 16: 245-59.
4. Krawczak M, Wacey A, Cooper DN. Molecular reconstruction and homology modelling of the catalytic domain of the common ancestor of the

haemostatic vitamin-K-dependent serine proteinases. Hum Genet 1996; 98: 351-70.
5. Wacey AI, Krawczak M, Kemball-Cook G, Cooper DN. Homology modelling of the catalytic domain of early mammalian protein C: evolution of structural features. Hum Genet 1997; 101: 37-42.
6. Souto JC, Almasy L, Borrell M, Garí M, Martínez E, Mateo J, Stone WH, Blangero J, Fontcuberta J. Genetic determinants of hemostasis phenotypes in Spanish families. Circulation 2000; 101: 1546-51.
7. Spence MA, Westlake J, Lange K. Estimation of the variance components for dermal ridge count. Ann Hum Genet 1977; 41: 111-5.
8. Hopper JL, Matthews JD. Extensions to multivariate normal models for pedigree analysis. Ann Hum Genet 1982; 46: 373-83.
9. Almasy L, Blangero J. Multipoint quantitative trait linkage analysis in general pedigrees. Am J Hum Gen 1998; 62: 1198-211.
10. Comuzzie AG, Blangero J, Mahaney MC, Haffner SM, Mitchell BD, Stern MP, MacCluer JW. Genetic and environmental correlations among hormone levels and measures of body fat accumulation and topography. J Clin Endocrinol Metab 1996; 81: 597-600.
11. Kendall MG, Stuart A. Advanced Theory of Statistics. Hafner Publishing Co., New York, NY, 1972.
12. Self SG, Liang K-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 1987; 82: 605-10.
13. Boehnke M, Lange K. Ascertainment and goodness of fit of variance component models for pedigree data. Prog Clin Biol Res 1984; 147: 173-92.
14. Weiss KM. Genetic variation and human disease. Principles and evolutionary approaches. Cambridge University Press. Cambridge 1995.
15. Hong Y, Pedersen NL, Egberg N, de Faire U. Genetic effects for plasma factor VII levels independent of and in common with triglycerides. Thromb Haemost 1999; 81: 382-6.
16. Orstavik KH, Magnus P, Reisner H, Berg K, Graham JB, Nance W. Factor VIII and factor IX in a twin population. Evidence for a major effect of ABO locus on factor VIII level. Am J Hum Genet 1985; 37: 89-101.
17. Lowe GDO, Rumley A, Woodward M, Morrison CE, Philippou H, Lane DA, Tunstall-Pedoe H. Epidemiology of coagulation factors, inhibitors and activation markers: The third Glasgow MONICA survey I. Illustrative reference ranges by age, sex and hormone use. Br J Haematol 1997; 97: 775-84.
18. Sakkinen PA, Cushman M, Psaty BM, Kuller LH, Bajaj SP, Sabharwal AK, Boineau R, Macy E, Tracy RP. Correlates of antithrombin, protein C, protein S, and TFPI in a healthy elderly cohort. Thromb Haemost 1998; 80: 134-9.
19. Furie B, Bouchard BA, Furie BC. Vitamin K-dependent biosynthesis of γ-carboxyglutamic acid. Blood 1999; 93: 1798-808.
20. Pechlaner C, Vogel W, Erhart R, Pümpel E, Kunz F. A new case of combined deficiency of vitamin K dependent coagulation factors. Thromb Haemost 1992; 67: 617.
21. Brenner B, Sánchez-Vega B, Wu SM, Lanir N, Stafford DW, Solera J. A missense mutation in γ-glutamyl carboxylase gene causes combined deficiency of all vitamin K-dependent blood coagulation factors. Blood 1998; 92: 4554-9.
22. Spek CA. Characterization of the human protein C gene promoter. Doctoral thesis 1998. Leiden University.
23. Soria JM, Almasy L, Souto JC, Tirado I, Borrell M, Mateo J, Slifer S, Stone W, Blangero J, Fontcuberta J. Linkage analysis demonstrates that the prothrombin G20210A mutation jointly influences plasma prothrombin levels and risk of thrombosis. Blood 2000; 95: 2780-5.
24. Souto JC, Almasy L, Borrell M, Blanco-Vaca F, Mateo J, Soria JM, Coll I, Felices R, Stone W, Fontcuberta J, Blangero J. Genetic susceptibility to thrombosis and its relationship to physiological risk factors: the GAIT study. Am J Hum Genet 2000; 67: in press.
25. Vlieg AH, van Der Linden IK, Bertina RM, Rosendaal FR. High levels of factor IX increase the risk for venous thrombosis. Blood 2000; 95: 3678-82.

Received December 24, 1999 Accepted after resubmission August 25, 2000

5.5 El análisis de ligamiento demuestra que la mutación G20210A en el gen de la protrombina influye a la vez en los niveles de protrombina y en el riesgo de trombosis

(*Blood 2000;95:2780-2785*)