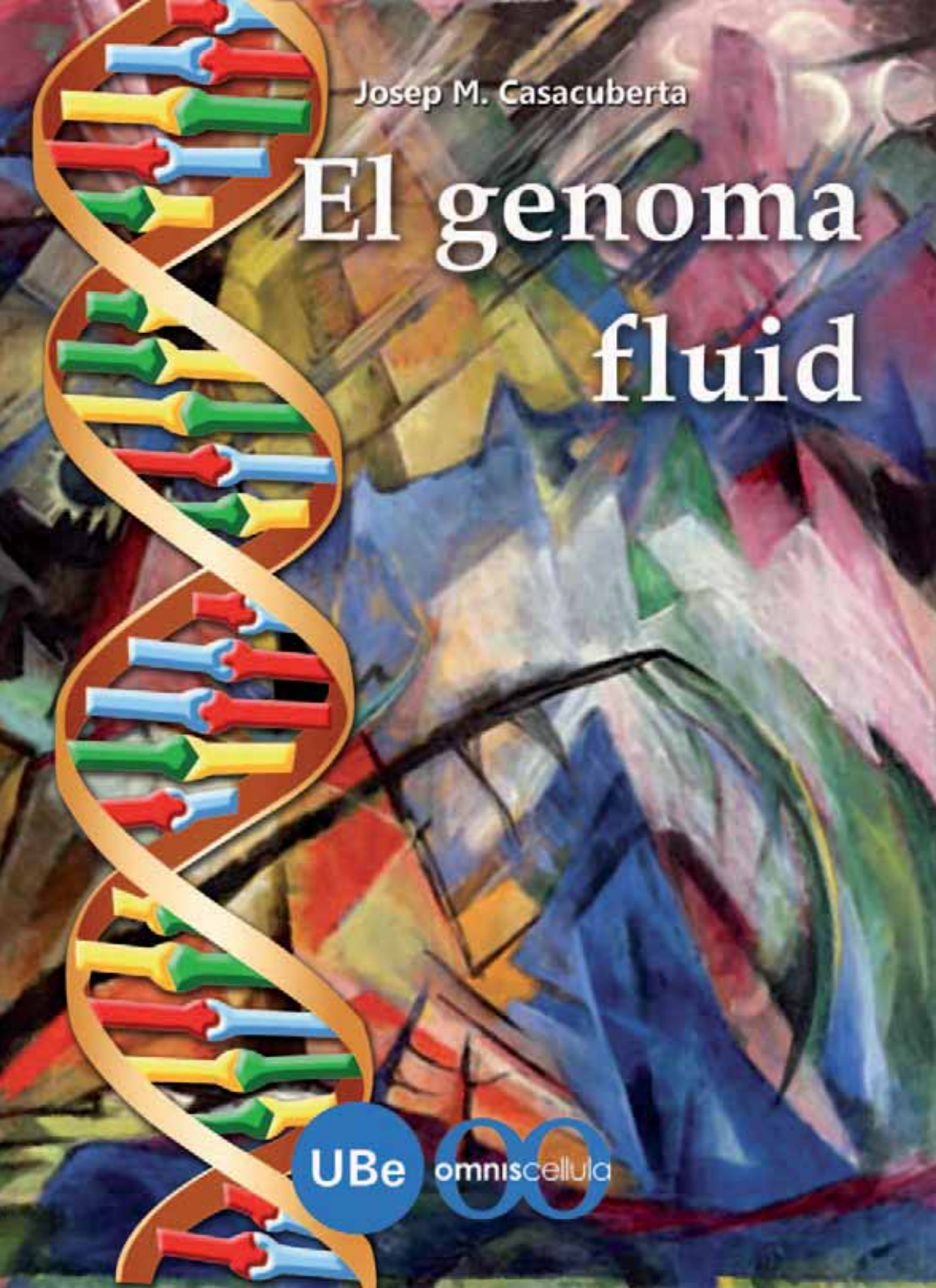


The background of the cover is a complex, abstract collage of various colors and textures, including shades of blue, green, red, and yellow. On the left side, there is a large, stylized DNA double helix structure. The helix is composed of two orange ribbons twisted around each other, with colorful blocks (red, blue, green, yellow) representing the base pairs. The title 'El genoma fluid' is written in a large, white, serif font, and the author's name 'Josep M. Casacuberta' is written in a smaller, white, sans-serif font above it. At the bottom, there are two circular logos: 'UBe' in a blue circle and 'omniscellula' in a blue circle with a stylized 'oo' symbol.

Josep M. Casacuberta

El genoma fluid

UBe

omniscellula

EL GENOMA FLUID

Col·lecció Catàlisi

Ciència Catalana Contemporània

Josep M. Casacuberta

EL GENOMA FLUID



Introducció	9
El genoma, aquella llarga cadena de DNA	13
• La naturalesa química del genoma	15
• El codi genètic: el llenguatge del genoma	18
• Llegint el genoma: la regulació gènica	20
• Traduint el genoma, el paper del missatger	21
• El DNA emmagatzemat: els cromosomes	22
Un únic llenguatge, moltes formes diferents	25
• Els genomes: tots iguals i tots diferents	27
• Mateixos gens, diferents proteïnes	30
• Mateixes proteïnes, diferents resultats	32
El genoma en moviment: l'evolució	35
• Mutació, variabilitat i canvi	36
• Fidelitat i diversitat: un equilibri difícil	38
• Variacions en el genoma humà: l'estratègia conservadora	39
• Viure al caire de l'abisme, l'agitada vida dels virus	42
• Buscant el primer genoma: l'origen de la vida	44
• De ratolins i homes	45
• Barrejant genomes: sexe i recombinació	47
Cap a on ens porta l'evolució?	51
• Un sol mecanisme, infinites direccions	53
• L'home seleccionador: la millora genètica de plantes i animals	56
• Domesticació de les plantes silvestres	58
• Domesticació del genoma, millora genètica	61

El genoma fluid	67
• Gens saltadors i patògens interns	69
• Gens que s'independitzen i estranys que se sotmeten	73
• Duplicar per divergir	76
• El genoma com a ecosistema	78
 Esprement el genoma: la utilització de la informació genètica	81
• Llegint el futur al DNA	83
• Comparar genomes per buscar gens malalts	85
• Reparant el genoma: la teràpia gènica	87
• Organismes genèticament modificats i biomedicina	89
• Modificació genètica de plantes i animals	90
• La clonació: el món després de Dolly	93
• Fabricant l'home perfecte	96
 Agraïments	101

EL GENOMA, AQUELLA LLARGA CADENA DE DNA

capítol 1

Des que Craig Venter, director de l'empresa PE Celera, va anunciar la primavera de l'any 2000 la finalització del primer esbós de la seqüència del genoma humà, els diaris no han deixat de publicar notícies i articles relacionats amb l'estudi del genoma. Les aparicions de científics i periodistes especialitzats en diferents mitjans de comunicació, abans tan escasses, s'han convertit en habituals i una plèiade de termes especialitzats que rarament escapaven dels laboratoris s'han convertit ja en tòpics de diaris i tertúlies radiofòniques. I tot indica que això continuarà així durant algun temps. I és que el genoma humà conté la informació que explica per què som com som, i aquesta informació, a més a més, pot tenir múltiples aplicacions.

El genoma humà podria comparar-se amb un immens manual d'arquitectura, d'enginyeria o de primers auxilis, amb instruccions per construir, connectar i reparar aquesta extraordinària màquina que és el nostre cos; un manual que podria ocupar més de mil volums de mil pàgines cadascun i que contindrien un total de 3.000 milions de lletres. El que Craig Venter va anunciar la primavera de l'any 2000 va ser el resultat d'una primera lectura d'aquells 1.000 toms de 1.000 pàgines. Quatre anys després, aquella anàlisi s'ha completat en bona part, però hi continuen quedant pàgines per llegir i, per suposat, volums sencers per entendre.

Ningú no dubta de la importància d'haver aconseguit determinar la pràctica totalitat de les lletres del nostre genoma, però el cert és que això només és el principi. S'ha seqüenciat el genoma però tardarem encara molts anys a desxifrar-lo correctament, ja que seqüenciar significa únicament llegir i transcriure sense entendre necessàriament allò que es llegeix, i entendre bé el que hem llegit no serà fàcil.

Imaginem que cau a les nostres mans una novel·la escrita en danès. La majoria de nosaltres, malgrat que coneixem pràcticament totes les lletres presents al text, no serem capaços de poder explicar-ne l'argument. Podrem "seqüenciar" el llibre però, a menys que ens apuntem a un curs accelerat de danès i aconseguim aprendre aquesta llengua, no serem capaços de comprendre allò que llegim.

Aquesta diferència entre llegir i comprendre és especialment important en el cas del genoma, perquè el genoma està escrit d'una manera una mica estranya: les paraules no estan separades, sinó que estan unides entre si, i no és fàcil saber on termina una i on comença la següent, però a més a més no totes les lletres del genoma formen paraules. Només una part del genoma, d'aquells

3.000 milions de lletres en el cas del genoma humà, conté informació; només una part forma frases amb sentit, i és una part molt petita.

Imaginem que finalment hem decidit apuntar-nos a classes de danès. Arribem a classe i el primer dia el professor de danès ens dona un text, suposadament en aquella llengua, consistent en una sola frase que emplena totalment el full, sense punts, comes ni espais. De dalt a baix i d'esquerra a dreta el full està farcit completament de lletres. Primer exercici: trobar-hi les paraules, i demà examen. Comencem malament. Doncs així és més o menys com es troba qui pretén llegir la informació continguda en un genoma, a l'humà per exemple. Una llarga frase sense punts, comes o espais, sense capítols ni lletres en majúscula i sense cap índex, faltaria més.

Un cop llegit el genoma cal buscar les paraules i les frases on està continguda la informació. Aquestes frases, com veurem més endavant, es coneixen amb el nom de gens. Cada gen conté una ordre discreta que permetrà a la cèl·lula realitzar una funció. Ja fa més de 30 anys que s'han anat descrivint gens humans i d'altres organismes. La recerca d'aquests gens coneguts a la seqüència del genoma humà no ha estat massa complicada. Però la tasca de definir nous gens, dels quals no se'n coneix res, pot ser tan complexa, en alguns casos, com la de buscar una paraula desconeguda entre el mar de lletres sense sentit aparent de l'exercici del professor de danès. El gran avenç que ha experimentat la informàtica aplicada a l'estudi dels genomes, la bioinformàtica, ha permès realitzar aquestes anàlisis aparentment tan complexes; però cal l'aportació d'informacions procedents de diferents camps de la biologia per confirmar les prediccions bioinformàtiques i desenvolupar nous algoritmes i tècniques d'estudi. I aquesta és una tasca laboriosa que necessitarà temps, perquè fins i tot quan haguem estat capaços de definir tots els gens, ens haurem de preguntar per la funció que desenvolupa cadascun d'ells. Trigarem, doncs, encara molts anys a poder desxifrar correctament el genoma humà. Tot i que tenir a les nostres mans el genoma suposa una fita important en la història de la biologia i en la de l'home, no s'ha d'oblidar que encara no entenem del tot el que veuen els nostres ulls.

La naturalesa química del genoma

Però, què és físicament el genoma? En quin material estan escrits aquests 3.000 milions de lletres que constitueixen aquest fabulós conjunt d'ordres?

Des dels anys 1950 sabem que el suport de la informació genètica és l'àcid desoxiribonucleic, o DNA. Les característiques químiques d'aquesta molècula i, en particular la seva estructura, li permeten complir tres condicions essencials: poder contenir la informació suficient per dirigir el desenvolupament i el funcionament d'un organisme, que aquesta informació pugui ser "llegida" per la maquinària cel·lular i, a més a més, que pugui copiar-se fidelment. Per què és indispensable que pugui copiar-se amb molta fidelitat?

La gran majoria dels organismes, incloent-hi l'ésser humà, no estan formats per una única cèl·lula, sinó per moltíssimes, i totes necessiten una informació genètica per poder ser com són i funcionar correctament. Tota aquesta informació genètica està ja present en la primera cèl·lula de l'organisme, aquella que prové de la unió dels dos gàmetes dels seus progenitors, de l'òvul i de l'espermatozoide en el cas de l'espècie humana. Per formar un cos humà, amb els més de deu bilions de cèl·lules que té, aquesta primera cèl·lula, el zigot, es dividirà moltes vegades i en cada divisió copiarà tota la informació genètica i donarà una còpia a cadascuna de les dues cèl·lules resultants. La informació genètica continguda al zigot s'haurà copiat milions de vegades abans que es formi una cèl·lula del nostre cor o de la nostra pell, per la qual cosa és essencial que existeixi un sistema eficaç i fiable de copiar el genoma.

Des dels anys 1940 ja s'havia suggerit que el genoma s'havia de poder copiar per un sistema simple de motlle i còpia, i ja s'havia vist que el nucli, que suposadament dirigia les funcions cel·lulars, contenia DNA. Però no va ser fins als anys 1950, quan James Watson i Francis Crick van proposar el seu model de l'estructura del DNA, que va quedar clar com i per què el DNA era la molècula de suport de la informació genètica.

El DNA és una estructura polimèrica, és a dir que és una cadena formada per unitats més petites que es repeteixen de manera regular. Però Watson i Crick van proposar que el DNA també podia formar estructures en les quals aquest polímer podia enrotllar-se sobre un altre formant una doble hèlix. Dues llarguíssimes cadenes bastant simples amb un esquelet format per unitats d'un sucre alternades amb grups fosfat del qual penegen quatre molècules lleugerament més complexes: quatre molècules denominades nucleòtids i que s'abreugen amb les lletres A (adenina), T (timina), G (guanina) i C (citosina).

L'estructura proposada deixava les dues cadenes de sucre-fosfat a l'exterior, com dues cintes metàl·liques a les quals s'ancorarien els esglaons d'una

moderna escala de cargol. Cadascun dels esglaons estaria format per la unió de dos nucleòtids mitjançant interaccions febles i específiques, com dos trapezistes que, balancejant-se en els seus trapezis penjats dels dos pals oposats de la carpa, es troben en ple vol i es donen la mà. La interacció, l'abraçada dels trapezistes, es pot trencar i formar-se de nou al vol següent. En cada posició a cada cadena es podria trobar qualsevol de les quatre bases, però un cop fixada una base en una de les dues cadenes la base oposada queda definida automàticament. El trapezista no abraça qualsevol, sinó que té les seves preferències i són estrictes. Les parelles de bases en un esglaó són sempre les mateixes: una A solament comparteix esglaó amb una T i una C amb una G. Les dues cadenes són complementàries, una és el motlle de l'altra i viceversa. Es pot imaginar fàcilment un sistema senzill de duplicació basant-se en aquesta estructura. Només cal separar les seves dues meitats i, utilitzant cadascuna d'elles com a motlle, sintetitzar les seves noves parelles. Imaginem una moderna exposició de fotografies en la qual l'artista exposés alhora el negatiu i el positiu de cada imatge. Cada meitat de l'obra podria servir per obtenir l'altra meitat, donant a qualsevol comprador la possibilitat de duplicar indefinidament l'obra. D'una manera anàloga, el DNA no necessita cap motlle per copiar-se; ell mateix és el seu propi motlle.

Aquesta estructura explica, amb una gran simplicitat i bellesa, una de les característiques essencials del suport de la informació genètica: la capacitat de copiar la seva informació fidelment. Aquesta informació no és res més que la pròpia estructura de la molècula, la seqüència dels quatre nucleòtids que la constitueixen i que formen els esglaons de l'escala de la seva estructura.

Tot i que solament són quatre nucleòtids diferents, quatre lletres, l'extraordinària longitud de la molècula permet moltes combinacions i, per tant, moltes paraules diferents. En el cas de l'ésser humà, la longitud del seu genoma és de 3.000 milions de lletres unides en una llarguíssima frase. Una frase que si estigués escrita amb un tipus de lletra normal, el mateix que s'ha utilitzat en aquest llibre o el que s'utilitza habitualment als diaris, mesuraria 5.000 km: una llarga cinta que podria unir Barcelona i Nova York, creuant l'Atlàntic per damunt de milions de peixos impossibles enfront de la nuesa d'una frase que defineix totes i cadascuna de les nostres característiques biològiques.

El codi genètic: el llenguatge del genoma

Quan es diu que el genoma conté els plànols d'un organisme, en realitat el que es vol dir és que el genoma conté la informació necessària per construir algunes de les peces d'aquest organisme, així com la informació necessària per construir "màquines" que puguin fabricar les peces restants. Aquestes "màquines", que denominem enzims, són les encarregades de mantenir l'activitat de cada cèl·lula i, amb això, la de l'organisme sencer. Els enzims estan especialitzats a accelerar les reaccions químiques que es donen en la cèl·lula i que sense la seva ajuda no podrien realitzar-se. Cada enzim fa possible, catalitza, en llenguatge tècnic, una reacció, i ho fa d'una manera altament específica. Existeixen enzims que ajuden, per exemple, a digerir els aliments a l'estómac, uns altres que faciliten la respiració, uns altres que són essencials perquè les cèl·lules puguin fabricar les seves parets i altres que són necessaris per llegir o copiar les molècules de DNA.

Tant les unitats que formen l'estructura de les cèl·lules com els enzims que permeten la seva activitat tenen la mateixa naturalesa química: totes són proteïnes. Tota la informació genètica continguda al genoma de l'home, de la patata o del cuc de seda serveix per fabricar proteïnes. O el que és el mateix, l'única cosa que el genoma porta escrit és com i quan sintetitzar un gran nombre de proteïnes estructurals o enzimàtiques. D'alguna manera el genoma conté les instruccions per fabricar un cargol i, al mateix temps, les instruccions per fabricar un tornavís que ens permeti fixar-lo i formar, amb diversos cargols, una estructura més complexa. Uns i altres, cargols i tornavisos, són proteïnes, i totes les proteïnes estan formades al seu torn per unitats més petites que anomenem aminoàcids.

Amb només 20 aminoàcids diferents es poden fabricar la gran majoria de les proteïnes de tots els organismes, com si fos un sofisticat mecano que permet construir un nombre gairebé il·limitat d'estructures possibles a partir d'unes poques peces.

Cada proteïna està caracteritzada per una seqüència d'aminoàcids que determina les seves qualitats i la seva funció. I aquesta és la informació que conté el genoma: en ell està escrita la seqüència d'aminoàcids de totes i cadascuna de les proteïnes que són necessàries per formar un organisme i per aconseguir que funcioni correctament.

L'extraordinària diversitat d'estructura de les proteïnes així com les seves habilitats per realitzar multitud de funcions diferents eren conegudes des

de principis del segle XX, cosa que va dur molts investigadors a pensar que elles, i no el DNA, podien ser el suport molecular de l'herència. Semblava lògic que fos una molècula complexa, composta de 20 unitats bàsiques diferents, i no una molècula simple formada per només quatre monòmers diferents, la que permetés emmagatzemar la informació necessària per construir un organisme. El fet que el nucli de les cèl·lules contingués proteïnes, a més de DNA, va reforçar moltes creences i va fer que la hipòtesi del DNA com a base de l'herència trigués un cert temps a ser acceptada. Aquesta hipòtesi s'enfrontava, a més, amb un problema: el DNA es compon de només quatre nucleòtids diferents, mentre que els aminoàcids, la unitat estructural de les proteïnes, són 20. Com podia el DNA codificar l'estructura de les proteïnes amb tan poques lletres diferents?

Cadascun dels nucleòtids podria designar un aminoàcid, però com que solament tenim quatre nucleòtids, amb aquest llenguatge solament podrem definir quatre aminoàcids. Perquè es puguin designar 20 aminoàcids amb només 4 nucleòtids, amb 4 lletres, és necessari agrupar-los formant paraules. De la mateixa manera que, encara que el nostre abecedari tingui poc més de 20 lletres, l'agrupació d'aquestes en paraules permet a la llengua catalana designar un nombre indeterminat de conceptes, l'agrupació dels nucleòtids en "paraules gèniques" permet la codificació en el DNA de tots els aminoàcids. L'agrupació dels nucleòtids en grups de tres, formant paraules de tres lletres a les quals anomenem "triplets", permet que existeixi una paraula per a cada aminoàcid. En realitat, atès que existeixen 64 maneres diferents d'agrupar les nostres 4 lletres en grups de 3, tenim 64 paraules diferents per designar els 20 aminoàcids que formen les proteïnes. Abans ens faltaven paraules però ara ens en sobren. De fet, n'hi ha diverses d'aquestes paraules que són equivalents, que tenen el mateix significat, són sinònims, tot i que alguns organismes prefereixen utilitzar-ne unes i altres organismes, altres de diferents. Aquestes paraules de tres lletres s'agrupen en frases més o menys llargues per formar un paquet d'informació completa que permetrà sintetitzar una proteïna. Cada frase genòmica, cada ordre completa, és coneguda com un gen, mentre que el producte de la seva traducció, la seva expressió executiva, és una proteïna.

L'ordre dels nucleòtids a la cadena de DNA determinarà l'ordre d'acoblament dels aminoàcids a les proteïnes corresponents. Una informació seqüencial traduïda a una altra seqüència; d'una llarga seqüència de DNA, molècula que emmagatzema informació però que no té capacitat d'exercir cap funció, a

moltes seqüències de proteïnes, molècules capaces d'exercir multitud de funcions cel·lulars però poc útils per emmagatzemar i transmetre informació.

Tots els organismes, dels bacteris a l'ésser humà, utilitzen el DNA per emmagatzemar la informació que els permet construir les seves proteïnes. I a totes les espècies aquesta informació està escrita en un mateix llenguatge. En tots els casos, la correspondència entre triplets de nucleòtids i aminoàcids és la mateixa. És per això que, atès que una cèl·lula d'un organisme pot interpretar perfectament la informació d'un gen d'un altre organisme, podem, en principi, introduir gens humans en bacteris, de bacteris en plantes, o de plantes en animals, i aconseguir que funcionin correctament. Això, com veurem més endavant, obre possibilitats amplíssimes per a l'aplicació en biotecnologia.

Llegint el genoma: la regulació gènica

A la majoria dels organismes, totes les cèl·lules de l'individu provenen d'una única cèl·lula inicial formada de la unió de les cèl·lules sexuals dels seus progenitors. Aquesta cèl·lula es divideix i dona lloc a cèl·lules que es dividiran de nou, seqüencialment, per formar un individu de bilions de cèl·lules. A cada divisió, la cèl·lula copia el seu genoma i dona una còpia idèntica a cada cèl·lula filla, per això cadascuna de les cèl·lules d'un organisme té una còpia idèntica del genoma. Per què existeixen tantes diferències entre els diferents tipus de cèl·lules d'un organisme si totes elles posseeixen la mateixa informació genètica? Per què les cèl·lules que formen les nostres ungles són diferents de les de la nostra retina, i les de nostre intestí són clarament diferents de les del cervell? Malgrat que totes les cèl·lules d'un organisme posseeixen la mateixa informació genètica, cada cèl·lula només utilitza una petita part de aquesta informació, una part diferent en cada cas. Cada tipus cel·lular llegeix solament una fracció del genoma, permetent-li expressar unes característiques particulars i fent possible la formació d'òrgans diferents. Aquesta informació, que permet controlar que cada cèl·lula utilitzi únicament la fracció d'informació correcta a cada moment, també està codificada al genoma, encara que el llenguatge sigui en aquest cas diferent.

La informació que determina on, quan i en quina quantitat ha de sintetitzar-se una proteïna es troba a les regions del gen pròximes a la seqüència codificant, davant i darrere d'aquesta. La regió anomenada promotor, situada normalment davant de la zona traduïble, indica a la maquinària cel·lular on

comença el missatge per traduir i és, a més a més, la responsable de regular el moment i el lloc on s'ha d'activar el gen. A l'altre extrem del gen, després de la zona traduïble, se situen els senyals que marquen el límit en el qual la maquinària cel·lular ha d'acabar la seva lectura, i que reben el nom de terminadors. El conjunt de la zona codificant i de les zones reguladores que determinen quan i de quina manera cal llegir aquesta informació constitueix un gen.

La manera de codificar les seqüències de les diferents proteïnes és, com diem, molt similar a tots els genomes. No obstant això, la informació que determina on, quan i en quina quantitat cal produir cada proteïna està escrita en un llenguatge diferents i menys rígid que l'utilitzat per codificar les proteïnes, i difereix entre bacteris i organismes superiors, entre plantes i animals, i fins i tot difereix lleugerament entre dues plantes com el tomàquet i el blat. Així, tot i que la informació continguda en un gen d'una carxofa pot ser entesa perfectament per una cèl·lula d'un gos o d'una panerola, i que aquestes cèl·lules no tindran cap problema per fabricar la proteïna codificada per aquest gen, molt probablement no s'apiguen quan ni en quina quantitat fabricar-la.

Traduint el genoma, el paper del missatger

Als organismes superiors, aquells que coneixem com a eucariotes (els fongs i llevats, les plantes i els animals), el genoma es troba al nucli cel·lular, separat de la resta de la cèl·lula, és a dir, del citoplasma, per una membrana que l'aïlla i el preserva de l'activitat frenètica que allí es realitza. Al nucli, com al despatx d'un capatàs, s'emmagatzema la informació sobre allò que el citoplasma ha de fabricar. Però a la cèl·lula no existeix un sistema de megafonia que permeti al capatàs comunicar-se directament amb el taller on es construeixen peces i màquines. Es necessita un missatger que comuniqui al citoplasma allò que el nucli determina a cada moment que la cèl·lula ha de fer.

Aquest missatger, com en una fàbrica, ha de ser capaç d'entendre's amb el capatàs i amb els treballadors del taller, ha d'entendre els seus dos idiomes, ha de saber com estan codificades les ordres en els plànols de la direcció i com han de ser traduïdes aquestes ordres al taller perquè es fabriqui allò que s'ha de fabricar quan s'ha de fabricar. La millor manera d'entendre's amb el capatàs és parlar la seva mateixa llengua i assemblar-s'hi. Aquest missatger és, de fet, una molècula molt semblant al DNA que es coneix com a RNA, RNA missatger.

Els RNA missatgers són també cadenes d'un àcid nucleic, àcid ribonucleic en aquest cas, i també estan formats per nucleòtids, per lletres pràcticament idèntiques a les del DNA. En el moment en què un gen s'expressa i la seva informació ha de portar-se al citoplasma, les dues cadenes de DNA se separen, les parelles de trapezistes que es donaven les mans fa només uns paràgrafs ara se separen per agafar les mans d'altres trapezistes, semblants però no idèntics, els ribonucleòtids, les lletres de l'RNA. L'especificitat dels nous trapezistes és la mateixa. Les lletres T del DNA, que es donaven la mà amb les A, solament s'aparellaran amb les A de l'RNA; les G, que s'unien fortament a les C del DNA, només agafaran les mans de les C de l'RNA.

La doble cadena de DNA de 3.000 milions de lletres del nostre genoma se separa localment perquè es copiï en RNA un missatge d'uns pocs milers de lletres, la informació corresponent a un gen, i aquesta molècula, molt més petita que el genoma, surt del nucli, travessa la membrana nuclear i viatja al citoplasma on l'ordre que porta escrita podrà traduir-se en una proteïna. Cada gen, doncs, s'haurà de copiar prèviament en RNA perquè pugui ser finalment traduït en proteïna.

El DNA emmagatzemat: els cromosomes

Encara que a cada moment només s'estigui fabricant un nombre limitat de productes a cada cèl·lula, totes les cèl·lules d'un organisme tenen una còpia exacta de tot el genoma, una còpia exacta d'aquest text de 3.000 milions de lletres que, escrit en caràcters normals, podria arribar de Barcelona a Nova York. I tota aquesta informació està guardada a cadascuna de les nostres cèl·lules! Per sort per nosaltres, les lletres del genoma són bastant més petites que les d'impremta, i el nostre text genòmic no és una molècula de 5.000 km. Malgrat això, si bé les lletres genòmiques, els nucleòtids, són molt petites, el DNA contingut en una cèl·lula humana és un filament de gairebé dos metres. Amb el contingut total de DNA que té una persona, amb els seus 10 bilions de cèl·lules, es podrien fer més de deu voltes al món. Si tenim en compte la longitud i el gruix del DNA, la dificultat d'acomodar el DNA al nucli de la cèl·lula equival a voler ficar un fil de cosir de 160 quilòmetres de longitud en un didal. I, no obstant això, cadascuna de les nostres cèl·lules aconsegueix compactar aquest DNA al nucli.

El DNA està extraordinàriament empaquetat al nucli. Les dues cadenes de DNA, entrelaçades com els brins d'un cordill, s'enrotllen sobre una sèrie de

proteïnes estructurals formant allò que coneixem com a nucleosomes. Cada cert nombre de bases es forma un nucleosoma amb el DNA enrotllat en dues voltes i mitja sobre aquest, de la qual cosa resulta una fibra estesa de DNA que sembla un rosari. Aquest rosari de DNA i proteïnes s'enrotlla al seu torn sobre si mateix formant unes fibres compactes que en alguns moments són visibles al microscopi i que coneixem com a cromosomes. Tots els genomes complexos estan organitzats en cromosomes, encara que el seu nombre varia entre les diferents espècies. En el cas de l'home existeixen 23 parells de cromosomes diferents, 23 calaixos plens d'informació en els quals es guarda tot el nostre genoma.

L'existència dels cromosomes, amb el seu extraordinari grau de compactació del DNA, explica com pot emmagatzemar-se la informació genètica, però planteja altres preguntes: com pot funcionar un genoma tan compactat? Com es pot llegir una informació genòmica en aquesta barreja de DNA? Qui seria capaç de trobar una marca feta amb bolígraf en aquest fil de 160 quilòmetres sense treure'l del didal?

Encara que sembli mentida, les nostres cèl·lules ho fan contínuament. La maquinària cel·lular pot trobar la marca del fil i, una vegada fet això, pot separar els dos brins que el formen, llegir la informació que hi contenen, copiar-la i tornar a enrotllar de nou els dos brins, l'un sobre l'altre. Així doncs, l'elevat grau d'empaquetament del genoma, necessari per poder acomodar tota aquesta informació a cadascuna de les nostres cèl·lules, no compromet, però, la seva capacitat de funcionar correctament, de ser llegit i copiat tantes vegades com calgui al llarg de la vida de l'organisme.

En realitat, el grau d'empaquetament del DNA al nucli no és uniforme, sinó que existeixen zones del DNA més compactades, mentre que d'altres, que solen correspondre a les que tenen gens, és a dir, informació que ha de ser llegida, es mantenen en general més obertes. D'altra banda, el grau d'empaquetament de les diferents zones del nucli canvia també amb el temps; per exemple, és diferent en les diverses fases de desenvolupament d'un individu o bé varia en funció de canvis ambientals. Dades relativament recents mostren que, de fet, bona part de la regulació de l'expressió dels gens, és a dir, la decisió de quins gens han d'expressar-se a cada moment, es realitza modificant el grau de compactació del DNA i exposant o ocultant d'aquesta manera la informació genètica a la maquinària que ha de llegir-la, com una sàvia institutriu que obre i tanca segons la seva voluntat les diferents ales d'una biblioteca per tenir cura que les nostres lectures siguin les més convenients a cada moment.