

L'impacte dels projectes post-genòmics en ciències de la salut

Roderic Guigó

L'any 2001, el president dels Estats Units, Bill Clinton, i el primer ministre del Regne Unit, Toni Blair, anunciaven conjuntament els primers esborranys de la seqüència del genoma humà, un projecte finançat majoritàriament pels National Institutes of Health (NIH) dels Estats Units. Aquests esborranys foren presentats com una fita en la història de la ciència que canviaria per sempre més la manera com enteníem la biologia humana i, en particular, la manera de lluitar contra les malalties. Hom passaria d'una medicina que «detecta i tracta» les malalties a una medicina que «preveu i evita» que les malalties es manifestin.

Aquelles expectatives aviat es demostraren exagerades. A final de la dècada dels noranta, hom acceptava que la seqüència del genoma tenia un valor limitat a l'hora de predir les malalties. La seqüència del genoma d'una espècie específica els trets biològics que defineixen uns organismes, en el cas dels éssers humans, el color de la pell, dels ulls, del cabell, l'alçada, el tipus sanguini, etcètera, i també, efectivament, la

predisposició a patir determinades malalties. Tanmateix, la manera com aquestes instruccions estan codificades en el genoma és molt més complexa d'allò que anticipàvem a principis del mil·lenni, i encara ens és majoritàriament desconeguda en entrar en la seva tercera dècada.

Per tal de maximitzar el valor mèdic de la seqüència del genoma humà, els NIH, en concret el National Human Genome Research Institute (NHGRI), va iniciar projectes de recerca amb dos objectius genèrics principals. El primer d'aquests objectius era caracteritzar la diversitat genòmica humana. Els esborranys publicats a principis dels anys 2000 corresponien a la seqüència d'un únic individu, però persones diferents tenim genomes diferents i es per això que exhibim trets biològics diferents, incloent-hi la predisposició diferent a patir determinades malalties. L'obtenció de la primera seqüència del genoma humà va involucrar centenars de científics, va durar més de 10 anys i va tenir un cost de 3000 milions de

dòlars. Amb la tecnologia disponible a principis dels anys 2000, obtenir la seqüència d'un segon genoma era impossible. Tanmateix, a finals dels anys 2000 es va produir una revolució tecnològica en el camp de la seqüència dels àcids nucleics (la matèria que conforma el genoma) i el cost i l'esforç necessari per obtenir seqüències genòmiques es va abaratir exponencialment. Era possible, per primer cop, obtenir la seqüència del genoma de molts individus.

Un projecte pioner en aquest sentit va ser «el projecte dels 1000 genomes», que es va desenvolupar entre el 2008 i el 2015, i que tenia com a objectiu la caracterització de la variabilitat genòmica de l'espècie humana. Aquest projecte, que no estava associat a l'estudi de cap malaltia en concret, va ser seguit pel International Cancer Genome Consortium (ICGC), el qual tenia com a objectiu obtenir la seqüència del genoma dels tumors en diferents tipus de càncer. Espanya hi va participar amb la seqüenciació del genoma de la leucèmia limfàtica

crònica. Aquests projectes van permetre identificar variants genètiques que ocorren preferentment en diferents tipus de tumors.

L'ICGC és només un exemple de molts altres projectes de seqüenciació genòmica que han identificat variants genètiques associades a malalties particulars, com ara la diabetis, el Parkinson o moltes malalties coronàries. Aquests estudis, que es coneixen com a estudis d'associació al llarg de tot el genoma (GWAS, *genome-wide association studies*), han posat de manifest que la base genètica de les malalties és molt més complexa del que s'esperava. La majoria de malalties més prevalents estan associades a centenars de variants genètiques diferents, cada una de les quals té, sovint, un efecte gairebé imperceptible. A finals de la dècada dels 2010, s'havien descrit gairebé 140.000 associacions entre variants genètiques i malalties. Aquesta informació, tanmateix, és útil per predir la predisposició de cadascú de nosaltres a patir determinades malalties i l'eventual resposta al tractament.

És per això que, a mida que el cost de la seqüenciació genòmica continua reduint-se, molts països han iniciat els anomenats Projectes Nacionals de Medicina de Precisió. L'objectiu d'aquests projectes és incloure la seqüència del genoma en la història clínica dels ciutadans, per tal que sigui utilitzada en la diagnosi

i tractament de les malalties. El Regne Unit és el país que es troba més avançat en aquest sentit. S'hi han seqüenciat més de 100.000 genomes de pacients de càncer o de malalties rares i l'objectiu és arribar a 500.000 genomes l'any 2025. El fet que moltes dades genòmiques siguin obtingudes en el marc dels sistemes nacionals de salut dificulta la recerca científica. Els estudis d'associació (GWAS) són més potents quan més gran és la cohort de pacients que s'estudia, però els sistemes de salut, per motius de confidencialitat, són molt reticents a distribuir les dades dels pacients, encara més fora de les fronteres dels seus països. És per això que han sorgit iniciatives per tal de fer compatible la confidencialitat de les dades dels pacients amb el seu ús en la recerca científica. La més important d'aquestes iniciatives és la Global Alliance for Genomics and Health (GA4GH). La GA4GH desenvolupa estàndards tècnics que permeten compartir dades mantenint-ne la confidencialitat. Una altra iniciativa important, en la qual el nostre país hi juga un paper molt rellevant, és el European Genome-phenome Archive (EGA), una base de dades sofisticada mantinguda conjuntament pel Centre de Regulació Genòmica a Barcelona i l'Institut Europeu de Bioinformàtica a Cambridge, en la qual es deposita tota la informació genòmica que s'utilitza en la recerca biomèdica.

La informació genòmica és particularment rellevant quan està

associada a informació fenotípica. La iniciativa més important en aquests moments és l'anomenat UK BioBank al Regne Unit. Aquest projecte ha reclutat 500.000 voluntaris entre els anys 2006 i 2010, en edats compreses entre els 40 i els 69 anys, i en farà un seguiment durant trenta anys: anàlisis de sang i d'orina, mesures de pes, pressió arterial, ritme cardíac, imatges de ressonància magnètica de la caròtida, ossos, cor, cervell i articulacions, informació sobre l'estil de vida, etcètera. Aquestes dades es creuen amb la informació en la seqüència del genoma per tal d'identificar com la informació genètica afecta el fenotip dels individus al llarg del temps. Es tracta d'una base de dades d'una importància excepcional i ha estat molt rellevant, per exemple, per identificar la base genètica de la resposta diferencial a la infecció amb SARS-CoV-2.

En paral·lel a la caracterització de la variabilitat genòmica en l'espècie humana i de com aquesta afecta els trets biològics dels individus, el segon gran objectiu de la recerca post-genòmica ha estat intentar esbrinar quins són els mecanismes biològics mitjançant el quals la seqüència del genoma determina aquests trets. Els estudis de GWAS són meres associacions estadístiques; simplement ens diuen que una determinada variant genètica està associada a un tret biològic, per exemple el color dels ulls, però no ens diuen la raó perquè

la variant fa que els ulls siguin d'un color o d'un altre. Els estudis encaminats a identificar quines regions són funcionals en el genoma i com funcionen s'anomenen precisament de Genòmica Funcional. Un dels principals projectes en aquesta àrea és l'anomenada *Encyclopedia of DNA Elements* (ENCODE), la qual, iniciada l'any 2003, ha tingut una durada de 20 anys. Les dades produïdes per ENCODE han estat crucials per identificar regions en el genoma que tenen rellevància en càncer, malalties autoimmunes i al·lèrgies, malalties neuronals i psiquiàtriques, cardiovasculars, etcètera. Un altre projecte de genòmica funcional molt important és el projecte *Genotype-Tissue Expression* (GTEx), en el qual s'ha caracteritzat com funciona el genoma en òrgans i teixits diferents,

mitjançant l'estudi de l'expressió dels gens.

En resum, des de la finalització del projecte original del genoma humà, ara fa més de vint anys, els avenços en la comprensió de la informació genòmica han estat extraordinaris. Hem descobert centenars de milers de variants genètiques implicades en la manifestació i el desenvolupament de les malalties, i hem catalogat els elements en el genoma que confereixen funcionalitat i els òrgans, teixits i condicions en els quals funcionen. Hi ha, tanmateix, una tercera línia de recerca post-genòmica, que fins fa poc semblava, erròniament, més allunyada de l'interès biomèdic. L'espècie humana és només una dels milions d'espècies que habiten la terra, i una àrea de recerca emergent en aquests

moments, la medicina evolutiva, intenta identificar factor genètics implicats en malalties, mitjançant la comparació de la seqüència del genoma de moltes espècies diferents. Per exemple, recentment, mitjançant la comparació del genoma de 400 espècies de vertebrats, hom ha determinat el risc d'aquestes espècies de ser hostes del SARS-CoV-2. En aquests moments, hi ha en marxa una iniciativa, el *Earth Biogenome Project* (EBP), que té com a objectiu l'obtenció de la seqüència del genoma de les aproximadament dos milions d'espècies eucariotes conegudes. El nostre país participa en aquest projecte mitjançant la Iniciativa Catalana per l'EBP (CBP), a través de la qual volem contribuir significativament a un dels projectes més importants de la història de la biologia.