

ENTREVISTA AMB EL DR. RODERIC GUIGÓ I EL DR. RORY JOHNSON

1. INTRODUCCIÓ:

Durant el segon curs de Batxillerat, els alumnes de Catalunya desenvolupem un treball de recerca a través del qual aprenem a cercar informació i a treballar-la, tot aportant el nostre punt de vista. Gràcies a una notícia que es va publicar el mes de Setembre de 2012 que exposava noves troballes referents a una part del DNA coneguda com "escombraria" vaig decidir centrar el meu treball en aprendre i aprofundir més en aquest tipus DNA. El treball de recerca es va titular "*Els misteris del DNA escombraria*" i constava de dues parts, una teòrica i una pràctica. Com a part pràctica del treball vaig entrevistar a un dels científics espanyols que ha contribuït a l'estudi del DNA no codificant i sobretot a l'estudi del RNA. Aquest científic és el doctor Roderic Guigó qui va accedir molt amablement a entrevistar-se amb mi. El doctor Guigó és membre del Comité Rector del projecte ENCODE, liderant el grup d'anàlisi de RNA, i va ser l'únic científic espanyol signant de la primera seqüència del genoma humà. Actualment és coordinador del programa de Bioinformàtica i Genòmica del Centre de Regulació Genòmica (CRG) de Barcelona i professor a la Universitat Pompeu Fabra.

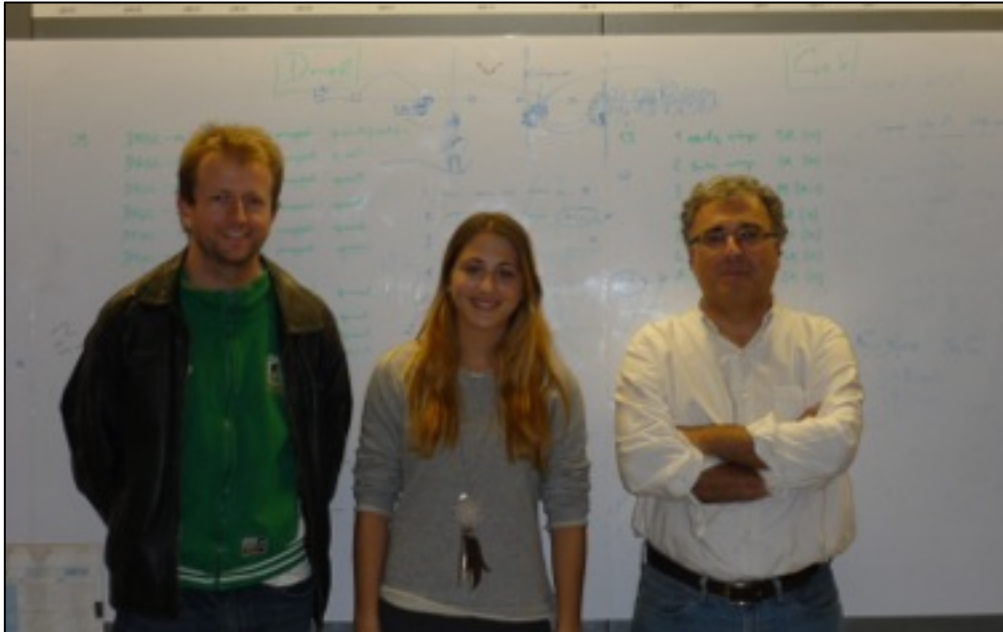
Els investigadors del CRG han col·laborat en dos dels articles publicats a *Nature* (autors principals en un d'ells), en dos més a *Genome Biology* i en quatre dels de *Genome Research* (autors principals en tres d'ells). El projecte ENCODE també ha permès als investigadors del CRG treballar i cooperar amb altres científics de tot el món.

Els propòsits d'aquesta entrevista van ser diversos. En primer lloc, rebre de primera mà informació sobre l'estudi que estan desenvolupant i tractar amb un científic amb reconeixement internacional. En segon lloc, aclarir dubtes sobre el projecte ENCODE i obtenir resposta a les meves preguntes sobre el DNA escombraria.

El dia 14 de novembre va tenir lloc la "trobadura" amb el Dr. Guigó al CRG (Centre de Regulació Genòmica) de Barcelona. Vaig tenir l'honor de passar prop de tres quarts d'hora parlant amb ell i, a més a més, vaig poder entrevistar a un altre investigador del seu grup de recerca, en Rory Johnson, especialista en recerca sobre el RNA

Data: 14 de novembre de 2012

Lloc: Grup de Bioinformàtica i Genòmica (Centre de Regulació Genòmica-CRG). Parc de Recerca Biomèdica de Barcelona



Imatge presa al CRG després de l'entrevista (Dr. Johnson, esquerra; Dr. Guigó, dreta)

Nota a la transcripció de les entrevistes. S'han intentat mantenir les expressions i construccions originals utilitzades durant l'entrevista. El Dr. Guigó va realitzar esquemes i dibuixos per aclarir les seves explicacions que no s'han inclòs. L'entrevista amb el Dr. Johnson es va desenvolupar en castellà, encara que la seva llengua materna és l'anglès. S'ha intentat mantenir el contingut original encara que en alguns casos les construccions no siguin gramaticalment correctes. Els noms tècnics utilitzats en anglès durant l'entrevista s'han traduït al castellà la primera vegada que s'utilitzen, i es recullen entre parèntesi.

2. TRANSCRIPCIÓ:

Sobre el projecte ENCODE...he vist que es va dividir en dues parts. En què va consistir cada part?

El projecte ENCODE té un objectiu que és el mateix objectiu a la primera part que es deia fase pilot, que a la fase que ha acabat ara que es deia la fase de *Scat scaling phase*. L'objectiu del projecte ENCODE és identificar en el genoma humà aquelles regions que tenen funcionalitat, les que no són *junk-DNA*, les que no són DNA escombraria. Saber quines són en el genoma les que no són. Llavors en la primera fase, la fase pilot es va iniciar el 2003 i va acabar el 2007. El que es va fer és seleccionar 30 regions de diferents cromosomes del genoma humà que més o menys representaven l'1% només del genoma humà, 30 milions de bases en total, i el que es va fer és veure quines tecnologies eren les més apropiades, quins mètodes, quins experiments eren els més apropiats per identificar precisament aquestes regions que eren funcionals, que no eren *junk-DNA*, que no eren DNA escombraria, que tenien alguna funció biològica.

A la fase d'escalat que va començar el 2007, aquestes tecnologies s'han aplicat a tot el genoma humà. La diferència entre la fase pilot i la fase següent és que a la fase pilot només es va mirar l'1% del genoma humà i a la fase següent s'ha mirat tot el genoma humà, i també s'han utilitzat tecnologies, clar la tecnologia ha canviat, ha evolucionat, hi ha mètodes millors, més precisos, més ràpids, més eficients. També han canviat els mètodes, es clar.

Aquest 1% que es va mirar és el que forma els gens, o és un 1% a l'atzar.

És un 1% a l'atzar. Es va intentar agafar un 1% que fos representatiu de la seqüència del genoma humà.

I com es va decidir?

Es va decidir... un grup de gent es va reunir i van agafar regions... no era ben bé a l'atzar. Es van agafar regions del genoma que se sap que son molt riques en gens, on hi ha molts gens, regions que hi ha molt pobres en gens, regions que eren molt similars al genoma del ratolí, que en aquell moment es coneixia, regions que eren més diferents del genoma del ratolí, regions que tenien un contingut en HGC. Es va intentar escollir una mica regions que fossin representatives dels diferents dominis, dels diferents tipus de seqüències del genoma que hi ha. O sigui no és que s'anés a l'atzar i s'agafés... I després hi va haver un conjunt, no sé si aproximadament la meitat d'aquests seqüències que si que es van agafar a l'atzar.

He vist que ets bioinformàtic, i que en tot això ha pres molta importància la part d'informàtica i tot de programes informàtics... Ha sigut més important això de la informàtica que no pas la de laboratori, no?

Sí,

O sigui s'ha avançat molt més amb...

Si, en tots aquests projectes genòmics pensa que la part experimental inicial és molt petita, és només obtenir la seqüència... un cop has obtingut la seqüència aquesta tot el que fas a continuació és purament computacional. A veure, després no es així, no és així... perquè després en la fase d'escalat s'apliquen moltes tècniques també experimentals per determinar quines són les regions funcionals. Es fan servir... ara es fa servir, no se si ho has sentit, ara tot el que es fa servir és el que es diu ultraseqüenciació... es fan servir unes màquines que tenen una gran capacitat de seqüenciar els àcids nuclèics. Per exemple, el que s'ha fet es utilitzar aquestes màquines per seqüenciar no només el genoma perquè ja es coneixia, tot i que no es coneixia de les cèl·lules que hem investigat per igual... per seqüenciar el contingut de RNA de diferents tipus cel·lulars. A ENCODE el que s'ha estudiat es com funciona el genoma en tipus

cel·lulars diferents. És a dir, totes cèl·lules del teu cos tenen el mateix DNA però les cèl·lules de la pell són diferents de les del cervell...

Diferents en com s'expressen, no?...

Si, com s'expressa el DNA, com funciona el genoma. I el que estem mirant és, en el projecte ENCODE, quins són els factor genètics i epigenètics que funcionen diferent en les diferents línies cel·lulars. Quins són en particular, per exemple, els gens que s'expressen diferent. Per tant el que fem es mesurar el RNA, tu ja saps que el primer que passa es que el DNA es copia a RNA... els gens... i el que fem es mirar el contingut en RNA d'aquests diferents tipus cel·lulars. No són teixits del cos humà, són línies cel·lulars que s'han mantingut en cultiu durant molt de temps, que s'originen... originalment venen de teixits però que es mantenen en condicions de laboratori. Es com si tinguéssim seqüenciades cèl·lules del cervell, de la pell, dels ossos, i veiem quin és el RNA diferent que hi ha en cadascuna d'aquestes cèl·lules.

El RNA missatger, oi?

Sí, l'RNAm i també hem mirat l'RNA primari i moltes altres coses. Però en principi diguem que l'important és "l'RNA missatger i mRNA de l'RNA missatger". El que també hem mirat són els canvis epigenètics del DNA. No se si tu has sentit parlar de les modificacions de les histones...?

Sí.... això també era una de les preguntes que et volia fer i la relació una mica...

El DNA s'embolcalla al nucleosoma, i els nucleosomes estan fet d'histones que son proteïnes, tenen una cua, i els aminoàcids d'aquesta cua poden estar modificats... metilats, etc, i el que se sap és que l'estatus d'aquestes modificacions de les histones té a veure amb l'estat transcripcional dels gens. Per exemple hi ha modificacions com la trimetilació de la lisina 36 de l'H3, no? Dins del nucleosoma tenim diferents histones, H2, H3, H4... l'H3, per exemple té una lisina en la posició 36, doncs bé aquesta lisina pot estar sense metilar, monometilada, bimetilada, trimetilada... Doncs bé, en els gens que s'expressen aquesta histona està trimetilada. Això està relacionat directament amb l'expressió dels gens. Això es una modificació epigenètica que està associada a l'expressió dels gens.

Per entendre ben bé el que és el DNA escombraria, no tinc molt clar com separar-ho del DNA no codificant. Està dintre, o...?

No. Bé, a veure, és que depèn. Aquestes paraules no s'utilitzen d'una manera molt precisa. Moltes vegades s'entén per DNA codificant només aquell DNA que codifica per proteïnes, el que dóna lloc als gens. Aleshores si només entenem com a DNA codificant

aquest, hi ha molta part del DNA no codificant que no és DNA escombraria. Per exemple, totes les regions promotores dels gens...

...serien no codificants....

... són no codificants però són funcionalment molt importants perquè allà és on s'uneixen els factors de transcripció, causen o associats a aquestes modificacions epigenètiques i fan que s'iniciï la transcripció d'un gen. Per tant, aquesta regió promotora no és codificant però no és DNA escombraria. Sabem que dins dels introns... saps que els gens estan... introns hi ha senyals que regulen l'*splaising*, precisament l'*splaising* diferencial, *splaising* alternatiu dels gens. Saps que de vegades els exons s'inclouen o no s'inclouen. Això està fet per regions dels introns. Aquestes regions dels introns no són codificats però no podem dir que siguin DNA escombraria. Tenim la regió 3' dels gens, tenim estructures tridimensionals que de vegades són reconegudes per micro RNAs, que es regula també l'estabilitat dels RNAs. Aquestes regions no codifiquen per proteïnes, no formen part, no es tradueixen a proteïnes... però són funcionals.

DNA escombraria seria tècnicament o estrictament aquella part del nostre DNA que si l'eliminéssim del nostre genoma no tindria cap efecte en l'organisme.

Pel que he entès d'això no tindria cap funció llavors el DNA escombraria?

A veure, el DNA escombraria, com a tal efectivament no tindria cap funció. No tindria cap funció... però això no es així perquè el que estem veient, el que nosaltres estem veient... El que trobem és que abans tot el que no era DNA codificant per proteïnes es considerava DNA escombraria i ara el que estem veient és que una gran part d'aquest DNA escombraria, per exemple, té funció. Per exemple una gran part d'aquest DNA està compost d'elements repetitius, de transposons... Aquests es pensava que eren totalment inactius. Ara estem veient que en realitat aquests també es transcriuen.. el que es diu pseudògens, que són gens que s'han copiat, que han perdut la pota de lectura, que s'han mutat i que sembla que siguin còpies inactives en realitat veiem que són actives. Per tant, una gran part del que pensavem que era inactiu ara estem veient que té activitat.

És més un problema del llenguatge, no? Perquè es tenia el DNA codificant i a tota la resta se li va dir escombraria, però que en realitat és DNA no codificant però que no tot aquest és DNA escombraria.

Exactament, una gran part del DNA que no codifica per proteïnes no es DNA escombraria. I mai ha estat considerat realment DNA escombraria perquè des del principi saben que hi ha la regió promotora dels gens que és essencial per a què aquests gens funcionin. Sense la regió *upstrain* dels gens els gens no es poden expressar. Si tu treus aquella regió promotora aquell gen potser no funcionarà mai. Entens? Perquè aquella regió controla que aquell gen s'expressi específicament en un determinat teixit, en una determinada condició, després d'un determinat estímul, etc, etc.

Llavors els trossos que s'han estudiat eren del genoma humà, oi?

En el projecte ENCODE vols dir?

Sí

En projecte ENCODE sí, s'ha estudiat el genoma humà. Ara hi ha un altre projecte que es diu modENCODE, que s'ha desenvolupat paral·lelament al projecte ENCODE, que *mod* vol dir d'organismes models, organismes models on s'han estudiat el genoma de dos organismes model d'una manera molt similar a la del projecte ENCODE, que són *Drosophila* i el cuc de *Caenorhabditis elegans* (*C.elegance*). Però el projecte ENCODE és específicament humà.

Llavors s'esperarien resultats similars al que s'ha trobat del genoma humà?

En altres espècies?

Sí.

No necessàriament. Perquè per exemple en els bacteris el genoma està molt més comprimit i els gens estan comprimits, no hi ha introns, per exemple. Per tant els introns molta gent considera que són dispensables. En realitat si la longitud d'un intró molt llarg disminueix, potser no passa res en el gen, no? No tots els organismes tenen la mateixa mida del genoma. Per exemple hi ha organismes que tenen el genoma més compactat. Per exemple la proporció de DNA que codifica proteïnes en el genoma humà és només del 3 %. Però en el genoma de *Drosophila*, possiblement és del... Ara mateix no sé el número exacte però potser és del 15 o del 20 %. Es a dir que les parts aquestes que podrien ser DNA escombraria són molt més petites en altres genomes d'organismes diferents de l'home. En realitat això és així tot i que... en *Drosophila* hi ha... no sé quans gens té però potser 15 mil o 16 mil gens, i el genoma humà només en té 20 mil. Són molts pocs gens més. Però el genoma nostre són 3.000 milions de bases, 3.000 milions de bases i el genoma de *Drosophila* només té 120 milions de bases, és a dir, és 30 vegades més petit, però el nombre de gens no és 30 vegades menys sinó que el nombre de gens és pràcticament idèntic.

I la part no codificant és molt menor...

Molt menor en el genoma de *Drosophila*, i en el de *C. elegance* crec que encara és més petit, i també té un nombre similar de gens.

És a dir, que mentre en el genoma humà tu tens per exemple aquí un gen... això és un gen amb els exons... per exemple llavors aquí tens una regió que no hi ha res...aquí tens un altre gen... en el genoma de *Drosophila*, el mateix... diguem tu tens aquí un gen de *Drosophila*, potser està fet així amb els exons molt més a prop, i l'altre gen a continuació també molt més a prop. És a dir el que aquí ocupa posem jo que sé, 100 megabases i

aquí potser ocupa només 10 megabases... i tenim en principi la mateixa informació. També per això es diu que aquest DNA és escombraria... perquè aquí, en la *Drosophila*, tenim la mateixa informació molt més comprimida que aquí. Ara bé, està clar que nosaltres no som mosques, no?. Per tant no està clar que nosaltres puguem comprimir el nostre genoma a la mateixa dimensió que s'ha comprimit el DNA de *Drosophila*.

Val. Llavors en el nostre genoma, el tant per cent que seria escombraria... quin seria... seria el que no s'ha trobat encara que tingui cap funció?

A veure, això és difícil. Segurament si mires la Wikipedia trobaràs estimacions. A veure, la part del genoma que està en introns, els introns no codifiquen proteïnes, és el voltant d'entre el 2 i el 3 %. Això es la part del genoma que codifica per proteïnes. Clar, la resta pot ser moltes coses diferents. Però per exemple, si nosaltres mirem la part del genoma en la qual trobem molècules d'RNA aquesta part és del 60%. Molts d'aquests RNA no sabem que fan, perquè només una part molt petita d'això, sobre el 5 % seria més o menys d'aquest, realment forma part dels exons que codifiquen per proteïnes. La major part de RNA que es produeix no sembla que tingui una funció com a RNA missatger. El que passa es que no saben si té funció o no. Saben que es produeix, no?

Es produeix un RNA que no codifica per a proteïnes però això no vol dir que no tingui funció, no?

... no vol dir que no tingui funció. Potser té una funció com a RNA per si mateix. Els RNAs poden tenir funció com a RNAs.

I què faria un RNA d'aquests?

Què farien?... Jo et dono un camp que potser hauries de mirar, *long non-coding RNA* que ara estan emergint que pràcticament es desconeixien fa cinc anys, que són RNAs que son iguals que els RNAs que codifiquen per proteïnes, que tenen exons i introns, però... no codifiquen per proteïnes, no donen lloc a cap proteïna. Hi ha exemples molt famosos d'aquests long non-coding RNA, hi ha un que té a veure amb la inactivació del cromosoma X. Són RNAs que ... es posen al promotor de l'RNA del cromosoma X i l'inactiven. Tenen una funció que no es codificar proteïnes però que tenen una funció... nosaltres estem descobrint cada vegada més que aquests *non-coding RNAs* que no codifiquen res però que la funció es activar l'expressió d'altres gens que estan al costat. Aquí tenim, per exemple, un RNA que no codifica per proteïnes, que té tres exons, aquí al costat tenim un gen que sí que codifica per proteïnes, que té quatre exons, el que veiem es que l'activació d'aquests *long non-coding RNA* (lnc RNA) activa l'expressió d'aquest *protein coding gen*.

Llavors aquests RNAs es formen a partir de... de regions de DNA que no codifica proteïnes, no?

Exacte, de regions del DNA no codifiquen. És més complicat que això però deixem-ho així. Tu tens el genoma, aquí tens un gen, aquest gen codifica per a proteïnes (protein coding gen), aquí tens al costat un altre gen que dona lloc a un non coding RNA que no codifica per proteïnes. Aquesta idea que teníem que tots els gens.. tu quan estudies... que tots els gens codifiquen per a una proteïna no es certa. Veiem que hi ha molts gens que no donen lloc a proteïnes.

És a dir, dóna lloc a un RNA però aquest RNA no dona lloc a una proteïna, no?

Exactament. Dóna lloc a un RNA, que fins i tot pot tenir *splaining* que eliminarà els introns però aquest RNA no va, o si que va, o bé es queda al nucli, o bé va al citosol o al citoplasma i allà no és traduït pels ribosomes. I en aquests moments... en el genoma humà hi ha uns 20 mil *protein coding gen*, doncs hi ha uns 12 mil... és un nombre molt alt de *long non-coding RNA*. Fins fa molt poc aquests gens eren totalment desconeguts.

Tots els gens... aquest gen sempre donarà un RNA que codificarà per a proteïnes, o potser que algun gen, segons com sigui activat, produeixi un RNA per proteïnes...

Aquesta és una molt bona pregunta. El que estem veient ara es que hi ha gens que... això ho pots parlar amb el Rory que ho està estudiant.

El que estàvem veient és que el mateix gen pot donar, depèn de com, un gen que codifica per proteïnes i al mateix temps pot donar un altre RNA que no codifica per proteïnes. I de fet el que veiem és que la major part de gens que codifiquen per proteïnes produeixen transcrits, dos tipus de transcrits, uns que són els RNA missatgers i uns altres que no es tradueixen a proteïnes.

Llavors... què regula que es faci una cosa o que es faci l'altra?

No ho sabem, no ho sabem. No sabem ni si es produeixen simultàniament, no sabem si hi ha... seria molt interessant saber si hi ha una regulació, si hi ha unes condicions en les quals es produeix... aquesta és una bona pregunta... si hi ha condicions en que només es produeix el gen que codifica per a proteïnes i en quines condicions només el long non-coding RNA. Seria molt interessant realment.

O sigui, seria... potser, jo ho trobo molt més lògic que es regulés d'alguna manera. Perquè si... quan vol fer una cosa fa l'altra, seria com si...

Però el que també hem de pensar... això és cert... però també hem de pensar que hi ha molt comportament estocàstic.

Què és això?

Estocàstic vol dir que hi ha moltes coses que passen que no tenen cap funcionalitat, que passen una mica a l'atzar. Que allò es va transcrivint, i es produeixen uns transcrits i llavors uns d'aquests codifiquen i uns altres no. Hi ha una part que està regulada però també hi ha una part que no està massa regulada, que passa una mica perquè passa. La cèl·lula és molt resistent a coses que passen de forma gairebé molt atzarosa, *random*, estocàstica.

Què guai, és interessant... Pel que havia trobat...jo havia entès que s'havia trobat funció pel DNA escombraria i que aquesta era la d'interruptor...

Si, el que passa és el següent. Tornem-hi. L'interruptor segurament és la regió que regula l'expressió d'un gen. Tu tens un gen, i estàs en una cèl·lula, posem A, i aquí tens el mateix gen i estàs en una cèl·lula B. Aquest gen posem que aquí està actiu, i aquí està inactiu. Això fa que aquesta cèl·lula sigui per exemple una cèl·lula del múscul i aquesta sigui una cèl·lula que no sigui del múscul, posem que sigui de l'os. Val?. Quan parlem d'interruptors ens estem referint a les regions del DNA que fan que aquest gen estigui *on* o *off*, són els interruptors. La regió típica és la regió promotora. La regió que hi ha davant del gen, que és on s'uneixen els factors de transcripció. Quan aquests factors de transcripció s'uneixen el gen s'expressa. El que passa aquí és que segurament per la raó que sigui aquí no es poden unir possiblement perquè aquí no s'han produït les modificacions de la cromatina necessàries perquè s'uneixin, i per tant aquí no hi ha expressió. Hem vist... això es el que es diuen els promotors. Però en realitat si això és un gen i això és el seu promotor també hem vist que hi ha promotors... la majoria es que hi ha promotors que estiguin directament *upstrain* directament davant del gen...

... quan es plega la

Sí. Llavors hem vist que en realitat també hi ha promotors que no es diuen promotors sinó que es diuen *enhancers*, són promotors, fan el mateix però com tu dius no estan al costat. En realitat podria ser fins i tot que un enhancer d'aquest gen estigues en un altre cromosoma, i per la forma com els cromosomes es pleguen dins s'acosten aquí... tot això d'aquí son els interruptors aquests. Depèn de quina sigui la combinació per exemple aquest està on, aquest està on i aquest està a 0, doncs això produeix zero còpies. Si aquest està a 1, encès, aquest està a 0, i aquest està a 0, això produeix diguen deu còpies; si aquest està a 1, aquest està a 1 i aquest està a 1, això produeix 100 còpies. Entens? Depenent de la combinació de l'estatus, llavors aquests enhancers que estan lluny del gen abans es pensava que eren junk-DNA entens?...

Sí.

I ara es veu que una gran part d'aquesta regió ideogènica que està entre els gens que està lluny dels gens, el que fa és... conté regions que es diuen enhancers que el que fan és contribuir a regular específicament l'expressió d'un gen. De manera que per un gen

normalment tens set o vuit enhancers. Depenent de l'estatus d'aquests enhancers tu tens més o menys expressió d'aquest gen.

Això és el que ens pensem que és ara, vull dir que... d'aquí a deu anys potser això canvia. Però és el que ens pensem.

I això com es troba?, vull dir, sabeu quins enhancers actuen en quin gen? Com heu arribat a saber els d'un cromosoma...

Bé això ho sabem per molts pocs casos. Això... els *enhancers* no s'han descobert en el projecte ENCODE. Ja se sap des de fa molt temps que hi ha gens que estan regulats per regions distants del genoma. Com se sap això? Suposo que de moltes maneres, però segurament en algun cas s'ha vist, possiblement per casualitat algú estava afectant un promotor d'un gen que estava aquí, es pensava que... i va veure que aquest promotor afectava a aquest gen d'aquí que no es troba al costat. Per exemple estava mesurant l'expressió d'aquest i... jo muto aquest... canvio aquest promotor i en lloc de canviar-se el gen que estava al costat...o a més a més també es canvia l'expressió per alguna cosa. Amb el projecte ENCODE el que si sabem és que aquestes regions que es diuen enhancers tenen unes modificacions de cromatina A característiques i a més tendeixen a transcriure's, a generar RNA elles mateixes. El que nosaltres hem trobat és que hi ha molts enhancers, perquè tenen les marques de cromatina dels enhancers.

Saber a quin gen regulen aquests enhancers és més complicat. Hi ha tècniques que s'estan fent... una hipòtesi és que podem dir per exemple...si aquesta és una regió que creiem *enhancer*... veure en quines altres regions del mateix cromosoma o d'un altre cromosoma, aquesta regió interacciona. Tenint això, pensem que és un *enhancer*, podem determinar amb no molta precisió, però si amb certa precisió, quines són les altres regions del genoma amb les quals aquest *enhancer* interacciona. Llavors podem pensar que els gens que estan a prop d'aquestes regions estan regulats per aquest *enhancer*. Entens una mica la idea?

Si... Llavors el que passa és que el problema és com estan relacionats també entre si els *enhancers*...

... sí és molt complicat... un enhancer aquí, que toqui un altre enhancer aquí...aquí tot una sèrie de gens que estiguin d'alguna forma regulats... és difícil establir una relació unívoca entre un enhancer i el gen que regula perquè a més a més pot ser que un *enhancer* reguli molts gens també...

D'acord. O sigui, un gen té diferents *enhancers* i un *enhancer* pot regular diferents gens.

Exactament.

Llavors el fet que ho faci dependrà... per exemple si aquí està l'*enhancer* i aquí el gen... aplegar-se quedarà així i llavors regularà a aquest... però si es plegués d'una altra manera regularia un altre?

Exactament. Però depèn... encara no sabem totes les maneres de veure com es pleguen... com es plega el DNA... però els cromosomes és una cosa que només es veuen a la metafase. Quan no estem a la metafase tot el DNA està descompactat i llavors hi ha moltes interaccions entre les diferents parts del DNA. Llavors jo diria que una cosa que ajuda en la regulació diferencial de l'expressió dels gens té a veure precisament amb aquesta topologia diferencial del DNA a les cèl·lules.

O sigui no se sap ben bé el plegament del DNA, no?... Se sap el plegament dels cromosomes però no el del DNA?

No. És molt difícil. Se saben algunes coses. Hi ha unes tècniques ara que et permeten detectar llocs d'interacció entre el DNA. Aquests llocs d'interacció entre el DNA podrien... és una hipòtesi, eh?... podrien apropar els *enhancers* amb els gens que regulen.

També he trobat referències a l'RNA quimèric.

L'RNA quimèric és el següent. Nosaltres tenim la idea típica de com s'explica això... Aquí tens el DNA... aquí tens un gen... posem per cas... aquí tens un altre gen... Llavors tu aquí tens un RNA que es produeix, un altre RNA que es produeix, un altre RNA que es produeix, no? Llavors aquest RNA té l'*splaining*, i al final queda convertit en una molècula de RNA, molècula de RNA... La idea és que aquí s'alliberen els introns, s'alliberen els introns aquí, s'alliberen els introns aquí... Però aquesta mol·lècula de RNA és col·lineal, es diu col·lineal amb el DNA, o bé que l'ordre d'aquí, aquí, aquí, es manté.... no hi ha coses entremig. Nosaltres el que veiem és que en ocasions, i no sabem quin és el mecanisme, resulta que hi ha molècules de RNA que estan fetes... una part ve d'aquí i una part ve d'aquí...

Dels introns?

No, no, dels exons, o dels introns. Per exemple agafa dos exons d'aquí i tres exons d'aquí. I es fa una nova molècula d'RNA que no es col·lineal amb el DNA però que se salta entremig... Entens? Això és el que diem quimèric-RNA, o *fugent transcripts*. Tenen molts noms. I aquests es veuen molt... sembla que són molt importants en el cas dels càncers.

O sigui... la formació del RNA quimèric seria un comportament normal o seria anormal o atzarós?

És una bona pregunta. Nosaltres el que veiem que passa en ocasions és que veiem dos gens que estan junts i cadascun genera el seu RNA... en ocasions... doncs aquest té

l'*splicing* i genera un RNA més petit... Però en ocasions hi ha una única unitat de transcripció en *redtrough*... no es passarà la polimerasa... o sigui aquí dintre és curt, no es para i llavors es genera un transcrit molt més gran. Llavors quan hi ha *splicing* llavors agafa exons d'aquí i exons d'aquí... Això es quan aquests gens estan contigus i es pot originar. En els casos en els quals això és més distant no tenim una idea molt clara de quin és el mecanisme. I també sabem que es produeix en condicions fisiològiques, però tampoc sabem si té una funcionalitat o no... I sabem que es produeix en molts càncers... El que passa es que en els càncers és difícil saber si això és una conseqüència del reordenament dels cromosomes o no.

Nosaltres normalment no seqüenciem el genoma de les coses, fem servir un genoma de referència. Tu imaginat que això es el meu genoma de referència. Jo se que hi ha aquest gen aquí i aquest gen aquí, hi ha aquest gen aquí al mig... A, B i aquest C. Pot ser que en el genoma del càncer hagi passat el següent... hagi passat el que es diu un reordenament cromosòmic i aquí tinguem A, aquí tinguem C i aquí tinguem D. Això ha canviat així. El que semblava una quimera distal que semblava feta d'A i C, en realitat es una quimera pròxima perquè en realitat els cromosomes estan junts. Pot ser que sigui nomes part d'A i part de C... Llavors ens dona la impressió que sigui un transcrit quimèric que ve de llocs diferents, però en realitat... el que ha passat és que en el genoma del càncer s'ha produït aquesta reordenació

La Laura es presenta al Dr. Johnson, explicant-li el motiu de l'entrevista, el seu interès en l'estudi del DNA escombraria i en què consisteix un Treball de Recerca de Batxillerat a Catalunya.

Es mucho más complejo de lo que pensaba. Al ser todo nuevo... la información se publicó en septiembre.... y todo lo que he leído son resúmenes... muy por encima... y ahora el Dr. Guigó me ha estado explicando muchas cosas...

Lo que no sabemos es el 99% de todo... Lo que estamos viendo, lo que sabemos es como un iceberg de todo... sólo sabemos una parte muy pequeña de todo. Tu pregunta ¿cuál ha sido?

El Dr. Guigó me había explicado que había RNAs que podían producir proteínas o que podían no producir proteínas, y quería saber qué regulaba que hiciera una cosa o la otra.

Antes creíamos que los mRNA tenían la secuencia de *open reading frame* que tiene la proteína dentro que empieza con *methionine* (metionina) y termina con *stop*. Todo era muy sencillo porque encuentras el mRNA y encuentras la *open reading frame* y luego sabes qué proteína tiene dentro. Pero ahora creemos que no es tan sencillo. Hay ejemplos donde tienes algo que... esto es muy claro que es largo, aquí es como unos doscientos aminoácidos, es muy largo y es muy claro que tiene una proteína. Es muy

obvio. Pero hay otros casos que tienen un *open reading frame* que es bastante pequeño o que se expresa de forma muy diferente de todas las otras proteínas que conocemos. No se si es algo al azar o es algo de biología.

Lo que yo había pensado es que si un RNA podía codificar una proteína o podía no hacerlo, tendría que haber algo que lo regulara, porque si solo fuera al azar...

Depende de muchas cosas. Creemos ahora que hay señales dentro de la secuencia... aquí... que muestra si debe traducir o no. También hay procesos que controlan si el mRNA se queda dentro del núcleo o si se va fuera. Últimamente nos damos cuenta de que hay RNA que son no codificantes pero que van al ribosoma, pero que controlan el proceso del ribosoma pero que no dan proteínas. Es todo muy complicado. Además hemos encontrado ejemplos donde antes creíamos que eran no codificantes (porque no podíamos encontrar ningún *open reading frame*)... pero al final cuando miras datos de *mass spectrometry* (espectrometría de masas)... ¿sabes qué es *mass spectrometry*? Cuando leen trozos de proteínas... es que sacas una proteína de una célula, cortas las proteínas en trozos pequeños y luego lees todos los fragmentos de proteína, y usando computadores puedes encontrar de dónde han venido estas proteínas, de qué secuencia de RNA. Porque sabes que la secuencia de RNA va a proteína con el *genetic code* (código genético)

Con esta secuencia puedes saber o adivinar cuál era la secuencia de dónde vino. Usando estos fragmentos de *mass spectrometry* a veces encuentran fragmentos que deben de venir de algunas secuencias que creíamos antes que era no codificante.

O sea, unas proteínas vienen de no codificante...

Algunas, pocas. Pero algunas vienen de algo que antes creíamos que era no codificante. La razón para esto es que cuando tienes un mRNA que es muy convencional, es muy sencillo de encontrar cuál es el *open reading frame*. Porque el *open reading frame* significa simplemente que empiezas con un *methionine* y terminas con un *stop*.

Si imaginas que es muy fácil encontrar una secuencia al azar... Si buscas en mil secuencias encontraras algo que tiene una región de una *methionine* y un *stop*, al azar, por azar. No hay una regla de oro que puedas usar para diferenciar entre los dos, entre algo que no es codificante y algo que sí es codificante. Toda secuencia tiene algún *open reading frame*, y por eso es muy difícil en estos casos saber si es codificante o no. Por eso ahora creemos que hay proteínas escondidas entre todos los no codificantes que nadie había encontrado antes. Además creen que hay proteínas muy cortas, *mono peptides*, y con nuestros métodos de ahora no encontraremos estos.

Es superinteresante...

Al principio de la Biología en los años 50, creían que todo era muy sencillo, que tenías ADN, ARN, proteínas... Cada vez que investigamos más vemos que las reglas no son tan sencillas.

Hay muchos factores que influyen en las expresiones, en las regulaciones.

Exacto. Antes no teníamos las herramientas para mirar tan profundo. Ahora sí.

Muchas de las herramientas son computacionales...

Computacionales y también secuenciación. Sabes, las máquinas de *high through sequencing*.

Los genes codificantes producen una cantidad mayor de ARN, mucho mayor. Con los medidores de antes, que eran muy poco sensibles, sólo se ven los codificantes. Ahora que tenemos estas máquinas... Ahora vemos más y más profundo y vemos todas esas cosas que no veíamos antes. Ahora vemos los no codificantes, que están allí pero de nivel mucho más bajo. Eso es una pregunta muy grande porque los no codificantes tienen un nivel tan bajo....

¿Qué harían los RNA no codificantes?

Pueden ser muchas cosas. Pero creen que... si esto es el genoma, indicamos los genes con flechas así... puede ser... Sabes que los genes producen mRNA, y muchas veces tienes todos los *transcription factors* (factores de transcripción) aquí, que se unen al ADN y regulan el gen. Lo que creemos es que si esto es un *non* codificantes, algo que es parecido al mRNA... puede ser dos mil bases o algo así... y esto en vez de ser traducido forma algunas estructuras... sabes... es algo complicado... con *base pairing* o algo así y estos se unen quizás con otras proteínas que son reguladores de algo, y que se vuelve al genoma y regula otros genes. Puede ser. Esta es la idea, pero no lo sabemos. Pero hacen muchas cosas. Otras se unen con mRNA codificante para regular a través de *base pairing* o pueden formar estructuras muy grandes en la célula, pero que todavía no se sabe qué hacen, o puede ser que las células quitan los ADN como para hacer señales a otras células.

La verdad es que sólo podemos observarlos... y ahora es el paso de saber, de investigar qué hacen

Por lo que me ha explicado el Dr. Gugió, toda la regulación está muy relacionada. Hay muchas cosas que...según si esto varia se regulará de una manera o de otra, hay muchos factores.... es muy difícil... porque si varias una cosa entonces...

Sí, sí, sí. Sólo lo que quedas hacer es quitar una cosa a la vez y ver qué pasa pero eso es muy sencillo y muchas veces no captura las relaciones muy sutiles entre cosas más grandes, complejas.

Pero lo que es muy interesante es que ahora sabemos que tenemos más de 10 mil de estos y hace diez años pensábamos que no teníamos. Y ahora tenemos 10 mil, y probablemente hay más. Tienes que saber que tenemos 20 mil proteínas y hemos descubierto 10 mil que no transcriben y lo probable es que hay muchos más.... lo probable es que tengamos más de estos que de estos. Que es una revolución.

Estos serían los no codificantes... o sea vienen de una parte del RNA que es no codificante...

Si, pero normalmente vienen de otra parte del genoma. No es que sea un gen que es codificante que tiene un trozo *non* codificante que es separado... son genes distintos. Hay un par de casos donde tienes un mRNA codificante que también funciona como *non* codificante. Lo que digo aquí es que son separados en el genoma. Es muy complicado.

Sí, es muy complicado. No se sabe todavía.... **Sí, al ser todo nuevo, al hacer preguntas un poco más allá te encuentras que todavía no se sabe...**

Sí, sí, falta la nueva generación para entender cómo funciona. Entonces tienes que escribir un ensayo, un proyecto...

(La Laura torna a explicar en què consisteix el treball de recerca)

¿Por qué has elegido el tema?

Quando nos explicaban el DNA siempre nos decían que el 2% eran los genes y que el resto era basura. Quería saber.... si teníamos todo ese DNA basura... por algo tenía que ser. Entonces justo salió la noticia del proyecto ENCODE y me pareció perfecto.

(Agradecimientos y despedida).