

Bases Genètiques de l'Osteoporosi: Estudi del gen *LRP5*.

Lídia Agueda Calpena

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tesisenxarxa.net) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tesisenred.net) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tesisenxarxa.net) service has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading and availability from a site foreign to the TDX service. Introducing its content in a window or frame foreign to the TDX service is not authorized (framing). This rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

BASES GENÈTIQUES DE L'OSTEOPOROSI: ESTUDI DEL GEN *LRP5*

Memòria presentada per
Lídia Àgueda Calpena
Per optar al grau de
Doctora per la Universitat de Barcelona

Tesi dirigida per la Dra. Susana Balcells Comas i pel Dr. Daniel Grinberg Vaisman
Al departament de Genètica de la Facultat de Biologia
de la Universitat de Barcelona
Programa de Genètica 2004-2006

Dra. Susana Balcells Comas

Dr. Daniel Grinberg Vaisman

Lídia Agueda Calpena
2010

CONTINGUTS

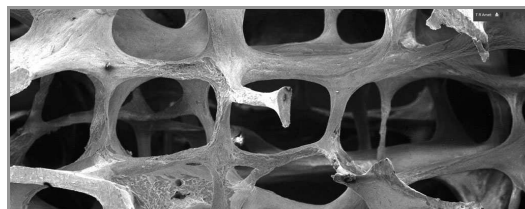
INTRODUCCIÓ	6
1. El teixit ossi.....	8
1.1. Funció i estructura del teixit ossi.....	8
1.2. Les cèl·lules òssies i l'osteogènesi	9
1.3. El remodelat i el control de l'homeòstasi òssia.....	12
2. Elements clau per a la diferenciació i funció dels osteoblasts	15
2.1. El factor de transcripció Runx2.....	16
2.2. La via de Wnt.....	20
3. Exemples de fenotips ossis.....	24
3.1. Dos fenotips ossis monogènics oposats: la síndrome de l'osteoporosi i pseudoglioma (OPPG) i el <i>High Bone Mass</i> (HBM)	24
3.2. L'osteoporosi postmenopàusica	26
3.2.1. Osteoporosi i fractura osteoporòtica: definició i tipus	26
3.2.2. Factors de risc.....	29
3.2.3. Diagnòstic i tractament de l'osteoporosi postmenopàusica.....	29
4. L'estudi genètic d'una malaltia complexa: l'osteoporosi com a exemple.	33
4.1. Conceptes bàsics d'epidemiologia genètica.....	33
4.1.1. Malaltia monogènica <i>versus</i> malaltia complexa	33
4.1.2. La variabilitat genètica	34
4.1.2.1. Tipus de variabilitat genètica	34
4.1.2.2. El desequilibri de lligament i els blocs haplotípics	36
4.1.2.3. Les tècniques de genotipat	38
4.1.2.4. La funcionalitat dels polimorfismes	39
4.2. Mètodes d'estudi de les malalties complexes. Resultats en el camp de l'osteoporosi.	41
4.2.1. L'heretabilitat	42
4.2.2. Els estudis de lligament (linkage)	42
4.2.3. Els estudis d'associació.....	44
4.2.4. Les meta-anàlisis	49
4.2.5. Els estudis d'expressió gènica	50

4.2.6. Els estudis en animals	51
OBJECTIUS	52
RESULTATS	55
<i>CAPÍTOL 1 : A haplotype-based analysis of the LRP5 gene in relation to osteoporosis phenotypes in Spanish postmenopausal women.</i>	57
<i>CAPÍTOL 2: Large-scale analysis of association between LRP5 and LRP6 variants and osteoporosis.</i>	58
<i>CAPÍTOL 3: Functional relevance of the BMD-associated polymorphism rs312009. Novel implication of Runx2 in LRP5 transcriptional regulation.....</i>	63
<i>CAPÍTOL 4: Analysis of three functional polymorphisms in relation to osteoporosis phenotypes: replication in a Spanish cohort</i>	93
DISCUSSIÓ	95
1. Evolució dels estudis d'associació en el camp de l'osteoporosi	99
1.1. El primer pas: les estimes d'heretabilitat dels fenotips osteoporòtics.....	99
1.2. Els estudis d'associació directa en els gens candidats clàssics	100
1.3. Nous gens candidats i més polimorfismes. Les estratègies d'associació indirecta.	101
1.4. L'era dels GWAs	103
1.5. El paper de les CNVs en l'osteoporosi	105
2. La força de la mida mostral: les meta-anàlisis.....	106
3. La replicació: Una solució o més preguntes?.....	109
4. "The missing heritability"	112
4.1. El replantejament de la hipòtesi variant comuna / malaltia comuna.....	112
4.2. El paper de les interaccions	115
5. El per què de les associacions trobades: els estudis funcionals	117
6. Aplicacions dels estudis d'associació i perspectives de futur	119
CONCLUSIONS	123
BIBLIOGRAFIA	129
ANNEX	139

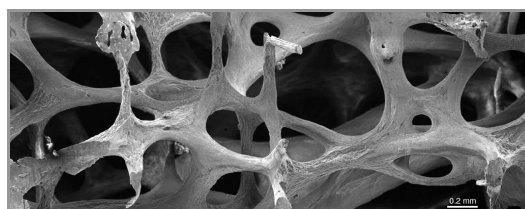
LLISTAT D'ABREVIATURES:

BARCOS	cohort de dones potsmenopàusiques de l'àrea de Barcelona
BMPs	proteïnes morfogenètiques de l'os
CBFA1	<i>core binding factor α 1</i>
CCD	displàsia cleido-cranial
CD/CV	<i>common disease / common variant hypothesis</i> ; malaltia comuna / variant comuna
CD/RV	<i>common disease / rare variant hypothesis</i> ; malaltia comuna / variant minoritària
CNVs	<i>copy number variants</i> , variants de nombre de còpia
COL1A1	cadena α 1 del col·lagen tipus I
CHIP	immunoprecipitació de cromatina
DMO/BMD	densitat mineral ossia, <i>bone mineral density</i>
DXA	Absorciometria dual de rajos X
EMSA	<i>Electrophoretic mobility shift assay</i> , assaig de canvi en la mobilitat electroforètica o estudi de retenció en gel
ESR1 ESR2	receptors d'estrògens 1 i 2
FGF	factor de creixement fibroblàstic
FN	<i>femoral neck</i> , coll del fèmur
GEFOS	Genetic Factors for Osteoporosis, projecte europeu
GENOMOS	Genetic Markers for Osteoporosis, projecte europeu
GWA	<i>genome wide association</i> , associació a nivell de tot el genoma
HapMap	Mapa d'Haplotips del Genoma Humà, projecte internacional
HBM	<i>high bone mass</i> , fenotip d'alta massa òssia
HeLa	línia cel·lular d'adenocarcinoma humà
Kb	quilobase
LCT	lactasa
LD	desequilibri de lligament
LRP5	proteïna 5 relacionada amb el receptor de lipoproteïnes de baixa densitat
LRP6	proteïna 6 relacionada amb el receptor de lipoproteïnes de baixa densitat
LS	<i>lumbar spine</i> , columna
MAF	freqüència de l'al·lel minoritari
M-CSF	factor estimulador de colònies de macròfags
MSC	cèl·lules mare mesenquimals
MTHFR	metilentetrahidrofolat reductasa
OPG	osteoprotegerina
OPPG	síndrome de l'osteoporesi-pseudoglioma
OR	<i>odds ratio</i> , oportunitat relativa, raons d'oportunitat
OSE	element específic d'osteoblast
P	nivell de significació
pGL3	vector amb el gen reporter Luciferasa (de Promega)
PTH	hormona paratiroidea
QTL	<i>quantitative trait locus</i> , locus de tret quantitatiu
RANK	membre 11A de la superfamília de receptors del TNF
RANKL	ligand del receptor RANK
RFLP	polimorfisme de fragments de restricció de longitud polimòrfica

RUNX2	<i>runt related transcription factor 2</i> , factor de transcripció 2 relacionat amb el domini runt (Cbfa1)
SAOS-2	línia cel·lular d'osteosarcoma humà
SNP	<i>single nucleotide polymorphism</i> , polimorfisme de canvi puntual de nucleòtid
TCF/LEF	<i>lymphoid enhancer binding factor</i> / factor de transcripció específic de cèlules T
TNFs	factors de necrosi tumoral
U-2 OS	línia cel·lular d'osteosarcoma humà
VDR	receptor de la vitamina D ₃



DISCUSSió



Els resultats presentats en aquesta tesi s'emmarquen en el camp de l'estudi de les bases genètiques de l'osteoporosi. L'osteoporosi és un clar exemple de malaltia complexa, on els factors genètics tenen una elevada participació en la seva etiologia. Des de mitjans dels anys 90, el nostre grup d'osteoporosi de la Universitat de Barcelona treballa en col·laboració amb la unitat URFOA de l'Hospital del Mar en la creació d'una col·lecció de mostres d'ADN de dones postmenopàusiques de l'àrea geogràfica de Barcelona, la cohort BARCOS. En aquesta es recullen, paral·lelament a les mostres biològiques, dades antropomètriques de les participants, valors de densitat mineral òssia i la ocurrència de fractures osteoporòtiques, així com d'altres informacions sobre l'estil de vida o hàbits alimentaris. Tot aquest material ens permet realitzar estudis d'associació genètica per trobar variants polimòrfiques que puguin explicar la variabilitat en els valors de DMO. La DMO s'utilitza com a fenotip quantitatiu per estudiar l'osteoporosi. Idealment, el coneixement d'aquestes variables associades ha de tenir un valor predictiu de la susceptibilitat de cada individu a patir osteoporosi o a presentar fractura osteoporòtica.

En aquest sentit, el principal objectiu d'aquesta tesi ha estat la realització d'un estudi d'associació en el qual es varen testar un grup d'SNPs seleccionats per tal de capturar la màxima variabilitat comuna en el gen *LRP5* amb la DMO lumbar i femoral i amb la presència/absència de fractures osteoporòtiques. Es varen obtenir resultats positius per diversos polimorfismes: rs312009 associat amb DMO lumbar, rs2508836 associat amb DMO lumbar i femoral, i rs729635 i rs643892 associats amb fractura (Agueda i col., 2008).

D'entre les associacions trobades, es va decidir prosseguir a l'estudi funcional d'una d'elles: el polimorfisme rs312009. Aquest polimorfisme es troba localitzat en la regió 5' del gen *LRP5* i els programes de predicció indicaven que en la seva posició s'hi podria unir el factor de transcripció osteoblàstic Runx2. Els assaigs de retenció en gel (EMSAs) van permetre confirmar aquesta unió. La següent hipòtesi va ser que la presència d'un o altre al·lel de rs312009 (C o T) podria estar afectant la capacitat transcripcional del promotor d'*LRP5*, i que aquest mecanisme molecular podia ser en part responsable de l'associació trobada. Tanmateix, es va voler aprofundir en

l'anàlisi de la implicació del factor de transcripció Runx2 en la transcripció d'*LRP5*. Aquestes noves preguntes es varen respondre experimentalment per mitjà d'assaigs de gen reporter (luciferasa) en dues línies d'osteosarcoma. La construcció amb la regió promotora que contenia l'al·lel T va presentar una capacitat transcripcional significativament superior a la de l'al·lel C. També es va descriure un efecte estimulador de Runx2 sobre el promotor d'*LRP5*, observat en els estudis de cotransfecció.

Seguint en la mateixa línia de recerca de variants genètiques implicades en la determinació dels fenotips osteoporòtics, però a diferent nivell experimental, la cohort BARCOS ha continuat participant en les meta-anàlisis realitzades pel consorci internacional GENOMOS. En el treball concret presentat en aquesta tesi, dues variants de canvi d'aminoàcid del gen *LRP5* i una del gen *LRP6* (Val667Met, Ala1330Val i Ile1062Val, respectivament) varen ésser interrogades en relació a la seva associació amb DMO i fractura en 37.534 individus de 18 grups europeus i nord-americans (van Meurs i col., 2008). En aquest treball varem trobar associacions consistents amb DMO i fractura per als polimorfismes d'*LRP5*, però amb efectes molt modestos.

Finalment, tres polimorfismes prèviament associats amb DMO en altres cohorts, i amb evidències experimentals de la seva funcionalitat biològica [(Ala222Val (rs1801133) del gen *MTHFR*, -13910C>T (rs4988235) del gen *LCT* i Ile1062Val (rs2302685) del gen *LRP6*], es varen testar en la cohort BARCOS. El principal objectiu d'aquest treball era la replicació dels resultats previs en una cohort independent, una aproximació actualment indispensable per confirmar els resultats trobats en els estudis d'associació (Agueda i col., 2010).

1. Evolució dels estudis d'associació en el camp de l'osteoporosi

1.1. El primer pas: les estimes d'heretabilitat dels fenotips osteoporòtics

De la definició oficial de l'*osteoporosi* presentada per l'NIH (malaltia sistèmica de l'esquelet caracteritzada per una massa òssia baixa i una alteració de la microarquitectura de l'os, amb el conseqüent augment de la seva fragilitat i una major susceptibilitat a les fractures (WHO Report, 2001) podem extreure'n que els principals elements de diagnòstic d'aquesta malaltia són la quantitat de massa òssia i la presència de fractures. L'observació de l'agregació familiar d'aquests fenotips va posar de manifest la implicació de factors genètics en la malaltia i en va permetre la classificació com a malaltia complexa. La quantificació d'aquest component genètic, mitjançant els estudis d'heretabilitat, ha fet possible el plantejament de la cerca de gens, i de variants concretes en aquests, que arribin a explicar aquesta agregació. Els valors d'heretabilitat per als diferents fenotips osteoporòtics obtinguts per mitjà d'estudis de bessons i de famílies són força variables: entre el 25 i el 85% [revisat per Ferrari (2008)]. Els valors més alts corresponen a l'heretabilitat de la DMO en diferents llocs esquelètics, en canvi, per al fenotip fractura, aquests valors són inferiors. Aquest fet il·lustra com d'altres factors no genètics (o no directament involucrats en l'òs) guanyen més pes en la determinació del fenotip fractura. Certament, els factors no genètics (entre ells els factors nutricionals, el sexe, la menopausa o els hàbits de vida) són molt importants, i fins i tot podem arribar a emascarar completament el component genètic en alguns casos. Actualment, gràcies a poder disposar de grans mides mostrals i de dissenys experimentals adequats, podem controlar al màxim aquests factors no genètics i estudiar el component genètic d'una malaltia complexa, en el nostre cas l'osteoporosi. La identificació d'aquests factors genètics permetrà entendre millor la malaltia; desenvolupar millors eines de diagnòstic, pronòstic i tractament; i per altra banda, analitzar els factors no genètics.

1.2. Els estudis d'associació directa en els gens candidats clàssics

Com ja s'ha comentat prèviament, una de les principals eines d'estudi de les malalties complexes són els estudis d'associació. El primer estudi d'associació de variants genètiques amb l'osteoporosi va ser el presentat per Eisman i col·laboradors el 1992 (Morrison i col., 1992). En aquest treball es varen estudiar tres variants polimòrfiques del gen del receptor de la vitamina D (*VDR*). Ràpidament van aparèixer a la literatura molts altres estudis d'associació amb fenotips osteoporòtics. La majoria d'ells es centraven en l'estudi d'un determinat gen del qual ja es coneixia una implicació biològica en el metabolisme ossi, els anomenats gens candidats, i amb un nombre limitat de polimorfismes, ja que aquests calia identificar-los prèviament. En ells s'avaluava l'associació directa del polimorfisme amb el fenotip, ja que s'assumia que aquest era el causant de l'associació (o se'n tenia una evidència experimental). Mitjançant aquesta estratègia es van obtenir resultats positius per els anomenats *gens candidats clàssics* de l'osteoporosi, entre ells: *VDR*, *ESR1*, *ESR2* i *COL1A1* i que es recullen extensament en la revisió de Li i col. (2010).

Entre les principals limitacions d'aquesta estratègia experimental destaca la necessitat d'un coneixement previ de la implicació en el metabolisme ossi del gen seleccionat per a l'estudi. A més, en el nombre de variants testades per cada gen és molt limitat, quedant-se sense avaluar molts SNPs i gens. Tot i aquestes limitacions, aquests estudis tenen un gran valor, ja que varen posar de manifest fets molt importants que varen marcar l'evolució dels posteriors estudis d'associació. En primer lloc, demostraren la presència de diferències en les associacions trobades entre llocs esquelètics, entre sexes i ètnies, o segons els grups d'edat o d'estatus menopàusic de les poblacions estudiades. I en segon lloc, varen il·lustrar la necessitat de dissenys experimentals extremadament acurats pel que fa a la definició del fenotip, les consideracions sobre la mida mostral, la població escollida per a l'estudi i els mètodes estadístics utilitzats en l'anàlisi.

La curta llista inicial dels gens candidats clàssics és va anar ampliant amb l'aparició de nous gens. Els avenços en els estudis de lligament en famílies, els estudis en animals i l'estudi de fenotips ossis monogènics, va permetre la

identificació de nous gens implicats en el metabolisme ossi, com ara *LRP5*, *OPG* o *RUNX2*, que presumiblement també contindrien polimorfismes associats a la variació normal de la DMO. Tanmateix, la manera d'analitzar aquests nous gens també anava evolucionant.

1.3. Nous gens candidats i més polimorfismes. Les estratègies d'associació indirecta.

El següent pas en l'evolució de les anàlisis d'associació va estar determinat per la finalització del Projecte Genoma Humà el 2003 i l'entrada en l'anomenada "era post-genòmica". L'avenç tecnològic, sobretot en tècniques de seqüenciació, va permetre la identificació de gran part de les variants polimòrfiques del genoma. Va néixer també, l'anomenat Projecte HapMap, que a part de catalogar aquestes variacions, també va pretendre determinar quines eren les seves relacions en termes de desequilibri de lligament.

A nivell dels estudis d'associació, això es va traduir en un gran augment en el nombre de polimorfismes candidats a testar. Els SNPs a analitzar ja no es limitaven a aquells identificats pels propis grups i amb sospites de funcionalitat (SNPs de canvi d'aminoàcid, sobretot), sinó que s'ampliava molt el nombre de variants descrites susceptibles a ser estudiades. Els estudis d'associació directa (variant testada, variant funcional) van ser desplaçats pels d'associació indirecta, en que les relacions de desequilibri de lligament permetien la selecció d'un determinat subgrup de polimorfismes com a marcadors de molts altres. Aquesta mena d'estudis van permetre reduir molt els costos de genotipat i es podia prescindir d'hipòtesis inicials sobre la funcionalitat dels SNPs a testar. Paral·lelament, s'avançava en el coneixement de nous mecanismes moleculars en que variants localitzades en qualsevol lloc del genoma podien esdevenir candidates a ser les variants funcionals responsables d'una associació. I, també, degut al *boom* tecnològic de les eines de genotipat, es disparava el nombre de SNPs i de mostres a genotipar a preus assequibles.

Aquests fets van possibilitar la realització d'estudis d'associació a més gran escala, en els quals es cobria la major part de la variabilitat comuna d'una regió genòmica. N'és un clar exemple el treball presentat en aquesta tesi (Capítol 1 de

Resultats); en el qual es va optimitzar la cobertura de la variabilitat genètica en la regió del gen *LRP5*. La selecció de la regió analitzada continuava estant limitada per l'elecció d'un gen candidat (*LRP5*), però els SNPs a testar ja no es limitaven als tradicionals canvis d'aminoàcid, sinó que de manera directa o indirecta, és testaven més variants, sense cap hipòtesi inicial respecte la seva funcionalitat. La mateixa estratègia va ser utilitzada en l'estudi d'associació amb *OPG* (Jurado i col., 2010), dut a terme en el nostre grup de recerca (veure l'Annex), i per aprofundir en l'estudi de gens candidats com ara *VDR* (Fang i col., 2005), entre d'altres.

És important destacar que, tot i el gran pas que suposen les estratègies d'associació indirecta, aquestes tenen certes limitacions. En primer lloc, les freqüències al·lèliques dels polimorfismes estudiats són claus per assegurar l'èxit del disseny experimental. Les eines de selecció dels SNPs marcadors basades en el LD utilitzen la informació de les variants comunes (MAF>5%), per tant, aquests no seran capaços de capturar satisfactòriament la informació de les variants més minoritàries. Aquest fet no semblava molt limitant inicialment, ja que s'esperava que les variants responsables de les malalties comunes fossin també comunes (*common disease/common variant hypothesis*), però com veurem més endavant, pot haver estat determinant en el no assoliment de les expectatives inicialment generades per aquesta mena d'estudis. No volem deixar de mencionar tampoc, l'existència de treballs crítics amb en els mètodes d'avaluació dels patrons de LD generats per HapMap i de l'aplicabilitat dels tagSNPs que se'n deriven en d'altres poblacions, sobre els quals es basen molts dels treballs d'associació indirecta [revisat per Guo i col·laboradors (2009)].

En segon lloc, la possibilitat d'analitzar l'associació d'un elevat nombre de polimorfismes comporta un replantejament dels llindars estadístics de significació usats per avaluar la seva validesa. La incidència de falsos positius entre els resultats és proporcional al nombre de tests realitzats i per tant, els criteris estadístics de significació tradicionals s'han de veure modificats quan es realitzen comparacions múltiples. La manera concreta de fer-ho ha engegat un intens debat entre els experts per tal de trobar criteris ni massa laxos ni massa conservatius. D'entre les correccions de tests múltiples més habituals destaquen la de Bonferroni, les permutacions de Westfall i Young o els càlculs de la taxa de falsos descobriments

(*False Discovery Rate*) de Benjamini i Horcheberg (ordenats de major a menor severitat, de menys a més permissius amb els falsos positius). En el cas concret del treball presentat per *LRP5*, és va optar per utilitzar el càlcul del *False Discovery Rate*, i es va emfatitzar en la necessitat d'acompanyar les associacions descrites amb estudis de replicació en cohorts independents i/o amb estudis que demostrin la funcionalitat de les associacions trobades (veure més endavant en la discussió, replicació i funcionalitat).

Tal i com ja s'ha comentat en la discussió de l'article presentat en el primer capítol de Resultats, tot i no considerar-se un dels gens candidats clàssics per a l'osteoporosi, *LRP5* ha esdevingut en els últims anys un dels protagonistes en la recerca dels factors genètics implicats en aquesta malaltia. Per al seu estudi s'ha emprat des d'anàlisis d'associació de variants individuals i estratègies d'associació indirecta que maximitzen la seva cobertura genètica fins als GWAs, com es discutirà en el següent apartat; en tots els casos s'ha obtingut resultats positius per a aquest gen i així es demostra la seva importància en la determinació dels fenotips osteoporòtics.

1.4. L'era dels GWAs

El desenvolupament continuat de noves tecnologies de genotipat d'alt rendiment (*throughput*) i el treball continu d'identificació i catalogació dels SNPs, així com els esforços per establir les seves relacions de desequilibri de lligament, han conduït al nivell màxim actual dels estudis d'associació indirecta: els *Genome-Wide Association Analysis* (GWAs, anàlisis d'associació de cobertura genòmica). La major avantatge d'aquesta aproximació és que tota (o gairebé tota) la variabilitat genètica comuna pot ser analitzada respecte el fenotip en qüestió, sense necessitat d'identificar ni regions genòmiques ni gens candidats. Aquest fet ha obert la porta a la identificació de variants en gens o posicions intergenòmiques que, donat el coneixement actual de la biologia òssia, mai haguessin estat considerades *a priori* com a candidates.

Actualment, hi ha publicats més d'una desena de GWAs de fenotips osteoporòtics [revisat per Li i col. (2010)], veure taula 3 de la introducció). El primer GWAs per a factors genètics comuns i el risc d'osteoporosi va ser publicat per Kiel i

col·laboradors el 2007, on es varen analitzar més de 100.000 marcadors en 1.141 individus. Es varen trobar 40 SNPs potencialment associats amb diferents fenotips ossis, però cap de les associacions superava els tests de significació a nivell de *genome-wide*. Com ja hem discutit a l'apartat anterior, la multiplicitat de tests que es realitzen en aquesta mena d'estudi comporta un replantejament dels criteris de significació estadística, que en el cas dels GWAs han de ser encara més restrictius que en el cas dels estudis *gene-wide*, ja que el nombre de comparacions és molt superior (valors de p de fins a 10^{-7} per tal d'evitar una elevada presència de falsos positius).

En els dos següents estudis de GWA el nombre d'SNPs genotipats ja superava els 300.000 i es varen dur a terme en cohorts molt més grans, fet que es va reflectir en la significació a nivell de *genome-wide* dels resultats obtinguts (Richards i col., 2008; Styrkarsdottir i col., 2008). En poc temps varen aparèixer més estudis d'aquest tipus (Cho i col., 2009; Guo i col., 2009; Liu i col., 2009; Timpson i col., 2009; Xiong i col., 2009), realitzats en diverses poblacions europees o no (Taula 3 de la introducció).

De manera molt resumida, els GWAs han validat o han proporcionat fortes evidències de la implicació en osteoporosis dels gens següents: *LRP5*, *ESR1*, *SOST*, *OPG*, *RANKL*, *RANK* i *SP7* [extensament revisat per Li i col. (2010)]. Tot i així, la majoria de les associacions que han superat els estrictes llindars de significació *genome-wide* corresponen a nous gens, com ara *FAM3C* (membre C de la família amb similaritat de seqüència 3) o *MARK3* (*MAP/microtubule affinity-regulating kinase 3*) entre d'altres (Styrkarsdottir i col., 2008 i Cho i col., 2009). Aquest gens no tenen actualment una connexió coneguda amb l'os i, una vegada hagin estat validats, el seu descobriment pot desemmascarar nous mecanismes biològics importants per al metabolisme ossi.

Amb els resultats d'aquests treballs a la mà, juntament amb els de les meta-anàlisis de GWAs (Rivadeneira i col., 2009), ens trobem en el mateix escenari que s'ha observat amb d'altres malalties complexes: s'han trobat les associacions en els gens candidats clàssics que ja eren conegudes i algunes associacions en gens nous, però conjuntament aquestes comporten efectes fenotípics de molt petita magnitud (veure apartat 4 de la Discussió).

Com ja s'ha comentat anteriorment, la possibilitat de realitzar aquesta mena d'estudis a nivell de *genome-wide* implica la utilització d'eines d'associació indirecta. Aquesta gran avantatge, a nivell d'estalvi en termes de genotipat i minimització de la obtenció d'informacions redundants, té per altra banda una gran limitació. Quan es troba una variant associada, no podem saber *a priori* si es tracta de la variant causal o bé del marcador de la variant responsable no identificada. Aquest fet afecta igualment als estudis a petita i mitjana escala, i només podem intentar minimitzar-lo quan els polimorfismes a genotipar es seleccionen en base a coneixements de la seva funcionalitat biològica (com per exemple els SNPs avaluats en el tercer treball presentat en aquesta tesi: *MTHFR*, *LRP6* i *LCT*). Com es discutirà en apartats posteriors, la identificació de les variants (o fins i tot gens) responsables de les associacions trobades per mitjà dels GWAs pot ser una tasca realment complicada, ja que moltes vegades és extremadament difícil formular hipòtesis de funcionalitat. Per altra banda, d'aquí neix el seu principal avantatge: en no partir de cap hipòtesi inicial de funcionalitat, s'obre la porta al descobriment de nous gens i vies mai implicats en el metabolisme ossi.

El resultat final dels GWAs és la identificació de regions cromosòmiques associades que, depenent dels patrons de desequilibri de lligament entre marcadors, poden cobrir varies kilobases i contenir diversos, un o cap gens candidats. En el cas de localitzar-s'hi un gen candidat, un estudi més profund podrà ésser completat segurament per mitjà d'estudis d'associació a nivell de gen o d'una variant polimòrfica en concret.

1.5. El paper de les CNVs en l'osteoporosi

L'interès per avaluar la implicació d'altres tipus de variabilitat genètica, a més dels SNPs, també s'ha traslladat en els últims anys al camp de l'osteoporosi. Les variants en el nombre de còpia (CNVs) en el DNA són també una important font de variabilitat genètica entre individus. Gràcies a les millores en la identificació i en les tècniques de genotipat basades en *arrays*, s'ha pogut descriure la implicació de les CNVs en diversos fenotips humans complexos, com són l'autisme (Sebat i col., 2007 i Glessner i Hakonarson 2009) o la progressió de la infecció per VIH (Gonzalez i col.,

2005). Actualment, dos treballs publicats per Yang i col·laboradors i Deng i col·laboradors (Yang i col., 2008 i Deng i col., 2010), han estudiat l'associació de CNVs amb fenotips osteoporòtics a nivell *genome-wide* en poblacions xinesa i caucàsica, respectivament. Dos nous gens candidats per l'osteoporosi han estat identificats mitjançant aquesta estratègia: *UGT2B17*, que codifica per a un enzim catabolitzador de les hormones esteroïdees, i *VPS13B (COH1)*, mutacions en el qual causen la Síndrome de Cohen que es caracteritza en part per anomalies esquelètiques (OMIM#216550).

S'ha proposat que els efectes funcionals de les CNVs són via l'alteració de la dosi d'expressió gènica, ja que s'ha vist que globalment les CNVs són responsables del 18% de la variació en l'expressió gènica (Stranger i col., 2007).

A mesura que la seva identificació i genotipat ha anat optimitzant-se tecnològicament, el seu ús en estudis d'associació s'ha pogut generalitzar. De manera similar als SNPs, han aparegut iniciatives com la Base de Dades de Variants Genòmiques (*Database of Genomic Variants*, <http://projects.tcag.ca/variation>) que pretenen proporcionar un resum exhaustiu de la variació estructural del genoma identificada fins al moment en poblacions control sanes. Una descripció més detallada de la seva estructura i contingut pot trobar-se en el treball publicat per Zhang i col·laboradors (Zhang i col., 2006).

2. La força de la mida mostral: les meta-anàlisis

Un altre mètode àmpliament usat en els estudis d'associació són les meta-anàlisis. Aquestes han posat de manifest la necessitat de crear consorcis internacionals d'investigadors per facilitar l'intercanvi d'informació i l'estandardització en la definició i recol·lecció del fenotip i les covariables. La seva importància rau en la necessitat de generar grans cohorts amb poder suficient per detectar variants amb efectes petits, i a la vegada, propiciar l'entorn adequat per facilitar els estudis de replicació.

La principal limitació de les meta-anàlisis, i que afecta principalment als estudis retrospectius, és la presència de biaixos en les dades emprades. Aquests poden ser deguts sobretot, al biaix de publicació (publicació de resultats positius, no

inclusió d'estudis no publicats en anglès, fonts d'informació ...) (revisat per Kavvoura i Ioannidis, 2008).

En aquest sentit, el nostre grup d'investigació de les bases genètiques de l'osteoporosi, participa en el consorci internacional GENOMOS creat l'any 2003 per diversos grups europeus per coordinar esforços en l'estudi de l'osteoporosi. Actualment, aquest consorci es nodreix de la col·laboració de 18 grups internacionals que conjuntament sumen una cohort de més de 37.000 individus amb informació fenotípica rellevant per a l'estudi de l'osteoporosi (dades extretes de l'última meta-anàlisi d'SNPs individuals publicada per GENOMOS, (van Meurs i col., 2008). El treball presentat en aquesta tesi sobre dos SNPs *missense* en *LRP5* i un en *LRP6* (Capítol 2 de Resultats) és un clar exemple de la mena d'estudis realitzats per aquest consorci. Sota el mateix disseny experimental s'han analitzat variants funcionals en diversos gens clàssics per l'osteoporosi, com són el *ESR1*, *COL1A1*, *VDR*, *TGF β -1*, *LRP5* i *LRP6* (molt breument resumits en la Taula 4 de la Introducció), tots ells amb la participació de la nostra cohort BARCOS.

En aquesta tesi s'han presentat dos treballs sobre la implicació d'*LRP5* en la variació dels valors de DMO o la incidència de fractura osteoporòtica emprant: un estudi a nivell de *gene-wide* en la cohort BARCOS (Capítol 1 de Resultats) i un altre on s'han estudiat dos SNPs de canvi d'aminoàcid a nivell de meta-anàlisis on les dades de BARCOS s'integren dins del projecte GENOMOS (Capítol 2 de Resultats). Hem posat de manifest que ambdues estratègies poden donar resultats diferents, en la cohort BARCOS individualment varem ser capaços d'identificar l'associació de l'SNP rs312009, una variant que no es va incloure en la meta-anàlisi de GENOMOS ja que no es tractava d'una variant no-sinònima amb alta probabilitat *a priori* de ser funcional. Per altra banda, GENOMOS va identificar associacions significatives per als SNPs p.V667M i p.A1330V amb uns efectes de magnitud molt petita; i possiblement per aquesta raó la cohort BARCOS no va tenir prou potència estadística per identificar-los (veure discussió del treball presentat en el Capítol 1 de Resultats).

El projecte GENOMOS com a tal, fou rellevat l'any 2008 pel projecte GEFOS (*Genetic Factors for Osteoporosis Consortium*) en el marc del 7è Programa Marc Europeu (*European Union Seventh Framework Package*). Es poden trobar més

detalls sobre el nou projecte GEFOS a la seva pàgina web: <http://www.gefos.org>. GEFOS és un clar exemple de l'impacte dels estudis d'associació *genome-wide* traslladat al nivell de les meta-anàlisis, ja que s'ha passat de l'estudi de variants individuals a la cobertura global del genoma [revisat per Zeggini i Ioannidis (2009)]. Aquest fet ha modificat el paper de les cohorts participants en GENOMOS/GEFOS, ja que un nombre menor de cohorts de gran mida són les que realitzen un GWA inicial, mentre que la resta de cohorts menors passen a centrar-se en estudis de replicació i/o confirmació funcional de les variants associades identificades en els primers.

Recentment GEFOS ha publicat la primera meta-anàlisi de cinc GWAs de DMO lumbar i femoral en 19.195 individus de descendència nord-europea (Rivadeneira i col., 2009). Aquesta meta-anàlisi ha confirmat 8 loci prèviament implicats amb la DMO a nivell de significació *genome-wide* (*LRP5*, *ESR1*, *OPG*, *RANK*, *RANKL*, *SP7*, *ZBTB40*, and *FOXC2*), i ha identificat 12 loci nous (mostrats a la Figura 22).

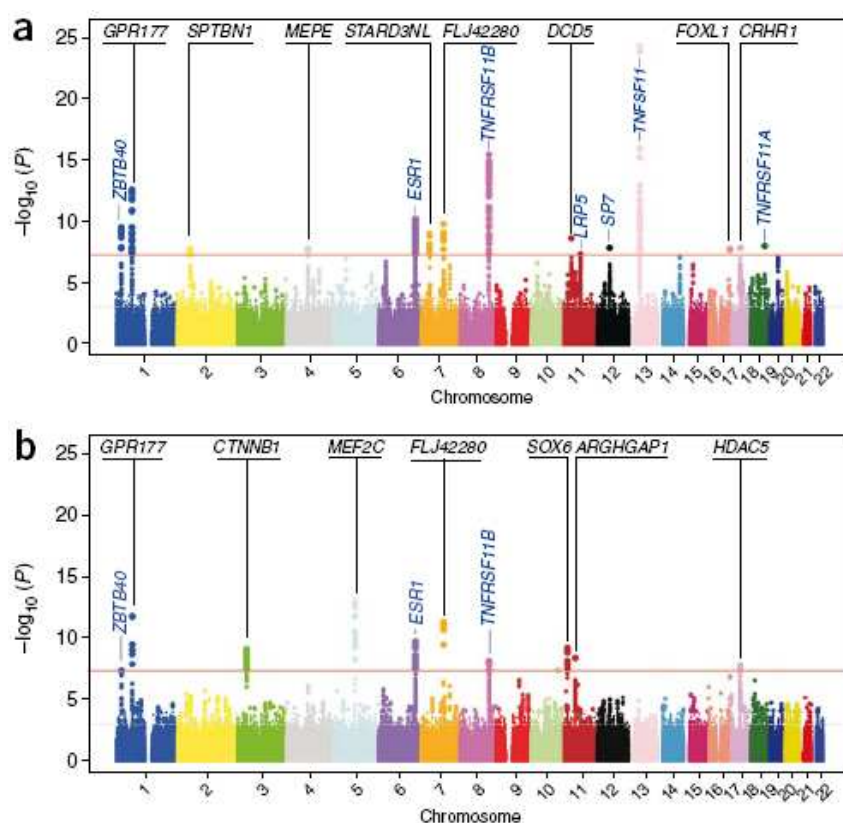


Figura 22. Manhattan plots resultants de la meta-anàlisi de GWAs de Rivadeneira i col. (2009). Els nous loci associats es mostren subratllats, i les associacions prèviament descrites en blau. En el panell (a) trobem els loci associats a DMO lumbar, i el (b) per DMO femoral.

De manera similar a les altres aproximacions experimentals comentades anteriorment, les meta-anàlisis de GWAs també es poden veure afectades per la presència de falsos positius generats a causa de la multiplicitat de tests realitzats o per la presència d'estratificació poblacional, tot i que existeixen mètodes per minimitzar-ne l'impacte. Igualment, cal tenir presents les limitacions respecte la mida dels efectes identificables per una potència estadística donada (al voltant del 0,2% de la variància de les DMOs en aquest cas concret), la impossibilitat d'avaluar les interaccions gen-gen o gen-ambient o la no captura de variants rares en els panells d'SNPs utilitzats fins al moment.

Però augmentar la mida mostral no sempre és la solució. En principi, per al mateix model experimental, com més gran sigui la cohort, millor. Però també cal considerar altres opcions. Una d'elles és la creació de cohorts amb membres seleccionats de tal manera que se'n minimitzi al màxim l'heterogeneïtat genètica, com ara per selecció de fenotips extrems o per tractar-se de poblacions aïllades. Per exemple, Sims i col·laboradors varen confirmar l'existència de dues associacions conegudes de polimorfismes en *LRP5* amb una cohort de només 320 dones postmenopàusiques seleccionades per presentar valors extrems de DMO de maluc (Sims i col., 2008).

3. La replicació: Una solució o més preguntes?

Durant la última dècada el nombre d'estudis d'associació per als fenotips osteoporòtics ha augmentat ràpidament. S'han trobat desenes de gens i polimorfismes implicats, però moltes vegades els resultats per una mateixa variant al·lèlica són discordants i molt freqüentment els resultats no es repliquen. La replicació, és a dir, traslladar un resultat d'una associació trobat en certa cohort a una nova població d'estudi independent, ha estat postulada per molts com la clau per reafirmar associacions reals i identificar els resultats espuris. Però de nou, la replicació afronta diversos dels problemes dels estudis d'associació en general, que posen de manifest que la no replicació de resultats no sempre indica que la associació inicialment trobada no sigui real. Per exemple, cal tenir molt en consideració les mides mostrals i la mida de l'efecte que cerquem, ja que aquests

afecten directament la potència estadística de trobar l'associació (**figura 23**). En osteoporosi s'ha vist que existeixen diferències entre els factors genètics que determinen la DMO segons el lloc esquelètic estudiat, segons l'ètnia, el sexe o l'edat dels individus estudiats, de l'estat menopàusic si es tracta de dones, etc. Tots aquests factors han de estar controlats al màxim a l'hora de realitzar un estudi de replicació, ja que sinó es fàcil trobar-se amb resultats inconsistents entre ells. A més, l'ús de diferents fenotips a l'hora d'avaluar les associacions per a un gen donat suposa un altre factors de complicació [revisat per Shen i col. (2005)].

En els treballs presentats en aquesta memòria tots aquests aspectes han estat avaluats i minimitzats al màxim dins de les nostres possibilitats. Els nostres resultats sempre han anat acompanyats d'**estimacions de la potència estadística** de l'estudi en qüestió, que ens serveixen per delimitar quina mena d'efectes podem arribar a determinar i la fiabilitat dels resultats obtinguts. La potència estadística, a part de dependre de la mida mostral, també està afectada per molts d'altres factors com són: la heterogeneïtat genètica, el grau i extensió del desequilibri de lligament entre els marcadors testats i les variants causals, o les diferències de freqüència al·lèlica entre elles [revisat per Shen i col., (2005)]. De fet, les eines per fer estimacions de la potència estadística es basen en valors de freqüències al·lèliques, valors de desequilibri de lligament entre SNPs o de mides d'efectes esperats que realment no coneixem *a priori*. L'únic que podem millorar en aquest sentit, és la mida de la cohort BARCOS, prosseguint en el reclutament i genotipat de més participants i amb la participació en consorcis de grans projectes col·laboratius.

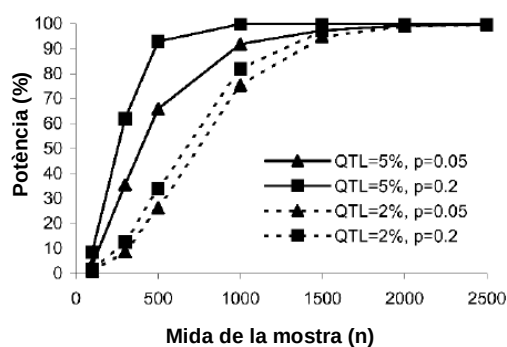


Figura 23. Potència estadística per trets quantitius en una mostra d'individus no emparentats. Assumint que els QTLs són responsables del 2% o 5% de la variabilitat fenotípica i que les variants responsables tenen una freqüència al·lèlica (p) de 0.05 o 0.2. La potència s'ha estimat considerant la situació ideal de que el marcador estudiat és el responsable del QTL per ell mateix i que el QTL presenta un patró d'herència additiu. El nivell de significació es va establir a $\alpha=0.001$ (adaptat de Shen i col., 2005).

En el nostre cas d'estudi d'associació basat en població, és indispensable un control òptim per la presència d'**estratificació poblacional**, ja que aquesta pot conduir a resultats espuris tant per la presència de falsos positius com de falsos negatius. Les estratègies d'associació basades en famílies (p.e. TDT) permeten evitar els efectes de l'estratificació poblacional, però en el nostre cas, d'estudi poblacional d'individus no emparentats, l'únic que podem fer és aplicar mètodes per detectar-la i controlar-la. En aquest sentit, la cohort BARCOS es recluta d'una àrea geogràfica molt concreta i només s'inclouen dones d'origen espanyol. Igualment, la presència d'estratificació es descarta mitjançant la metodologia adequada utilitzant informació genotípica de marcadors no lligats (veure el Capítol 1 de Resultats).

Pel que fa al **fenotip** estudiat, sempre hem treballat amb els valors de DMO, amb un clar component hereditari i àmpliament utilitzats en aquest camp. El fenotip qualitatiu fractura/no fractura també és dels més utilitzats, ja que l'obtenció de models predictius sobre aquest fenotip tindran un gran impacte en la pràctica clínica. Altres fenotips com paràmetres de mida i resistència òssia, paràmetres de formació o recanvi ossi, etc... també són importants predictors de la fractura osteoporòtica. El fenotip qualitatiu osteoporosi/no osteoporosi no l'hem aplicat en cap dels estudis presentats, ja que considerem que convertir els valors de DMO quantitativs a una variable qualitativa comporta certa pèrdua d'informació que no ens podem permetre amb el nostre mida mostral. Cada fenotip pot tenir determinants genètics únics, per tant, avaluar un mateix gen o variant candidata per a diferents fenotips pot donar lloc a diferents resultats. El treball presentat en aquesta tesi, en el capítol 4 de resultats, exemplifica la importància de la elecció del fenotip a estudiar. En aquest cas concret, es varen estudiar tres polimorfismes en relació a la DMO i amb fractura. Degut al nombre relativament reduït de dones amb fractures osteoporòtiques en la cohort BARCOS, inicialment es va optar per l'agrupació dels diferents tipus de fractura osteoporòtica en una sola categoria; quan aquest grup es va avaluar contra el grup de dones no fracturades no es va obtenir cap associació significativa, tot i que es va observar una lleugera tendència per al SNP d'*MTHFR*. Posteriorment, es va realitzar l'estudi pel subgrup de dones amb fractura vetebral, que tot i reduir a la meitat el nombre de casos va posar de manifest una associació significativa per a aquest polimorfisme.

Els estudis d'associació han d'incloure un rigorós **control de qualitat**, tant per detectar errors en el genotipat com en el fenotipat, ja que aquests afecten notòriament a la potència estadística [revisat per Shen i col. (2005)].

A part dels factors relacionats amb el disseny de l'estudi comentats anteriorment, els resultats inconsistents entre diferents estudis poden ser deguts a diferències biològiques o ambientals reals, que poden incloure elements com l'epístasi, la interacció de gens amb l'ambient, la penetrància incompleta, l'expressivitat variable o la pleiotropia. Aquest factors poden ser diferents entre poblacions per a un mateix tret o malaltia estudiada, donant lloc a la no replicació dels resultats. En aquest sentit, cal centrar esforços per maximitzar la homogeneïtat de les mostres, en el reclutament individus del mateix grup ètnic i controlant rigorosament els factors ambientals, per maximitzar les possibilitats d'èxit. Una de les grans limitacions, clarament posada de manifest en la publicació presentada en el Capítol 4 de Resultats, és la manca de certs valors de **covariables tant ambientals com nutricionals** que afecta actualment a la cohort BARCOS. En aquest sentit treballem per poder-les incorporar en futurs estudis, ja que la implicació d'interaccions entre gens i variables ambientals o nutricionals s'està fent molt evident en l'actualitat.

Idealment, un bon estudi d'associació ha d'incloure ja des del seu plantejament inicial, un primer estudi d'associació en una cohort (descobriments de l'associació) i en una segona fase, la realització de l'estudi de replicació en una segona cohort independent (replicació de l'associació). Actualment alguns treballs ja segueixen aquesta planificació, sobretot els que es realitzen a nivell de *genome-wide* (per exemple Koller i col., 2010 entre d'altres).

4. "The missing heritability"

4.1. El replantejament de la hipòtesi variant comuna / malaltia comuna

Actualment ja disposem dels resultats de centenars d'estudis GWA per diverses malalties complexes, aquests han conduït a la identificació de variants genètiques significatives estadísticament (Manolio i col., 2008). Malgrat l'èxit en la

identificació d'aquestes variables, totes elles col·lectivament expliquen només una petita fracció de la component genètica d'aquestes malalties. Aquesta mena d'estudis han donat lloc a associacions amb típicament 30-50 variants de seqüència implicades, cadascuna d'elles amb una penetrància relativament baixa, amb riscos relatius al voltant del 1,2 (**figura 24**). Excepte alguns casos minoritaris, més del 90-95% del component genètic de la malaltia ha quedat sense resposta després d'extensius GWAs. Aquest fet suggereix que les variacions comunes individuals no expliquen la majoria de la prevalença de les malalties comunes i llença la pregunta sobre quina és la natura dels factors genètics restants que contribueixen a la malaltia, designada amb el terme anglès “*missing heritability*” (Maher 2008).

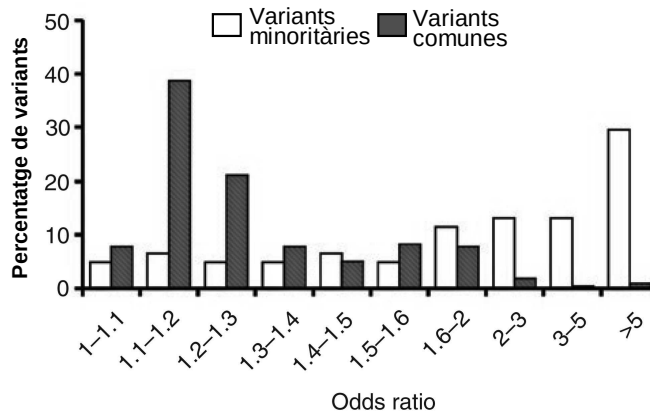


Figure 24. Distribució dels efectes (en termes d'odds ratios) per les variants comunes i les minoritàries. L'anàlisi inclou informació de 217 variants comunes i 61 variants minoritàries (extret de Bodmer i Bonilla 2008).

L'osteoporosi és un clar exemple d'aquesta situació, perfectament il·lustrada pels resultats de la meta-anàlisi presentada per Rivadeneria i col·laboradors per als fenotips osteoporòtics (Rivadeneira i col., 2009). Les 15 millors associacions per DMO lumbar i les 10 millors per a DMO femoral només expliquen un 2.9% i un 1.9% de les variàncies en DMO d'aquests dos llocs esquelètics respectivament.

Els estudis de GWAs estan fonamentats en la hipòtesi de la *common disease/common variant* (CD/CV), que postula que la variabilitat en les malalties humanes complexes es atribuïble en gran part a les variants al·lèliques presents amb freqüències superiors al 1-5% en la població humana (Reich i Lander 2001). Però les evidències experimentals que acabem de comentar, han despertat un nou interès per les variants minoritàries (o rares, *rare variants*), que tenen freqüències a mig camí entre les de les mutacions i les dels polimorfismes comuns (MAF entre 0,1

i 2-3%), i que escapen dels dissenys actuals dels GWAs. La “*rare variant hypothesis*” (CD/RV) proposa que una proporció significativa de la susceptibilitat a una malaltia humana comuna heretable és deguda al sumatori dels efectes d’una sèrie de variants de baixa freqüència que actuen independentment i que es troben localitzades en diferents gens, cadascuna d’elles conferint uns efectes moderats però detectables en l’augment del risc relatiu (Figura 25). Aquestes variants seran probablement específiques de població a causa de l’efecte fundador resultants de la deriva genètica. La revisió de Bodmer i Bonilla (2008) proposa que cada individu pot tenir 10 o més variants minoritàries. En general, quan parlem de variants rares ens referim a SNPs, però poc a poc, les variants estructurals minoritàries també van entrant en joc.

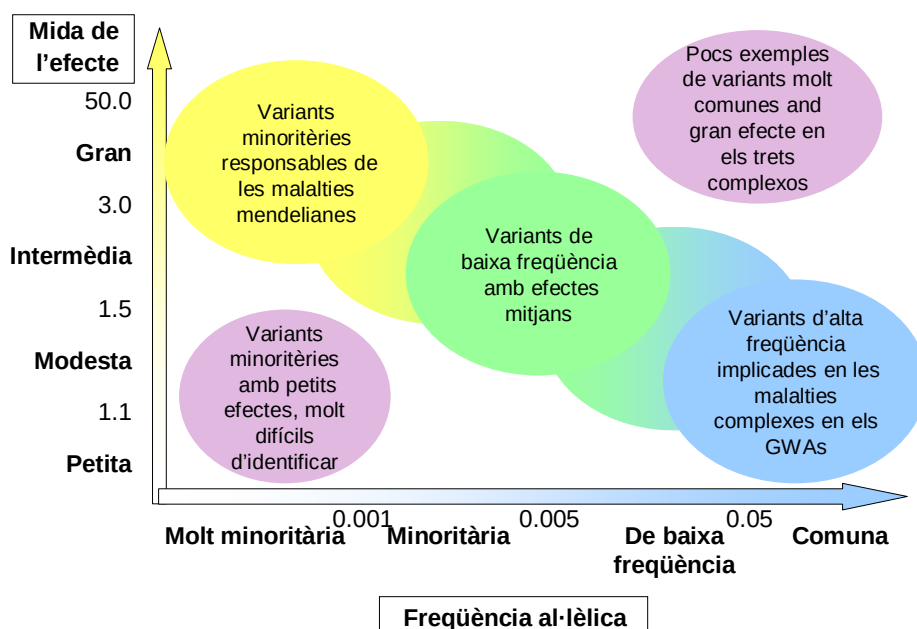


Figura 25. Viabilitat de la indentificació de variants genètiques segons la seva freqüència al·lèlica i la mida del seu efecte genètic (*odds ratio*) (adaptat de Manolio i col., 2009).

Els principals reptes que suposa la incorporació de les variants minoritàries en els estudis d’associació són bàsicament: i) la selecció dels grups d’estudi poblacionals adequats, ii) la necessitat de reseqüenciació extensiva d’un gran nombre de gens i regions gèniques, iii) necessitat de genotipar major nombre de mostres, iv) noves i millors metodologies per a l’avaluació de les conseqüències funcionals de les variants implicades (revisió Bodmer i Bonilla 2008).

Per a les variants rares s'espera trobar valors d'OR superiors, al voltant de 2. Si es així, el sumatori dels efectes de diverses variables podrà contribuir significativament a la risc atribuïble a la població (PAR) i tenir una penetrància suficientment elevada com per justificar intervencions preventives.

També cal considerar la interacció entre els estudis de les variants comunes i les minoritàries, ja que per una banda, les variants comunes poden actuar com a modificadores de les minoritàries, i les associacions de variants comunes trobades en determinats gens, els converteixen en els gens candidats per estudiar-hi la variabilitat minoritària.

Actualment, els *arrays* de genotipat de polimorfismes de nova generació ja van incorporant SNPs de menor MAF (1-2%), com per exemple la sèrie de *microarrays* de genotipat Omni d'Illumina.

Certament, ambdues hipòtesis, CD/CV i CD/RV, són plausibles i múltiples variants comunes de baixa penetrància juntament amb variants minoritàries d'efecte moderat o elevat contribueixen en l'expressió i freqüència de les malalties comunes humanes. La revisió d'Schork i col·laboradors (2009) recull evidències experimentals d'ambdues hipòtesis.

Per tal de facilitar la cerca de variants rares en múltiples gens, o fins i tot per tot el genoma, usant les tecnologies de seqüenciació de nova generació, s'ha iniciat el projecte dels "1000 Genomes" (www.1000genomes.org). Aquest projecte pretén caracteritzar les variants de seqüència de 1000 individus per tal de proporcionar una base per a futurs estudis de seqüenciació orientats a l'estudi de les malalties, així com desenvolupar protocols apropiats i eines bioinformàtiques.

4.2. El paper de les interaccions

Part d'aquesta *missing heritability* pot ser deguda al paper de les variants minoritàries, però igualment s'han de considerar altres factors. Entre ells, la implicació de variants estructurals comunes o minoritàries, la presència d'interaccions entre gens i ambient, interaccions epistàtiques entre gens i/o la presència de fenòmens epigenètics.

En general, en els estudis d'associació publicats fins al moment no s'ha valorat adequadament la implicació dels factors ambientals, o no han tingut suficient poder estadístic per plantejar-se l'estudi d'interaccions entre gens. Aquest fet posa novament de manifest la necessitat de realitzar grans estudis, normalment multicèntrics, per recollir els nombre suficient de participants i realitzar estudis amb suficient potència estadística. En aquests, la necessitat d'un fenotipat extremadament acurat serà evident, així com l'estandardització dels paràmetres d'exposició a factors ambientals a recopilar. Com es recull en la revisió de Manolio i col·laboradors (2009), els GWAs futurs hauran de fer l'esforç d'incorporar investigacions sobre les interaccions entre gens i les interaccions gen-ambient.

Estadísticament, l'estudi de les interaccions suposa nous reptes, ja que les tècniques de regressió estàndard emprades normalment estan molt limitades per la mida mostral i per altra banda, per avaluar interaccions, cal crear subgrups de dades cada vegada més petits. En la revisió de Montana (2006) es descriuen alguns dels mètodes alternatius.

Com es recull en la revisió de Moore i Williams (2009), la credibilitat d'un resultat estadístic rau més en la seva interpretació biològica i en l'evidència experimental que no pas en la significació del resultat en si. En aquest sentit, diversos autors han demostrat que utilitzar el coneixement biològic a l'hora de guiar un estudi d'interaccions dona lloc a molts més resultats amb sentit. Per tant, serà més interessant avaluar les interaccions entre gens d'una mateixa via metabòlica que no pas avaluar interaccions sense cap hipòtesi prèvia. Noves aproximacions bioinformàtiques permeten incorporar el coneixement sobre la funció gènica, xarxes de gens (*gene networks*) o rutes metabòliques. I sempre recórrer a la biologia experimental per tal de validar els models genètics proposats.

En el camp de l'osteoporosi l'estudi de les interaccions encara és molt limitat [revisat per Li i col. (2010)]. Només a tall d'exemple citarem el treball de Fang i col·laboradors (2009), en el que s'explorava la interacció de variants de dos gens, *DBP* (proteïna d'unió a la vitamina D, *vitamin D binding protein*) i *VDR*, ambdós implicats en la via endocrina de la Vitamina D, en la determinació del risc a per osteoporosi. Tanmateix, també varen testar la interacció amb la ingesta de calci amb els polimorfismes de *DBP* en una cohort de més de 6000 participants. Els seus

resultats varen mostrar que l'efecte del gen DBP en el risc de fractura només estava present en combinació amb altre factors de risc per al metabolisme ossi, tant genètics com ambientals.

Avui en dia, aquesta mena d'estudis no són abordables en la cohort BARCOS, bàsicament per les limitacions en la mida mostral. Amb el continu reclutament de més dones i de més informació sobre variables ambientals o nutricionals, plantejem aquesta mena d'estudis per a un futur. Per exemple, es podrien replantejar els estudis d'associació per els SNPs funcionals estudiats en els gens *MTHFR* o *LCT*, en els que altres estudis han posat de manifest la rellevància d'interaccions amb factors no genètics (veure Capítol 4 de Resultats).

5. El per què de les associacions trobades: els estudis funcionals

A més de les evidències estadístiques d'associació entre una variant genòmica i el fenotip d'una malaltia (ja siguin variants comunes o minoritàries, identificades en un estudi inicial o després d'una replicació), és molt important avaluar el significat biològic de la variant en qüestió, ja sigui via mètodes computacionals, fent assaigs de laboratori o emprant sistemes model. Això és encara més important, si cap, quan l'estudi d'associació és indirecte, ja que la variant associada s'ha usat com a marcadora d'un grup més gran, d'entre les quals caldrà buscar la variant funcional.

Les variants polimòrfiques situades en les regions codificants han estat tradicionalment les més estudiades, ja que si aquestes donen lloc a un canvi d'aminoàcid en la seqüència proteica els efectes sobre la proteïna poden arribar a ser notables, però sempre més subtils que els de les mutacions de les malalties monogèniques. Les variants reguladores però, els hi van guanyant terreny, ja que treballs com els de Stranger i col·laboradors (2005 i 2007) posen de manifest que les variants reguladores en *cis* tenen una contribució més gran en la determinació del fenotip de les malalties complexes que les variants codificants. Les variants reguladores poden trobar-se en promotors, *enhancers*, LCRs (*locus control regions*) i regions d'unió a la matriu nuclear, tots ells elements clau en el control transcripcional. Actualment s'estan realitzant grans esforços en la catalogació de

tots els elements funcionals del genoma i així poder identificar la funcionalitat de variants fora de regions codificants, bàsicament en el marc del projecte ENCODE (Birney i col., 2007).

La caracterització funcional d'una variant genòmica representa un gran repte experimental. En primer lloc, moltes vegades no es pot formular una hipòtesi concreta de funcionalitat, i s'ha de recórrer a eines bioinformàtiques que prediguin l'afectació d'un lloc d'*splicing*, d'un *enhancer* o d'un lloc d'unió d'un factor de transcripció, per exemple (com és el cas de l'SNP rs312009, Capítol 1). Les eines experimentals per confirmar aquestes hipòtesis a vegades resulten poc sensibles per a la detecció dels efectes subtils que esperem per una variant associada, i sovint massa allunyades del context *in vivo* real. Pel que fa als estudis funcionals de variants localitzades en regions promotores, les tècniques més usades són el northern blot, RT-PCR, els assaigs de protecció a la RNAasa, patrons de digestió amb DNAase I, assaigs de retenció en gel i mutació selectiva i transfecció transitòria de construccions reporteres en cèl·lules en cultiu, entre d'altres. Algunes d'aquestes aproximacions s'han emprat durant el desenvolupament d'aquesta tesi per avaluar la possible funcionalitat d'un SNP de la regió promotora d'*LRP5* (Capítols 1 i 3 de Resultats).

Pel que fa a l'estudi de la funcionalitat de les variants estructurals (CNVs) els seus efectes estan més relacionats amb l'alteració de la dosi gènica, l'augment de la distància dels elements reguladors, la creació de nous elements reguladors, la interrupció gènica o l'afectació d'elements de regulació epigenètics.

Els estudis de funcionalitat són potencialment el següent pas després de la gran onada de GWAs que han aparegut en el camp de l'osteoporosi i moltes altres malalties complexes. A tall d'exemple, el treball de Xiong et al 2009 ja inclou un EMSA (Xiong i col., 2009).

De totes maneres, després de descriure una associació per un determinat polimorfisme i haver pogut demostrar experimentalment que aquest presenta un efecte funcional, mai podrem otorgar-li la total causalitat d'aquesta associació per ell mateix, ja que no descartem la implicació d'altres variants correlacionades o que es trobin en el seu mateix bloc haplotípic, que també contribueixin en la determinació del fenotip.

6. Aplicacions dels estudis d'associació i perspectives de futur

La identificació de variants genètiques de susceptibilitat a una malaltia complexa sempre ha tingut com a objectiu final la translació a la pràctica clínica, en els seus diferents aspectes esquematitzats a la figura 26.

En primer lloc, i a nivell bàsic, la identificació de variants associades permet el coneixement de noves vies biològicament implicades amb el fenotip en qüestió, en les quals es podran trobar noves dianes terapèutiques per potencialment desenvolupar nous tractaments. En són dos clars exemples la via de senyalització de Wnt/ β -catenina i la de RANKL/RANK/OPG, que en gran part gràcies als estudis d'associació, s'han dibuixat com a elements clau en la homeòstasi òssia. Per exemple, existeix un nou fàrmac, el *denosumab* (anticòs monoclonal humà contra RANKL), que s'ha vist que redueix significativament el risc de fractura en dones osteoporòtiques (Cummings i col., 2009); i se n'està desenvolupant un altre contra SOST, un dels components de la via de Wnt, que podria esdevenir un important agent anabòlic ja que ha donat resultats esperançadors en els estudis en animals (Li i col., 2009). És important destacar que d'aquesta manera es poden identificar també nous biomarcadors, que permetran la millora en la predicció i la monitorització de la progressió o la resposta al tractament de la malaltia. A vegades, els descobriments en epidemiologia genètica poden desemascarar l'efecte d'importants elements ambientals que contribueixen al fenotip, facilitant posteriors mesures preventives a nivell de salut pública. En aquest sentit, serà cada vegada més necessari incorporar la informació de noves covariables en els estudis de GWAs per avaluar a aquest nivell les interaccions.

En aquest sentit, i de manera molt primerenca, els nostres resultats fruit de l'estudi funcional de l'associació descrita per rs312009 (capítols 1 i 3 de Resultats) han identificat la regulació d'LRP5 per part de Runx2, que afegeix evidències experimentals de la interacció de dues vies de senyalització indispensables per a l'osteoblasts i la determinació de la DMO: la via de senyalització de Wnt i la de Runx2.

En segon lloc, una altra via de translació a la clínica és la que ha de permetre

la utilització del coneixement dels patrons individuals de predisposició a una malaltia per tal de desenvolupar aproximacions més personalitzades per al seu maneig. En aquest aspecte, la principal limitació és que, per la majoria dels trets complexes, les variants associades identificades fins al moment només expliquen una petita proporció de la variabilitat individual al risc de patir la malaltia. Per tant, per la gran majoria dels individus, el perfil genètic usant els marcadors coneguts actualment dóna informació molt limitada vers al risc de patir la malaltia. Com hem pogut veure en apartats anteriors, l'osteoporosi és un clar exponent d'aquesta situació, ja que tant els primers estudis en gens candidats com els GWAs actuals, han mostrat com les variants associades no tenen efectes majoritaris i que encara queda per determinar quines són les variants que conjuntament puguin arribar a predir el risc a la malaltia i tenir un valor diagnòstic real.

Una visió resumida de l'estat actual de la farmacogenòmica en el camp de l'osteoporosi es pot trobar en la revisió de Marini i Brandi (2009).

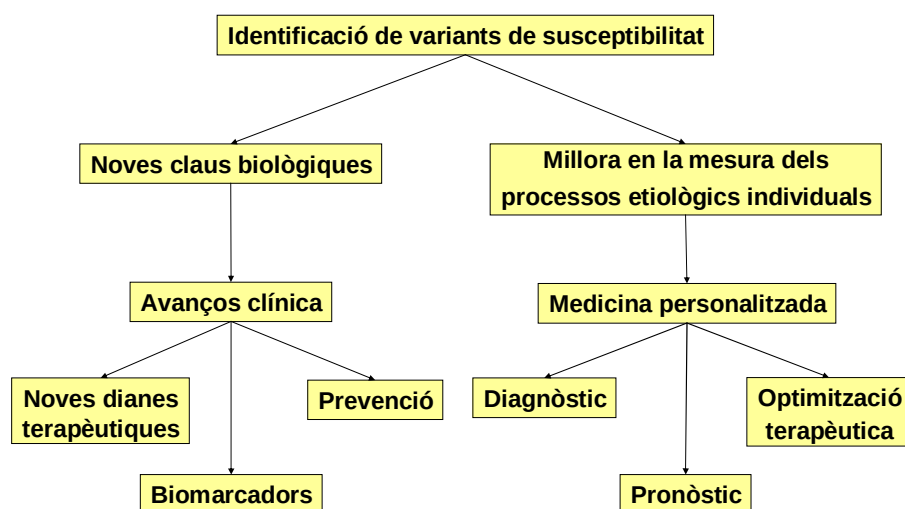


Figura 26. Translació clínica dels resultats trobats en els estudis d'associació (adaptat de McCarthy i col., 2008).

Per poder aplicar habitualment el perfil genètic individual a la pràctica clínica encara ens queda feina per fer. Cal millorar en la precisió en la predicció dels riscos

per mitjà de la identificació de variants de susceptibilitat addicionals; cal demostrar, per mitjà d'estudis prospectius, que aquest perfil genètic pot resultar en modificacions beneficioses de l'atenció mèdica i/o de la responsabilitat personal. A la vegada, cal establir un context regulatori adequat per a l'ús d'aquesta mena d'informació [revisat per McCarthy i col. (2008)].

La seqüenciació de nova generació, com a nou avanç tecnològic, sembla que marcarà també l'evolució de l'estudi de les bases genètiques de l'osteoporosi. Replatejarà en primer lloc, l'estudi de les malalties monogèniques òssies, descobrint-se noves mutacions i nous gens, que podran traslladar-se com a nous gens candidats en els estudis d'associació. Tanmateix, els projectes de reseqüenciació de més individus identificaran molta de la variabilitat genètica que pel fet ser més minoritària, encara no s'havia estudiat. Actualment, els *arrays* per al genotipat d'SNPs i CNVs ja plantegen la seva adaptació incloent cada vegada més variants de baixa freqüència al·lèlica i fins i tot, exclusives de certes poblacions.

Però ens estem trobant en una situació en la que les eines tecnològiques de genotipat i seqüenciació avancen més ràpidament que la nostra capacitat d'anàlisi i interpretació dels resultats. La gran quantitat d'informació disponible està forçant la necessitat de noves eines estadístiques per avaluar-ne la significació o per determinar interaccions, per exemple. Tanmateix, el nostre coneixement sobre els elements funcionals del genoma està tot just començant.

Pel futur dels estudis d'associació serà molt important el coneixement i la combinació de diverses disciplines, incloent l'epidemiologia genètica, l'estudi dels perfils d'expressió gènica, la proteòmica, etc... per complementar-se entre elles i arribar a aclarir els elements genètics determinants d'un tret complex com és l'osteoporosi. Quants i quins gens determinen la susceptibilitat a patir osteoporosi, la variabilitat en els valors de DMO o el risc de fratura continua èssent una pregunta oberta.