

Disseny d'una eina d'edició genètica CRISPR per corregir la progèria

Disseny d'un ARN guia i uns iniciadors de PCR per aplicar aquesta tècnica



Pseudònim: Liti

Abstract

El trabajo ha consistido en un análisis de las propiedades y límites de la técnica de edición genética CRISPR para curar una enfermedad. Para conseguirlo se han diseñado los elementos moleculares reales que se utilizarían para hacer una corrección genética en células en cultivo de una enfermedad real. Se ha realizado una búsqueda de las posibles enfermedades que tengan las características necesarias para poder ser corregidas con la técnica CRISPR, es decir, enfermedades genéticas y causadas por mutaciones en un solo gen. De todas las posibilidades se ha seleccionado la progeria, una enfermedad muy grave que produce el envejecimiento prematuro de las personas que la sufren. Para conseguir la corrección de la mutación y de esta manera curar la enfermedad a nivel celular, se ha identificado y localizado el error puntual en el ADN. Después, se ha diseñado un ARN guía específico para la enfermedad de la progeria, que conjuntamente con la proteína Cas9 forman la herramienta CRISPR. También se han diseñado dos iniciadores para obtener por PCR un fragmento del gen sano que contiene la zona de mutación. Todos estos elementos moleculares, ARN guía, Cas9 y el fragmento amplificado, si se introdujeran dentro de la célula corregirían la mutación en el ADN y como consecuencia se curaría la progeria. También se ha dedicado un espacio para describir las bases de la genética molecular, las bases de la técnica CRISPR y los datos de la enfermedad de la progeria desde un nivel clínico hasta un nivel molecular.

The research project consists of an analysis of the characteristics and limits of the CRISPR gene editing technique to cure a disease. In order to do it, the real molecular elements that would have been used to make genetic corrections in cultured cells have been designed for the cure of a real disease. A research of the possible illnesses that can be cured with the CRISPR technique has been done, i.e. genetic diseases caused by mutations in a single gene. Progeria has been selected from all possibilities, a very serious disease that provokes the sudden and premature

aging. To correct the mutation, and cure the illness in cells, the error in DNA has been identified and located, then a specific guide RNA has been designed for the progeria disease that along with the Cas9 protein form the CRISPR tool. Also, two primers have been designed to obtain a fragment of the healthy gene that contains the mutation area, by use of PCR technique. These molecular elements, the guide RNA, the Cas9 and the piece of homologous DNA, if they were introduced into the cell they would correct the mutation in DNA and consequently cure progeria. Also, in the project there are sections with information about the bases of molecular genetics, the bases of the CRISPR technology as well as progeria information from a clinical level to a molecular level.

Índex

1. Introducció	5
Presentació del tema	5
Motivacions	5
Objectius	6
Parts del treball	6
Mètode utilitzat	7
Fonts utilitzades	7
Dificultats	7
Agraïments	8
2. La genètica molecular	9
2.1 Transcripció d'ADN a ARN	11
2.1.1 Procés de la transcripció	12
2.1.2 Els introns, exons i el procés de splicing	12
2.1.3 Splicing alternatiu	13
2.1.4 Seqüència Codificadora (CDS)	14
2.2 Traducció d'ARN a proteïna	15
2.2.1 El codi genètic	15
2.2.2 Procés de la traducció	16
2.3 Mutacions	18
3. Reacció en cadena de la polimerasa (PCR)	21
4. La tècnica seleccionada: tecnologia CRISPR	22
4.1 CRISPR en els bacteris	22
4.2 La tècnica CRISPR com a eina biotecnològica i d'edició genètica	25
4.3 Mecanismes de reparació d'una ruptura en la doble cadena d'ADN	27
5. La malaltia seleccionada: Síndrome de la progèria de Hutchinson-Giford	29
5.1 Dades clíniques i epidemiològiques de la progèria	29
5.2 Dades metabòliques i genètiques	29
5.3 Gen responsable	31
5.3.1 Portal NCBI	31
5.3.2 Maduració de la Proteïna Lamin A	37
5.4 Base de dades OMIM	37
5.5 Variant al·lèlica	41
5.5.1 Maduració de la progerina	42
6. La cura de la malaltia: aplicació de la tècnica CRISPR en la progèria	44
6.1 Disseny de l'ARN guia	44
6.2 Disseny dels iniciadors de PCR	46

7. Conclusions	48
8. Bibliografia	50
Webgrafia	50
Bases de dades	52
Articles	52
Documents de docència	52
Annex	54
Seqüència del gen LMNA	54
Localització de la mutació 0.22 de la progèria en la seqüència del gen.	72

1. Introducció

Presentació del tema

La tècnica d'edició genètica CRISPR és una tècnica que permet editar el genoma de qualsevol cèl·lula en cultiu tallant a un lloc molt concret de l'ADN. La pròpia cèl·lula executa, com a resposta, uns processos de reparació que aconseguiran arreglar el dany i així curar la malaltia.

Aquesta tecnologia té el seu origen en la resposta immunitària que generen els bacteris contra les invasions virals. A partir d'aquí la biotecnologia la va redissenyar per poder-se utilitzar com a eina d'edició genètica. Emmanuelle Charpentier i Jennifer Doudna han sigut les dues científiques que van desenvolupar al 2012 la tècnica CRISPR i aquest any 2020 han guanyat el Premi Nobel de Química.

La tècnica es basa en l'acció de dues molècules, l'àcid nucleic ARN guia i la proteïna Cas9, que formen el complex ARNGuia-Cas9. Aquest, quan s'introdueix dins una cèl·lula, els seus components començaran a realitzar les seves funcions, l'ARN guia complementarà amb la seqüència d'ADN diana i la proteïna Cas9 tallarà pel lloc on l'ARN ha indicat. L'impressionant d'aquesta tècnica és que en escollir una seqüència diana es pot dissenyar un ARN guia complementari que permet controlar el lloc concret del genoma on la Cas9 tallarà.

El sistema CRISPR-Cas9 s'està investigant molt perquè té un gran potencial en el camp biosanitari, en un futur potser es podran curar malalties genètiques en humans i també s'està treballant en utilitzar el sistema com a fàrmac per curar càncers. A més a més es podrà editar el genoma de bacteris perquè deixin de presentar resistència a algun tipus d'antibiòtic.

Motivacions

Vaig decidir fer aquest treball perquè m'interessen molt els temes relacionats amb la biologia, la genètica i el funcionament de l'ADN. També em va cridar molt l'atenció quan vaig sentir a parlar per primera vegada de la tècnica CRISPR i em va semblar molt impressionant totes les coses que pot arribar a fer. En el moment que em van proposar el tema vaig pensar que era una cosa molt complicada i massa difícil però al final vaig decidir tirar endavant amb aquesta idea ja que aprendria molt sobre la matèria que m'agrada i a més la tècnica és molt innovadora i actualment està causant un gran impacte en el camp de la biotecnologia. També em vaig acabar de decidir per aquest treball ja que tenia l'oportunitat de que m'ajudés un expert en la matèria, l'Agustí Fontarnau que m'ha ajudat molt durant tot el procés. Realitzant el treball he après moltes coses noves i tot hi haver sigut un treball difícil m'ha agradat molt fer-lo.

Objectius

La meva intenció en aquest treball és descriure les bases de la tecnologia CRISPR i usar-les de manera teòrica per tal de curar una malaltia humana. Pretenc justificar la correcció de la malaltia en dos passos. Un és el disseny de l'ARN guia, que forma part de l'eina CRISPR i s'utilitzarà per indicar a la proteïna Cas9 en quin lloc del genoma ha de tallar. El segon és el disseny dels dos iniciadors de PCR que s'utilitzaran per sintetitzar la seqüència homòloga necessària perquè el gen es recomбини i es corregeixi el dany en l'ADN aconseguint així la cura de la malaltia.

Parts del treball

El treball està dividit en quatre parts, la primera explico la genètica molecular ja que és la base per entendre les altres seccions del treball. En aquest apartat s'hi troba

l'estructura de l'ADN, els processos de transcripció i traducció i els tipus de mutacions gèniques que existeixen. Seguidament també hi ha un petit apartat on s'explica la PCR ja que serà important per entendre el funcionament dels iniciadors de PCR que s'utilitzaran per crear l'ADN plantilla. En la segona part explico la tècnica CRISPR, la seva funció en la natura, on forma part del sistema immunitari dels bacteris, i la seva funció com a eina d'edició genètica. A la tercera, és on s'explica la malaltia seleccionada, que en aquest cas és la progèria, aquesta secció inclou les seves dades clíniques i epidemiològiques, dades genètiques, el gen responsable de la malaltia i la variant al·lèlica que s'estudia. A més a més també explico el funcionament de les bases de dades NCBI i OMIM ja que les he utilitzat molt en el procés de realització del treball. Finalment, en l'última part realitzo de forma teòrica el procediment simplificat que se seguiria per curar la progèria que consisteix en el disseny d'un ARN guia i d'uns iniciadors de PCR.

Mètode utilitzat

He començat el treball coneixent una mica el tema de la tecnologia CRISPR per posar-me en context. Després he buscat la informació bàsica de l'ADN i els seus processos ja que són la base per entendre la genètica molecular. Seguidament amb l'ajuda de la meva tutora em vaig informar i vaig entendre el CRISPR en bacteris i com a eina biotecnològica. En escollir la malaltia vaig haver de buscar-ne una amb els requisits necessaris perquè el CRISPR la pugui curar. En aquest pas vaig utilitzar molt el portal de la NCBI i la base de dades OMIM per tal de trobar tota la informació relacionada amb la malaltia i el component genètic que la provoca. El meu mentor m'ha anat ajudant en totes les etapes del procés però especialment en la d'escollir la malaltia i en dissenyar el CRISPR . En aquesta última etapa a partir de tots els coneixements que ja tenia vaig dissenyar l'ARN guia i els iniciadors de PCR. Durant tot el procés he buscat informació de webs, vídeos i articles fiables.

Fonts utilitzades

Les fonts d'informació més utilitzades en el treball han sigut el portal de la NCBI, la base de dades OMIM i un document escrit per l'Agustí Fontarnau on hi ha les pautes per realitzar un estudi i edició de l'ADN. També s'ha consultat molts documents didàctics de la genètica molecular per tal de fer una síntesi entenedora del tema.

Dificultats

La dificultat principal que he tingut realitzant el treball ha sigut entendre alguns temes de genètica molecular i biotecnologia. Comprendre tots els conceptes començant des d'un nivell de coneixement baix en biotecnologia i genètica molecular ha sigut un repte i ha portat molt temps i esforç. Però a poc a poc he anat comprnent tota la informació i he aconseguit realitzar el treball. També m'ha sigut difícil redactar de manera entenedora tot el que he après fent el treball ja que, com he dit anteriorment, és un tema difícil. Altres dificultats han sigut la informació en anglès ja que hi havien algunes pàgines que estaven en anglès i no totes les paraules eren fàcils de comprendre. Finalment m'he trobat amb la dificultat d'utilitzar bases de dades i sobre tot el portal de la NCBI que al principi va ser una mica difícil de saber fer servir.

Agraïments

Finalment m'agradaria agrair l'ajuda de la meva tutora per guiar-me durant la realització del treball i també al meu mentor i expert Agustí Fontarnau que m'ha ajudat molt a comprendre molts conceptes difícils, a utilitzar el portal de la NCBI i la base de dades OMIM i a validar-me que l'ADN guia i els iniciadors estiguessin correctament.

2. La genètica molecular

La genètica molecular és un camp de la biologia que estudia l'estructura i la funció dels gens a nivell molecular.

Diversos estudis al llarg de la història han demostrat que l'ADN és la molècula que conté la informació genètica de tots els organismes formats per cèl·lules. L'ADN és un àcid nucleic, que juntament amb els glúcids, lípids i proteïnes és un dels quatre tipus de macromolècules que formen la vida. Les unitats estructurals dels àcids nucleics són els nucleòtids, en el cas de l'ADN els nucleòtids que el componen estan constituïts per un monosacàrid anomenat desoxiribosa, un àcid fosfòric i una base nitrogenada que pot ser l'adenina (A), la timina (T), la citosina (C) o guanina (G).

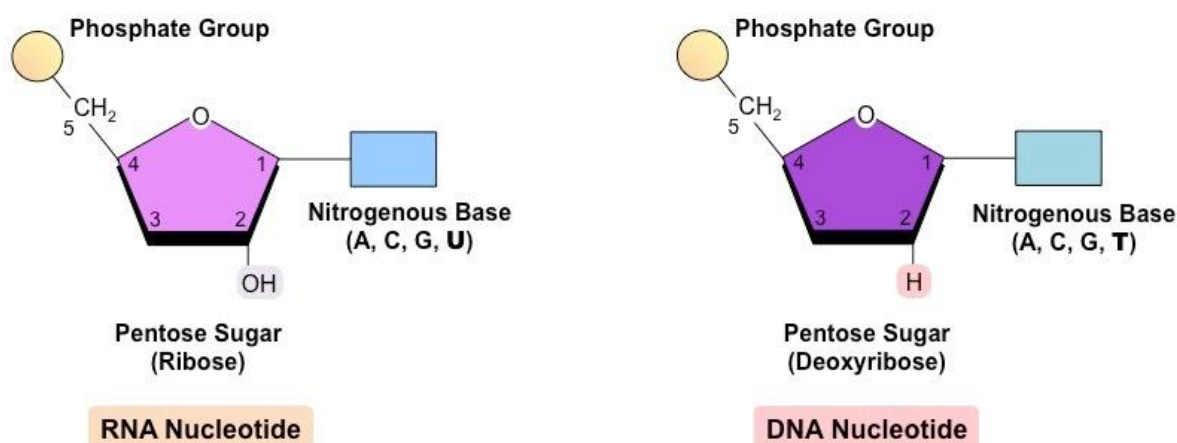


Figura 1: Diagrama d'un nucleòtid d'ADN i un d'ARN

Font: <https://ib.bioninja.com.au/standard-level/topic-2-molecular-biology/26-structure-of-dna-and-rna/nucleotides.html>

L'ADN està format per dues cadenes de nucleòtids enrotllades entre elles formant una doble hèlix. Els nucleòtids d'una cadena i de l'altre estableixen enllaços de complementarietat, les bases adenina (A) complementen amb les timines (T) i les guanina (G) amb les citocines (C). Aquestes dues cadenes són antiparal·leles, una està orientada en sentit 5' → 3' i l'altre 3' → 5'.

En els cèl·lules eucariotes el nombre de molècules d'ADN ve determinat per l'espècie, en el cas dels humans tenim 46 molècules d'ADN per tan 46 cromosomes. Dins d'aquestes molècules s'hi troben els gens, unitats hereditària dels éssers vius constituïts per un segment d'ADN que codifica per una proteïna o un ARN.

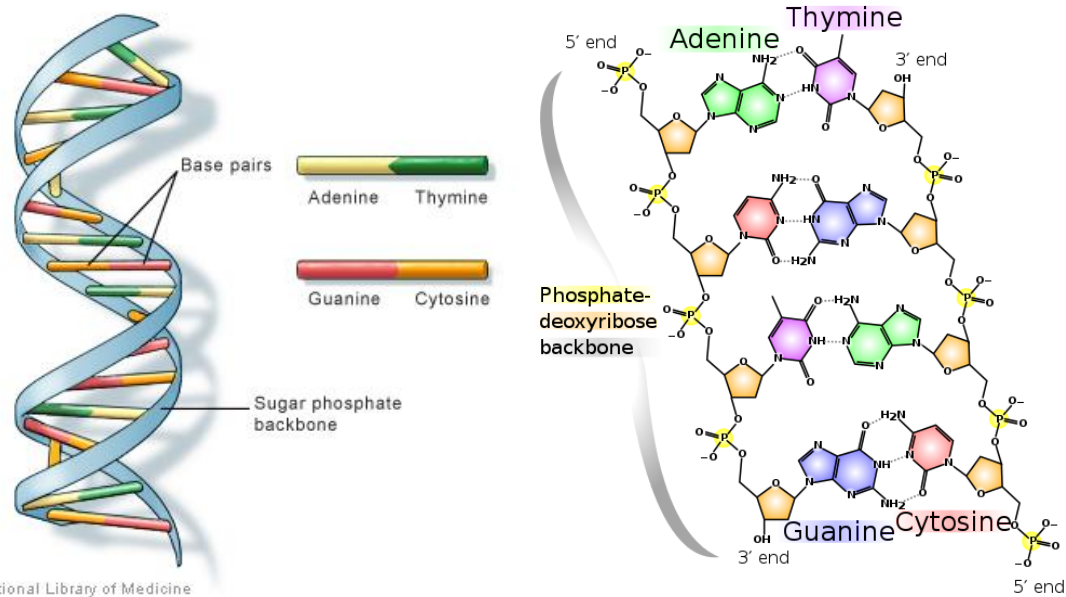


Figura 2: Estructura de l'ADN

Font: [https://www.news-medical.net/life-sciences/What-is-DNA-\(Spanish\).aspx](https://www.news-medical.net/life-sciences/What-is-DNA-(Spanish).aspx)

Font: https://commons.wikimedia.org/wiki/File:DNA_chemical_structure.svg

L'ARN és un àcid nucleic d'una sola cadena que intervé en l'expressió de la informació genètica. L'ARN està format pels nucleòtids adenina, citosina, guanina i uracil. Aquests també poden formar enllaços de complementarietat entre ells, l'adenina complementa amb l'uracil i la citosina amb la guanina, a més de amb els corresponents de l'ADN. Els nucleòtids de l'ARN estan compostos per una monosacàrid anomenat ribosa, un àcid fosfòric i una de les bases nitrogenades. En la figura 1 s'observa l'estructura d'un ribonucleòtid.

Les proteïnes són macromolècules formades per aminoàcids. Aquestes executen pràcticament totes les funcions del nostre organisme i són les que ens fan ser qui som. Les proteïnes desenvolupen la funció estructural, de protecció, transportadora,

nutritiva, de reserva, reguladora, hormonal, de moviment, de defensa, de reconeixement, homeostàtica i enzimàtica. Per tan són unes molècules molt importants pel nostre organisme i per fabricar-les es necessita extreure la informació dels gens a través dels processos de transcripció i traducció. La proteïna que es forma al final d'aquests processos està determinada per la seqüència de nucleòtids del gen corresponent. El fet de que tots els éssers vius tinguem cadenes de nucleòtids diferents fa que existeixi una gran diversitat i continuï l'evolució.

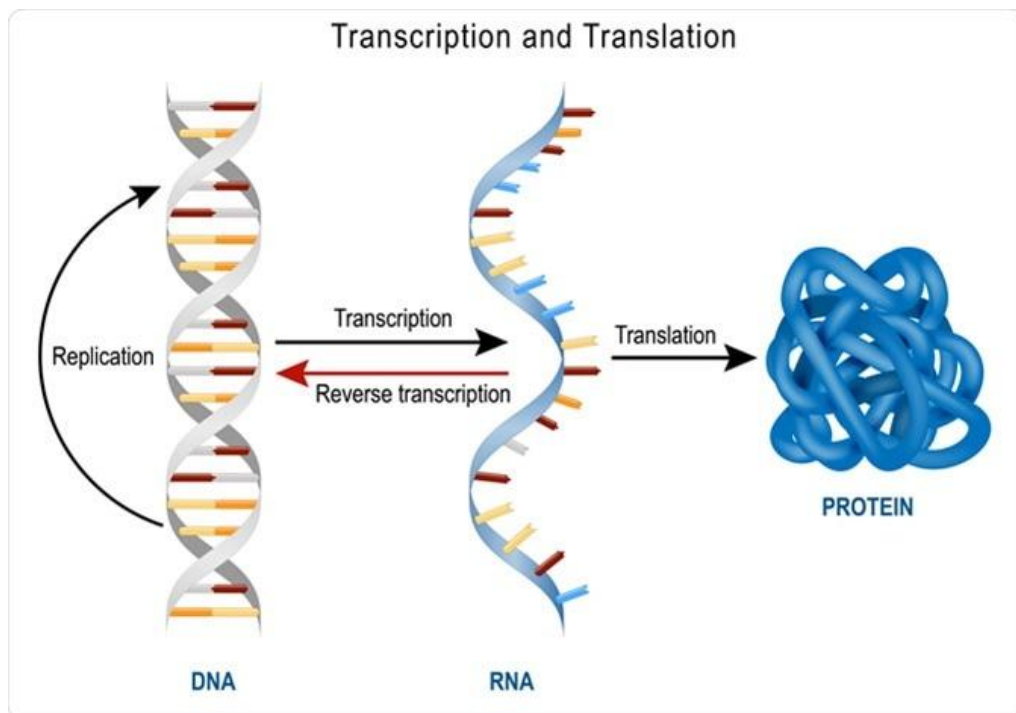


Figura 3: Procés de traducció i transcripció

Font: [https://www.news-medical.net/life-sciences/DNA-Interactions-with-Proteins-\(Spanish\).aspx](https://www.news-medical.net/life-sciences/DNA-Interactions-with-Proteins-(Spanish).aspx)

2.1 Transcripció d'ADN a ARN

La transcripció és el procés pel qual la informació d'un gen és copiada en una altra molècula, un ARN. D'aquesta manera es sintetitza una nova molècula amb la mateixa informació que la primera però capaç de sortir del nucli per així traduir-se.

Generalment l'ARN que es forma codifica per proteïna i l'anomenen ARN missatger. Existeixen altres tipus d'ARN que no codifiquen per proteïna per exemple l'ARN ribosòmic que forma part dels ribosomes o l'ARN transferidor que és el responsable d'activar i col·locar els aminoàcids corresponents durant la síntesi de proteïnes. Aquest procés es dona dins el nucli cel·lular i el porta a terme l'enzim ARN polimerasa.

2.1.1 Procés de la transcripció

La transcripció consisteix en quatre etapes: la iniciació, la síntesi de la cadena d'ARN, la terminació i la maduració de l'ARN missatger.

La iniciació comença quan l'enzim ARN polimerasa reconeix el centre promotor i s'hi uneix. El centre promotor és una regió del gen on es regula la seva expressió.

Seguidament l'ARN polimerasa inicia la síntesi de l'ARN missatger col·locant ribonucleòtids a l'extrem 3' de la cadena motlle d'ADN, construint així una única cadena en direcció 5' → 3'. La cadena motlle és la que serveix com a patró a l'ARN polimerasa per sintetitzar l'ARN missatger. En canvi la cadena codificadora coincideix en l'orientació i la seqüència de l'ARN produït per la motlle, excepte per la substitució de timines per uracils. L'híbrid que s'està formant entre l'ARN missatger i l'ADN és inestable per això es va desenganxant de la cadena d'ADN a mesura que avança la transcripció.

En la fase de terminació l'ARN polimerasa reconeix la seqüència de terminació del gen i allibera l'ARN missatger.

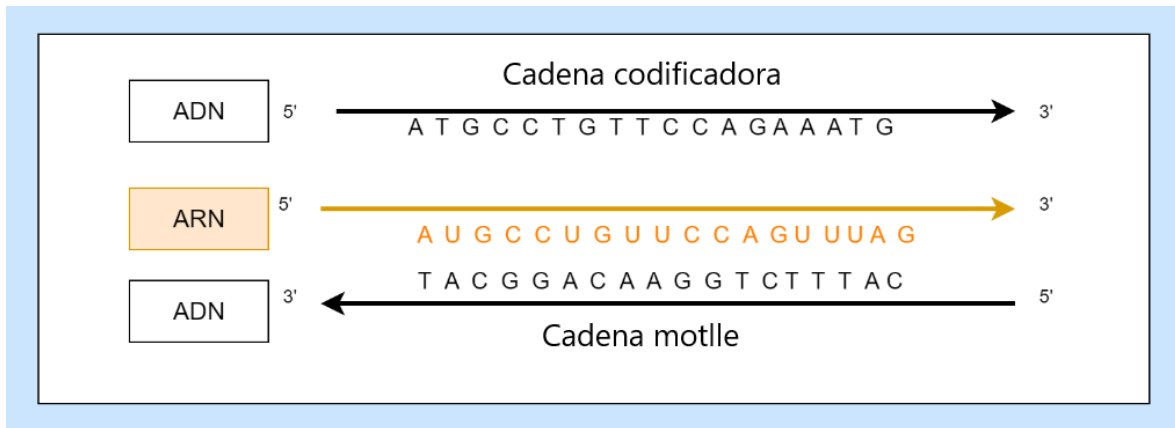


Figura 4: Transcripció d'ADN a ARN missatger
 Font: Pròpia

2.1.2 Els introns, exons i el procés de splicing

En la última fase de la transcripció, la maduració de l'ARN missatger, es retiren seqüències de la molècula que no s'utilitzaran en la síntesi de la proteïna. Els fragments que es retiren s'anomenen **introns** i les que s'uneixen per formar l'ARN missatger madur són els **exons**. Aquest procés s'anomena splicing. Perquè el procediment funcioni en l'ADN es troben unes senyals que determinen el principi i el final dels introns, d'aquesta manera quan la cèl·lula les detecta pot tallar l'ARN missatger en el lloc correcte. La senyal que es troba al principi de l'intró són els dos nucleòtids **gt** i s'anomena splicing donor site. La senyal del final és **ag** i s'anomena splicing acceptor site.

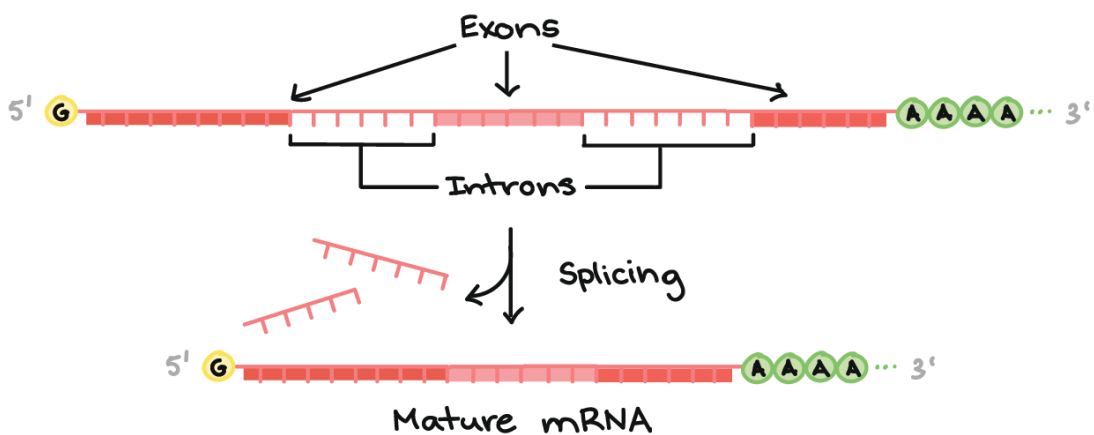


Figura 5: Procés de splicing

Font: <https://www.khanacademy.org/science/biology/gene-expression-central-dogma/transcription-of-dna-into-rna/a/eukaryotic-pre-mrna-processing>

2.1.3 Splicing alternatiu

És molt curiós que el genoma humà contingui entre 20.000 i 25.000 gens però es fabriquin 90.000 proteïnes, com és possible això? S'ha vist que alguns gens fan més d'un procés de splicing, o sigui que no sempre s'ajunten tots els exons per formar l'ARN missatger madur. Per tant el gens amb aquestes característiques podran sintetitzar diferents proteïnes depenent de com combinin els exons per formar l'ARN missatger madur. Aquest tipus de splicing s'anomena **splicing alternatiu** i té com a resultat diferents ARNs missatgers madurs generats per un sol gen, és a dir diferents proteïnes.

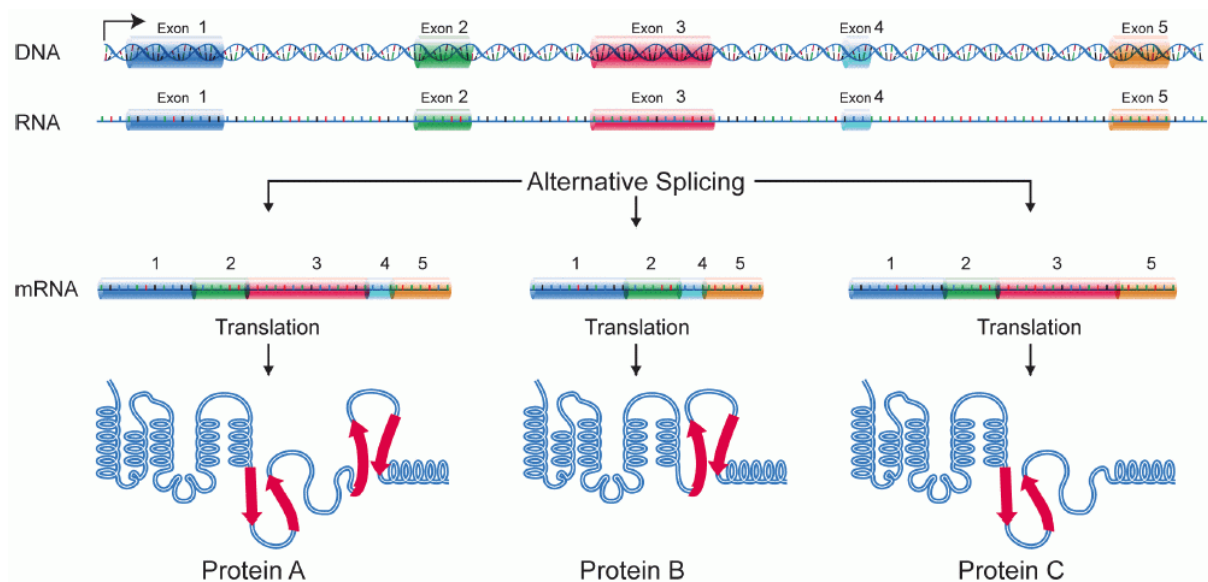


Figura 6: Procés de splicing alternatiu

Font: <https://www.khanacademy.org/science/biology/gene-expression-central-dogma/transcription-of-dna-into-rna/a/eukaryotic-pre-mrna-processing>

2.1.4 Seqüència Codificadora (CDS)

Dins de l'ARN missatger madur existeix la seqüència **CDS** (CoDifying Sequence) que conté els triplets del codi genètic que codifiquen per un aminoàcid. És a dir, la CDS té la seqüència de nucleòtids que serà traduïda a proteïna. Les altres regions de l'ARN missatger madur tenen altres funcions. L'extrem 5' de la cadena s'anomena UTR (UnTranslated Region) i serveix perquè el ribosoma es pugui unir a

l'ARN i començar la traducció. També hi ha una seqüència bastant llarga d'adenines a l'extrem 3' anomenada PolyA, aquesta es creu que determina el nombre de vegades que un ARN missatger pot ser traduïda per diferents ribosomes.

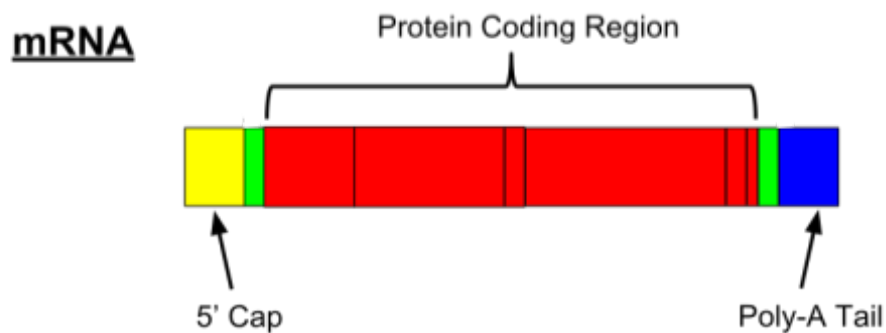


Figura 7: Parts de l'ARN missatger on es mostra la seqüència codificadora
Font: https://en.wikipedia.org/wiki/Non-coding_DNA

2.2 Traducció d'ARN a proteïna

Quan l'ARN missatger ja ha passat per el procés de maduració surt del nucli i es dirigeix al citoplasma on s'unirà als ribosomes. Allà serà on es sintetitzaran les proteïnes mitjançant el procés de traducció.

La traducció és el procés pel qual la informació codificada de l'ARN missatger és descodificada i utilitzada per sintetitzar una cadena d'aminoàcids.

La informació que porta l'ARN està codificada en un llenguatge format per les lletres A, C, G i U que representen les quatre bases nitrogenades mentre que la seqüència de les proteïnes es construeix amb 20 aminoàcids diferents. Les normes que determinen la correspondència entre bases nitrogenades de l'ARN i aminoàcids d'una proteïna s'anomena codi genètic.

2.2.1 El codi genètic

		Second letter					
		U	C	A	G		
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G	
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G	
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G	
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G	

Figura 8: Codi genètic

Font: <https://www.khanacademy.org/science/biology/gene-expression-central-dogma/central-dogma-transcription/a/the-genetic-code-discovery-and-properties>

El codi genètic es caracteritza perquè cada tres nucleòtids d'ARN codifiquen per un aminoàcid. Aquests grups de tres s'anomenen codons. Per tan en el codi genètic hi ha 64 codons que corresponen a 20 aminoàcids, això significa que un aminoàcid no sempre és codificat per un mateix codó, en canvi un codó sempre codifica per el mateix aminoàcid. Per exemple l'aminoàcid treonina (Thr) pot ser codificat pels codons ACU, ACC, ACA i ACG però aquest codons sols poden codificar per l'aminoàcid treonina.

El codi genètic conté un codó d'inici de proteïna, aquest sempre és AUG el que codifica per metionina (Met), i els codons de stop, que finalitzen la síntesi, són UAA, UAG i UGA. El codi genètic és universal, tots els éssers vius utilitzen el mateix.

2.2.2 Procés de la traducció

Les etapes de la traducció són l'activació dels aminoàcids, la iniciació, la elongació, la terminació i la maduració de la proteïna.

En la primera etapa els aminoàcids que es troben al citoplasma s'uneixen a l'ARN de transferència que els correspon i es dirigeixen cap al ribosoma.

En la iniciació l'ARN missatger arriba al ribosoma i aquest reconeix el primer codó AUG que sempre codifica per metionina.

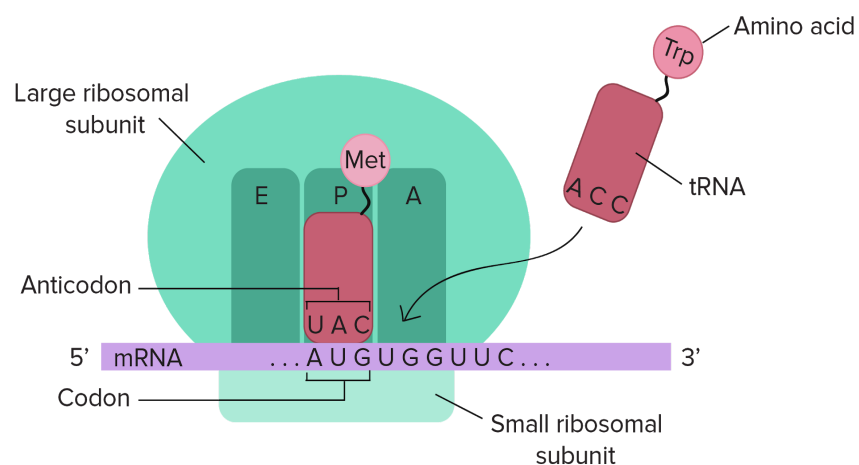


Figura 9: La iniciació, etapa dos de la traducció

En la fase d'elongació els ARN de transferència que contenen l'anticodó complementari a l'ARN missatger van entrant al ribosoma i dipositen l'aminoàcid que porten formant la cadena polipeptídica.

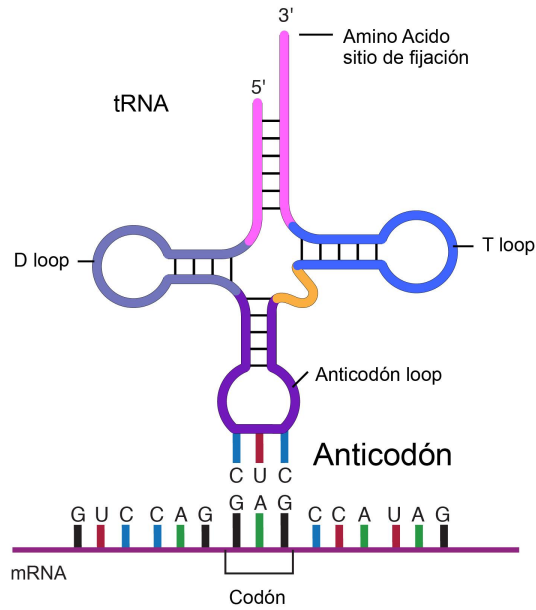


Figura 10: Un anticodó complementant amb el codó de la seqüència d'ADN
 Font: <https://www.genome.gov/es/genetics-glossary/Anticodon>

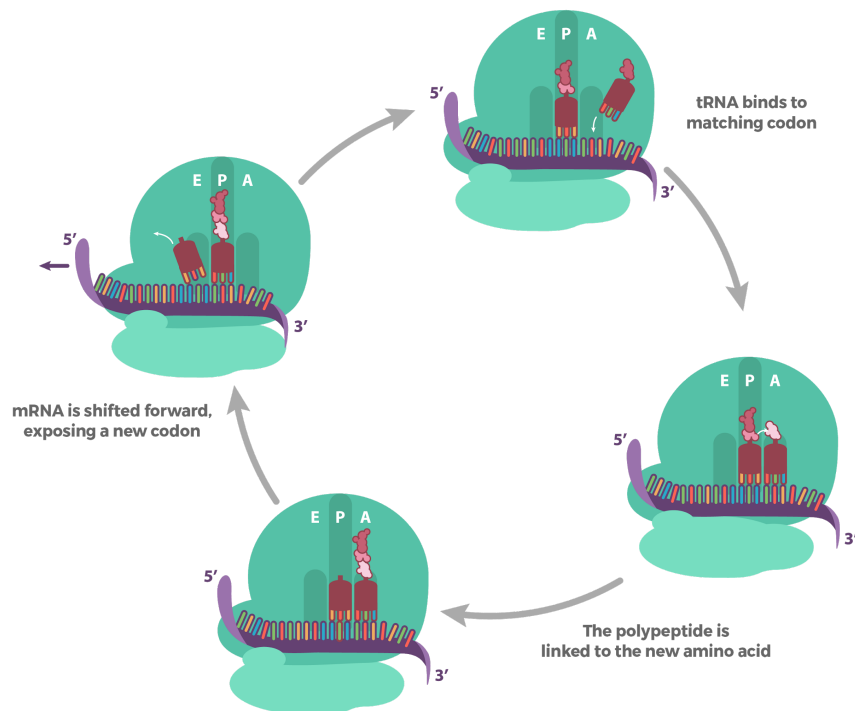


Figura 11: La elongació, fase tres de la traducció

Font: <https://www.khanacademy.org/science/biology/gene-expression-central-dogma/translation-polypeptides/a/translation-overview>

Al final de l'ARN missatger hi ha el codó STOP, quan es detecta aquest codó entra el ribosoma el factor de terminació i provoca l'alliberament de la cadena polipeptídica.

En l'última etapa de la traducció la cadena polipeptídica es replega adoptant l'estructura que li donarà la funcionalitat.

2.3 Mutacions

Una mutació és una alteració a l'atzar en el material genètic que aporta variabilitat a la població i permet l'evolució de les espècies. Les mutacions poden ser de diferents tipus segons l'extensió del material genètic afectat. Existeixen les mutacions gèniques que només afecten la seqüència de nucleòtids d'un gen, les cromosòmiques que afecten una seqüència de gens d'un cromosoma i genòmiques que alteren el nombre de cromosomes. De les mutacions gèniques n'existeixen dos tipus: les mutacions per substitució de bases i les mutacions per pèrdua o inserció de nucleòtids.

Les mutacions per substitució de base només és un intercanvi d'una base nitrogenada per una altre. Aquestes alteren un sol triplet i pot ser des de molt greu a totalment inofensiva. Si el triplet afectat pertany a un exó pot provocar que la proteïna que codifica sigui afuncional, de la mateixa manera que si el triplet afectat codifica per STOP i no ho hauria de fer, provocarà l'aturada de traducció i la proteïna quedarà incompleta. En canvi si el triplet afectat no és cap dels dos casos anteriors la mutació serà silenciosa i inofensiva.

Consecuencias de una sustitución: Cambio de un aminoácido por otro.

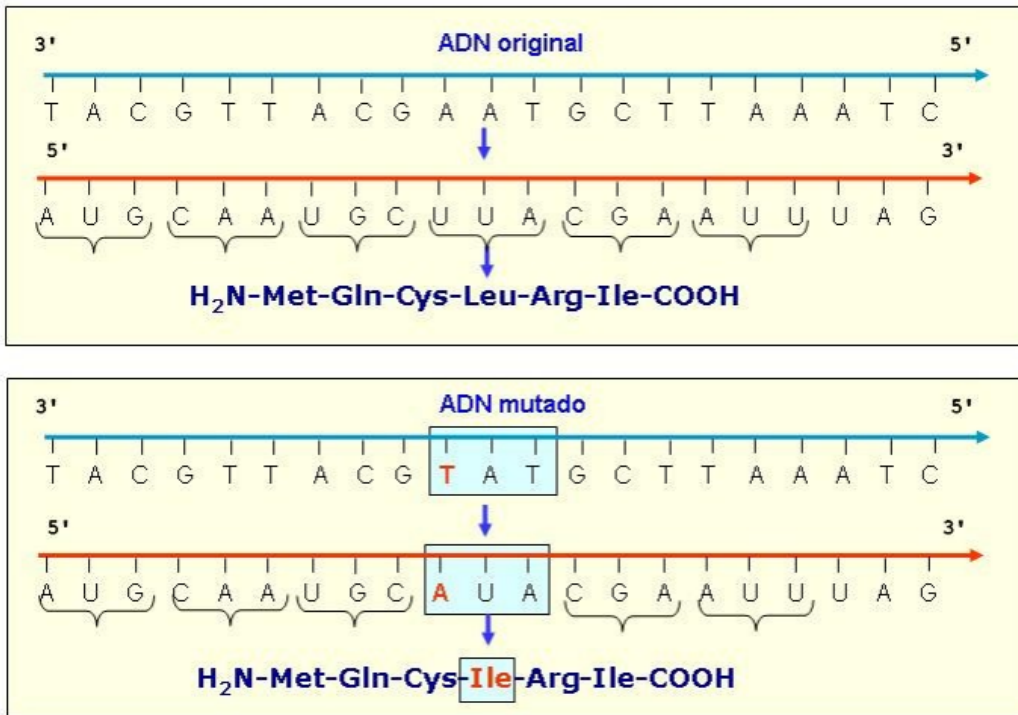


Figura 12: Mutació gènica de substitució que té com a conseqüència un canvi d'aminoàcid.

Consecuencias de una sustitución: Generación de una tripleta de stop.

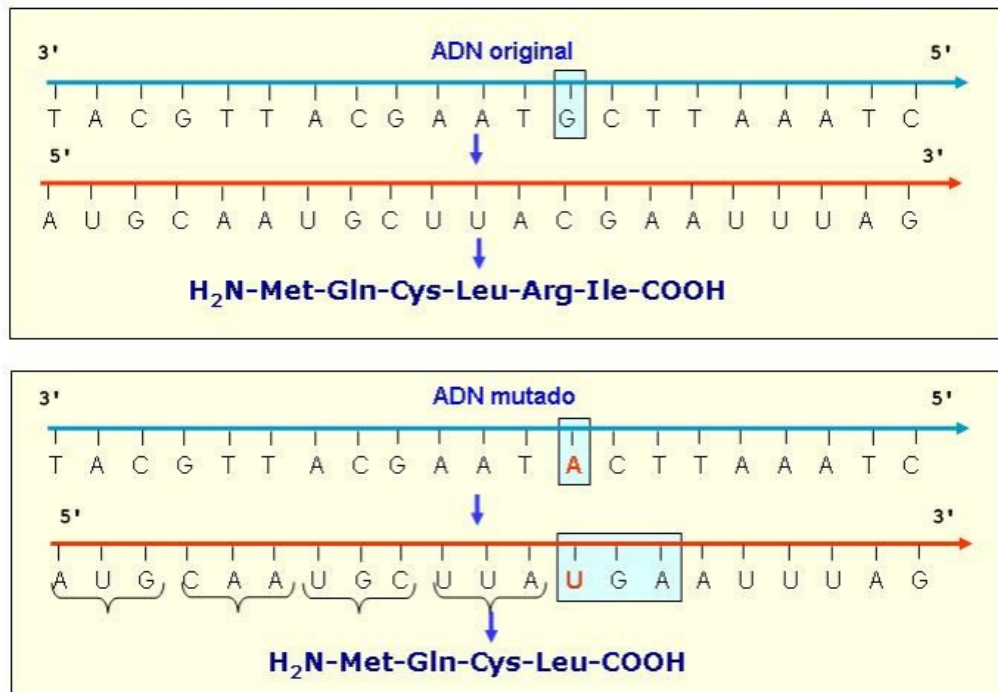


Figura 13: Mutació gènica de substitució que té com a conseqüència la generació d'un codó de stop.

En les mutacions per pèrdua de nucleòtids s'eliminen nucleòtids en la seqüència això significa que tots els nucleòtids correran un espai en el seu ordre i s'alteraran tots els triplets que venen després de la mutació. Aquestes mutacions són molt greus perquè la proteïna no es forma correctament.

En les mutacions per inserció de nucleòtid s'afegeixen bases nitrogenades en la seqüència, les conseqüències són les mateixes que en el cas de pèrdua de nucleòtid.

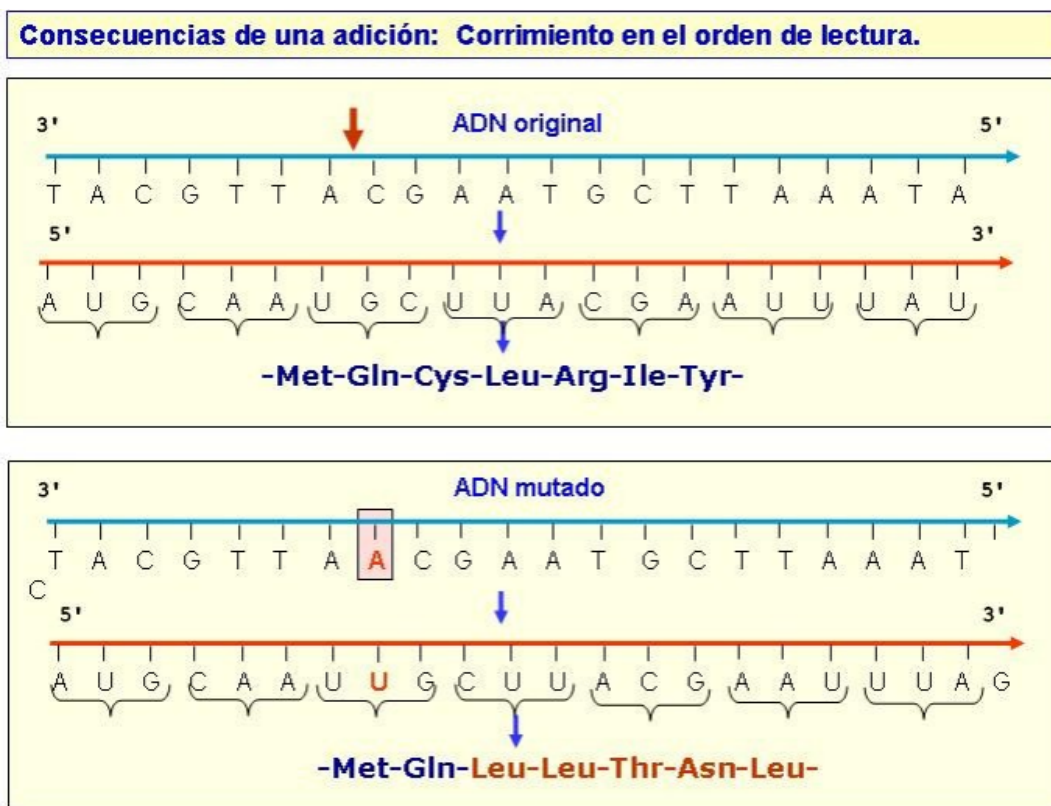


Figura 14: Mutació gènica de incesió o adició de nucleòtid.

Font: http://cosmolinux.no-ip.org/recursos_aula/BIO1erBAT/Genetica_molecular/Unitat_11_Les_mutacions1.pdf

3. Reacció en cadena de la polimerasa (PCR)

La reacció en cadena de la polimerasa, també dita PCR, és una tècnica que permet amplificar exponencialment un segment d'ADN d'interès fins obtenir milions de còpies.

Per fer una PCR es necessiten els següents elements: un fragment d'ADN, nucleòtids, l'enzim ADN polimerasa i uns **iniciadors** o primers. Aquestes molècules es posen en un tub d'assaig per després introduir-se dins un termociclador que permet realitzar els cicles d'amplificació repetits en l'ADN. Cada cicle és compost per tres passos: la desnaturalització, l'alineació i l'elongació.

En el primer s'eleva la temperatura fins als 94 graus centígrads durant un minut, en aquest temps i temperatura les dues cadenes d'ADN es desnaturalitzen, es separen. En la segona etapa es baixa a una temperatura d'entre 50 i 60 graus per més o menys un minut i mig, on els iniciadors s'uneixen als segments d'ADN complementari, un a cada cadena. Finalment es torna a pujar la temperatura a 72 graus durant un o dos minuts. En aquesta última part les molècules d'ADN polimerasa s'adjunten en els iniciadors i comencen a replicar l'ADN a partir d'afegir nucleòtids en sentit 3' → 5' complementaris a la cadena original d'ADN. Per cada cicle les molècules d'ADN es dupliquen per tan després d'uns 30 o 40 cicles ja hi haurien milions de fragments d'ADN iguals entre si. En aquest procés és molt importat l'elecció dels iniciadors ja que determinen la regió de l'ADN que s'amplificarà.

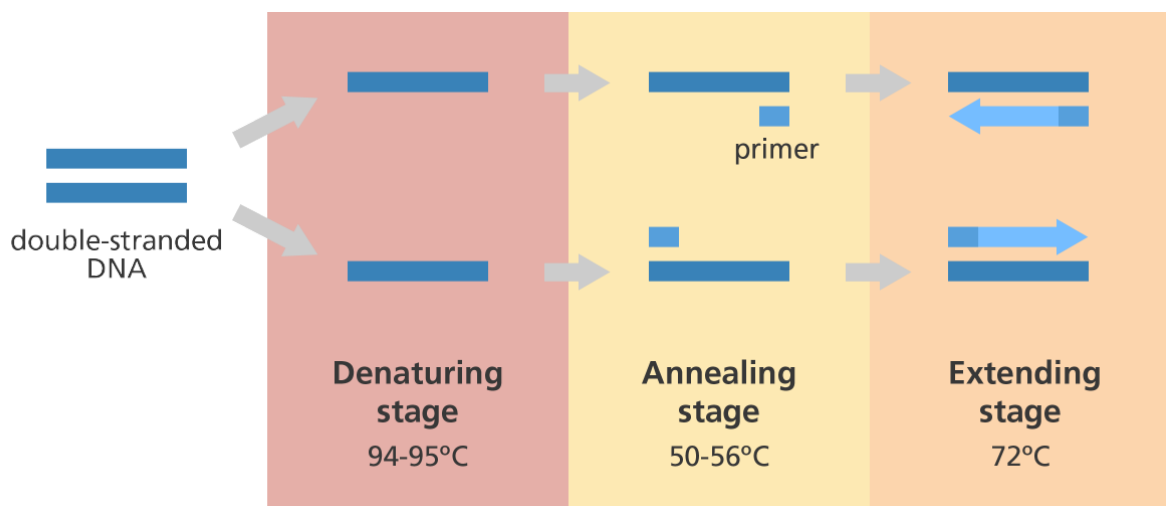


Figura 15: Passos d'un cicle de PCR

Font: <https://www.yourgenome.org/facts/what-is-pcr-polymerase-chain-reaction>

4. La tècnica seleccionada: tecnologia CRISPR

La tecnologia **CRISPR** és una tècnica d'edició genètica que s'utilitza per modificar i editar el genoma de qualsevol cèl·lula inhabilitant o corregint un gen per aconseguir que la cèl·lula realitzi o deixi de realitzar una funció. Aquesta tècnica actualment és la més ràpida, barata, precisa i eficient que existeix, tot i així de moment només es pot posar en pràctica en cèl·lules en cultiu. El realment innovador i sorprenent d'aquesta tècnica és que permet tallar l'ADN en un lloc molt concret i la mateixa cèl·lula és capaç de reparar la ruptura, aconseguint així, que la cèl·lula adquireixi les propietats desitjades amb la finalitat de curar d'arrel una malaltia genètica, curar càncers o fins i tot eliminar la resistència als antibiòtics que tenen alguns bacteris. En aquest treball s'ha descrit el funcionament de l'eina CRISPR en base a la correcció d'una malaltia.

4.1 CRISPR en els bacteris

Les CRISPR provenen del sistema immunitari que utilitzen els bacteris per fer front a les invasions de virus i plasmidis. Les seves sigles signifiquen *Clustered Regulatory Interspaced Short Palindromic Repeats* (Agrupacions Regularment Intercalades de Repeticions Palindròmiques Curtes). Les Repeticions Palindròmiques Curtes fan referència a uns fragments curts i iguals entre si d'ADN que formen una doble hèlix de dues cadenes palindròmiques¹, i que es van repetint una després de l'altre. Les Agrupacions Regularment Intercalades són unes altres seqüències, però aquestes són diferents entre si ja que són fragments d'ADN que provenen dels diferents virus que han atacat el bacteri al llarg de la seva vida. Aquests fragments són incorporats en el CRISPR intercalats entre les repeticions palindròmiques curtes i s'anomenen espaiadors.

¹Palindròmica: Seqüència d'àcids nucleics que es llegeix igual en sentit 3' → 5' que 5' → 3'

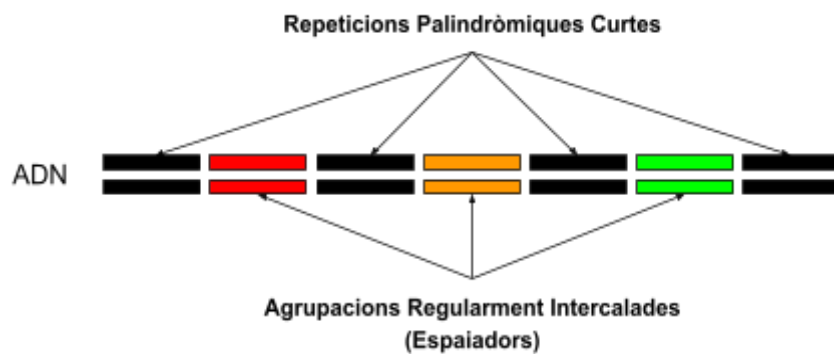


Figura 16: El CRISPR dins el genoma bacterià Font: Pròpia

Si els bacteris no tinguessin sistema immunitari el virus dipositaria el material genètic dins del bacteri i aquest es replicaria fins a matar-lo. Però com que aquests tenen el sistema CRISPR, poden defensar-se dels virus invasors. El sistema funciona de la següent manera: quan entra per primera vegada el material genètic d'un virus desconegut dins del bacteri, aquest produeix les endonucleases² i helicases³ Cas1 i Cas2 que tallen per unes regions curtes de l'ADN víric anomenades **PAM** (Protospacer Adjacent Motif). Després el bacteri captura el tros d'ADN localitzat després del PAM, anomenat protoespaiador, i l'incorpora enmig de les seqüències repetides per tal de recordar aquest virus si torna a entrar. Un cop el protoespaiador està incorporat dins el genoma bacterià passa a dir-se espaiador.

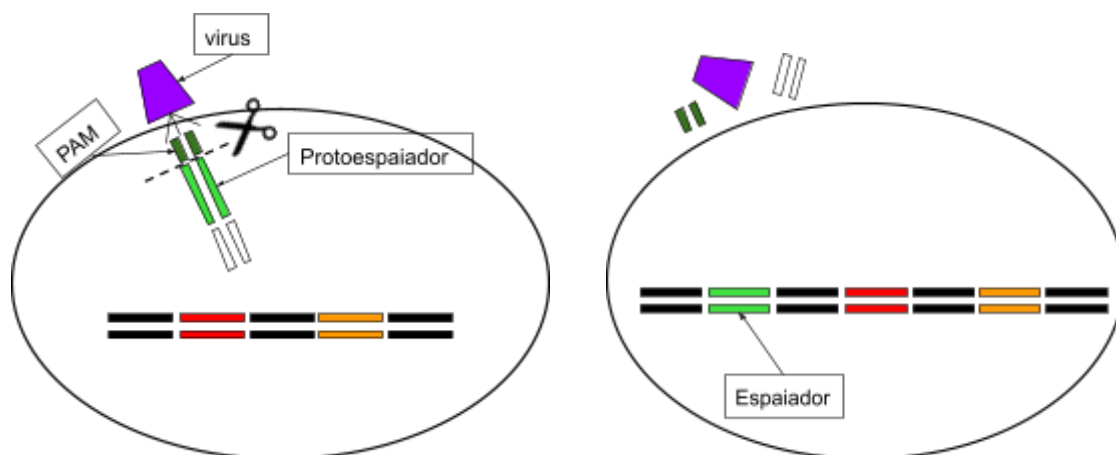


Figura 17: La introducció de l'ADN viral dins un bacteri i aquest introduir-lo dins el seu CRISPR. Font: Pròpia

² Endonucleasa: Enzim capaç de tallar àcids nucleics

³ Helicasa: Enzim capaç de desenrotllar una doble cadena d'àcid nucleics.

Quan el virus diposita el seu material genètic dins el bacteri per segona vegada, aquest és capaç de reconèixe'l i generar una resposta immunitària d'interferència. Primer el bacteri transcriu tot el CRISPR en un sol ARN. Després aquesta cadena es divideix en fragments que contenen un espaiador i una seqüència repetida. Aquest fragment d'ARN s'anomena **ARN crispr** (ARNcr) i porta la seqüència que reconeixerà i complementarà amb un ARN viral ja conegut. El bacteri també transcriu el **tracrARN** que exerceix la funció estructural ja que uneix l'ARNcr i la proteïna Cas9. Finalment el bacteri sintetitza l'endonucleasa i helicasa Cas9 (proteïna associada al CRISPR) encarregada de tallar en el lloc que l'ARN crispr li ha indicat. El complex d'aquests tres elements és **capaç de reconèixer l'ADN invasor i tallar-lo**, destruint d'aquesta manera el virus.

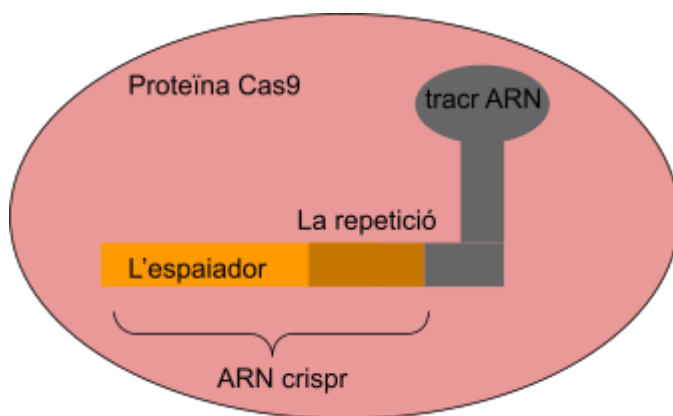


Figura 18: Esquema del complex format per l'ARN crispr, el tracrARN i la proteïna Cas9
Font: Pròpia

L'endonucleasa Cas9 a l'hora de tallar l'ADN viral es podria confondre i tallar l'ADN del propi bacteri ja que el protoespaiador del virus i l'espaiador en el genoma bacterià tenen la mateixa seqüència. Això no passa gràcies a la seqüència PAM. Quan la Cas9 la reconeix sap que és el lloc correcte per on tallar.

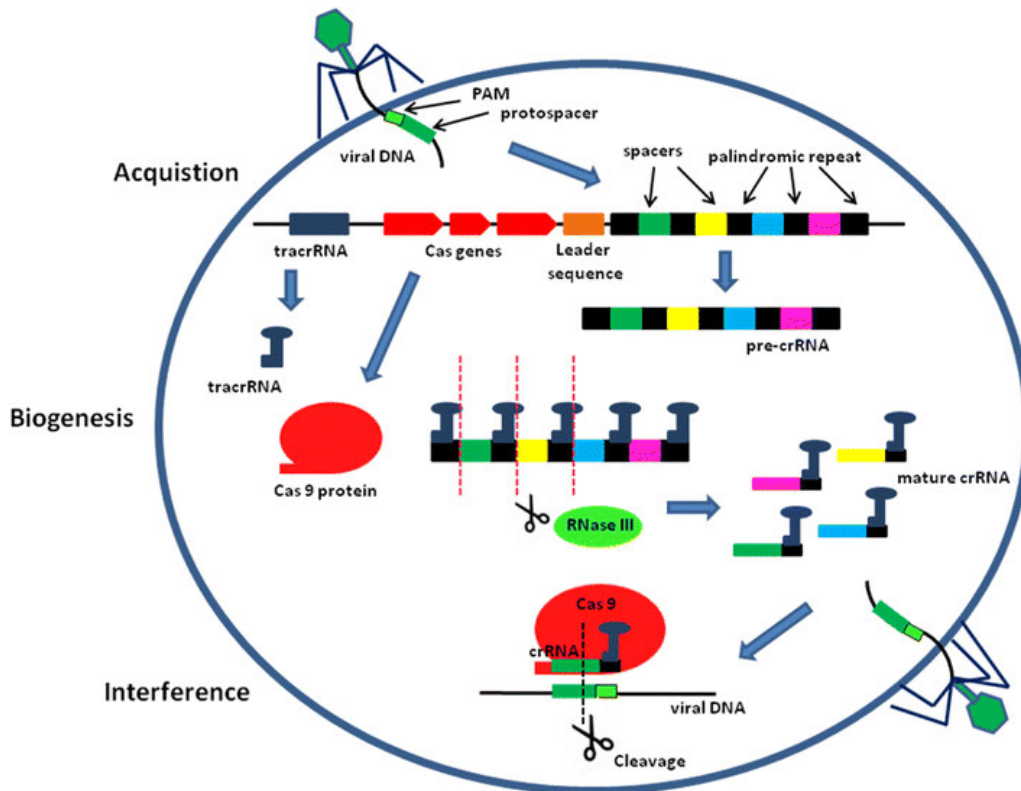


Figura 19: La reacció del sistema immunitari bacterià

Font: https://www.researchgate.net/figure/The-three-stages-of-the-CRISPR-Cas-bacterial-adaptive-immune-system-acquisition-crRNA_fig2_308870153

4.2 La tècnica CRISPR com a eina biotecnològica i d'edició genètica

A partir de l'estudi del sistema immunitari bacterià s'ha vist que la **resposta immunitària** que dona el bacteri quan entra un virus conegut es pot adaptar per utilitzar-se en la modificació del genoma de les cèl·lules, d'aquesta manera s'ha pogut desenvolupar la tècnica biotecnològica CRISPR. Aquesta està composta per dos elements moleculars, la endonucleasa i helicasa **Cas9** i un **ARN guia**. La molècula d'ARN guia està formada per la unió de l'ARNcr i l'ARNtracr per tal de simplificar l'eina, les dues molècules segueixen fent la funció que feien quan formaven part del sistema immunitari bacterià però ara estan unides.

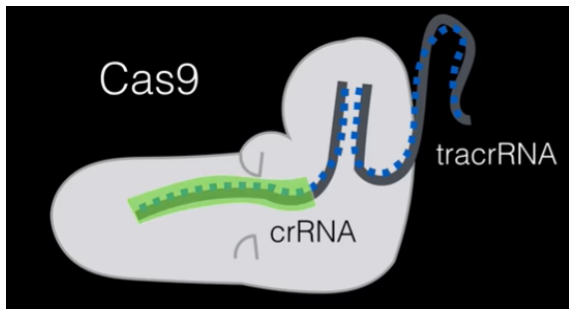


Figura 20:

Figura 20: Esquema del complex proteïna Cas9-crRNA-tracrRNA que actua de manera natural en la resposta immunitària dels bacteris contra una invasió viral.

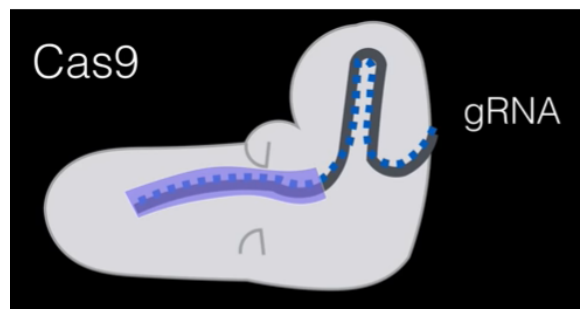


Figura 21:

Figura 21: Esquema del complex ARN guia-Cas9 que ha estat adaptada per actuar com a eina d'edició genètica.

Font: <https://www.youtube.com/watch?v=MnYppmstxIs>

La part de l'extrem 5' de l'ARN guia (seria la part de l'ARNcr en els bacteris) està formada per una seqüència de 23 nucleòtids complementaris a la **seqüència diana**⁴ de l'ADN. Són 23 perquè s'ha arribat a la conclusió que 23 nucleòtids és una mida on la probabilitat de trobar dues seqüències iguals en tot el genoma és molt baixa i si la seqüència és més gran que 23 ja és més difícil de fer.

La molècula de l'ARN guia és molt important ja que es pot modificar la seva seqüència segons la diana que es decideixi complementar, aconseguint així canviar el gen o la part del gen que es vol editar.

⁴Seqüència diana: Seqüència específica de l'ADN reconeguda per una molècula, en aquest cas l'ARN guia

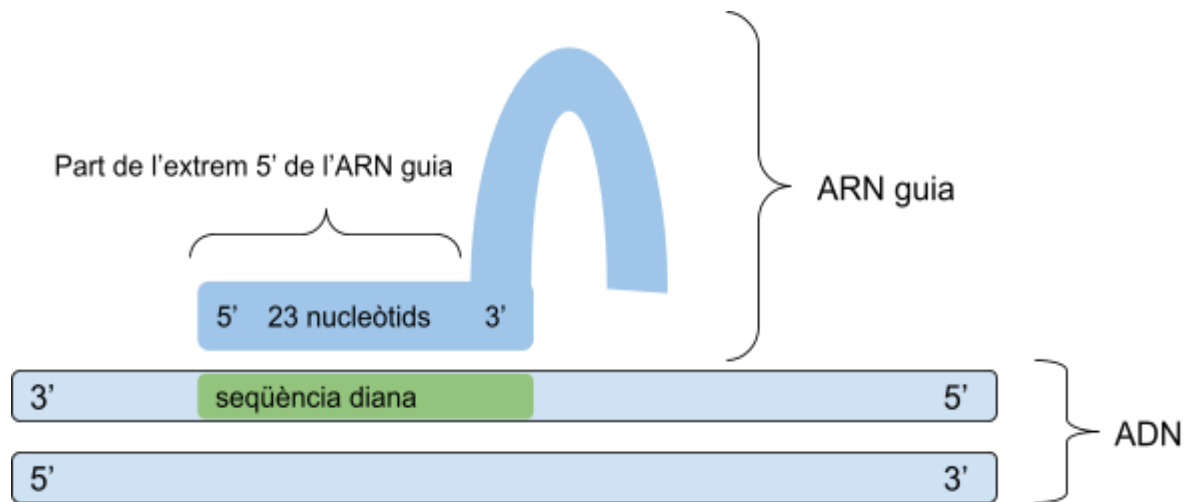


Figura 22: Esquema de l'ARN guia on la part de l'extrem 5' està complementant amb la seqüència diana

Font: pròpia

La part de l'extrem 3' (seria l'ARNtracr en els bacteris) és una seqüència fixa que subjecta l'ARN guia amb la Cas9.

La proteïna Cas9 fa la funció de desenrotllar, obrir i tallar l'ADN en el punt concret on indica l'ARN guia. També ajuda a l'ARN guia a trobar la seqüència diana més ràpidament i efectivament a través de localitzar la seqüència PAM més propera a la mutació. Aquesta seqüència d'entre dues i sis bases, ve determinada per la proteïna que s'utilitza, en el cas de la proteïna Cas9 de *S. pyogenes*, una de les més utilitzades en el CRISPR, localitza la seqüència NGG en sentit downstream (5'-NGG-3') sent N una base nitrogenada qualsevol. La localització d'aquesta seqüència orienta al complex cap a la seqüència diana.

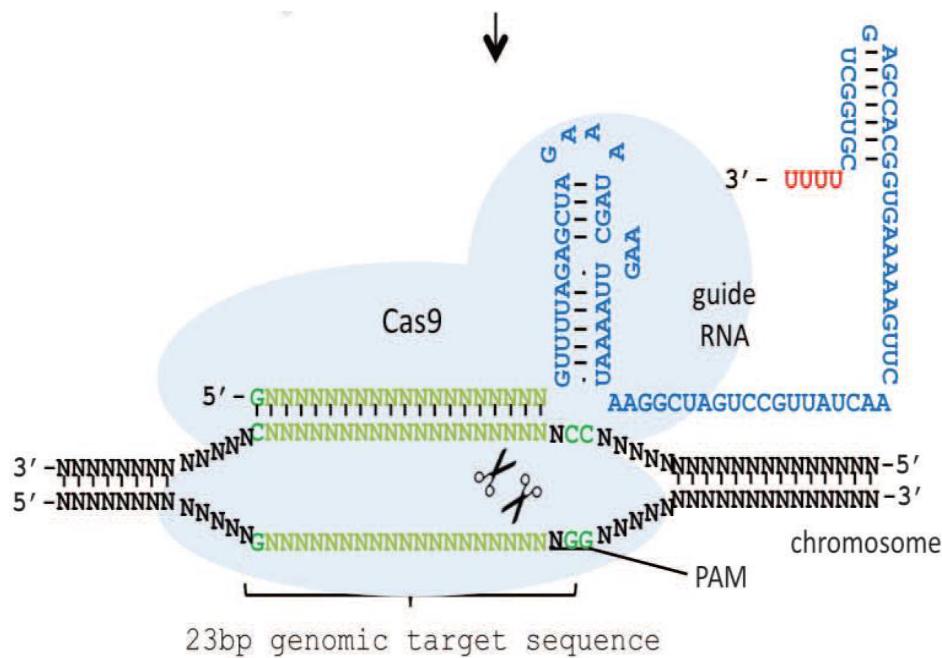


Figura 23: L'ARN guia i la proteïna Cas9 actuant sobre l'ADN

Font: Estudi de l'ADN i edició genètica. A. Fontarnau

Quan el complex entra dins del nucli cel·lular aquest comença a fer les seves funcions. La proteïna Cas9 localitzarà la seqüència PAM i orientarà l'ARN guia cap a la seqüència diana d'ADN, on es troba la mutació. Quan el complex hagi detectat el segment la Cas9 obrirà les dues cadenes d'ADN i els nucleòtids de l'ARN guia complementaran amb els de l'ADN, formant un híbrid. Finalment la proteïna Cas9 tallarà les dues cadenes d'ADN en un extrem de la seqüència diana.

4.3 Mecanismes de reparació d'una ruptura en la doble cadena d'ADN

Quan es forma una ruptura en la doble cadena la cèl·lula es troba en perill ja que el material genètic no està en la seva totalitat sinó que està trencat, això podria provocar la mort cel·lular. En aquests casos extrems la cèl·lula activa un dels mecanismes de reparació d'una ruptura de doble cadena en l'ADN.

Els dos mecanismes més comuns alhora de reparar ruptures de la doble cadena d'ADN són la recombinació homòloga (homologous recombination) i recombinació

no homòloga (non-homologous end joining) . La recombinació homòloga consisteix en agafar un tros d'ADN igual al danyat però sense mutació i utilitzar-lo com una plantilla per reconstruir l'ADN. En aquest cas els propis enzims de la cèl·lula eixamplen la ruptura així les dues molècules s'entrecreuen i restauren l'ADN danyat amb parts de la cadena homòloga. Si en la molècula d'ADN plantilla s'introdueix la base nitrogenada correcta la mutació quedaria arreglada.

La recombinació no homòloga no és un mètode tan precís com el primer ja que no hi ha una cadena homòloga per reconstruir la ruptura, en aquest cas els enzims intenten eliminar unes quantes bases nitrogenades del voltant de la ruptura i n'afegeixen d'altres per aconseguir tornar ajuntar la cadena. Aquest mecanisme normalment crea més mutacions i fa disfuncional el gen. Això pot ser bo quan l'objectiu és desactivar gens i que deixin de fer les seves funcions.

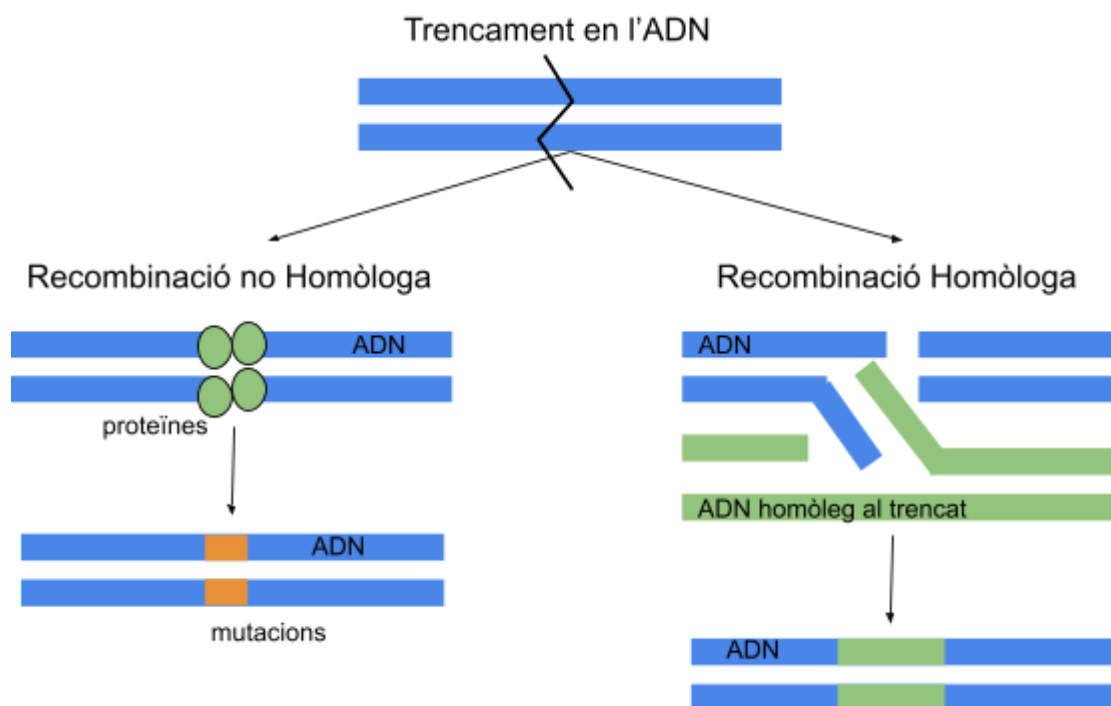


Figura 24: Esquema del funcionament dels dos mecanismes de reparació que porta a terme la cèl·lula quan es produeix un dany en la doble cadena: la recombinació homòloga i la recombinació no homòloga.

Font: pròpia

5. La malaltia seleccionada: Síndrome de la progèria de Hutchinson-Giford

5.1 Dades clíniques i epidemiològiques de la progèria

El síndrome de la **progèria** de Hutchinson-Giford és una malaltia genètica que es caracteritza per l'envelliment brusc i prematur. Les persones que la pateixen presenten baixa estatura, baix pes, caiguda de cabells prematura, poca mobilitat en les articulacions, pèrdua de greix a la pell, problemes cardiovasculars, trets facials que recorden a persones envellides, etc. S'ha vist que la progèria tot i afectar a molts sistemes no afecta al desenvolupament intel·lectual i les habilitats motores del malalt.



Figura 25: Un infant amb progèria.
Un nucli cel·lular sa i un deformat
Font: <https://en.wikipedia.org/wiki/Progeria>

Aquest síndrome sorgeix en la infantesa i és molt poc freqüent, només apareix en 1 de cada 4 milions d'infants acabats de néixer. Actualment hi ha 127 casos en tot el món. La majoria dels pacients tenen la malaltia sense haver-la heretat dels seus progenitors, tot hi així sí que és una malaltia genètica. Els afectats normalment moren de manera natural abans dels 20 anys, encara que hi ha hagut algun cas que ha sobreviscut fins als 40. No existeix ni cura ni tractament per la progèria tot i així s'està estudiant perquè es produeix aquest envelliment brusc i com funciona el procés d'envelliment. (Hennekam RCM. 2006. Extret OMIM)

5.2 Dades metabòliques i genètiques

El gen **LMNA** sintetitza les proteïnes lamin A i lamin C per splicing alternatiu. Però només algunes mutacions que afecten a la proteïna **lamin A** acaben provocant la

progèria, la lamin C no es veu involucrada en la malaltia. Les dues proteïnes formen part d'un grup de lamines anomenades lamines tipus A. També existeix un altre grup anomenat lamines tipus B, aquest està format per les proteïnes lamin B1 i lamin B2 sintetitzades per dos gens diferents el LMNB1 i el LMNB2. El conjunt d'aquestes proteïnes, també denominades lamines nuclears, formen una estructura que recobreix l'interior de la membrana nuclear anomenada làmina nuclear.

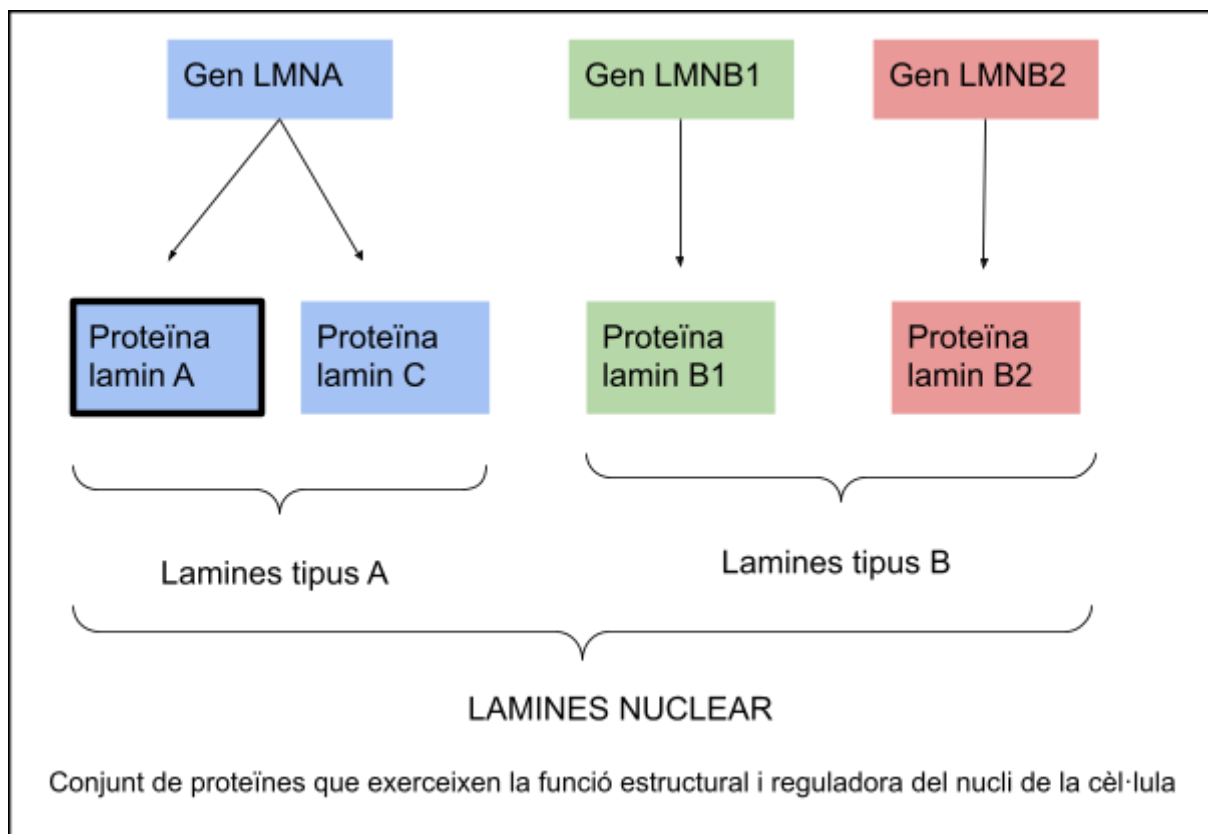


Figura 26: Esquema dels tipus de proteïnes Lamines que existeixen i quin gen les sintetitza.
Font: Pròpia

Les lamines nuclears són les encarregades de **mantenir la forma del nucli així com la seva estabilitat**. Juntament amb altres proteïnes ajuden a encaixar el reticle endoplasmàtic amb el nucli i interactuen amb el citoesquelet del reticle endoplasmàtic. També regulen el cicle cel·lular, aquestes proteïnes fan una funció important en la profase per desfer el nucli i en la telofase en tornar-se a formar l'embolcall. A més a més regulen l'organització del genoma i l'expressió d'alguns gens.

En la progèria com que es produeix una proteïna lamin A no funcional l'embolcall nuclear és inestable i va danyant la cèl·lula progressivament, produint així que la cèl·lula mori prematurament. (Dittmer, T.A., Misteli, T. 2011) (B.M. Paschal, J.B. Kelley 2013)

5.3 Gen responsable

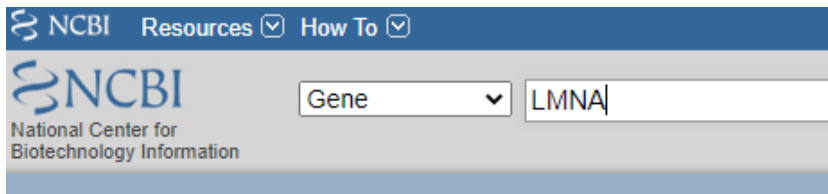
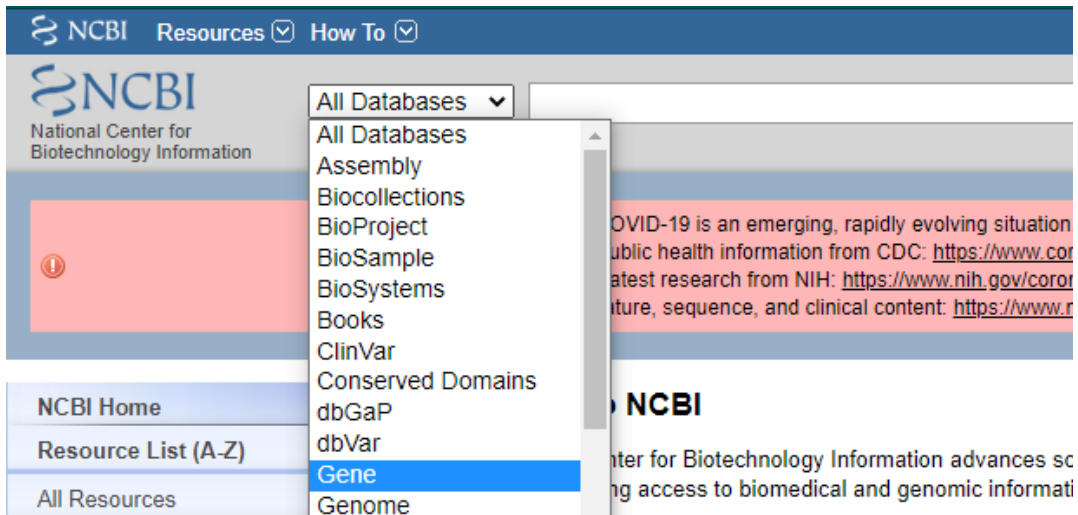
La informació del gen responsable de la progèria s'ha extret del portal NCBI se'n explica el seu funcionament a continuació juntament amb les dades d'aquest gen.

5.3.1 Portal NCBI

La **NCBI** (*National Center for Biotechnology Information*) és un portal o conjunt de bases de dades d'accés públic on s'emmagatzemen dades rellevants per la biotecnologia i la biomedicina. Dins de l'NCBI hi ha una base de dades anomenada GenBank que és on s'hi troba la informació referent a les seqüències genòmiques i s'encarrega d'emmagatzemar i actualitzar les dades de tots els gens coneguts de diferents espècies. La NCBI i sobretot la base GenBank són les bases que s'han utilitzat per buscar la informació del gen LMNA a més de l'ARNm i les proteïnes que sintetitzen.

Com funciona la NCBI

Per trobar un gen en la base es comença per entrar a la web de l'NCBI (<https://www.ncbi.nlm.nih.gov/>) i en el buscador seleccionar *Gene*, per tal d'acotar la cerca a només gens, i en el costat escriure el gen que es vol buscar. En el cas del gen estudiat s'ha d'introduir LMNA en el buscador.



Seguidament s'obrirà una taula on hi ha els gens LMNA de diferents organismes, però només s'està interessat en gens humans, s'ha de clicar sobre *d'Homo sapiens*.

Search results

Items: 1 to 20 of 1665

<< First < Prev Page 1 of 84 Next >> Last >>

[See also 9 discontinued or replaced items.](#)

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> LMNA ID: 4000	lamin A/C [<i>Homo sapiens</i> (human)]	Chromosome 1, NC_000001.11 (156082573..156140081)	CDCD1, CDDC, CMD1A, CMT2B1, EMD2, FPL, FPLD, FPLD2, HGPS, IDC, LDP1, LFP, LGMD1B, LMN1, LMNC, LMNL1, MADA, PRO1	150330
<input type="checkbox"/> Lmna ID: 16905	lamin A [<i>Mus musculus</i> (house mouse)]	Chromosome 3, NC_000069.6 (88481148..88509932, complement)	Dhe	
<input type="checkbox"/> Lmna ID: 60374	lamin A/C [<i>Rattus norvegicus</i> (Norway rat)]	Chromosome 2, NC_005101.4 (187842884..187863552, complement)		
<input type="checkbox"/> lmna ID: 195815	lamin A [<i>Danio rerio</i> (zebrafish)]	Chromosome 16, NC_007127.7 (30520857..30569587, complement)	cb948, wu:fk66d12	

En la primera pàgina del gen LMNA ens dona informació del gen com per exemple que el es troba en el cromosoma 1, en la regió 1q22, té 17 exons, però com que fa splicing alternatiu no tots aquests formen part de les diferents proteïnes finals. El gen comença a la base 156.082.573 del cromosoma 1 i acaba a la 156.140.081, per tan està format per 57.509 parells de bases. Una altre informació molt important és el nombre d'identificació del gen en la base de dades NCBI, en el cas de LMNA és 4000. Per tan sols introduint el nombre 4000 en la NCBI es dirigeix directament a la pàgina d'informació del gen.

A partir d'aquí es pot entrar a la base de dades GenBank només clicant sobre el link *GenBank* que hi ha més avall en la pàgina.

001.11

[156212796 ▶

SLC25A44

Go to [reference sequence details](#)

assembly ▾

Go to nucleotide: [Graphics](#) [FASTA](#) [GenBank](#)

ATG

Tools ▾ | Tracks ▾ | Download ▾

156,120 K | 156,130 K | 156,140 K

Aquí és on es troba la seqüència d'ADN del gen, les coordenades dels ARNm que sintetiza i les seves CDS, i finalment apareix la seqüència d'aminoàcids de les proteïnes.

La cadena d'ADN es troba al final de la pàgina i comença sota la paraula *origin*. La cadena que ens dona la NCBI és la codificadora per tan coincideix en seqüència i orientació amb l'ARNm que produeix la cadena motlle. Les bases nitrogenades estan ordenades en grups de deu en deu i en files de 6 grups, per tant 60 bases per fila. En el costat esquerre de la fila està indicada la posició de la primera base per

facilitar la seva localització. Cal tenir present que la primera base del gen és en realitat la 156.082.573 de la cadena del cromosoma 1 i la 57.481, la última, és la 156.140.081.

ORIGIN

```

1 gacaaattcc ttgacccgag gaggataggg atgtggcctt cggcttttcc tcgcagctcc
61 ggggcaagct aggagtggga tggaaagtcga ggtccctaata tttttaaggg gaggggtgcgg
....

```

Veure la seqüència del gen completa al primer apartat de l'annex.

El gen LMNA que s'està estudiant fa splicing alternatiu per tan es pot observar que hi ha més d'un ARN missatger. L'ARN que s'ha utilitzat en el treball és la variant transcripcional 1 ja que és el més abundant en l'organisme.

La seqüència del ARNm madur ve indicada per les coordenades dels exons, això significa que en la base 32.139 comença l'exó 1 i acaba en la base que fa 32.2702 i així amb tots els altres exons. Els espais del mig són els introns. Les coordenades de l'ARN es basen amb la seqüència d'ADN que dona la NCBI ja que és la cadena codificadora, les bases són les mateixes excepte les timines canviades per uracils.

```

join(32139..32702,48045..48201,51831..51956,52233..52403,
52615..52740,53329..53549,53642..53864,54349..54456,
54541..54660,55082..55171,55916..56185,56508..57509)
gene="LMNA"
transcript variant 1

```

Veure els exons localitzats dins la seqüència del gen completa al primer apartat de l'annex

Exons

CDS

Senyal que indica el principi i el final dels introns

(32139..32702)

```

32101 aggacctatt agagcctttg ccccgcgctc ggtgactcag tggtcgcggg agcgcgcac
32161 ctacaccagc caaccagat cccgaggtcc gacagcgcc ggcccagatc cccacgcctg
32221 ccaggagcaa gccgagagcc agccggccgg cgcactccga ctccgagcag tctctgtcct
32281 tggacccgag ccccgcgccc ttccggggac cctgccccg cgggcagcgc tgccaacctg
32341 ccggccATGg agaccccgctc ccagcggcgc gccaccgca gcggggcgca ggccagctcc
32401 actccgctgt cgcccaccg catcaccgg ctgcaggaga aggaggacct gcaggagctc
32461 aatgatoget tggcgggtcta catcgaccgt gtgcgctcgc tggaaacgga gaacgcaggg
32521 ctgcgccttc gcatcaccga gtctgaagag gtggtcagcc gcgaggtgtc cggcatcaag
32581 gccgcctaag aggcagagct cggggatgcc cgcaagacct ttgactcagt agccaaggag

```

32641 cgcgcccgcc tgcagctgga gctgagcaaa gtgcgtgagg agtttaagga getgaaagcg
32701 cggtaggttc gccacggtgg ctgcgtgcct ggcggggagt ggagagggcg gcgggcccgc
32761 gcccctggcc ggccgcagga agggagtggag agggcctgga ggccgataac tttgccatag

(48045..48201)

48001 cacagactcc ttctottaaa tctactctcc cctctcttct ttagcaatac caagaaggag
48061 ggtgaactga tagctgtca ggctcggtg aaggacctgg aggtctgtct gaactccaag
48121 gaggcgcac tgagcaactgc tctcagtgg aagcgcacgc tggaggggcga gctgcatgat
48181 ctgcgggggcc aggtggccaa ggtgaggcca ccctgcaggg cccaccatg gccccaccta
48241 acacatgtac actcaactct ctacctaggc cctcccccat gtggtgctct gctgacctg

(51831..51956)

51781 ttctgtgttt ctgtgacccc ttttctctat ctctgcctgc ttctcacacg cttgaggcag
51841 cctaggtgga ggccaagaag caacttcagg atgagatgct gggcggggtg gatgctgaga
51901 acaggctgca gaccatgaag gaggaaactgg acttcagaa gaacatctac agtgaggtgg
51961 ggactgtgct ttgcaagcca gagggtctgg gctgggtgat gacagacttg ggctgggcta

(52233..52403)

52201 tggtttctgt gtccttctcc caacccttcc aggagctgct tgagaccaag cgccgtcatg
52261 agaccgact ggtggagatt gacaatggga agcagcgtga gtttgagagc cggctggcgg
52321 atgcgtgca ggaactgcgg gccacgcatg aggaccaggg ggagcagtat aagaaggagc
52381 tggagaagac ttattctgcc aagggtgctt ctctcgattg gttccctcac tgctctgcc
52441 cttggcagcc ctacccttac ccacgctggg ctatgccttc tggggatcag gcagatggtg

(52615..52740)

52561 gtgcctccac cctcccagc caccacagtc ctaacccttt gtctcccct ccagctggac
52621 aatgccaggc agtctgtgta gaggaaacgc aacctggtgg ggctgcccga cgaggagctg
52681 cagcagtcgc gcatccgcat cgacagcctc tctgcccagc tcagccagct ccagaagcag
52741 gtgatacccc acctcacccc tctctccagg ggcctagagt ctgggccgga tgcaggctgg

(53329..53549)

53281 atacttaggg cccttgggag ctcaacaaac cctcccacc cccttcagct ggagccaag
53341 gaggcgaagc ttogagacct ggaggactca ctggcccgtg agcgggacac cagccggcgg
53401 ctgctggcgg aaaaggagcg ggagatggcc gagatgcggg caaggatgca gcagcagctg
53461 gacgagtacc aggagcttct ggacatcaag ctggcccctg acatggagat ccacgcctac
53521 cgcaagctct tggaggggcga ggaggagagg tgggctgggg agacgtcggg gaggtgctgg

(53642..53864)

53581 cagtgtctcc tggccggcaa ctggccttga ctagacccc acttggtctc cctctcccga
53641 ggetacgct gtccccagc cctacctgc agcgcagccg tggcgtgct tctctcact
53701 catcccagac acagggtggg ggcagcgtca ccaaaaagcg caaactggag tccactgaga
53761 gccgcagcag cttctcacag cagcagcga ctagcggcg cgtggccgtg gaggaggtgg
53821 atgaggagg caagtttgtc cggctgcgca acaagtccaa tgaggtaggc tctgctcag
53881 ggtctaaggg gatacagctg catcaggggag agagtggcaa gacagaagga tggcatgtgg

(54349..54456)

54301 agcctgggtg agcctcccgc accttctct tcctatctt cccggcagga ccagtccatg
54361 ggcaattggc agatcaagcg ccagaatgga gatgatccct tctgactta ccggttccca
54421 ccaaagttca cctgaaggc tgggcagggtg gtgacggtga gtggcagggc gcttgggact

(54541..54660)

```
54481 ctggggaggc cttgggtggc gatgggagcg ctggggtaag tgtccttttc tctctccag
54541 atctgggctg caggagctgg ggccaccac agcccccta cggacctggt gtggaaggca
54601 cagaacacct ggggctgchg gaacagcctg cgtacggctc tcatcaactc cactggggaa
54661 gtaagtaggc ctgggcctgg ctgcttgctg gacgaggctc cccctgatgg ccaacatcgg
```

(55082..55171)

```
55021 gacatgctgt acaacccttc cctggccctg acccttgac ctggttccat gtccccacca
55081 ggaagtggcc atgogcaagc tggtgogctc agtgactgtg gttgaggacg acgaggatga
55141 ggatggagat gaacctgctc atcaccacca ctggagtggg agccgccgct gaggccgagc
55201 ctgcactggg gccaccacgc caggcctggg ggcagcctct cccagcctc cccgtgcaa
```

(55916..56185)

```
55861 ctgagtggtc agtcccagac tgcgcgtccc gctgagcct tgtctcctt cccaggctc
55921 ccaactgcgc agctcggggg acccogctga gtacaacctg cgtcgcgca ccgtgctgtg
55981 cgggaacctg gggcagcctg ccgacaaggc atctgccagc ggctcaggag ccaggtggg
56041 Cggacccatc tctctggct cttctgctc cagtgtcacg gtcactcga gctaccgcag
56101 tgtggggggc agtgggggtg gcagctcgg ggacaatctg gtcaccctc cctacctct
56161 gggcaactcc agccccgaa cccaggctgag ttgtctctgc tttgtctcca aatcctgcag
```

(56508..57509)

```
56461 gtccaggccc tgctgctcac acctctctcc tctgttttct ctcttagagc cccagaaact
56521 gcagcatcat gtaactctggg acctgccagg caggggtggg ggtggaggct tctgcgtcc
56581 tctcaacctc atgcccacc cctgcccctg acgtcatggg agggggcttg aagccaaaga
56641 aaaataaacc tttggttttt ttctctgtga ttttttttc taagagaagt tattttctac
56701 agtggtttta tactgaagga aaaacacaag caaaaaaaaa aaaaagcatc tatctcatct
56761 atctcaatcc taatttctcc tcccttctt ttccctgctt ccaggaaact ccacatctgc
56821 cttaaaacca aagagggctt cctctagaag ccaagggaaa ggggtgcttt tatagaggct
56881 agcttctgct tttctgccc ggctgctgcc cccaccccgg ggacctgtg acatgggtgcc
56941 tgagaggcag gcatagaggc ttctccgcca gctcctctg gacggcaggc tcaactgccag
57001 gccagcctcc gagagggaga gagagagaga gaggacagct tgagccgggc ccctgggctt
57061 ggctgctgtg gattocacta cacctggctg aggttctct gctgccccg ccccagctcc
57121 ccaccctgc ccccagcccc ggggtgagtc cattctcca ggtaccagct gcgcttctt
57181 ttctgtattt tatttagaca agagatggga atgagggtgg aggtggaaga agggagaaga
57241 aaggtgagtt tgagctgctt tccctagctt tagaccctgg gtgggctctg tgcagtact
57301 ggaggttgaa gccaaagtgg gtgctgggag gaggagagag gaggctactg gaaaggggag
57361 agcctgctgg caccaccctg ggaggaggaa ggcaagaggg ggtggagggg tgtggcagtg
57421 gttttggcaa acgctaaaga gcccttgcct cccatttcc catctgcacc ccttctctcc
57481 tccccaaatc aatacactag ttgtttcta...
```

La seqüència de CDS (Codifying Sequence) que li correspon a l'ARNm de la variant transcripcional 1 dona unes coordenades que un cop juntes formaran la seqüència de triplets que serà traduïda pels ribosomes. Per tant el primer triplet de la CDS serà AGT i l'últim serà un triplet de stop.

La seqüència CDS està inclosa dins l'ARN missatger que hi ha anteriorment.

Veure la CDS localitzada dins la seqüència del gen completa al primer apartat de l'annex

```
join(32347..32702,48045..48201,51831..51956,52233..52403,  
      52615..52740,53329..53549,53642..53864,54349..54456,  
      54541..54660,55082..55171,55916..56185,56508..56534)  
gene="LMNA"  
isoform A is encoded by transcript variant 1
```

Finalment després de les dades de la CDS es troba la seqüència de la proteïna de 664 aminoàcids que concorda amb la variant estudiada.

Seqüència de la proteïna sana:

```
/translation="METPSQRRATRSGAQASSTPLSPTRITRLQEKEDLQELNDRLAV  
YIDRVRSLTENAGLRLRITESEEVVSREVSGIKAAYEAE LGDARKTLDSVAKERARL  
QLELSKVREEFKELKARNTKKEGDLIAA QARLKDLEALLNSKEAALSTALSEKRTLEG  
ELHDLRGQVAKLEAALGEAKKQLQDEMLRRVDAENRLQTMKEELDFQKNIYSEELRET  
KRRHETRLVEIDNGKQREFESRLADALQELRAQHEDQVEQYKKELEKTYSAKLDNARQ  
SAERNNSNLVGA AHEELQQSRIRIDSLSAQLSQLQKQLAAKEAKLRDLEDSLARERDTS  
RRLLAEKEREMAEMRARMQQQLDEYQELLDIKLALDMEIHAYRKLLEGEERLRLSPS  
PTSQRSRGRASSHSQTQGGGSVTKKRKLESTESRSSFSQHARTSGRVAVEEVDEEGK  
FVRLRNKSNEDQSMGNWQIKRQNGDDPLLYRFPFKFTLKAGQVVTIWAAGAGATHSP  
PTDLVWKAQNTWGCNSLR TALINSTGEEVAMRKLVR SVTVVEDEDEDGDDLLHHHH  
GSHCSSSGDPAEYNLRSRTVLCGT CGQPADKASASGSGAQVGGPISSGSSASSVTVTR  
SYRSVGGSGGGSFGDNLVTRSYLLGNSSPRTQSPQNC SIM"
```

5.3.2 Maduració de la Proteïna Lamin A

Quan el gen LMNA d'una persona sana ja s'ha transcrit i traduït donant lloc a la proteïna Lamin A, aquesta ha de passar per unes modificacions postraduccionals que habilitaran la seva funció.

```
1 metpsqrrat rsgaqasstp lsptritrlq ekedlqelnd rlavyidrivr sletenaglr  
61 lriteseevv srevsigikaa yeaelgdark tldsvakera rlqlelskvr eefkelkarn  
121 tkkegdliia qarlkdleal lnskeaalst alsekrleg elhdlrgqva kleaalgeak  
181 kqlqdemlrr vdaenrlqtm keeldfqkni yseelretkr rhetriveid ngkqrefesr  
241 ladalqelra qhedqveqyk kelektysak ldnarqsaer nsnlvgaah eelqqsririd  
301 slsaqlsqlq kqlaakeakl rdledslare rdtsrrllae keremaemra rmqqqldeyq  
361 elldiklald meihayrklle egeeerlrls psptsqrsrg rasshssqtq gggsvtkkrk  
421 lestesrssf sqhartsgrv aveevdeegk fvrlrnksne dqsmgnwqik rqnqddpllt  
481 yrfppkftlk agqvvtiwa gagathsppt dlvwkaqntw gcgnslr tal instgeevam  
541 rklvrsvtvv eddededgdd llhhhhgshc sssgdpaeyn lrsrtvlgcgt cgqpadkasa  
601 sgsgaqvggp issgssassv tvtrsyrsvg gsgggsfgdn lvtrsylgn ssprtqspqn  
661 csim (1) (2)
```

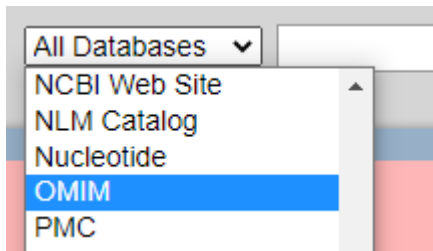
El primer procés pel qual passa la proteïna és l'agregació d'un grup farnesil en la posició 661 de la proteïna (1), en una cisteïna (c). També els tres últims aminoàcids després de la cisteïna s'eliminen. Aquests dos canvis marquen la proteïna perquè pugui entrar al nucli de la cèl·lula sense cap problema. Quan la proteïna ja està dins del nucli es produeix l'últim procés perquè sigui funcional. Un enzim talla entremig de l'aminoàcid tirosina (y) i leucina (l) de les posicions 646 i 647 (2) eliminant els 15 aminoàcids que venen a continuació, incloent el grup farnesil. Perquè es produeixi aquesta reacció els aminoàcids del voltant de la tirosina i la leucina han de ser els correctes ja que sinó és així l'enzim no pot tallar la proteïna. Aquest últim canvi permet a la proteïna integrar-se a la làmina nuclear on farà la seva funció juntament amb altres lamines. (González Morán, M. 2013)

5.4 Base de dades OMIM

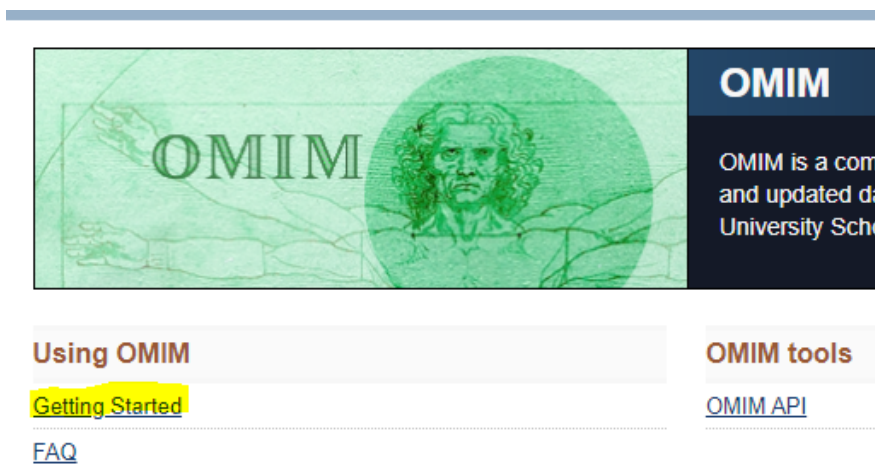
La informació sobre la malaltia de la progèria s'ha extret de la base de dades OMIM que està explicada a continuació.

La OMIM (Online Mendelian Inheritance in Man) és una base de dades que cataloga totes les malalties conegudes amb un component genètic i s'enfoca amb la relació entre gen i fenotip. Aquesta base de dades s'ha utilitzat per buscar dades sobre la progèria, com els seus símptomes, òrgans i teixits que afectats, la població que la pateix... També s'ha consultat per veure les variants al·lèliques de la malaltia, per tant el lloc exacte on hi ha la mutació escollida.

Per accedir a la base de l'OMIM s'hi pot entrar des de la NCBI o directament per la seva web (<https://omim.org/>). Per accedir-hi des de la NCBI només s'insereix en el buscador OMIM i es clica buscar.



Seguidament apareix una pàgina on clicant “getting started” ja entra directament a la pàgina oficial de l’OMIM.



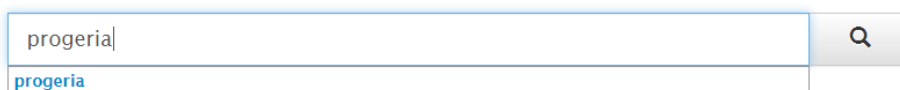
Quan s’arriba a la pàgina principal de l’OMIM hi ha un buscador on es pot introduir el nom de la malaltia, el fenotip i fins i tot el gen que es vol buscar, en anglès, i es clica buscar. En aquest cas s’ha buscat *progeria*.

OMIM[®]

Online Mendelian Inheritance in Man[®]

An Online Catalog of Human Genes and Genetic Disorders

Updated August 5, 2020



De tots els resultats que surten es clica el que correspon amb la malaltia que es necessita, la progèria de Hutchinson-Gilford.

- 1: [# 176670. HUTCHINSON-GILFORD PROGERIA SYNDROME; HGPS](#)
[PROGERIA SYNDROME, CHILDHOOD-ONSET, INCLUDED](#)
 Cytogenetic location: 1q22
 Matching terms: included, progeria
 ▶ Phenotype-Gene Relationships ▶ Phenotypic Series ▶ ICD+ ▶ Links
- 2: [# 614008. NESTOR-GUILLERMO PROGERIA SYNDROME; NGPS](#)
 Cytogenetic location: 11q13.1
 Matching terms: included, progeria
 ▶ Phenotype-Gene Relationships ▶ Phenotypic Series ▶ ICD+ ▶ Links
- 3: * [150330. LAMIN A/C; LMNA](#)

#176670

Table of Contents

- Title
- Phenotype-Gene Relationships
- Clinical Synopsis
- Phenotypic Series
- Text
 - Description
 - Clinical Features
 - Biochemical Features
 - Inheritance
 - Cytogenetics
 - Population Genetics
 - Molecular Genetics
 - Genotype/Phenotype Correlations
 - Pathogenesis
 - Animal Model
 - History
- See Also

ICD+

176670

HUTCHINSON-GILFORD PROGERIA SYNDROME; HGPS

Alternative titles; symbols

PROGERIA

Other entities represented in this entry:

PROGERIA SYNDROME, CHILDHOOD-ONSET, INCLUDED

Phenotype-Gene Relationships

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key	Gene/Locus	Gene/Locus MIM number
1q22	Hutchinson-Gilford progeria	176670	AD, AR	3	LMNA	150330

Clinical Synopsis ▾
Phenotypic Series ▾
PheneGene Graphics ▾

En la pàgina principal de la malaltia es troba informació sobre les dades clíniques i epidemiològiques, dades metabòliques i genètiques, la descripció de la malaltia, entre altres dades.

La malaltia de la progèria té diferents variants al·lèliques que la provoquen, per veure quines són en el mateix lloc web es clica sobre el nombre d'identificació del gen en la base de dades OMIM. Seguidament en l'índex de l'esquerra hi ha una

entrada on diu variants al·lèliques, allà és on hi ha la informació de totes les variants, quina mutació provoquen, en quina posició, el canvi d'aminoàcid...

PROGERIA SYNDROME, CHILDHOOD-ONSET, INCLUDED

Phenotype-Gene Relationships

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key	Gene/Locus	Gene/Locus MIM number
1q22	Hutchinson-Gilford progeria	176670	AD, AR	3	LMNA	150330

*150330

Table of Contents

Title

Gene-Phenotype Relationships

Text

Description

Cloning and Expression

Gene Structure

Mapping

Gene Function

Molecular Genetics

Genotype/Phenotype Correlations

Animal Model

Allelic Variants

Table View

* 150330

LAMIN A/C; LAMIN A/C

Other entities represented in this gene

LAMIN A, INCLUDED
LAMIN C, INCLUDED
PRELAMIN A, INCLUDED
PROGERIN, INCLUDED

HGNC Approved Gene Symbol

Cytogenetic location: 1q22

Gene-Phenotype Relationships

Location	Phenotype
----------	-----------

5.5 Variant al·lèlica

La tecnologia CRISPR és capaç, de moment, de corregir les mutacions **monogèniques** de **substitució** de poques bases, això vol dir que no es podran corregir totes les mutacions que provoquen la progèria de cop.

La mutació del gen LMNA pot provocar moltes altres malalties a part de la progèria com per exemple la cardiomiopatia diluïda, diferents tipus de distròfia muscular.... De les 58 diferents mutacions en el gen LMNA que provoquen una malaltia sol 8 produeixen progèria, això vol dir que la progèria és una malaltia amb 8 variants al·lèliques conegudes. El CRISPR que es dissenya podrà corregir una d'aquestes variants, tot i que si les mutacions es produeixen molt juntes es podrien corregir diverses de cop. D'aquestes 8 mutacions en el treball s'estudia només la variant al·lèlica **0.22** del gen LMNA perquè és la més comú i la que afecta a més gent. Corregint la variant 0.22 també es corregeixen la variant 0.23 i la 0.40 ja que les mutacions es troben molt a prop entre elles. A la taula es pot observar que la variant 0.23 es troba al mateix codó que la variant 0.22, i la 0.40 es troba al codó anterior.

Variant al·lèliques de la progèria					
Nombre (OMIM)	Mutació	Posició en la CDS	Canvi d'aminoàcid	Posició codó	Exó
0.22	C→ T	1824	Gly→ Gly	608	11
0.23	G→ A	?	Gly→ Ser	608	11
0.24	?	?	Glu→ Lys	145	2
0.31	T→ G	834	Leu→ Arg	140	
0.40	G→ A	1821	Val→ Val	607	11
0.55	G→ A	1968	Gln→ Gln	656	11
0.56	G→ A	1968	Gly→ Gln	567	11
0.57	A→ G	899	Asp→ Gly	300	

La mutació 0.22 és un canvi d'una **citocina** a una **timina** en la posició 56041 del gen LMNA (1824 de la CDS i 608 de la proteïna). Curiosament el nou codó crea la mutació que produeix el mateix aminoàcid que el sa, vull dir que el codó sa (GGC) codifica per l'aminoàcid glicina i el mutat (GGU) també codifica per una glicina.

ADN (sa)

```
55981 cgggacctgc gggcagcctg cgcacaaggc atctgccagc ggctcaggag ccaggtgag
                                         A S G S G A Q V G
56041 cggaccctac tcctctggct cttctgcctc cagtgtcacg gtcactcgca gctaccgcag
      G P I S S G S S A S S V T V T R S Y R S
56101 tgtggggggc agtgggggtg gcagcttcgg ggacaatctg gtcaccgct cctacctcct
      V G G S G G G S F G D N L V T R S Y L L
56161 gggcaactcc agccccgaa ccaggtgag ttgtctctgc tttgtctcca aatcctgcag
      G N S S P R T Q
```

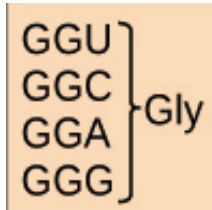


Figura 27: Els codons que codifiquen per la glicina

ADN (mutat)

```
55981 cgggacctgc gggcagcctg cgcacaaggc atctgccagc ggctcaggag ccaggtgag
                                         A S G S G A Q V G
56041 cggaccctac tcctctggct cttctgcctc cagtgtcacg gtcactcgca gctaccgcag
      G P I S S G S S A S S V T V T R S Y R S
56101 tgtggggggc agtgggggtg gcagcttcgg ggacaatctg gtcaccgct cctacctcct
      V G G S G G G S F G D N L V T R S Y L L
56161 gggcaactcc agccccgaa ccaggtgag ttgtctctgc tttgtctcca aatcctgcag
      G N S S P R T Q
```

Tot i així, sí que provoca la progèria. Al principi veient que no canvia l'aminoàcid es pot pensar que la mutació és silenciosa, però no és així. El canvi de nucleòtid crea una senyal que indica el principi d'un intró, per tan el procés de splicing no es fa correctament. La mutació crea una senyal errònia de principi d'intró, per tant quan un enzim talli per aquesta senyal es perdran 150 bases nitrogenades i 50 aminoàcids. Aquesta petita mutació d'una base és la culpable d'aquests conjunt de processos erronis que provoquen la progèria. (Eriksson, M., Brown, W., Gordon, L. et al. 2003)

5.5.1 Maduració de la progerina

La proteïna que es forma defectuosa s'anomena **progerina** i intenta seguir el procés de maduració de la lamin A. La primera reacció, en la qual s'afegeix un grup farnesil en la posició 661 (1) i s'eliminen els tres nucleòtids finals, si que es porta a terme correctament ja que aquesta part de la proteïna segueix intacte. Per tant la proteïna és capaç d'entrar al nucli cel·lular. Però en la segona reacció (2) química com que la tirosina (y) i leucina (l) no hi són per culpa de l'error en el splicing la proteïna no es podrà unir a la làmina nuclear i no exercirà la seva funció. (Eriksson, M., Brown, W., Gordon, L. et al. 2003) (González Morán, M. 2013)

```
1 metpsqrrat rsgaqasstp lsptritrlq ekedlqelnd rlavyidrvr sletenaglr
61 lriteseevv srevsgikaa yeaeldgark tldsvakera rlqlelskvr eefkelkarn
121 tkkegdliaa qarlkdleal lnskeaalst alsekrteleg elhdrlrgqva kleaalgeak
181 kqlqdemlrr vdaenrlqtm keeldfqkni yseelretkr rhetriveid ngkqrefesr
241 ladalqelra qhedqveqyk kelektysak ldnarqsaer nsnlvgaah eelqgsririd
301 slsaqlsqli kqlaakeakl rdledslare rdtsrllae keremaemra rmqqqldeyq
361 elldiklald meihayrkll egeeerlrls psptsqrsrg rasshssqtq gggsvtkkrk
421 lestesrssf sqhartsgrv aveevdeegk fvrlrnksne dqsmgnwqik rqnqddpllt
481 yrfppkftlk agqvvtiwa gagathsppt dlvwkaqntw gcgnslrtal instgeevam
541 rklvrsvtvv eddededgdd llhhhhgshc sssgdpaeyn lrsrtvlgct cgqpadkasa
601 sgsgaqvGgp issgssassv tvtrsyrsvg gsgggsfgdn lvtrsylngn ssprtqspqn
661 csim (es perdran aquests 50 aa) | (2)
    ↑ (1)
```

6. La cura de la malaltia: aplicació de la tècnica CRISPR en la progèria

La part de recerca del treball és dissenyar els elements moleculars de la tècnica CRISPR per curar la malaltia de la progèria en cèl·lules en cultiu. El procediment que s'ha seguit per aconseguir aquest objectiu consisteix de dues parts. La primera part és la construcció del complex ARNg/Cas9. Per fer això es necessita la proteïna associada al CRISPR, Cas9, i un ARN guia dissenyat especialment per la mutació que es vol curar, en aquest cas la variant 0.22 de la progèria. Aquest procés de disseny és molt important pel funcionament de la tècnica ja que l'ARN guia és la molècula que indica el lloc exacte de l'ADN per on ha de tallar la Cas9.

El segon pas és la selecció d'un segment d'ADN que serveixi com a plantilla per la cèl·lula en el moment de reparar el dany que s'ha produït al genoma. Per tal de disposar de moltes còpies del fragment i augmentar la probabilitat de que alguna entri dins les cèl·lules i es recombinï amb l'ADN danyat, s'haurà de realitzar el procés de PCR. Per fer-lo es dissenyaran uns iniciadors de PCR ja que l'enzim ADN polimerasa necessita uns petits trossos d'àcid nucleic per començar la síntesi. Els iniciadors també són els encarregats d'indicar la regió de l'ADN que s'amplificarà, per això estan dissenyats de manera específica, per determinar l'amplificació del segment d'ADN homòleg.

6.1 Disseny de l'ARN guia

El disseny de l'ARN guia és essencial pel funcionament del complex ARNg/Cas9 i per portar a terme la tècnica. L'ARN guia és la molècula que guia i indica a la proteïna Cas9 al lloc exacte del genoma on ha de tallar, gràcies a que complementa amb la seqüència d'ADN de la cèl·lula. Aquest està format per exactament 23 nucleòtids, llargada que assegura que només s'enganxi al lloc desitjat del genoma.

La part interessant d'aquest àcid nucleic és que només canviant la seva seqüència perquè complementi en un altre diana es podria corregir una mutació diferent.

Per dissenyar l'ARN guia específic per la mutació 0.22 de la progèria es requereix saber la ubicació exacte de la mutació en l'ADN. Per fer això es necessita anar a la base de dades OMIM i buscar la posició de la qual s'està fent la simulació de cura. En el cas de la variant 0.22 de la progèria, la mutació és la substitució d'una citosina per una timina que es troba a l'exó 11, concretament a la posició 1824 de la CDS i en el codó 608, que també és la posició en la proteïna.

L'OMIM pot proporcionar la informació de diverses maneres i cal adaptar-se segons cada cas. Si es proporciona directament la localització de la mutació en l'ADN sols s'haurà de buscar aquesta en la seqüència del gen. Si no és així i es proporciona en la CDS s'haurà de localitzar la posició de la mutació en la CDS a partir de la seqüència del gen. En el cas que es doni la posició en la proteïna s'haurà de passar d'aminoàcid a seqüència de nucleòtids, trobar la posició en la CDS i finalment la posició en l'ADN.

En la variant al·lèlica 0.22 de la progèria l'OMIM només dona les dades de la proteïna i la CDS, per tant el més senzill de fer per trobar la posició de la mutació en el gen és localitzar la CDS a partir de les seves coordenades directament a la seqüència del gen. (També es pot localitzar l'ARNm però no és necessari)

Quan estan tots els fragments de la CDS localitzats, es pot procedir a ajuntar-los i comptar les bases fins a la 1824.

Veure al segon apartat de l'annex els fragments de la CDS i l'ARNm, localitzats, ajuntats i el recompte.

La posició 1824 de la CDS, on hi ha la citosina que muta, correspon a posició és la 56041 del gen.

```
55981 cgggacctgc gggcagcctg cggacaaggc atctgccagc ggctcaggag cccagggtggg
56041 cgggaccata tctctgggt cttctgcctc cagtgtcacg gtcactcgca gctaccgcag
```

El següent pas és agafar el fragment de la seqüència de l'ADN mutat i construir-li, a la cadena codificadora, la cadena motlle complementaria.

Cadena codificadora
Cadena motlle
Extrem 5' de l'ARN guia

```
5' 55981 cgggacctgc gggcagcctg cggacaaggc atctgccagc ggctcaggag cccaggtggg
3'-----tagacggtcg ccgagtctc gggtcaccc

56041 Tggaccatc tctctggct cttctgcctc cagtgtcagc gtcactcgca gctaccgcag 3'
Acctgggtag aggagaccga gaagacggag----- 5'
```

Seguint amb el passos per trobar l'ARN guia que cal per l'aplicació del CRISPR, es busca una seqüència PAM (5'-NGG-3') a prop de la mutació en la cadena codificadora. La PAM és important ja que ajuda al complex a orientar-se cap a la seqüència diana, on es vol que s'enganxi l'ARN guia i que coincideix amb on hi ha la mutació. El segment d'ADN PAM que s'ha escollit és 5'-TGG-3' que casualment està inclòs dins la mutació. Finalment s'ha seleccionat unes 23 bases en direcció 3' de la cadena motlle prop de la mutació i s'ha creat un ARN complementari a aquest segment.

```
5' atctgccagc ggctcaggag cccaggtggg Tggaccatc tctctggct cttctgcctc 3'

5' gc ggcucaggag cccagguggg U...3'

3' tagacggtcg ccgagtctc gggtcaccc Acctgggtag aggagaccga gaagacggag 5'
```

La seqüència 5'- **cgggcucaggagcccaggugggu** -3' és l'extrem 5' (23 nucleòtids) de l'ARN guia dissenyat únicament per la variant al·lèlica 0.22 de la progèria.

6.2 Disseny dels iniciadors de PCR

Els iniciadors són seqüències d'ADN conegudes d'entre 18 i 25 nucleòtids que determinen on s'inicia la replicació de l'ADN. En una PCR es necessiten dos iniciadors complementaris a l'extrem 3' de cada una de les cadenes senzilles de l'ADN original, l'iniciador que replica cap endavant, anomenat *Forward*, i el que el replica cap enrere, el *Reverse*. La importància d'aquests iniciadors està en el fet que delimiten la regió d'ADN que es vol amplificar i donen pas a la síntesi d'aquest segment, que sense ells no es podria dur a terme. La part amplificada ha de tenir una llargada d'entre 100 i 300 nucleòtids, es tracta de que no sigui ni molt llarga ni molt curta, i pugui fer correctament la seva funció. La importància del fragment sa d'ADN és que servirà com a plantilla a la cèl·lula en el procés de la recombinació homòloga, d'aquesta manera la pròpia cèl·lula curarà la malaltia.

Per començar a dissenyar els indicadors primer es necessita conèixer quin serà el fragment amplificat i tenir en compte que l'ADN que es vol amplificar és el sa.

El segment, que inclou la posició de la mutació, que s'ha escollit va des de la base 55961 fins a la 56180, que són 220 nucleòtids.

Seguidament en cada un dels extrems 3' de les cadenes es seleccionen unes 20 bases i es crea un tros d'ADN complementari, que seran els iniciadors.

Iniciadors de PCR
Segment d'ADN amplificat

```
5' ccaactgcagc agctcggggg accccgctga gtacaacctg cgctcgcgca ccgtgctgtg  
forward 5' cgctcgcgca ccgtgctgtg 3'  
3' ggtgacgtcg tcgagcccc tggggcgact catgttggac gcgagcgcgt ggcacgacac
```

```
55981 cgggacctgc gggcagcctg ccgacaaggc atctgccagc ggctcaggag cccaggtggg  
gccctggacg cccgtcggac ggtgttccg tagacggctc ccgagtcctc ggggccaccc
```

```
56041 Cggaccatc tcctctgget cttctgcctc cagtgtcagc gtcactcgca gctaccgag  
Gcctgggtag aggagaccga gaagacggag gtcacagtgc cagtgagcgt cgatggcgtc
```

```
56101 tgtggggggc agtgggggtg gcagcttcgg ggacaatctg gtcaccgct cctacctct  
acacccccg tcacccccac cgtcgaagcc cctgttagac cagtgggcca ggatggagga
```

```
gggcaactcc agccccgaa cccaggtgag ttgtctctgc tttgtctcca aatcctgcag 3'  
3' cccgttgagg tcgggggctt 5' reverse  
cccgttgagg tcgggggctt ggtccactc aacagagacg aaacagaggt ttaggacgtc 5'
```

Amb aquests iniciadors, el *Forward* **5'cgctcgcgca ccgtgctgtg 3'** i el *Reverse* **3'cccgttgagg tcgggggctt 5'** ja es podrà iniciar el procés de PCR i amb uns quants cicles s'obtidran milions de còpies del fragment.

L'últim pas de la simulació seria introduir l'ARN guia, la Cas9 i el segment amplificat dins la cèl·lula perquè de manera natural corregeixi la mutació.

7. Conclusions

Amb aquest treball s'ha aconseguit descriure de manera sintetitzada les bases de la tecnologia CRISPR, com una tècnica que prové de la reacció natural del sistema immunitari bacterià i que la biotecnologia l'ha adaptat per tal de que es pugui utilitzar per editar el genoma de qualsevol cèl·lula.

S'ha trobat una malaltia amb les característiques necessàries perquè sigui possible la seva correcció amb el CRISPR, és a dir una malaltia monogènica i de substitució de pocs nucleòtids. La malaltia seleccionada ha sigut la progèria, una malaltia molt greu que provoca l'envelliment brusc i prematur dels infants que la pateixen.

S'ha trobat que les mutacions de les variants al·lèliques 0.23 i 0.40 de la progèria estan molt a prop de la mutació per la qual s'ha dissenyat el CRISPR, la variant 0.22. Això significa que un sol complex ARNguia-Cas9 dissenyat per la variant 0.22 també podria servir per curar la progèria produïda per la variant 0.23 i 0.40.

S'ha posat en pràctica de manera simplificada i en base als fonaments teòrics del CRISPR una simulació de cura de la variant al·lèlica 0.22 de la progèria que requereix una correcció d'una mutació gènica de substitució d'una citosina per una timina en el gen LMNA. La cura de la progèria s'ha aconseguit dissenyant un ARN guia de seqüència 5'-cgggcucaggagcccagguggu-3' de 23 nucleòtids que permet al complex ARNguia-Cas9 complementar amb la seqüència on es troba la mutació on seguidament la Cas9 la tallarà. A més a més, per dissenyar l'ARN guia s'ha escollit la seqüència PAM 5'-tgg-3' en la cadena codificadora per tal de guiar al complex cap a la seqüència diana. També s'han dissenyat els iniciadors de PCR *Forward* de cadena 5'cgctcgcgca cctgtctgtg 3' i *Reverse* de cadena 3' cccgttgagg tcgggggctt 5', per dur a terme l'amplificació del fragment de 220 nucleòtids d'ADN homòleg a la cadena danyada.

S'ha comprovat que és possible el disseny de l'ARN guia específic per la mutació 0.22 de la progèria i també el disseny dels iniciadors de PCR per amplificar el segment homòleg, que va de la base 55961 fins a la 56180 del gen LMNA sa. Això significa que també és possible la correcció de la mutació 0.22 de la progèria de manera teòrica. Si es sintetitzessin els elements anteriors reals podria ser possible la cura de la progèria en cèl·lules en cultiu.

Les eines que s'han utilitzat per realitzar aquest treball han sigut totes gratuïtes i d'ús públic, com la base de dades OMIM i el portal de la NCBI. Per tant el disseny d'un ARN guia i uns iniciadors de PCR per corregir de manera teòrica una malaltia humana està a l'abast de tothom.

8. Bibliografía

Webgrafía

Anatomy of a Gene. Learn.genetics. Recuperat 2020, <https://learn.genetics.utah.edu/content/basics/geneanatomy/>.

Andersen, P. (2016). *What is CRISPR?*. Recuperat 2020, <https://www.youtube.com/watch?v=MnYppmstxls>.

Antama, Francis. “Aplicaciones y Regulación Del CRISPR.” *Fundacion Antama*, 27 Aug. 2018, <https://fundacion-antama.org/aplicaciones-y-regulacion-del-crispr-por-francis-mojica/>

Álvarez Gómez, E. (2016). *¿Cómo funciona la PCR?*. Empíreo Diagnóstico Molecular. Recuperat 2020, <https://www.empireo.es/como-funciona-pcr/>.

Coll-Bonfill, N. (2020). “Investigo para curar la progeria, el envejecimiento prematuro”. *La Vanguardia*. Recuperat 2020, <https://www.lavanguardia.com/lacontra/20200229/473829374947/investigo-para-curar-la-progeria-el-envejecimiento-prematuro.html>.

CRISPR: Gene editing and beyond. Nature. (2017). Recuperat 2020, <https://www.youtube.com/watch?v=4YKfw2KZA5o>.

Eukaryotic pre-mRNA processing | RNA splicing. Khan Academy. Recuperat 2020, <https://www.khanacademy.org/science/ap-biology/gene-expression-and-regulation/transcription-and-rna-processing/a/eukaryotic-pre-mrna-processing>.

Garrido Ramos, M. *Reacción en Cadena de la Polimerasa (PCR)*. Universidad de Granada. Recuperat 2020, <http://www.ugr.es/~mgarrido/PCR.htm>.

Gene Structure. California State Polytechnic University, Pomona. (2014). Recuperat 2020, <https://elearning.cpp.edu/learning-objects/gene-structure/>.

Genome Editing with CRISPR-Cas9. McGovern Institute. (2014). Recuperat 2020, <https://www.youtube.com/watch?v=2pp17E4E-O8>.

HUTCHINSON-GILFORD PROGERIA SYNDROME; HGPS. OMIM. (2018). Recuperat 2020, <https://omim.org/entry/176670>.

Hutchinson-Gilford progeria syndrome. Medlineplus. (2020). Recuperat 2020, <https://medlineplus.gov/genetics/condition/hutchinson-gilford-progeria-syndrome/>.

Lamin. Wikipedia. (2020). Recuperat 2020, <https://en.wikipedia.org/wiki/Lamin>.

Les mutacions. Recuperat 2020, http://cosmolinux.no-ip.org/recursos_aula/BIO1erBAT/Genetica_molecular/Unitat_11_Les_mutacions1.pdf].

Les proteïnes. Xtec. Recuperat 2020, <http://www.xtec.cat/ieslabisbal/salut/lesproteines.htm>.

Lo, E. (2017). *What is the PAM?*. Innovative Genomics Institute. Recuperat 2020, https://www.youtube.com/watch?v=iSEEW4Vs_B4.

Menesini, M. (2015). *What happens when your DNA is damaged?*. Ted-Ed. Recuperat 2020, https://www.youtube.com/watch?v=vP8-5Bhd2ag&feature=emb_title.

Morán, A. (2015). *¿Qué es la tecnología CRISPR/Cas9 y cómo nos cambiará la vida?*. Dciencia. Recuperat 2020, <https://www.dciencia.es/que-es-la-tecnologia-crispr-cas9/>.

Nucleic acids. Khan Academy. Recuperat 2020, <https://www.khanacademy.org/science/ap-biology/gene-expression-and-regulation/dna-and-rna-structure/a/nucleic-acids>.

Online Mendelian Inheritance in Man. Wikipedia. (2020). Recuperat 2020, https://en.wikipedia.org/wiki/Online_Mendelian_Inheritance_in_Man.

Primary transcript. Wikipedia. (2020). Recuperat 2020, https://en.wikipedia.org/wiki/Primary_transcript.

What are genome editing and CRISPR-Cas9?. Medlineplus. (2020). Recuperat 2020, <https://medlineplus.gov/genetics/understanding/genomicresearch/genomeediting/>.

What is PCR (polymerase chain reaction)?. Yourgenome.org. (2016). Recuperat 2020, <https://www.yourgenome.org/facts/what-is-pcr-polymerase-chain-reaction>.

Bases de dades

Portal NCBI: <https://www.ncbi.nlm.nih.gov/>

Base de dades OMIM: <https://omim.org/>

Articles

B.M. Paschal, J.B. Kelley, in Encyclopedia of Biological Chemistry (Second Edition), 2013, secció Lamin Genes <https://www.sciencedirect.com/topics/neuroscience/lamin>

Dittmer, T.A., Misteli, T. The lamin protein family. *Genome Biol* 12, 222 (2011).
<https://doi.org/10.1186/gb-2011-12-5-222>

Eriksson, M., Brown, W., Gordon, L. *et al.* Recurrent *de novo* point mutations in lamin A cause Hutchinson–Gilford progeria syndrome. *Nature* 423, 293–298 (2003).
Extret de OMIM, inheritance i molecular genetics.

González Morán, M. (2013). *Síndrome de Progeria de Hutchinson-Gilford. Causas, investigación y tratamientos farmacológicos*. Recuperat 2020,
<https://www.sciencedirect.com/science/article/pii/S0187893X14700631>.

Hennekam RCM. 2006. Hutchinson–Gilford progeria syndrome: Review of the phenotype. *Am J Med Genet Part A* 140A:2603–2624. Extret de OMIM, descripció.

Documents de docència

Estudi de l'ADN i edició genètica. A. Fontarnau. Desembre 2019

Document de la maduració de la progerina i la lamin A de l'Agustí Fontarnau

Annex

Seqüència del gen LMNA

En aquest annex es mostra la seqüència completa del gen LMNA on estan localitzats els segments que junts formen l'ADN missatger i els que formen la CDS. A més també estan marcades les senyals que indiquen el principi i el final dels introns i la base mutada en la posició 56041 del gen.

Coordenades de l'ARN missatger de la variant transcripcional 1:

```
join(32139..32702,48045..48201,51831..51956,52233..52403,  
52615..52740,53329..53549,53642..53864,54349..54456,  
54541..54660,55082..55171,55916..56185,56508..57509)  
gene="LMNA"  
transcript variant 1
```

Coordenades de la CDS corresponent a la variant transcripcional 1 de l'ARN missatger

```
join(32347..32702,48045..48201,51831..51956,52233..52403,  
52615..52740,53329..53549,53642..53864,54349..54456,  
54541..54660,55082..55171,55916..56185,56508..56534)  
gene="LMNA"  
isoform A is encoded by transcript variant 1
```

GEN LMNA

mRNA (exons)

CDS

Senyal que indica el principi i el final dels introns

Base mutada

ORIGIN

```
1 gacaaattcc ttgaccogag gaggataggg atgtggcctt cggctcttcc tgcagctcc  
61 ggggcaagct aggagtggga tggaagtcga ggtccctaat tttttaaggg gagggtgcg  
121 ggagaagggg tagtatgagg aaacagagcg ggtatgaagc tggctaaccg cgcgcgccc  
181 ctcccaggac cgcctcctgc cccgcgcccg ccggctcctgg gggcccgcct ttttatggaa  
241 atgaggaggg ggggcccggg cggggggcgg ggagccggga gccgtccccg ctgcctcact  
301 geggetttct ctctcgcctc cctctccccg ctccctgccg cgctcactct ccgcttcccc  
361 ctcccgcctc ccagagtcg atcccggagc ccggcccgcg ggagaggttc tcggcagaga  
421 agacaaagcc cgcagcagcg atgggggggag agctgggctc tgcgtgtgtg gggggccagg  
481 aaaggttgcc aggctggggc tggaaccccc tggcaaagga tgggggtcccc tcatccctaa  
541 acagcaagcc atctcccctc gcccgcccc cccccccca gtctcggaga tctcagaggc  
601 accgactggg aggtgagtta gttcacgtcc ttctgctcgg tggagggggg cgggagggcg  
661 gtggagtgat gaagtgcaga ggttgaagga gtgcagggac agaactgggg gtcggacgga  
721 gagtgggcag gcaggaaaag tcttaaagc ctgggggtgac ttaaccggtt ggggtgaggg  
781 ccttgagctt tggatatgac aggagggagg agagaccagg atggactggg aaccaggct  
841 tccttcagge ctgggctgtg aatacagctt ggggtgatct gtcctttagc actcatcgta  
901 atttccactc tgtttcctcc acctctctcc cattgaaaac ctattcaggg cggggcgcg  
961 tggetcaocg ctgtaatccc agcacttcgg gaggcagagg cgggtggatc acctgaggtc  
1021 aggagttoga gaccagcctg accaacaagg tgaaacctcg tctctactaa aaatacaaaa
```

1081 attagccgag catggtggca gacgcctgta gtcccageta ctcaggaggc tgagacagga
1141 gaattgcttg aaccggggag gcggaagtgt cagtgagtct agatcgcaga gtgagactcc
1201 gtcccccca caaaaaaca aaaacaaaac aaacaaacaa aaaaacccat tcagtcccg
1261 ttgactccgg aaattcgggtg agaatgaaaa taagttttta aggcttggag agttaggcac
1321 ttttagagtt tcaggaagggt ggagtgtaat gcaatttttt cttctccctt taccatata
1381 tttctgcatt tttttatat cacccttttt atgtctggat ctttcccaag ggaagaagg
1441 atggcactgc agcttgagggt ctcgagggtca gacttgaaga agaacttcca taaagagttt
1501 tataatagtg ggtgggtgag ccacagtgggt aagggtgcct tttctggagt gacttga
1561 acaaaaatct gaactagttt gtaagataaa gtccctgctga gaggcaagga aacagatgaa
1621 ttgacttcac agacttagca atgtggggcc ttgagggctg agctgttttag tccagttccc
1681 taaactgttt cttggaaaga gtgaatttgc acactaatgg gagtgtgga acctgagctg
1741 aggatctcag aaggctcgggg ggtgggtgggg gcagtgggca cactgcaggg attttttcc
1801 tttgctaggg agaagatcca ggggaaggaga accttgactg aggaacccgg gagggtcctg
1861 gaggttgggga agtttgcccc ctggactggg gatttaacat aagatcaaca gtaaacctta
1921 ttgtgggatt tgatgctact gccaccaagc aagagtgaaa gaggaagaat tagagagaca
1981 gggccctgga gagtgggcta tttctgcctc ttgtcccctc cgtctgcccc aggaacctgg
2041 attgcagaga ctgtggtact ggggacctat ggtctctttt gggagaaaagt ttttgcctt
2101 tggaccaggt gactatcttt cacataaata gaaatagcaa acagttgcat aggtcttatt
2161 gtgtgctagg catttttgtg gggccaccag ctggctgagc cttcctgtcc cactgaaat
2221 gagattcccc ccagactccc accctgaact tgcgctgcct ctgctttcct cctcactaag
2281 ctaattaagg caggagggtta atgaccaaag aagccttggg gatttggcca gggaggaaaa
2341 caaagggcag aaacaagtat gtgtcagtgg ggagggtcca gctgctgaaat ggggaagggg
2401 ctggctgaaa ggtgggtggg gcgagggcag gtcccactct gaccacttc gtcccactg
2461 caccaacaag aatgaatcag cctctgggt aagttaatgt ccaccctcag ggtcccactg
2521 ggactccaga caaatttctc tcacagcaga tgtgagcctg gcacctggag ggcaaaggat
2581 atgtgtgcca atgcagttag gtgtgtggca tgggtgacat gtgagagtgt gtgctgcaga
2641 tcatgtagga gagtgccag gtgcagagga ggcccagggt gggcaccttg aggagctgga
2701 ttctcagagt gtttttaacg aatccttgcg tctctcccgg aggtgtgtgg cctcctttgt
2761 tcaggagacc cctcctca actgagaagg tagatttctg tatacaggt cctcctttcc
2821 ctgcttgctc gatccagaat tatggcaatg agagcagaag gtcttactta aacaatactc
2881 cccactctct gccttatgac cagaaagggt tagagggtggc atacttctcc ctttcaggac
2941 tttggaggac agcactatgt ctctcatctt catcatggtc cccaacccta gctgcccaga
3001 acctagcatg cagccctgcc cagagcagag gctcaatagg ggtatgttga gtgtaactgc
3061 atttttaaaa cttcatggtg tcaggaggaa agagcaggtc tccaaagata aacatgctgc
3121 ttattctgcc actttgggga gtcatttatt gctaagcctc agtttctta tctctcaaat
3181 ggaaataata agccaaggtc tggcagggca tgggtggctc ttctgtaat cccagcactt
3241 tgggaggcca aggtgggtga atcacttgag gccacgagtt ccagaccagc ctgggcaaca
3301 tggcaaaaacc ctgtctctac taaaaataca aaaattagct gggcatgggt gtgcacacct
3361 gtaatogcag ctacttggga ggctgaggca tgagagttgc ttgaaccagg aagcagaggt
3421 tgcagccaag attgtgccat tgcactccag cctgggcaac agagggagac tccgtctcaa
3481 aaaataaaat aaaataagcc aaggctcctaa cacagtgcct ggctatctta ggtgttcaat
3541 cagttttagt ttctggccc tggttgttct ccccttgggt ttgactttag ccagattgtg
3601 gtotctoatga attttcatte attogttcat tcattcagcc agccacttcc tattgagctc
3661 ctgggatgtg ccaggcactg gaggagattt attctgacct atcataggtc acttgaagtc
3721 acttctaggt ctagactcgc tgaataaagg tcattctaaa caaacagcaa ataattctaa
3781 aaataattcc tgttaagccc tggaatgcag gtgacctctg actctctctc tctctctctc
3841 tctctctctc tctttgtct ttcttacaca agccaggctt ctggcctcat gcctggggaa
3901 caaaggaccc tgttgagttg tttagtacat acacacacac aagccaggct gcagggcagg
3961 aaaggtgaca cagtggagag tccaggattc ttgtggaggt tgggggagga gtgggggtgga
4021 tgatccagac tgtgtgcttg tctgtgtctg gaggttgtatc ccttcttctt cttttttttt
4081 ccttttgaga cagagtctcg ctctgttgcc aggttggagt gcagtggcac gatcttggcc
4141 cactgcaacc tccgctctct gggttcaagc tattctcctg cctcagcctc ctgagtagct
4201 ggcattacag catgcaccac cactcctggc taattcttgt acttttagta gagacgggggt
4261 ttcaccatct tgaccaggct ggtctcaaac tctgacctc aggtgatccg cccccacctt
4321 ggcctoccaa agtactggga ttacaggcgt gagccaccgc gcccggccag gaattatctc
4381 ccttctttcc tctgagcact tctgggcacc tgttatcact gtctgaacac tcagtaccaa
4441 gggatggggg tgcacaccta acacttgcct ctccttagcc ctattctcag aacctgagg

4501 gagggagaca caggaggctg ccaccaaggc tgccttacca agcagactgg cccttaaaga
4561 ggtcatctcc cccatcttca tacttctctc aggatgcctt ctctatctgt cctagttagc
4621 tgagccttct ctcccatgta taatgagatt ttcccatttt cttttctttc cttttttttt
4681 ttttgtgggg ggacagagtc ttgctctgtc acccaggctg gaggtcagtg gcgccatctc
4741 ggctcactgc aacctccgcc tcccaggttc aagcagttct ctgcctcagc ctcttgagta
4801 gctgggatta cagatgcctg ccaccacgcc cggctaattt ttgtattttt agtagagaca
4861 gggtttcatc atgttgccca ggctgatctt gagctcctga ccttgtgatc caccacctc
4921 agcctcccaa agtgcctgaa ttacagtcgt gagccatcgc gcccggctcc cttttctttt
4981 tcttttcttt tctttttttg agacaaggtc ttgctctgtc acccagactg gaatgcagtg
5041 gcttgatcat ggctcactga agccttgatc tcccaggata aagtgatcct cccacctcag
5101 cctctccaat agctgaacca caggcacatg ctgccatacc cagccaattt ttaaattttt
5161 tgtagagatg agatctcgtc atgttgccca gtctggctct gaactcctgg cctcaggtga
5221 tcctcccact tcagcctccc aaagtgcctg gattatagga atgagccacc atgccagcc
5281 tttctttcct ttttttctt tctttctttt ttttttttga gatgggtgtc cgctcgggtca
5341 ccaggctggg gcgcagtggc accatctcag ctactgcaa cctccgcctc ctgggttcaa
5401 gtgattctcc tgcctcagcc tcctgagtag ctgggactac aggcgcgcac caccacaccc
5461 agctaatttt tgtattttta gtagagacag ggtttcaaca tgttgccag gatggtctgt
5521 atctcttgac ctcgtgatcc gcccgcctcg gcctcccaaa gtgctgggat tacaggtgtg
5581 agccaccagc ccggccccag cctttttttc cttaataacc ctttcttgtg ggaagtctgt
5641 gttgctatct aaccaccgtc ctttctgcta agagtccaga cagttttctc tcctttgata
5701 gaatgcactc aacacattaa ttaatttaat ttaatttttt tgttttcttt tttttaacct
5761 tataccagca gtccaaacct cactagcatt aattaatttt aatcaatgag aatatggaat
5821 aaaattctgt aggccacaca aaagaaatgt gagacttgat ccccaccctc ggtaagacac
5881 agaaaactggc ctgcatgacc cagagaacat gggtaagcag gattaaatgt aaggacaaaa
5941 tggctcttgg ggtagagcac ggtggctcct gcctataatc ccagcacctt ggggccaagg
6001 tgggaaggatt gcttgagccc aggagtttga gaccagcttg ggcaacaaag tgagacctca
6061 tctaatttat ttttatttta ttttatatat tttttgagac ggagtctcgc tctgttgcct
6121 aggctggagt gcagtagcac aatctcagct cactggagcc tctgcctcct ggattcaagc
6181 aattctctct cctcagcctc ccaagtagct gggattacag gcatgtgcca ccacaactgg
6241 ctacttgttt ttgttatttt tagtagagat ggggtttcac cgtgttagcc aggatggtct
6301 caatctctct acctcatgat ccacctgcct cagtctccca aagtgcctgat attataggcg
6361 tgagccaccg caccagcctt attttatttt attgattaat ttatttgttt gtttgtttat
6421 tttgagccag ggtcttgctc tgtcgtccag gctataatac tgcggcatga tcttgactca
6481 ctgaaaactc gaactcccag gctcaagtga gcctcccacc tcagcttctc gagttgtttg
6541 gtctacaggc atgcgcaccc tgcctggcta atttttgtat tttttgtaga gatggggttt
6601 tgccatgttg gccaggctgg tgtcaaactc ctgagctcaa gtgatccgcc tacctcagcc
6661 tctcaaagtg ctgggattgc aggcattgtac caccataccc aacccttgtt tatttattta
6721 tttatttatt tatttatttt tgagatgaag tcttgctctg ttgctcaggc tggaatgcag
6781 tgggtgtgatc ttggctcact gcaacctctg actcctgggt tcaagcgatt ctctgcctc
6841 agcctcctga gtacctggga ttacaggcgc gcacccccac gccagctaa tttttgtatt
6901 tttagtagag acggggtttt accatggttg tcaggctggg ctogaactcc tgacttcatg
6961 atctgcccgc cttggcctcc attgcactcc agcctgggca gcaagagcca aactccgtct
7021 aaaaaaaaaa aagaaagaaa gaaagaaaga aaggctcttg ggactcggag ggctacagag
7081 gttggggggag agaggctcact gtgagtcctt ctgttgttgt ggatctgatc tccaagggct
7141 ttgcaggctc agcctagggt tggagatgga caaacgtctc tgcttctccc ctgggactac
7201 actggcactt cctctgccag gtcccctgca ttgtctgatt tcccagggga gagtagctca
7261 gaggggctgg ggtggagtct gggcccctta ggacagctgg ttgatgagge ccctctggga
7321 cttagggctg gaccaggcg tccaggagge gggtgatact ggagctgtga cagttggctc
7381 ccagattoca caccagtggt cgcagctctt ggtttcagac tgcggtgagg gagggggagg
7441 ggcacctggg gccagccagg ggtggatgct tgtctggggg cagggagggg gaggcagtaa
7501 acatgtcctt cctaggttgt tgaggatttt ctaattgccc tctctgtgcc aggcactgtg
7561 caaggtacta acgagattct gcaatgcaag atgcagtctg gcacagatga attgacagca
7621 tacagtatag aattagaatg tcagagctgg actagaagtt gaatgaatga atgagtaggg
7681 gaatgagtga atgaatagtt catcagttac ctaatgcact cattttacag aggagactga
7741 ggcccagaga accogagtoc tgtgactccc ttctgctttc tgatctgtca tgcttcttt
7801 ttcaggaaaag ggaagcaaaa ggcagagaga ggggtgggaa gggacttgac gtttctctgag
7861 tgcattgttg gtgccaggct ctgttctggg cttttcagac ttgatgtaat tgccttca

7921 attctgtgaa gatggacttg gtgtctctgc ttcacatgaa gggacctgag gttcagagag
7981 gctaagtaac tttcctaggt cacagagctg gtaagtagag acttttagtgt tcaaaaccag
8041 agcagtctga cccagagacc tgggaccca gaggtagaca gtgatactca gtgagtgtctt
8101 ggagcccaca gagggtggag ggagtggttc ctggatgaag tagatctgaa ggggtgggtga
8161 ggttcatttg aaaggcgaga attggaggca tgtctggaag ttggtgctgg aggtgcccag
8221 ggagacttgc tggttgagca ggtggccggc ttgcatggat cttgggggagc aggggacagg
8281 ttcaaagaaa cacatcttta aacattggcc accaggtggc gccgcaggcc aattcctaga
8341 agcggctcca gcctcagcgc cctcactagg gcccaggtat gacttcgctg ctccagccagg
8401 gctccatgct cggcttgctt cctcaggtgt catctccagc ttccgcttgc ccctggcttc
8461 tggcatccaa cacatggggg ataattgatgc taccagtcgt ttctgactc gggacgaaca
8521 cccaggtagc tgagacctgg gaatccccgt gggctctgagc agctcccccg gtcacccgga
8581 gaccctcca gtcgcctaaa cctggcttca tgtccagcgc cttcttgagt ttctgggaga
8641 cactgttcag aggacaggct gaggaaacag gcattttgagt ccaactgacc tgggtctgaa
8701 gggaaataat aatatctacc ttgtaaggat ttaaaacatt tcatctatgg ccaggtgctgg
8761 tggctcacc ctgtaatccc agcactttgg gaggccgagg cgggcagatc acctgaggtg
8821 gggagttcga gaccagcctg accaacatgg agaaaaccgg tctctactaa aaatacaaaa
8881 ttagccgggc atggtggcgc atacctgtaa tcccagctac tcaggaggct gaggcaggag
8941 aatctcttga acctgggaga cagaggttgc ggtaagccaa gatcgcgcca ctgcaactcca
9001 gcctgggcaa taacagcaaa actccgtctc aaaaacaaac aaacaaacaa acaaaccttt
9061 catccaataa atatctatca aatacatgct gtgtcaggaa ctattctagg ccctgaatac
9121 atcagagAAC aagatagaca aggtccctac ttgcaaggag ctgatactct actggggagg
9181 aaggcagaca gtagtcaagt agataatgta ctttctggta gtgatcatca ctgcaaagaa
9241 agataagcca ggatatgggg gcagaaaggT ttgaggtggc taatttagaa aggatggtca
9301 tgctatctgg ccaggacaac atagtgagac cccatctcta caaaaaaaaa tttttttttt
9361 ttgagatgga gttttgtctt tgttgcccag gctggagtgc agtggcatga ttttggctca
9421 ctgcaacctc cgctcctgg gttcaagcca ttctcctgcc tcggcctccc tagtagctgg
9481 gattacaggt atgcatcacc acaccgggca aatttttgta ttttttagtag agatgggggtt
9541 togccatctt ggtcagactg gtctcaaaact cctgacctca ggtaatccac tcacctggc
9601 ctoccaaagt gctgggatta caggcgtgag ccacagtgcc cagccaaaaa aaattttttt
9661 taattagcca ggtgtggtga gttgtctctt tagtcttagc tacttaggag gctgagatag
9721 gaggagcaact tgagttcaag gttacagtga gctgtgatcg tgccactgca ctccagattg
9781 ggtgacagag cgagatccca tcactattca aaaatttaca aaaaggctgg gcacagtggc
9841 tctgtctgt aatccaagca ctttgggagg ccgaggcagg tggatcactt gaggtcggga
9901 gtttgagacc agcctgggca atatggcaaa acccctctc tacaaaaaat acaaaaaatta
9961 gctgggctg gtggtgggta cctgtaatcc cagctacttg ggaggctgag gcaggagaat
10021 tgcttgaagt tgagaggtgg aggttgcagt aagctgtgat tgcaccactg cactccggct
10081 ctagacaaca gagcgagact ccatctcaaa aaaaaaaaaa aaatgtgatc atgaaatgcc
10141 cctgataaaa gcaccatctt agttccccgt tgccatcaga ttcacactcc ttagtgtggT
10201 gatgatgtoc ctcttgtcct ggccagctcc tgcccacctc tccagcccat cctgtctctc
10261 cttcccttac tccagccaca tggactttct cttttctttc tttttctttc tttttttctt
10321 tttttttttt ttttgagaca gaggagtctt gctgtgttta tgctggagtg cagtgggtgcg
10381 atctoggtct actgcaactt ccgctcctg ggttcaagca attctcctgc ctccagctcc
10441 caagtagcta ggattacaga tgctgccac cacgctcagc taattttttg tattttttgt
10501 agagacagtg tttcaccatg ttggccaggc tggttcgaac tcctgacttc aggtgatctg
10561 cccaccttgg ccccccaaa gtgctgggat tacaggtgtg agccaccact cccagccac
10621 atggactttc tttcagggtct ttggcctgtg ctatcatcat ctgctattcc ctgacccttc
10681 acctggctaa tgtttgttca tcttgcaagga ctcaatttat atgtcaattt tttttttttt
10741 tttttttttt ttttagatagg ttctggctct gtcaccagg ctagagtgca gtggcacgat
10801 cttggcttac tgcacctct gctcctggc tcaagccatc ctctcacctc agcctcccaa
10861 gtagctggga ttacaggcgt gcgccaccat gctcagctaa tttttgtaat ttttttagag
10921 acagggtttc accatattgc ccaggctgat ctcaaactcc tgaactcaag tgatctgcc
10981 gccttggcct cccaaagtgc tgggattact ggtgtgagcc actgcgtcca actacatgtc
11041 actttttct agaaatctoc cctggctact tctggatata ccttacaatt catcatgtc
11101 tattgtaact gtatgtttac acttctgttt cccagtaca cgggaagctc tttgagaaca
11161 gggattggat ctagtctact cctgtattct cagcaccag cacaaagtag gtgcttaata
11221 aatgtttgct gttoggaaat tgtgtatgct catccatcct tccctcttga acaccctctc
11281 accatataat tgccaaatat tccacttctc cctctgcttt tctggcatac ctttctctcc

11341 cacatccagc tccatctctc ctctcccctt ccatcccccac tcctacacct gggccctcgc
11401 agcctgccct catctgatct ccaaggggct actgtctgca tgaggtttag tatactctgt
11461 cactgggttc ctccactctc cagcacctgg tctggggccg tacacagcag gaactccata
11521 gtgttgctaa ctcaactgaca tgacttggtg gatggctcag aagggatccc taagcatgca
11581 tacacctcct ctcaactcacc gctagtaaaa tgagtctatc acgagaagta tacatgggtc
11641 acatgctcct ggccaggccc tgtctatcct ctagggcttt gcacacactg ttccctctgc
11701 ctogaaccct ctctctcatt ttggctatct cccactcatc cttcagacct gagctgaagt
11761 ggaggagtca tgtcttctct ttttttgaga cagtcttget gtgtcaccca ggctggagtg
11821 cagtggtgca gtcttggtct actgcaacct ctgcctcccg ggttcaagtg attcttctgc
11881 ctcaactcct cgagtagctg ggactacaga ggcgtgccac catgcctggc taatttttgt
11941 attttttagta gagacagggt tttaccatat tggccagggc tttggcctgt gctatctcaa
12001 actcctgacc tcgtgatctg cccaccttgg cctcccaaag tgctgggatt acaggcgtga
12061 gccaccgcac ctggccaagg aggagtgtg tcttctgaga ggcttctctg accatgctga
12121 tggctcttaca tataccccc aagtctgctg ggcactcagt atttaccat ttgttagcac
12181 ttagctccta gtgctttctt ttttttgat ggggtcttgt gctgtcacc aggctggagt
12241 gcagtggcat gatctcggct cactgcaacc tccgcctcct gggttcaagt gattctcctg
12301 cctcagcctc ccaagtagct gagattacag gcgtttgtca ccacaaccag ctaatttttt
12361 tttttttttt tgagaacgga gtcttgetct gtcgccagg ctggagtaca gtggcgtgca
12421 ctcaactgaa cttccgcctt gctggttcaa gtgattctcc tgcctcagcc tcccaagtag
12481 ctgggactac aggcgcacgc caccatgccc ggcctaccgc taatttttgt atttttatta
12541 gagatgtggt tgcacatgt tggccaggct ggtcttgaac tcctgacctc aggtgatctg
12601 tcctcctcag cctcccaaag tgatgggatt acaggcgtta gccactgcac ccggccaact
12661 cccagtgcct tcaaagccag tttgtgtgtt catgttactc ggactgatcg ctctccatga
12721 ataagggatg tgtctctcca aaacctagca tataggaggc attcaataaa aatttattga
12781 gtaaattgatt gattgaatga atgaaaggaa taaaatgacc atgtgatact gaatgagtca
12841 ctcccatctc cagtttctct ttctcaagtg attctcctgc cttagcctcc tgagtagcta
12901 ggattacagg cacctgccac cacaccagc taatgctttt ttttttttct ttttttgtat
12961 ttttagtaga gacggggttt caccagggtg ggcaggctgg tctcgaactc ctgacctcag
13021 gtgatccacc cgctcggccc tctcaaagtg ctgggattac aggcgtgagc cactgcgccc
13081 ggtgtccaag acacttctaa cgccaacatc cttgtactag gccctcttcc agctccccc
13141 ggagtctggg ggttagaggc caggctgctc cccactctat tacctcccca cacttctaat
13201 taattcttgt gttcagtccc agaacctgga gagttacact caggctgcct ggtcctgctt
13261 attgcccata tcctagaggc acttgcctgc cctccaggc gggtaactgt acatgtctct
13321 cccaggcagc cttcacagac tactcagggg tatggattgg ctccagctct gacactcaga
13381 cctgagcaga ggcgtgatct actctggcct cctgtgtttc tgtgccccac cttcccacac
13441 tgcccagggc tcctgagag tgatgctgag tctagtctca gcttttatgt tgctttgctt
13501 ttgacctctg acccaacaca tgctccttt cctctgaaag tcttgatttt ctccctggct
13561 tttctccacc ttctgggtgt tcctttttgg ccttctttgc ccatctctcc tcctctgtct
13621 acacagccag gggctcccc ttccgtttcc tctgcctgga acacacttgc ttctcccttt
13681 ctgactggct attcaagcag gatcctgtag gtgggtggca gtgcactgtg tattacctgt
13741 gagccctatg caaattaccg cacatctctg aacctgtttt cccatctata aagtgaaget
13801 aatagtagca attacctcat gggattgttg tgcaattagt ggggaagaggc atgtaaagta
13861 ctgacatact gtctagtaca cagaaagtgc ttaataaatg ttggtgatta ttatccagga
13921 ctcaacttaa tcctcccttc ctcaaggcag cctgccacac aaccgcagac cgggttaggt
13981 gcgggttttt ctgctcccca gcgtttggag gcaactcatca cacattaatt actttatcca
14041 tgaagatgaa agctgtcttt ggcaccaga ttggtactgag taaatatctg tgaatgaatg
14101 aatgatgtgg aatgtctcct ctggaaggac acagtcaggc cctcctgcca ctcaaggacta
14161 gctgggatgc acccagaggg aaggagaagg tggggctggc ctccgctggg ccagcagggtg
14221 ggcagagggc ctgtgggtct ctccccttgg ccacctccct tgcgcttget catctcatgc
14281 ccaggcctg agggttctca gctggggcag gagagatggg gtcctgagge ctgggtgag
14341 ttccaagccc tggggctcgg gggcagactc tcgctcacc ctcccagaaa gatgttttcc
14401 taggggatgt ccagcctaac acggcaggga aggaagtggg ggtcacctgg gaaaggggccc
14461 aacctgagag gttgggggtc ctgatgaggc tgctattctc ttgaggacat gaaatgagag
14521 ggggtgaagt cagtcagtga gtcaaaagct gagatctagg aaccagacac aatgtgggag
14581 aagctgcggg gaaatcaatc agtgagagag gaaggggagg cccggtttcc atgacaacgg
14641 tagtgtttgt ggcaagctgg ggactcctgg gtcccaggg cccgcaagca gtgattcacc
14701 acaggaaagg gaaggctgca gagtgcctg gtgctgggga ccggatgcct tgtggctcct

14761 tctcaactgct gagcatgcat tccccagacc caggccttgg atctctgctc tgccagctac
14821 tggcagggaa actgagggat ggcagaactg agcccccaata aggggcaggc gctttgctcc
14881 acgccgtaca tgcacatata acttatataca cttattgctc acaacagtcc tatggaatta
14941 ccgttatccc tgttttatca aagaggaaac tctcagcctg gccacaggtg ccatgggcag
15001 gacatgagtg gcagagccag gattggaacc caggctagcc tgactcttta ccatgcttgg
15061 gaaggcccag aaagcagaaa aggaagagca ggagcatgag agttagagga tggttcagac
15121 cccagaggag atggacccca aggccctag cctccctgcc tgccctgccc tatgagagcc
15181 caacactgat acggaggaaa gctggccaac tgcatggcct aatgtgggcc agaacctggg
15241 cagccagcct acctctcact ctctccaag cccaagagag taggccccag ttccagtcca
15301 tccaactgga tggtagctcc cccaccgttt tctctctgcc ctggtcctag gggcaggggt
15361 gcgtgtgctg gtgtgctgct gtgtgtgtgt gtgtgtgatc atcatgtcta caccaccac
15421 caccatcacc cctatctctc agcttctgct atagcctggg agttcccagc cttttagtcc
15481 ctgtcaggat agtttcattc ttgttcagct ttggggccca cacaatttca gtacaaaagg
15541 ggtatgccc tctcaccctg gcccttgata gcatgcaaga gcatagttca ggcctcatc
15601 atctattccc ttgtagtaga agggatggag gttggaaaga ccttagaaaat cttccagtct
15661 ggttctcttg ttttgaagg gtgaaagctg aggcctatag aaggggagca atttttctga
15721 ggtcacagag caagttaatg gaagagtga ctagattccc tgtccacaga cttcaggttc
15781 tgtaatccca gaaaggactt cttgaatttg gcattgagtt cttaggatgg gagtcaggta
15841 gatgaacaga tggagtgaga cagatgaaat tggataacct agcataaaaat gccatgtgaa
15901 aagaaggtag gcaactcatg gcccgactgt ctaccatggt acatcaggaa taactgaggt
15961 gagatcatgg gaaagacttc tcacatatgt atgttagggg tacaagatct gtgggtcaaaa
16021 tgggctgtgt gatcccagac aaatcactta ccttttctga gccccagtgg taactcagga
16081 taccctacct gcaggcttgt ggagaggatg gagggggtgg tgggagagta cttgtgaact
16141 gtggggtgac atgaatgttg tcagttgtca gggaaaggcaa gtcatgtcta gtgggactga
16201 gatagagcta cctgcagtg caaaggtaag gggtaagcca tgatgctcac ccctcccacc
16261 cccatccttg ggagagaaaag ggaggaacat aaacttcag tatgcttgat acaagcaaca
16321 gttctccttg ggggtgggca gcaggttct gttagaccct ggggtgccc gaggtgtgg
16381 gaagctgagc actaattagc tctggctttc taatcgggct taattgcctc ctcatgatgg
16441 gcaattagga agctgcatca attagcacag ggaggtgaca actcagctgg ggttggtcag
16501 ggctccccgt tggcagggag ggaggcaggc aggcaaaggc tgccccactg tcagccccct
16561 ccttccccctt gaaggagagg gccctggagc cctctcctta gcaattaggc cttagttcct
16621 tagctoctag gagctactag ccttctagc ctggattgtg gcttttctgg ccttggttta
16681 ttcacctta ctggctctct gacttttctg cccagaggag acttgcatte tgggcctagc
16741 tccaccactg gctgctttgc aactttgggg ttgacagcta tggagtcaag gttgagaatg
16801 tggcctttga agctaggtct ctagggtttg aatcccagct ccaccgttta ttaattgtgc
16861 aacattggac cctttgtgct ttgttttctc acctgtaaaa tggaaactgat aataaaaagt
16921 ctgtoctagc tacacatgag gctggggcag gaggattgct tgagcccagg agttttaagg
16981 ctgcagtgag ggacaataga gtaagacacc atctctaaaa aataatttag gccaggcata
17041 gtggctocaa ccgtaatcct agcactttag gaggccgagg tgggtgatca tctgagcttg
17101 ggagttcaag accagcctgg ccaacatggc gaaaacccta cagagtctac taaaaataca
17161 aaaattagct gggcgtgctg gtgtgcacct gtaatctcag ctactcagga ggctgaggca
17221 aaagaattgc ttgagctcag gaggcagagg ctgcagtgag ccgagatggt gccactgcac
17281 tccagcctgg gtgacaaaaa aaaaaaaaaa aaaaagaaaa atgcccatgt gtcaccccag
17341 atcaattact aatcaaaaata tccaagactt tgggcatttg cagttttgaa agccccctc
17401 ctocaagcaa tgttgatgcc cagtgagggt tgagatcact agatgaatgg tctggagggc
17461 tccatccagc ctggcatagc acctccctgt cacataccac tcaggacccc tctgcatagc
17521 cctggagggtg ctgcactttg gaagagagct cattaccacc agtgtggctg gaagcaggtt
17581 cctctocact cacttagagg gagttctcct ctataaaatg agaataatga ccttgccatg
17641 ctcaacctcc agggctgccc agacataatg ttttactagt gataggacct gctttatagg
17701 taggcaatta tattgttctc tcccttgaaa aacttttatc gaggtctagt gcacaccat
17761 ctcttttaca catggtgact atggagttaa taatctcctg gagggacctg tcctgaacca
17821 tcagacaagt taggtcatag ggccaacccc ctttttttagc tcctgctggt tatgaaatga
17881 gccaaggaat ggaaagagcc tagaacaggt cctgacacat ggtgagtgct ctctaaacag
17941 tagctctcct tactattggt attgtgaccg tgccccaggc taggttggag ggaggcctgg
18001 ttgcagtgca cacagggagc tcagatgaag ggaggcagtg aaggtggggc ctgcaggagg
18061 aaggggagtt tagaatgagg tgctgaggag taaggtgtgt gggaaagagta gaggaaggag
18121 tagagcacat cgaggttgc tgggtgtgag ggggctatga tgaggatact ctactcaaac

18181 atgtattgag aaccttttgt gcaccaggca ctgtgccagg tgetggcaga gcaccagatc
18241 caccagctct ctacccttga ggagtccctc taatggggga agacacacat acacagacaa
18301 tgtcacagac agtgtgacaa gggatactgg tggcgagtgg ggacatttaa tccagactta
18361 gtggcacttt tagctgagcc tggatgatga gtaggagttg atcagagggg aaagttaaag
18421 ccatgtctgg aaaaattagt cagaaagcaa tgagagaggt tgagattgct ggaatttaga
18481 taaggctaga gatgcaagca gagcctaggt ctccggagggc attacaagtc aggtcaagga
18541 atttggactt atcctaaata caatgtggag ccactgttag tttaaagtct tctggaaaca
18601 aatgaggctt ttcattttga aaagatcgct gtgggctggg cgtggcacct cacgcctata
18661 atcccaacac ttttaggagg tcaagggtga aggatctctt tgagccccgc agttcgagaa
18721 tagcctgggc aacaaagtga gacccagct ctacaaataa aaataaaaaa aattagttgg
18781 ttgtggttgc atgaacctgg tcccagctac tcaggaggct gaggcaggag gatcacttga
18841 gcccatgaat tttggattgc agtgagctat gattgcacca atgcaactca gcctgctgga
18901 tgacaaagca agatcctgtc tcaaaaggag agaggggaagg aaggaaggga gggagggaaat
18961 cactgtggga aggagaaaca atgggagcag aggcaagggt ggaggcagag agactggttg
19021 ggaggctgct gcagttaccc aggaaggaag gacggaagga aggaaggag ggaggggagg
19081 agggaggggag ggagggagcg agggaatcac tatggtttat agaggagaaa ggtgaggagc
19141 agaggcaagg gtggaggcag agagactggt tgtgaggccg ctgcagttat ccagggtggg
19201 gttggtggtg agcagaccag agagctggag gtagaacata gaacagataa caattctgga
19261 ggtaggaacc acaaggactt gggggctcat tggccaaggg agatgaagga ggagaggctg
19321 aggaaggagg agcagagtga agctgtgtgt tgggagacat gaactccatc tgagatatcc
19381 aaggagtgat gatgagttat ggatgcttag gaatcactcc tggctctggga atacttggg
19441 tegtcttcta agaattctca agagtggag gccactcacc cttggcccag tgccaggccc
19501 tgccttccc tagtctctct gggacaggct gggaaattcct ggctcagaaa taccctgttg
19561 gggcaggccg gccttgatgc tcccagccc aaaggaggga atgtgcctgt gcacctgctc
19621 cgtgatacac gggggcggga cagtggagtg gaggtgagtg aattcagggg tgttcctcac
19681 tcccccatc ctggacttct aggtcagtt gtggccccag gaatggattc tgagtaagag
19741 gccagaaagc aagttctctg ggctggggg acagagggca gcaaggtaac gttgacatac
19801 cgcagacacc caggcagggc aagggcaccc aggtttcagt cctggttctg tcacctctgg
19861 ttaggagctg tgagcaccca agggctcatt cttgtcctaa aactgggtct tcaagatggt
19921 ttggtttagg tgggtgtaat gaaccttcc atctccacc tcgagagagc tgaattttcc
19981 cccaccogag gcgttccaca gtgggcagaa cgccgcagcc cagtctgccc tggcccacc
20041 agtgetoate tccatttcca gcccctctg ggcgcctctg gcctcctgca gccggtctt
20101 tgcgctcctt cattaagctt tggtttctgt gggaaatgta cttgaattaa ctcttgggag
20161 aaaaatcaca gaaatcagga aaggggggcc atggggaggg ggcagtgagt cacgtttcca
20221 tgtgtatttt tcaaaacttt cgatgggagt ttggaaaaca acagtctttt ggttgccagc
20281 ttgggcccag ctctccaagg gtcccatatg ctgtgtctat aggtcttccc gtaactgctg
20341 cgtcccggcc cacccttggg gtctgcctta tttcctggga tcccttccca ctctcctctg
20401 gggacttgtg ctgttgacaga ggcactgaga agtccatccc gtggtgcagc ttccgtctct
20461 cccctggttt catttcaacg ccaagccctc atctatcctt gcctcaaggt ggacccatgg
20521 gggctgcagc tgagctgggg actttgtccc tgcagctgct ctggccaacc ctccaaccct
20581 ggagaacctc cataggcacc aggaagtcct gggctctagg aacctccttc ctgccccac
20641 agcttctctt ttgtgtctct cctaccagtc cttcccggcc atctctgcat ccagtttcc
20701 agagggcctg tggctgaggg tgagaagcca gcgatcgacc cccagggaga agtcttgagc
20761 aggagggctc tcatoctaag gaggccgcca aggaagaggg ccccagatg gccctctctg
20821 tgtgctgggc cctgcactt cctctgtctg ggactcctct aacagctgtg gggaggggtg
20881 caggatcccc attttgcata cagtagctga ggctcagcca ggaaaatggg tgggaacctt
20941 aaggatgggt ggcagatggt ggcccccttg agggagagat gggagggcag cgcatacccc
21001 tagccaaccc ctocctgcaa gccctcattc gggcgggggag aaggaagggg tcttgacta
21061 gctctcctgt gtccccctct cgggggctgc cccctcccag tttctgacta atcctttcca
21121 toggcagtgg catgtctctg ccctgctgtg cctgtgcca caacagccca catcacaggg
21181 ccacatctgg ctocatttga acatacccca cctccctgc tctttatccc tcaacgcctt
21241 cccctgocca aggaccagag tagggagtgg cctctggac cccctctctt ggtccacca
21301 gccctgctcc cctgcccctt cgaagcctct agtggagtca ctctttctt tccccgaacc
21361 cggcctcagt tctgggaca tctggtctcc attcttcagc acacctccc tcatcatatc
21421 cacactcctt ggctcccccc ttacagccc ccaactgaagg agggattggg aaggggacat
21481 tttgcaagggt ctgagcccca agagatgtcc cagagtagag ggaaggcctg gcagcccagg
21541 gattttcct tagcccagct ctctggcata ttgccagctt gggcgtgttg ggcggcaggg

21601 gtcgggggtga tccaagagg tgcgtgtatg gaggggtata getcagcctc ccagctcggg
21661 tgggggagcgg tggctcaggc ctgtgtaggc tggcttttgt tggggaggag cctggaaggg
21721 cctgcagcta ctggcctccc tcctccttcc tccttgctta gcaactgttg tctgtctggta
21781 aatatttgcc ccaacaggat ctggggctgg agcactggcg tcagccgagg tagttgcccc
21841 ctctcaatt tatgagtctc ccctctgttc agttccctat tccagagccc cctggactgg
21901 attctaaatg tggctctcac cccctcccctc atggttgatt tcctcccaaa cctcttccct
21961 gtctcccttt tactctcctt catccccact catcctctgt gccaggtct ctctcgtcca
22021 tcctccactc ctggatccat tcaccaaggg gcctgagtga ggtcccaccc cctaagccac
22081 acaggctctc tgcttctccc ccgctacatc tcaactgggac cccagcctgg caacggctag
22141 tgcgtcgtcg gtgctcagtt agccccaaac tctctttcct tcctccaac aggaaacaca
22201 ccttccacct ctaaacctta gctccaccaa gctccaaggg gagagaagag agggcatatg
22261 ggaatgtttt gctggacccc atactgcaact cccaggccag gaagctctgc atcaggaagc
22321 cagcaccatt ttcacctcct ctggggtagg actgagggga ccatggccaa gaggaaacag
22381 atgccccctt agctcctccc tgggtagcct gagcgggcca gggcctgaga gcatgccagt
22441 tttagcctgt ctctctgtcc ttcaccagca gacccctctc atcctcctcc cccagtggtt
22501 ttctcattac ctgtcactga cggagagccc ctcaagagtc aaggccaaaag tgagggtggc
22561 tcttggcatg ctgtgcaagc ttggagctca cccggcaaaag agggccaatg ggggtctcct
22621 ctggggcggg cctgtgggct ggcttaggga tagcagacag gggaaatctgg ggagttggac
22681 taatgtggat ctagaaggga ggtggttggg ctgcacgatg ccttaagact cctcccagct
22741 ctgaagttct acttctaggc aatgaatgat gaaacgctgg cccaagtgca gtcccttccc
22801 catcccacag ggcggagacc ccagccatcc ctactctact tggcagcccc cctccccact
22861 cctcctgaga gtgccccctg ctccacccca gcggccagga aatctcctgc agattacca
22921 ccacctcctt ggctgggagt tcactttcct agttgacctc ctggcctgag ggccagagga
22981 gagctttcaa cggggacctt gaggagtgtg agggctgcag gggctgtgga ctggagtgga
23041 ttcaccaggg agcaaaggaa agtgaagttt cagggcctct cagttgcacg ggcccttca
23101 agaccogaag aggagccctg ccaaactttc cacatgatcc catgtccata acatttgcaa
23161 aaagatcatt ttgtatttct taaagagggc ctccccacaa aattgtgtgc actttaggta
23221 ggatgaacaa ctgtccctat tttagcatta aaagtccctt atctggccag gcacgggtggc
23281 tcacgcctgt aatcccagca ctttgggagg ccgaggcagg cggatcacga ggccagagga
23341 ccgagaccat cctggctaac acggtgaaac cccgtctcta cttaaaaaaaa aacagaaaac
23401 aaaaaacaaa aaaatagcgg ggcgtggtgg ctggctcctg tagtcccagg tactagggag
23461 gctgaggccg gagaatggcg tgaaccgggg aggcggggct tgcagtgagc cgagatcgcg
23521 ccactacact ccagcctggg agacacagcg agactccgtc tcaaaaaaaaa aaaaaaatc
23581 ccttatocta ggagactcct cagtccccgg caaacccagg gcaattgccc accctagttt
23641 caggctccac aaacatttgg cccacactgg gttgaggggtg agggagaaaag ggaatccttt
23701 tctgcctcac ttctggaaa tgtgtagggg cgtgagctat ggggcgaggg ggccctctcc
23761 tccagctoca gtctggagt tgtggcgtt ctgtttgttt ccttggcaac agggaaaagga
23821 aacccccatt taggaggccg gggaggcccc agcaaccaca cacacacaga gacacacaga
23881 ctcaacogcg gccgccaca cgcagtgcct tgcgtcagca ccagccgtca ggccctcget
23941 gccgcctgc ggggtgtgga gtggggagaa cagccccgcc tgaatccagt ccgactaccc
24001 gccctactg atttgcctgg gacgcccgg gccacatgg cccacccac gcctctgtcc
24061 ccaccccccc acgctctgt cccaccccc ccaccgcccc cggcagcgtt cccccctaa
24121 caccttcccc cagggtgtga ggggcaaaag ctccgggatg gacaagagaa aggaaggaga
24181 tgtcactgct gaatttctgc tgaatttctt tctgcctgct gcccgatttc tatttgactg
24241 gcaatagtct ggtgtgtatt gtgacagttc agggagcctg agactacaa gtgaggtggc
24301 gatttgggoc tggttgggtg ggggaggtgg tttggaatgc cgggaaatgga gctcctgggg
24361 agactgtagg attttgcctg tggattgggt tccttgggat ctgggtcgaa gagggaaaga
24421 gtatatttaa cctggaggca gaaacctggg gtttgggcct gcttctgtca cttgtcaget
24481 gtgtggcctg gagacaatca gttagcctgc ctaagcctct gtctcctcag ctgtaatatg
24541 cagataatgg tgcttgctg tgcacctcaa agaaatttga gaaggctaa gttagtaacct
24601 tgattgtgaa aagcttgtca ccggggggac tggggcagg gggaggcctt ggttactact
24661 tatccagctg cccctocagc cccgctggct cagtgcagccc agcagccact tcccacttt
24721 gtaggaaccc atgccagtc ctaagtggac ctgctctttg acatctgtcc agtgtggctg
24781 cctgtctctc catggcagcg tggcactgag ctgcagtgtc gtcttggagg ttttgacct
24841 goctggctgt gtgactttgg gcaaactttt cctcctctgg cctcagaatt ccactataa
24901 agaaaggggc ttattctca gagcccagg gctcctgga agtgggctgg ggggatcaac
24961 caggcagggg cctcaagctc cctaagtaga gaagctctgc ctctgtcatt tagtagattg

25021 tgctttttgg gtgagatttc atttgagtga gggggatctc tgactttaca aatagctgtc
25081 attcacttgc tccatctaag ggcccgttc tcacccggca ccacctatga gctgctca
25141 ttcccccttg ctgagcctga gaagaaatct actcagctgt gctacctcag gtcaattaca
25201 acctcgtgga gtttgggttg ttgttttttt tttctttttt ttcttttttt ttgagatgga
25261 gtctcactct gtcaccagg ctggagtgc gtgacatgat ctctgcctc agctccctga
25321 gtagctggga ttacagggtgc atgccaccac gccagctga ttttttgtat ttttggtaga
25381 gatggagttt taccatgttg gccaggctgg tcttgaactc ctgacctcaa gtaatccacc
25441 cacctcagcc tcccaaatg ctgagattac aggtgtgagc cactgctccc agccggaatt
25501 tgggttcttg acacataaaa tgcagataat aatacctacc tcataagatt gaggtgaaga
25561 ttaaattggga tcacctccat atagccccta gcttggttgc tggcacatgg cagcccctta
25621 aagagtaggc acggtcttcc actgtcaagg ttataactat cacctgcaca tgagtgactc
25681 agcacttcaa tcattgttta ataccctgta gtcaaaggat gaataagacc ctccctcaag
25741 aactcactat aaagagattc ccaaaccatg tcccagaatc tgtacatcac ctccacctcc
25801 tcagaccctc ctccaaagag tgcctgtct ctgactccta cctgtccagg gttctagcca
25861 cctgatggta agatcttggc tttccaattc aggcaaactc gtattaaaaat ttcagtctctg
25921 gctgggcgtg gtggctcatg cctgtaatcc cagcactttc ggaggctgag gtgaccagat
25981 cacctgaggt caggagttac atggtgaaat cccgtctcta ctaaaaatac aaaaaaatta
26041 gctgggctaa ttttttagtg gcgcatggtt gtagtcccag ctactcggga ggctgaggca
26101 ggagaatcgc ttgaattcag gaggtgcagg ttgcagtgag ctgagatcgt gccattgcac
26161 tccagcccag gcaacgggcg aaactccatc tcaaaaaaaaa aaaaatccag ttctgttact
26221 aaccatataga ccttagacaa gtgattctca ataaggggta attttatccc ccaggggatg
26281 ttaggcaatg cctgaagata tttttgattg tcataactgg gaatggaagt gctactggcc
26341 taaagtctgg agaggctagg gatgttacia accatctcag agtacaaaaga acctctctg
26401 acaacagaga attatctggc ccaaaattca gtagtgccaa ggttgagaaa cctgacctta
26461 aatgaataat ttaacctttc tgggcttcag tttctctatc tgtaaaaatag ggctattaat
26521 agtatctacc ttgtagagct attgtattgg tccaacaaaa taatgcaagt acctagtgcc
26581 taataataaa ggtagtccat aaatattccc tacacgaatg gttcttagca tgcctctgga
26641 atccctctg gcttcatcac tgggctgaac ccaccggacc ctagtttggc ttatttcttg
26701 ccaaggccag tgcctctgac tctttcttca ctggtcttga tgctttcctt ctctccttga
26761 cctcaaggcc tatcaagcag ctggaagcct ctccctcagg ccccatccca gcaccaggcc
26821 cctgcctcca cctcccactt ctcccttccc gtctctcttc cccctcttcc cagccttacc
26881 tctagtggct ccccatgccc cttggcagggt cagagcacac tcagcttttt cccacttcag
26941 cacctctgca cagacttctt ccagccagag tggcagaaga cagagaagca tgtccaggat
27001 ttggaccaga ctgcctggtc ccagttctgc actgggcctg ttacttacgg cctagtthta
27061 ggaaagtcag cctcatctgc aaagcaggga aataatttcc aacctatgag gttgtgtat
27121 ggaccagtt gagaatgtat gaaaacatgc aacaactata aagattgtta aaatgtaagt
27181 tttcactgt gtgagcatta ttctgctaaa tgtatgcatg tatgtatgtg tgtgtgtgtg
27241 tgtgtgcata tatatatata tatacacaca cacatatata tgtaatthtt tttttgagac
27301 agggthttct tcttgcctcag gctggagtgc agtagtgcaa tcacagctca ctgcagcctt
27361 gacctoccaa gctcagggtga tctctccatc tttagcctctt gggtagctgg gactagagge
27421 atgtgocacc acccccagct aathththtt ttaagagatg gggthttcact atgttggeca
27481 ggttgggtctt gaactcctgg gctcaagcaa tctgcaatct gcctgtctcg gactcccaaa
27541 gtattgggat tatgggtgtg agccaccatg ctacagcctt gctthtatatt thtatttcca
27601 tagctoctac cgtgttctaa caaactataa cthtgtcttt ggtthattga ctgtctcctt
27661 ccactagaat gtcagctcca taagggtagg gaththtgtc tgtthtattc acttctgcat
27721 cctgggtgtc tagaatagtt cctggtatat agtagatgth caataaataa ttgttgatg
27781 aatgaattct ctgtgtataa ggcagthctg taaagcccct aaatcttctt tctccaagaa
27841 gcoctcttcc aaatgtatgt aacctatttt agthattgca catccccatt ttacagatga
27901 tgaactaac gctggaaaaa cthtgcaga gtcactgacc tgaaaagtgg cagggecaga
27961 tgcatttgth tatccaagaa accctgagag ggctgggcat ggtggctggc accctgthc
28021 ccagcactth gggaggctga ggtgggagga tcactgact ccaggagthc aagaccaact
28081 tggcaacata gtgagacccc catctcttaa aaaaataaag aggtgggtg cgtggctca
28141 tgctgtaat ccagcactt tgagaggcca aggcaggcag atcatctgag gtcaggagth
28201 cgagaccagc ctggccaaca tggggaaacc ctgtgtctac taaaaatata aaaattagcc
28261 ggtgtgtgtg gtgggtgctt ataattcccag caaccctgga ggctgaggca ggcaaatcgc
28321 ttgaacatgg aaggcagagg ttgcagtgag ccgagattht gccactgcac tccagcctgg
28381 gcaacagagc gagactccaa ctcaaaaaaa attagaaaca ccgagtgcct gctgtgtgccc

28441 aggtgtagg gagacagatg accaagactt ggtcctttct gtccctcaca gcactttcag
28501 gaatgggggt cacagaggaa aactaggat gctataagaa cacagaggag cccccacc
28561 caatcaagggt cattggtcct caagataaag tgcctaggcc tgggtcagtg gctcatgcct
28621 gtaatctcac actttgggag gctaaggatg gggggcactt gaggtcagga gttcgagacc
28681 agcctggcca acatggtgaa acctcatctc tactcaaaaat acaaaaaatta gccggcgctg
28741 gtggcacacg cctgtaatcc cagctactca ggaggctaag gcaggagaag cgcttgaacc
28801 tgggaggcag aggttgagat gagctgagat tgtactgttg cacttcagcc tgggtgacaa
28861 gagcaaaaact ctgtctcaaa aaaaaaaaaa aaaaaattctc aaaccgtccg ttggacccaa
28921 ccactgcatg ctgatttcac tcatgcatgc tccagctccc agacttctgg gatccctttg
28981 gaattcatcc aattcagagt cctctactgc tctttaatca tttgtgcttg ttggttttgc
29041 ttctccctct gtctttccca ctggtgggag gctcccttgg gagcaagcat agtattgcta
29101 tcaccctgtg gcatagtgtg cagaactcct gcacaccag gagctcagtc cccgtctttg
29161 atagccactc taccatact acaggtaga aaactgaggt ccagagaatt cagtagcaga
29221 tctagcatca ctcagcagtt aagaggtagg attagaatgt ggtcaccctc tagccaccac
29281 tcccatcacc tctcacctcc agcaggctca gctcctctgg acctggctgc tgcctctggg
29341 ttggtgattt ttgccctggc aaaggagag cagacttatg acatctggag tccagagtta
29401 ggattgaata ttcaggataa agtttactt tgaagttcag atttaggatt gaggccattt
29461 tgatggtggt cgcagggaac ctttagaaag agcttggagt gaaatttggg gtgaaagccc
29521 tccttctttc tctgcacct tttccacaag atggggaggt ctggccagac acagtggctc
29581 acacctataa tccagcgtt ttggaaggct gaagtgagga tcatttgggg ccaggagttt
29641 tagaccaa at tgggcaacat agcaaggctc cgtctctatc atgatagcct atagtctcag
29701 ctactcagga ggctgagata aaagaatcac ttgaaaccag gaattcaagg ttgcagtgag
29761 ctatgatcct gccaccgcac tccagcctgg gcgacagacc aagattctca aaaaaagagc
29821 ggggaggctt aggacaggaa gggttaagtg tggtttagct ttcccagcct ggaaagggaag
29881 ccaggccaag cagctgggta ggggaaaagg gggcactgag tgccaggggag gggctgaggc
29941 agtgggaaca gcatgaaccc tccctcagc tttcactggc ttaccctcc tgcctctgct
30001 tctggtgtgg agaagaaca gactttttat tgtctgggaa gctgaggggt ggtgccacgc
30061 cctgggggag agaagccata ttctgggccc cctcatccgg gtgcatccct agtgetcacc
30121 atgctgcca ggcaccctca ctgagatgag aactttcctt gttcccctcc cctaataacc
30181 aaagtctctg ccttaatccc actccagctc ctcattccacc ccgaggccag gcttcagaac
30241 ccagaactga gggcctgtcc agccctgctt tcttgtgtt tgagggaagc cctgatatct
30301 tggagctgta caaggtagca acataatttg atttccctgg ttgctcatcg gaggataagc
30361 tgtgggaggg agttgggagc ggcctgcagg cagatgggac ccaggcctca atgctgtcac
30421 ctcttagaga ggataagggtg ggacgggcat ggtggctcac atctgtaate ccagaacttt
30481 ggggaagctga ggcaggcaga ttacctgagg tccgggagttt gagaccagtt tggccaacat
30541 ggtgaaaccc catctctact aaaaataaaa ataaaaaaa aattagctgg gcttgggtggc
30601 aggtgcctat aatccagct actcaggagg ctgaggcagg agaattctct gaacccggga
30661 ggtggagatt gcaatgagct gagatcatgc cactgcactc cagcctgggt gagagagcaa
30721 gacttaatct caaaaaaaca acaaaacaaa caaaaaaac aaacaaacag aggatgaggt
30781 ggatggtggg accaggatag gggccaggat gaggggaggg gagtgggaag cttatcaaaa
30841 gggtocttgg tgaggcctga ggatggaggc tccagagcct gaggtagcga accctgggga
30901 cctgagtgat ctogttttgt gagagagcct ggcccctccc cagagccagc tgcagaatgg
30961 acctggccag agaggaaagt agagatgagt atggtgctgg cctttgcagc cagcaaggct
31021 gtggggtcct tgctgttctt gtccccact accttcttgc cccccactac cttctttctg
31081 gctgaaacca ggatagagac ccaatattgg ctgtccagcc ccagccctg cttccccttt
31141 caggccctc tgggaaccac agaaatctgg gacctagtgt cttggcaacg taatgaatgc
31201 atgcacagct ctggtatctg ttttaaatta tccattaaa taagtacagt tctgggggaa
31261 aaaaataagt tgaactgggtg cggtaaagcag tagaaagggg ctgaaggggg aaggaacca
31321 ggcagtgctt gggtoaaaag gaaggggaca ggagatggaa ggggcagtc ctggtccta
31381 ttcttggctt tctttagggg acttcttttag gggactgtgg cttgttctt gggctaaaa
31441 acgaatgctt ggctttgaag agagatagat tggggcaaaa gaaagaaaa aagggacccc
31501 ccaaactcct tgatccctgg ccccaaactg ggggcataaa ggaactcag ttccagaact
31561 ttgctcccc cagggaaacc aggcattcct tctccacccc actcctggca cactgagatg
31621 cagctctgaa tgggctgccc acgtgtggag gggggttggg gtgactcact attactactg
31681 ggaggacag gggagccagt ggtggaagaa ggtgagtc cactgatggg caccagcctc
31741 agccctccc ccactttcct ggctccagc cctgcctacc tgaccctctc ccttgccttg
31801 cgccacttc cctctctttc tccccagccc ttttgcccac ccactctccc tcttggctc

31861 tgccctctag ccagaaaggt ctgaggcaat gggggcaagc ttggagccga cagtgtctgag
31921 caggcaggag ccaagagagg ggaagcttga gcctcacgca gttaggggtg cgctggagag
31981 ggtggggccc gactccgcca caccccaacg gtccctcccc ctccctacca ctcccgcccc
32041 ccccccaat ggatctggga ctgccctttt aagagtagtg gccctcctc ccttcagagg
32101 aggacctatt agagcctttg ccccgcgctc ggtgactcag tgttcgctgg agcgccgcac
32161 ctacaccagc caaccagat cccgaggtcc gacagcgccc ggcccagatc cccacgcctg
32221 ccaggagcaa gccgagagcc agccggcccg cgcactccga ctccgagcag tctctgtcct
32281 togaccogag ccccgcgccc tttccgggac ccctgccccg cgggcagcgc tgccaacctg
32341 ccggccATGg agaccccgtc ccagcggcgc gccacccgca gcgggggcga ggccagctcc
32401 actocgtgt cgcccacccg catcaccccg ctgcaggaga aggaggacct gcaggagctc
32461 aatgatogct tggcgtctc catcgaccgt gtgcgctcgc tggaaacgga gaacgcaggg
32521 ctgogccttc gcatcaccca gtctgaagag gtggtcagcc gcgaggtgtc cggcatcaag
32581 gocgocacg aggcagagct cggggatgcc cgcaagacc ttgactcagt agccaaggag
32641 ogogcccgcc tgcagctgga gctgagcaaa gtgcgtgagg agtttaagga gctgaaagcg
32701 cgtgagttc gccaggtgg ctgcgtgctt ggcggggagt ggagagggcg gcgggcccgc
32761 gcccctggcc ggccgcagga agggagttag agggcctgga ggccgataac tttgccatag
32821 tctcctccct ccccggaact gccccagcg ggtgactggc agtgtcaagg ggaattgtca
32881 agacaggaca gagagggag tgggtgtctc tgggagaggg tcggggagga tataaggaat
32941 ggtgggggta tcagggacaa gttggggctg gggccggcct gaattcggtc agattgggat
33001 ttgccaacta tttggagccg gggggagggg cttgagcaaa acagaactag ccctgccagc
33061 tcgaagaact ctgggcaccc aggcacacatc ggagtggcag aaaggtcctt gttagaactt
33121 tgtttagcgg cttggcactg tgctagcttt gcccaagctg gctctgaaca catgatgccc
33181 actaagacat aactctcaag ttggcatctg tccagcgtgt tggagcgagg tcaggaaggc
33241 agggcaatoc ccttttccc tcccagggc ttggcgggtg cccccctca gcatgacctt
33301 gtocctgggtt ctaagggttg ggaagttctc cctcactctg cactctcgc tgtctgggac
33361 cttocttggg ctctgacagg cccacaaaaa gagctccggg agatgagaga tcggctcccc
33421 cgcagctccc acagcccttg gcctgcttg cccaggaatg caagggaggg agggaggcag
33481 agggcagagg ctcccagctc aggaagttgt gttatgcca ggtctggccg cactcctccc
33541 ttggccctct gcctagtgtc ttogagggtt gggggcactg tccttccctc cttgggggtga
33601 gccactttoa ttttcccagc ggggcccaggc agtctttgct cgggcccac tccttagctg
33661 ctgacgtttt gatctttgtc ttattgaagt gctggaatac agtgacattt ttgaaatcca
33721 gocgttggaa gattcaggcc actcccactt taccacccc tgccccaccc taccacccc
33781 tactcaactg caccttcttc ttttctaaaa aagcctttgg gagcttggaa gtataggccc
33841 tctcttccag ccccatcaaa atttgtttcc cttcttctg ccttccctt ctctatgcag
33901 acccaggcca agagcactaa ggggtgcttg agatccgtaa agggctgttg gctttgactt
33961 cttctctctc ttttatcact tactccaaac ttctgctctt cctagaaccc tttgctaggt
34021 gtggttttgt tgcccaggct ggagtgcaat ggcacaatct cggctcactg caacctccgc
34081 ctcccagggt caagtgattc tctgcctca gcctcccga tagctgagat tacaggcatg
34141 tgccaccatg ccgggctaat tttgtatttc tagtagagat ggggtttctc catgttctgc
34201 aggctagtct tgaactccca acctcaggtg atccaccgc ctcagcctcc caaagtgcta
34261 ggattacagg catgagccac cacgctgggc ccatcaccct tctttctgaa gagtcaatgg
34321 aagttgtgtg taggaagaca ggcttaacgg tttttttttg agacagggtc ttactctgtc
34381 acccagactg gagtgaagtg gtgcgactct ggctcaccac aacctctgcc tcccagctc
34441 aaaagattct cctgcctcag cctcctgagt agctgggatt ataagtgtgt gccaccacac
34501 atggctatct tttttttttt tttttttttt aatttttagt agagatgggg tttcaccatg
34561 ttggctaggg tggctctcaa ctctgactt caaatgatcc acctgcctcg gcctcccaaa
34621 gtgctgggat tacagggtgt agctaccatg cccggccatc aacctttatt ttgtttttt
34681 gagacggagt cttgctttgt tgcccaggct ggagtacagt agtgtgacct caggctcactg
34741 caacctctgc ctcccagggt caagccatgc tctgcctca gcctcccaag tagctgggac
34801 tatagggtgc tgccaccaac ccggctact ttttatattt ttagtagaga cggggtttca
34861 ccattgtggc cagggtgatc togaactcct gacctcaagt gatctgctg cctcagctc
34921 ccaaagtgtt gggattagag acgggagcca ctgcgctgg cttctttttt tcttgagata
34981 gggtttccact ctgttaccca ggctggagtg cagtggcaag gtcattggctc actgcagctc
35041 ctacctctct ggctcaagcc atcctcccgc ctcagcctcc tgagtactg ggaccacagg
35101 caggcaccac caccacagc taatgttttt gtattatttt gtagagatgg ggttttgcca
35161 tgttgcccac agtcttgaac tctgggttcc attctgctga aagagaccac acctgtcctt
35221 ttctttatct ttattatatt tttcagagac agggccttgc cctgttctc aggctagagt

35281 gcaatggtac aatcataact tgctgcagcc tggaactcct cctggggetca agcgatccta
35341 ccgtctcacc ttccggaata gctgagacta agggcaggca ccaccacgct tggctaattt
35401 tttttttttt tttttttttt ttgctttttg tttgtaaaga tggaaaacttg ctatgtttgct
35461 cagctgggtc cgaagttttg gcctcaagca atcctcctgc ctccggcctcc ggaagcactg
35521 ggattacagg cataagccac caggcctgac gccaggcctg tcttttttct actagtgata
35581 tgaacaattt agttagcaag acagatagga agcaaggaag gggagacca gagaattcgt
35641 tgcattctaa actagtccac tcatctacca aagccctgtg aaggacattt ttagcagttt
35701 tagcagtttt ctggtcaaaa ctttgatcga gaaacagatt gagtggattc gatattctct
35761 tgctcaccga gccacgccag tttgtctcct ctgcctccta gtgcagctgt ccaggcctgg
35821 gacaccaggc gggatgtgct gcatgtgggg cagggcggag gtggtgtgtg tacttgttat
35881 attagccac ctccctctgt tctccccac tgatcctggc tggaaaggct gggcttccgg
35941 aaaagagagg tggatttgca cacctggatc ccaagctgat agaaagtggg gtgaagacaa
36001 aggggactca gactggggtg tctgtcctct tctatgccca cagtaggagg agccaggatt
36061 ggttactccc tgctgggtct gctgtgctca gagtgaggta gagaagtggg tagagtaaag
36121 aatttgggag aggaaaaaag gcattttccc aaccctccc accaaaagcct agagagaagg
36181 tgttgtctgg tttaatgttt aattagagct cagagttcag ggccagattt ggagttggga
36241 tggaaagtgg ttttaagac cctgtagcaa tttttgacc agcctgggta cctcaaccac
36301 actcaggagt ttgggggacc ttctgttggg ctggattata ggctccaaga agaaaccctt
36361 ttccccaata ctctctctct cttctttttt tgagacaggg tcttgttctg ttgccaggc
36421 tggggtgagc tggcatgac acagctcact gcaacgtcag cctcacaggc tctggtgatt
36481 ctccacctc agcctcctga gtagctggga ttacaagtgt gtgccaccat gccagctaa
36541 tttttttttt cttttttttt ttttgagacg gagtcttgtt ctggtgccag gctggagtgc
36601 agtgggtgca tctcggtcga ttgcaacctc cacctcccag gttcaagcga ttctcctgcc
36661 tcagcctccc gagtagctgg gactacaggc acatgccatc acgcccagtt aatttttgta
36721 ttttttagtag agttgggggt tcaccatgtt ggcaggatg gtcttgatct cttgacctcg
36781 tgatcogtcc acctggcct cccaaagtgc tgggattaca ggtgtgagcc accgtaccgg
36841 gccactaatt tttatatatt ttgtagagat ggggtttcac cgtgttgccc aagctggtct
36901 cgaactccta ggctcaagta atccacctgc cttggccttg gcctcccaaa gtgctgggat
36961 gtataggeat gagctaccgc acctggtacc cctgcccct tctctgtctc tttctagtct
37021 gtagcccaag ggatttggat acccaagtgc aggcagaatg ggaaggttgt aagcaccagg
37081 gaagcctgtc tggagtccag gcttgcagct gggccccacc ccaggcaagg cagctgggtg
37141 gatgactcag atgctgcccc cctccctccc accctggtgg ctttacagaa gacagcagga
37201 gacaggggtg agacagcagt tgtcttaaag ggaggagtgg tggctctgaat gtctacctct
37261 tctgcccccc tccccattgc atcctggagt cccttgctctg gctccttctt gagacctct
37321 ggtggtgtct ggacacatag ctctctctgg acaggtaaca tgcacaagta attagaatcc
37381 agagttgagt tcagagttat ggattgggct gcaggatagt gccagggtct gtgccttccc
37441 atgtgaaact gatggaggaa ggctgagtca gaagtgggga gatccgaggc ccacaaagca
37501 gaagcgttac ttccactcca aaaaggccct ggtgcttgac aacttctctg attgccact
37561 gttgcagccc cagtgtggac aggcaggggag atgcaggctc cagttcatgt aggctctgat
37621 caagacaaga acagcaaagg ccacagaggc acagatgctt gtcccatgct acacaataaa
37681 ggggtcagca cttgatcaca ggccattatga cttccagctg ggtgtgctct taccattaag
37741 cctcacttct ctagcttggg ggacaggttg gagggaggat ctagaggggt aggtaaaggtg
37801 aagtcaggta gctgaggctc acttctgcag cctggaaact ctgctctggg gccagtgaca
37861 ccttagtgct ctatggccat acttctgtggc tcatgcctgt aatcccagtg ctttgggagg
37921 ctaaggcagg aggatcactt gaggccagga gtttgagacc agtctgggca acatagcaag
37981 accccttct gtacaaaaaa attagccggt caacacctgt agtccagctg cttgggaagc
38041 tgaggcggga ggatcacctg aagccaggag tttgaggcta ttgtgagcta tgactgcact
38101 actgcactct agcctgggag agagaaagac cctgtctctg aaaaagaaaa aaacaaaaa
38161 aaactctgct gtctgcagg gcctgttagc atatgatcga tagcctttgc tccagcctat
38221 acctggacc aggaccctg ccagcccctc aatcgtgaga cggtcagagc tctgggagc
38281 tgggtgattct tgtcttgaga ctatcttgag acttgtcatg ggaattgtcc acccggattg
38341 aaaggaagct gtgccttttg gcagaccat taggttaatg gggttggaga ctttgagga
38401 tgcattgggc ctgggcttta tctgagggtg tctcctgggt ttacctctcc aaccctccac
38461 caccaaatoc attctttttt tttttttttt tttttttttg acagtctcgc tccctggccc
38521 aggctggagt gcagtggcat gatcttggct tactgcaatc tccacctccc aggctcaagt
38581 gatctocca cctcagcctc ccaagaagct gggactatag gcacgtgcca catgctcggc
38641 taatttttct attttttagta gagaccagggt ttcaccatgt tactcaggct ggtcttgaac

38701 tctggggcct aagcagtcca cccaccttga cctcccaaag tgctgagagc cactgagcct
38761 agcccaaatc cacgttctga ttcaaaggga aagaagaagg gtgcagctaa acctgggggg
38821 tgagaagtac ttaaaaagcc caagagaaac aaaagagaga ataattcctc actaggacct
38881 cctattgcct tccactatt ggtgcccttg cttggcactt cccctggcct ccaggagtct
38941 gagacttact cttccatgga tgtgccatt gcccccactt ccagggtccac cccccagtga
39001 ttcggtagct tagtgtctgc gctgaagccc aggacagctg gatggacaac tggtagatcc
39061 cttcacctac caactgtgct ttctgtccc cttccccttg cttccctcct cccagcccc
39121 tcgccacccc tagcagctgc agcagccaag accaagtctt cagagacca gacacaaggg
39181 cagggttcat tccattctca cctccttggg gtcccagtg actgataggc cgaactctaa
39241 tattatagga gatctctgga agattgcagg gtctcttata cctcaataag gggcaaggca
39301 agccgggccc agtggctcac gcctgtaatc ccagcacttt gagaagccga ggggaacaga
39361 tcaactcagg tcaggagtta agagaccagc ctggccaaca tggtgaaacc ctgtctctac
39421 taaaaataca aaaattaacc agaaatcgc tgaaccagg aggcagatgt tgcagtgagc
39481 cgagatcacg ccaactgcact ccagccaggg cgacagagca agattccgtc tcaaaaaaat
39541 aatactaata ataaataaat aaataagggg caaggtagtc caccaacaaa atgacaggca
39601 gtgtgatata gtggacaccc tagccctcgg tgcccttagt tctgtgtgtg gccctttcac
39661 taaattgctg tgtgacctg agcaaatcgc cttccccttc tggtcttctc tagctgtaaa
39721 agaaagggat tggagcggaa agtctccaga gaccttttag gttccaaagt agtacagtga
39781 cccacaaagt gagaaaacag tcttctaaaa taccaagtta ttaatagtaa aatcaaatat
39841 aaataatgtg aatatagtta atagctaata ttgttctcaa tagaaatgtt tcccacaagc
39901 tgtggaatta aacatactac cacatttctc tatttccccg tgaaagtttg ttgaaatgg
39961 ttaaattgtg acattaccct cttggcaaat gttttgtttt cattgtctact aggaaagggc
40021 aactcgtttt cgatgcctct ccttctgga cggtggaag ggctgtgtca tagagttaga
40081 acgggagatg cggcacagga atggctccca ttgaccggg ttgggggcta gggcgaaggc
40141 ctaggagagg cagaactgtt accttagagc tggccaggat tagagaacag tgcctggaac
40201 cggggggagg ggcacggtga ccttgggctg cccaccttct acccttccag caccatact
40261 ggctcccca acctgcggct gggctgggag gaggtcttgg ccctaccaa tcccttaagg
40321 aaggggaaag agtttgggaa ggggagtcct cccttcacc ctgcctcccc caagttgtga
40381 gagaggaagc cggaaatcctg cctgctgaag ccaggaataa ttctggctga gatcccaggc
40441 ccggcagggg cgctgagtca tggtagaggg cagagtggag agtggacagg agaccctaag
40501 cttgtccagt cagaaaagca gaggctgagg ggtggccttt tcttgagaac tacattcaag
40561 ttgcagcaag aaggacagtg gtctgaattt gacggggaca aatggaaggg agataggaca
40621 catgagttcc tttaggtctg gctcagggga gctagacttc atttcaaggg gtctaggttc
40681 tgggcagttg agaaggaggc ttttggggt caccaaggct cccctttctt cccaaagctc
40741 taacactgcc accttctgct ggctaggaga gagctgtgtc ttctgaggct agagctggaa
40801 tgcagtgaga ccagactgcc taggtcctcc ctcacttctt ctctgacct tggggtgtgg
40861 ctcccactct ctcccagtg cctcaggggt aataactatg tgccaccaga tagagagtta
40921 aggggctgct gaattggctt cttgtgaagg gaatccccta aatgtccctc gttttggtca
40981 ctggcctccc tcccgcctcc ttcaggacat tctactatct tcttaggcca tccctccctc
41041 ctocaggcac tacttctttt gctctatccc caagcccac ccctgcattt ttgtgacaac
41101 accggaatga tttctagaga gagaggccag gaagaaggaa agtggcactt ggcaggagac
41161 cttgcagggg gcggctggtg aggaagccag ccgcccattg tccaggacct cagtgcctg
41221 goctcgggoc tcaggcttct cctgcctctg tacaatgcca cgttgatagc cccagcagct
41281 gtgactcagg cctggccccc tgccaggccc agcacttcta ctggagtgtc gtctgaacat
41341 gtcaacaggc ttctatccc tctctcagca ccagttctcc ccacttcage cctccctct
41401 goctggaatt aaaacctggc tttgtcttag ggaaggacag ctgggagcct agtggctctg
41461 gtaggggatc tgagaggcct cagaccctag gcataatttg ctgtttggca ggtgtcacgc
41521 ccaagggaaag cgtgtggaag cagagccatg cctgctgtgg gtgcacatgc ccgctgagg
41581 gagtgggggt gtttcatcct ggggcacctg tgggcttttg aggtgtatga tattcagaac
41641 ttacagggtt ggggtttggg gaaggctcaa ggggcttcta agtccctgga acagctgcc
41701 cctcagttc ctctctctct ctctctttt ttttgagatg gagtctcgtc ctgttgcca
41761 ggctagaatg cagtggcgcg atcttggctc actgcaact ccgcctcctg gttcaagtg
41821 attctcctgc ctacgctcc caagtagctg ggactatagg tgcccgccac catgctggc
41881 taatttttgt atttttagta gaaatgggg ttcaccatgc tggccaggat ggtctcaaac
41941 toctgacctc gtgatccacc caccttggcc tcccaaagtg ctgggattac aggcgtgagc
42001 cactgcgccc agcctcagtt cctctcttta aggtctcctt tccagagagg gatagcact
42061 caaatgccag ggaggggaat tctccacatc ctgcccctac ccgagttgtg gcagaccac

42121 agactagcca agaaaccaag cagtggttac tttgccgggt tgggggggag gtaggggcta
42181 tcaaacctca tgattggccg cacacaaagg tgtgagtatg tgtataattg aggggtgggtg
42241 ggagtggcac tttcactagg cctccgtatc actctctgac tggggtatct cccagcaagc
42301 gagacagagg cagacacgct tcccagactg tcttactggg tctctctgtg ttattctctg
42361 cagtgtctgt gtgtatcgtg ccattttcta tgttttgcac caatctgctg tgagtgtcct
42421 caggtgacct gggggcagg ttttagtgcc tgagcctacc cgtctccagg ctttagtttc
42481 cccctgtaaa agtataggag ttggttcaag agaaggttcc tctagaagcc ttgagcctgt
42541 gaaccgtcta gtctccgggt atttgtggga cacacagaaa aagccccacg acccaacagg
42601 tagaacactg gctgaaatca gcagggcaga gctgagacag gctcaagtag gctgaggggt
42661 agggagggtt tgggtgaatg ggagggagg acagagagaa ggaggatata ttgcagtagg
42721 aggagtgtct ggaacaaaag gaggggtgg aggagtggct tgggggtggca gcagaagacg
42781 ccctgtcaca tggcgggaag tcagcctggg cagaggtcta ggtgtccagg aggggctggg
42841 tgtgggtggc cacgcctgta atcccaggac tttgggaggc tgatgcagga ggatcacgtg
42901 aggtcaggag ttcaagacca gcctggccaa catggcgaaa ccctatctct actaaaaatg
42961 ccaaaaatta gctgggtgtg gtggcaggcg cctgtaatcc cagctactct ggaggctgag
43021 gcacaagaat tgcttgaacc tgggaggtgg aggttgacag gagccgagat cgcgccactc
43081 tactctagcc tgggcaacac agtgagactc tgtctcaaaa ataataataa taggggctgg
43141 gcgcggtggc tcatgactgt aatcccagca ttttgggagg tggaggcggg tggatcacct
43201 gaggtcagga gtccgagacc agcctggcca acatggcaaa actccgtctc tactaaaaat
43261 agaaaaatta gctaggcatg gtggtgcagg cctgtaatcc agctactcgg gaggtgaga
43321 agcaggagaa tcaactgaac ctgggaggtg aaggttgacag tgagatcacc tgggcgacag
43381 aatgagactc cacctcaaaa taataataat agtaataata ataaatgaaa aattttaaaa
43441 ttaaacaatt aaaaatttta aattaaaatt aaacaaatta gatgccagg aggatacagg
43501 agagcatttg ccaccaggcg gactccctgt acccaccggg ccacaggggg cgatgttctc
43561 gggagacagg aaatgccag gggctgggag accctctgct cttctgctcc cttctgtgt
43621 gctgcctggc aatggggaac tctgagggtt ggtgagcagg gctgctgagg agtgggtcta
43681 aggagtccct gcagggctgg gccagctcct ccacctccc tttgtcttcc cctcccactt
43741 gttattttta gctacagtgt ctgtccctct tgccttctcc ccagattggg agaggaaacg
43801 gaggcctctc cctccgggcc tagcctgttg ccccagcaa cggggcccaa acaggcctgt
43861 ggcgggcctt ggcttccata tctggcatca gagttgggtt gagcaggggt actcagaggg
43921 tgggtcagcg cctggcccgg tgcccaccta gccctttgc tgtgctgggt cettctctcc
43981 ccaaacagcc ccaagggccc gggcctgctg cagctgggga gccggacttc cttgtcccac
44041 caggcacagc tcttcagacc cctgccttgg gtcacatttg caagtgccaa ctctcatttc
44101 taccttattc ttttctctc tgttcccctc cccaccccct ctcttcccctc tttctgagat
44161 cagatttgcc agtgatggga agagttagaa acaggatgcc cagcccttct cgcctcaaga
44221 ggccactggg atgcagccac tcctgtgctt ggggaacctg gaggatgcaa gggaaaggac
44281 tggcactctg ctggcacagc acccggcctg gggcaggaca cgggcgaagc cagggctctc
44341 cctgtgagca ctagaggatt tcccagcccc tgcccgggta ttgtgtgctt gagcatgagt
44401 caoctgaggg gcccaggttc ccaccttcc cagctcctct ggctgcccc accctgtcct
44461 cctgccaac ccagcacggg gacggcactc agcgtgtgct cagctttcct gatgccaac
44521 ccagtgagg tgggtgcac caccacctg ggaccgaatg cctggctagg gtctactttg
44581 gtccctgcta ggtctgagga cccctcctag gaaggaaatg gcacttgggg gcgggggag
44641 ggaggaggga ggagagacac tgggtcttac tgtacccta gtcactctct ggggtgtgag
44701 tgtggctccc tggccacaga gctcccagg tctgagtcag gagcccatgg gtgatatgg
44761 cttcttcccc gcagatggga gctcccctg cctaagaaaa ccacaaaggt tcttctcac
44821 ttccctctct gctcgtgggt tttctcatct gcagggtgtg tcttagtctt ttaatctct
44881 ctctttgcag tgctagtcaa aacctccacc agggaaagac aaataacccc cttactgttt
44941 tttttttttt tttttttttt ttttgagatg gagtctcgtc ctgtcaccca tgctgtagtg
45001 cagtggcaca atctcggctc actgcaacct ccgctccca agttcaagtg atcctctac
45061 ctcagcctcc tcagtagctg ggactacagg tgcacaccac cgtaccacg taaatttttt
45121 tttttttttt ttgagataga gtctcactct gtcaccagc ctggagtaca gtggtacaat
45181 ctcaactcac tacaatctcc gctcccagg ctcaagcaat tctcgtgtct cagcctccca
45241 agttgctggg actatggacg tgcaccacct tgcccacta atttttgtat ttttgataga
45301 gtcagagttt caccatgttg gcaggctggc ctcgaactcc tggcctcaag tgatccacct
45361 gocttggcct cccaaagtgc tgggattaca ggtgtgagcc accacactca gccagcccc
45421 ttactttcct tggagaccat aactgtggc ttgtgcaaaa gtggtacagc atggatttcc
45481 agctccccta tctacttgc tggggaccct agatatagct ttctgtgctt atttctcaa

45541 ttgcatagga atagcaccta tcgcataggg tagctgtgaa gatgacgtga gttaacataa
45601 tatttagagc agtgcttggg acctaataag ctctatataa gtgtttgcta ttatattatt
45661 attatcactg ccaccaccgc ttttgcaagc agcagaaggt gaagagggtta gactgaagaa
45721 aaaacttctg tgctcatcag cccataagct cgcagagcac agggatcatg catctatggt
45781 ttctcagtc agtgtctgcc aggcactggc aaggaaaggc tgttaccagg gggaaactcca
45841 ggaattcctc ctggcaccta agggaggctgg ggagacagga ctagggaaaa ggtgcccttg
45901 agacaccttc tgaatcadc ccattgcctt ccagcttctt tcagctcagg ctggctgggtc
45961 agggaaacgc tttgtgcat agtgtctgcc ctcttctctc tcttggcttc tccattctct
46021 ctggaacttg tggcttagga aagcagtggg gtggaggagg aggaacccta gatcagcagc
46081 tagaattgac tggaatgctg ctgctggctt tggtaattg aactggggcc attcaccttc
46141 ctcttttgca cctcagtttc ctcatctata aaaggagag ggttgagctg aatcaactct
46201 aagctccttc tagttctcta aattctgaga gcctcctagt acagccagca gcagccatta
46261 gccttcaggg tagagaggcc tcttctggga agccccagcc agcctggggg tcagcccaag
46321 gagctcggaa tctaagttgc cccagttgct tcactttacc agcggttttt cttcattttc
46381 cctcctcccc ctgcagctgc ttcagcttct gaaaagttct gaagtcattg aaagttgggg
46441 ctgtgctccc agccaggggc taggcccgat ggcagccaaa acctgagctg ggttttgact
46501 ttatttttag cttttctgac tgagacagag gaggaatac attctccggg tctggaaggg
46561 gctctttttt gcaggagaca gacacttaca ttaaacaact tgttctgagg tgtggccaga
46621 ggctgggact gagcaagtgt gcaggctggg ggagcttctt ctggcttctc atgtccttcc
46681 cctgcccctc tgagtgtcac tctatcctcc tccctgcctg gtgggggggag gtgggggtga
46741 ctcttttttt ggactctcct aagcagaaca ctgcctgggt ctctcctcc agagcttctg
46801 caaatctagc cttccctatc cctcttcaca gtgaattgct gggcctcttg gagtttagga
46861 cttttgtggg agaagaaaaa tgttggcagg gctgcttttc tctttccag gatagattht
46921 tcttctgcc cacgcttggg tttccttttt tccatctgct gtggtgggct catgcttaag
46981 cactgatgag ttacagatgg cagctggaac caggtcctct ggatcttctc ctccgctccc
47041 tgggtctgct gctttctctc accctatatt tgtgaagcaa ttgtaacatc tagaaagttc
47101 ttgggttctc tggaggtttt taagaaaata ggacctttct atttctccag tccactagca
47161 aaaataatca ggggcccaga aaaggtgagg gaggtggcag aggcagcgtc gttcagctgg
47221 ttatagctaa agctttacc actttgagga gcaggaggc ttaaagctgg ggcccagatg
47281 gaactggagg cctgggatcc acatctggaa ccagatgctg aggctatggt agatgggtag
47341 ggctcagcct tctcccaggg cacggatgag gcaggaggga gggaggcagg gaccctctg
47401 ttcagtgcag atcagggcac ccagactggg tctgagaaa ggaaagggtc aatattgtgc
47461 ctggctcatc ttgtctgagg tccctctgag ctctaaccag actttccttc cccacagtcc
47521 cacatgtgta aaagggacta ggagagggtg ccagtacctt tggggctcag atcgagaagt
47581 gctagggaca tgtgggcat gagcttagtt gtcaggctcc tcagaggggag ggaagcttgg
47641 ccaaagggaa gtgagtagag tccagggaga aggctaagta aggccctgtg tgggaagggg
47701 caggagacaa aggtaccctt gtctctttgg gaaagaatgg gaggagagag agggaaaagc
47761 attcatalca cggggtagag ctctgccctt ggcccaggc acgttctga gccctgagtc
47821 atgggaaggg tggagaagca ggaagggggg tttcaaggac cttggggagg tgggagcca
47881 gcccagagg caagcagatg caaaccaacc taatgcaagg atgccctctc ctggtaattg
47941 caggcatagc agcggcagcc cccatggctg acctcctggg agcctggcac tgtctaggca
48001 cacagactcc ttctcttaaa tctaactctc cctctcttct ttgcaatac caagaaggag
48061 ggtgaactga tagctgctca ggctcggctg aaggacctgg aggtctctgt gaactccaag
48121 gaggcgcac tgagcaactgc tctcagtggg aagcgcacgc tggaggggca gctgcatgat
48181 ctgogggggcc aggtggccaa ggtgaggcca cctgcaggg cccaccatg gccccaccta
48241 acacatgtac actcactctt ctacctaggc cctccccat gtggtgctg tctgacctg
48301 tcacctgatt tcagagccat tcacctgtcc tagagtcatt ttaccactg aggtcacatc
48361 ttatctaat ttggctgcca atgggatcta ccacagtga tttaaaataa tccaggaggc
48421 cgggcatggg ggttcacgcc tghtaatcca gcactttagg aggccgagg gggccgatca
48481 cgaggctcagg agatcgagat catcctgact aacatggtga aacccctct ctactaaaa
48541 tacaaaaaat tagcctggca tgggtggcgg cgctgtagt cccaactact cgggaggctg
48601 aggcaggaga atggcgtgag cctgcgaggc agagcttgca gtgagctgag atcatgccac
48661 tgcactocag cctgggcaac agagtgagac tccgtctcaa aaaaataata ataataataa
48721 taaaaataat ccaggccatg tgtgggtggct catgcctgta atcccagcat tttgggaggc
48781 caaggaggca ggattgcttg agtccaggag tttgagacca gcctgggcaa cacagacccc
48841 atctctagaa aataaaaatt taaagaaatt agctgggcat ggtggtgtgc acctatagtc
48901 ccagctactt gggaggctga ggcaggagga tggcttgaac ctgagaggtc gaggatacag

48961 tgagctgtga ttgcaccact gcacttcagc ctgggtgaca gagggaaacc ctgtctctac
49021 ataaataaat acataaaata aaataatcca caagccattt ctacttaact ttgcaatgaa
49081 ctgtacctga ccctagatcc ctcccagttt ggccctccgg tatacaaggg cctcctatag
49141 gcccttgtga tttctctggg gaaaaggagg actggagttg atcatttatt gaggccatca
49201 gaagcggatg gctaattaca tatgggacat gtgttaataa tgctttgtgt atatagagt
49261 gcctttactt tcaaaacact cttctccaat ttatcatgtt aaaagctagg aattgggctg
49321 ggtgcagtgg ctcacgccta taatcccagc actttgggag gccaaaggcg gtggatcatt
49381 tgaggtcagg agtttgagac cagtctgacc aacatggtta aactccgtct ctactaaaa
49441 tacaaaatta gccaggcgtg gtggcacaca cctgtagtcc caacaactac ttgtgaggct
49501 gaggcaggaa aatcatttga acccaggatc agaggttgtg gtgaaactgag attgcaccat
49561 tgcactccag cctgggcaac aagagcaaaa ctctatctca aaaaaataa aaaatagcca
49621 ggcacgggtg ctcatgcctg taatcctagc actttgggag gcagaggtgg gcagatcacc
49681 tgaggttagg agttcgagac tagcctggcc aacatggtga aaccccatct ctactacaaa
49741 tacaaaaatt agctaggcat ggtggcagcc acctgtaatc ccagctactt gggaggctga
49801 ggcaggagaa tcgcttgaac ccgggaggtg gaggttgacg tgagccaaga tcgggtcaca
49861 gcactccagc ctaggcaaca gagcgagact ccatctcaaa aaaacataaa taaataaaaa
49921 taaaaataaa taataataaa aagctaagaa tcaaagaagc agtttattcc taatttcaca
49981 gtctcatctg ttcatagtgg ggccaggatt agagtcagtg gccaaagctc catcctgggt
50041 tctttccctt cccaggccct accatcatag tataccaggg aaagacctgg agaagccagc
50101 aggttgacca ccgaaccaag gctgggccac cttcctcctg ggtctggtct ccagcctccc
50161 agttgtaccc ttccccagc cttcctgga tgcactgatc agcctgtgct tccttgcct
50221 gtttttcttt ataaatagag ccatgttctc ctctctctct ctctcttttt ttttttttt
50281 tttttgagat ggagtcttac tctgtcacc aggttgagat gcaatggcac gatctcagct
50341 cactgcaacc tctgtctccc aggttcaagc aattctcctg cctcagcctc ccgagtagct
50401 gggattacag gtgccacca ccatgccag ctactttttg gatttttagt agagacaggg
50461 tttcaccatg ttggtcaggc tggctctgaa ctctgacct taggtgttct gcccgctca
50521 gctcccaaaa gtgctgggat tacaggcgtg agccaccacg cctggcaaga cgtgttctct
50581 ctatgttgtt gaggttggtc ttgaaactct ggctgcaaga gatcttctct cctcagcctc
50641 ccaatgtgct gggattatag gcatgagcca ccacacttag ccagcctgt gctttcttaa
50701 atgaaaatct aagcatacgg ctgggtgtgg tggctcacgc ctgtaatccc agcattttgg
50761 gaggccaagg tgggcagatc acgaggtcag gagatcgaga ctatcctggc caacatgggtg
50821 aaacctgtc tctactaaag atacaaaaat tagctgggtg tgggtggcca tgtctgtagt
50881 cccagctact cgggagactg aggcaggaga atggcatgaa cctgggaggc agagcttgca
50941 gtgagctgag atcgcgccac tgctctccag cctaggtgac agagcgagac tccatctcaa
51001 aaaaaaata aaaaataaaa aaagaaaatc taagcgtggt gctcccctgc tcaaacatcc
51061 tcaggttctt ttcatggcag ataaggcat ctcttcatga gccagccct gcctactgac
51121 ccagccacct ctcccatccc ttcccacccc gtacttcagg cttcagcagt actgatcttt
51181 ccaaagaccc cagaacacac atgccttcat acctctgtgc ctgtacatgc ttgtttctgc
51241 ccttgaaatc atgacagtag ctctctgtag gccccgtag cctgtccctt gggctcttagc
51301 ctcttgaggg ccttcccaga gcccccaaaa agtaccocag gcatactttg gttcctctc
51361 tcatgtcccc tcagtaacttt gcacatacct cttttatagc agttgctatg ttgtgccaga
51421 gaagggagtc ctgtggctgg ggggcataata tcttttcttt ttgagacaga gtctagctgt
51481 gtcacccagg ctggagtgca gtagtgcgat ctcggtcac tgcaacctcc acctcctgga
51541 ttcaagogat tcttgtgcct cagcctcctg agtagctggg actacaggcg tgtgccacca
51601 tcatgcoctg ctactttttt gtattagata tataattttct ctcttagcac agtacctacc
51661 aagagtgagt gagtagatgt cctgaccct gcaggcatcc aaggccctcc ttcctggac
51721 ctgtttccac atgtgtgaag ggggtcacag gcagcagccc acctctcage ttcctccag
51781 ttctgtgtt ctgtgacccc ttttctctat ctctgctgc ttcctcacag **cttgaggcag**
51841 cctaggtgga ggccaagaag caacttcagg atgagatgct gggcggggtg gatgctgaga
51901 acaggctgca gaccatgaag gaggaactgg acttcagaa gaacatctac agtgagctgg
51961 ggactgtgct ttgcaagcca gagggctggg gctgggtgat gacagacttg ggctgggcta
52021 ggggggacca gctgtgtgca gagctgcct tcctgagtc cttgccctag tggacagggg
52081 gttgggggtg gccagcactc agctcccagg ttaaagtgg gctggtagt gctcatggag
52141 tagggctggg cagggagccc cgccctggg tcttggcctc ccaggaaacta attctgattt
52201 tggtttctgt gtccttctc caaccttcc **ag**gagctgcg **tgagaccaag cgccgtcatg**
52261 agaccogact ggtggagatt gacaatggga agcagcgtga gtttgagagc cggtggcg
52321 atgocgtgca ggaactgcgg gccagcatg aggaccaggt ggagcagtat aagaaggagc

52381 **tgagaagac ttattctgcc aagg**tgcttg ctctcgattg gttccctcac tgccctctgcc
52441 cttggcagcc ctacccttac ccacgctggg ctagccttc tggggatcag gcagatgggtg
52501 gcagggagct cagggtgcc caggacctgg ggctgtagca gtgatgccca actcaggcct
52561 gtgcctccac ccctcccagt caccacagtc ctaacccttt gtccctccct ccagctggac
52621 **aatgocaggc agtctgtga gaggaacagc aacctggtgg gggetgceca cgaggagctg**
52681 **cagcagtcgc gcattccgat cgacagcctc totgcccage tcagccaget ccagaagcag**
52741 **gtgatacccc acctaccccc tctctccagg ggcctagagt ctgggccgga tgcaggctgg**
52801 aagcccaggg ttgggggtgg ggggtggggg gggaggttcc tgaggaggag agggatgaaa
52861 agtgtcccca caaccacaga gaagggtcgc aggatgtgga gtcagatggc ctgtgtgctg
52921 tttctgtaca ctcttacctc accttacctt ctcagggtctt tggttttccc attcgaaaat
52981 ggaggctggt cttaatctcc ctaactcaga gttgccacag gactctgcaa tgtgaggtgt
53041 taaaagcatc agtatttttc tagttggctg tgctatttgt gacaggagaa aaagtctagc
53101 ctcagaacga gaggtttcag ttagacaagg ggaaggactt cccagttgcc agccaagact
53161 atgttttagag cttgtgatgt tcagagctgg ctctgatgag ggctctgggg aagctctgat
53221 tgcagatcct ggagagagta gccaggtgtc tctacaccg acccacgtcc ctcttcccc
53281 atacttaggg cccttgggag ctcaccaaac cctcccacc cccttcagct **ggcagccaag**
53341 **gaggogaagc ttogagacct ggaggactca ctggcccgtg agcgggacac cagccggcgg**
53401 **ctgctggcgg aaaaggagcg ggagatggcc gagatgcggg caaggatgca gcagcagctg**
53461 **gacagtagcc aggagcttct ggacatcaag ctggcccctg acatggagat ccacgcctac**
53521 **cgcaagctct tggagggcga ggaggagag**tgggctgggg agacgtcggg gagggtgctgg
53581 cagtgtctct tggccggcaa ctggccttga ctagaccccc acttggctct cctctcccc
53641 **ggtaocgct gtccccagc cctacctgc agcgcagccg tggccgtgct tctctcaet**
53701 **catcccagac acaggggtgg ggacagctca ccaaaaagcg caaactggag tccaetgaga**
53761 **gocgcagcag cttctcacag cacgcacgca ctagcggcgg cgtggcccgtg gaggaggtgg**
53821 **atgaggagg caagtttgtc cggtgcgca acaagtccaa tgag**gtaggc tctgtctcag
53881 ggtctaaggg gatacagctg catcaggggag agagtggcaa gacagaagga tggcatgtgg
53941 agagaggaac atccttgccc tcagaggggtg gaccaggggtg agcctgtata tctcctccac
54001 actctgggtc caggcctggc tcttgactc tttggctgtg agaccttgag caggttattt
54061 aacctctcag agcatcagtt tctcatctg taaaatgggg atgaatactg atcccctaagt
54121 ctttgagttg tcaggaagat gaaagataag gtatccgtgt gcctgggtgct gcgtatgtgt
54181 ccacagatca tggctattat ccccggggga agggcagtgca caggggtgtg ttagatgga
54241 aggagaggcc tcaattgcag gcaggcagag ggctgggcct ttgagcaaga tacaccaag
54301 agcctgggtg agcctccccg accttctct tccctatctt cccggcagga **ccagttocat**
54361 **ggcaattggc agatcaagcg ccagaatgga gatgatcct tgetgactta ccggttccca**
54421 **ccaaagttca cctgaaggc tgggcagggtg gtgac**gtga gtggcagggc gcttgggact
54481 ctggggaggc cttgggtggc gatgggagcg ctggggtaag tgtccttttc tctctcccag
54541 **atctgggctg caggagctgg ggccaccac agcccccta ccgacctggt gtggaaggca**
54601 **cagaacaact ggggtgcgg gaacagcctg cgtacggctc tcatcaactc cactggggaa**
54661 **gtaagtaggc ctgggcctgg ctgcttctg gacgaggctc cccctgatgg ccaacatcgg**
54721 agccagctgc cccaaccca agtttgcaa ttcagggcc ctttctagag ctctctgttg
54781 caggctocag acttctccac ccagtaggca aaccaaaaga tgcttctca acagcacaag
54841 ggggtggaagt tagacagtga ggattgttaa aggcagagcc atactcctac ccggagagct
54901 tgacagtgtc cctctggggg gaaatgagt tccttagctc catcaccaca gaggacagag
54961 taagcagcag gccggacaaa gggcaggcca caagaaaagt tgcaggtggt cactggggta
55021 gacatgctgt acaacccttc cctggccctg acccttgac ctggttccat gteccacc
55081 **ggaagtggc atgocaaagc tgggtgcgctc agtgactgtg gttgaggacg acgaggatga**
55141 **ggatggagat gaactgtcc atcaccacca ct**gagtggt agccgccgt gaggccgagc
55201 ctgcactggg gccaccagc caggcctggg ggcagcctct cccagcctc cccgtgcca
55261 aatcttttc attaaagaat gttttggaac tttactcgt ggctggcct tctctctct
55321 tctcctat acctgaaca gggaaaccag gtgtctgggt gccctactct ggtaaggag
55381 ggagtgggaa ctttctgatg ccatggaata ttctgtggg agcagtgac aagggtctg
55441 attgtcttc tgggaaagg agggaggac agacgtgggg catgcccgc ctgctctct
55501 cccattct tgttgcagc atactctctc atttccctca ttttctctc aagaatgtc
55561 tctctatc ctgaccgcc ctocactca attaatagt catgctgct gccctacaag
55621 cttgctocog ttctctctc ttttctctc aagctcagag tagctagaac agagtacag
55681 tcaactgtct ggttctctgt cccaagtct tctgagcct tctcccttt tatgtctcc
55741 ctctctct cgggcccct agcctccaa acccccattg cccgtggct ccttgggac

55801 agaaccacac cttcctgcct ggcggctggg agcctgcagg agcctggagc ctggttgggc
55861 ctgagtggtc agtcccagac tcgccgtccc gcctgagcct tgtctccctt cccagggtc
55921 ccactgcagc agctcggggg accccgctga gtacaacctg cgetcgcgca cctgtctgtg
55981 cgggaacctgc gggcagcctg ccgacaaggc atctgccagc ggetcaggag ccagggtggg
56041 cgggaacctgc tctctggct cttctgcctc cagtgtcacg gtcactcgcg getaccgcag
56101 tgtggggggc agtgggggtg gcagcttcgg ggacaatctg gtcacccgct cctacctcct
56161 gggcaactcc agccccgaa cccaggtgag ttgtctctgc tttgtctcca aatcctgcag
56221 gggggtccct ggtcatcgag gggtaggacg aggtggcctt gcagggggga gagcctgcct
56281 tctcttccgc agcccggggg agtgggagcc tcctccccac agcctgagtc ctagacagcc
56341 cacctctgca tctgcccct cttgtctgag cccagactg gagggcaggg gcagggctgg
56401 agtgtgaggg atgggggaga tgctacctc cttctagggg ccaggggagg gagggctgg
56461 gtccaggccc tgctgctcac acctctctcc tctgtttct ctcttagagc cccagaaact
56521 gcagcatcat gtaactctggg acctgccagg caggggtggg ggtggaggct tctgcgtcc
56581 tctcaacctc atgccacccc cctgccctgc acgtcatggg agggggcttg aagccaaaga
56641 aaaataacct tttggTTTTT ttctctgtga ttttttttcc taagagaagt tattttctac
56701 agtggTTTTa tactgaagga aaaacacaag caaaaaaaaa aaaaagcatc tatctcatct
56761 atctcaatcc taatttctcc tcccttctct tccctgctt ccaggaaaact ccacatctgc
56821 cttaaaacca aagagggctt cctctagaag ccaagggaaa ggggtgcttt tatagaggct
56881 agcttctgct tttctgcctt ggctgctgcc cccaccccgg ggacctgtg acatgggtgcc
56941 tgagaggcag gcatagaggc ttctccgcca gcctcctctg gacggcagge tcactgccag
57001 gccagcctcc gagagggaga gagagagaga gaggacagct tgagccgggc ccctgggctt
57061 ggctgctgt gattccacta cacctggctg aggttctctt gcctgccccg cccccagtcc
57121 ccacccctgc ccccagcccc ggggtgagtc cattctcca ggtaccagct gcgcttctt
57181 ttctgtatTT tatttagaca agagatggga atgaggtggg aggtggaaga agggagaaga
57241 aaggtgagtt tgagctgcct tccctagctt tagaccctgg gtgggctctg tgcagtcact
57301 ggaggttgaa gccaaagtggg gtgctgggag gagggagagg gaggtcactg gaaaggggag
57361 agcctgctgg caccaccgt ggaggaggaa ggcaagaggg ggtggagggg tgtggcagtg
57421 gTTTTggcaa acgctaaaga gcccttgcct cccatttcc catctgcacc ccttctctcc
57481 tccccaaatc aatacactag ttgtttcta...

Localització de la mutació 0.22 de la progèria en la seqüència del gen.

En aquest apartat s'observa el procés de localització la la base 1824 de la CDS, on es troba la mutació 0.22 que provoca la progèria, a partir de la seqüència del gen LMNA. En l'apartat anterior de l'annex s'ha localitzat les coordenades de la CDS per tant només falta ajuntar les seqüències i comptar les bases fins arribar a la 1824. En aquesta posició és on hi ha la citosina que muta a timina. Després del recompte s'observa que la posició 1824 de la CDS és la 56041 en l'ADN.

La CDS:

```
      ATGg agaccccgtc ccagcggcgc gccacccgca gcggggcgca ggccagctcc
32401 actccgctgt cgcccaccgc catcaccgg ctgcaggaga aggaggacct gcaggagctc
32461 aatgatoget tggcgggtcta catcgaccgt gtgcgctcgc tggaaacgga gaacgcaggg
32521 ctgcgccttc gcatcaccga gtctgaagag gtggtcagcc gcgagggtgc cggcatcaag
32581 gccgcctacg aggccgagct cggggatgcc cgcaagacct ttgactcagt agccaaggag
32641 cgcgcccgc tgcagctgga gctgagcaaa gtgcgtgagg agtttaagga gctgaaagcg
32701 cg
= 356 bases
```

```
      caatac caagaaggag
48061 ggtgacctga tagctgctca ggctcggctg aaggacctgg aggetctgct gaactccaag
48121 gaggccgcac tgagcactgc tctcagtgag aagcgcacgc tggagggcga gctgcatgat
48181 ctgcggggcc aggtggccaa g
= 157 b
```

```
      cttgaggcag
51841 ccctaggtga ggccaagaag caacttcagg atgagatgct gcggcgggtg gatgctgaga
51901 acaggctgca gaccatgaag gaggaactgg acttccagaa gaacatctac agtgag
= 126 b
```

```
      gagctgcg tgagaccaag cgccgtcatg
52261 agaccgcact ggtggagatt gacaatggga agcagcgtga gtttgagagc cggctggcgg
52321 atgcgctgca ggaactgcgg gccccagcatg aggaccaggt ggagcagtat aagaaggagc
52381 tggagaagac ttattctgcc aagg
= 171 b
```

```
      ct ggcagccaag
53341 gaggcgaagc ttogagacct ggaggactca ctggcccgtg agcgggacac cagccggcgg
53401 ctgctggcgg aaaaggagcg ggagatggcc gagatgcggg caaggatgca gcagcagctg
53461 gacgagtaac aggagcttct ggacatcaag ctggcccgtg acatggagat ccacgcctac
53521 cgcaagctct tggagggcga ggaggagag
= 126 b
```

```
      ct ggcagccaag
53341 gaggcgaagc ttogagacct ggaggactca ctggcccgtg agcgggacac cagccggcgg
53401 ctgctggcgg aaaaggagcg ggagatggcc gagatgcggg caaggatgca gcagcagctg
```

```
53461 gaogagtaacc aggagcttct ggacatcaag ctggcccttg acatggagat ccacgcctac
53521 cgcaagctct tggagggcga ggaggagag
```

= 221 b

```
gtaacgcct gtccccagc cctacctgc agcgcagccg tggccgtgct tcctctcact
53701 catcccagac acagggtggg ggcagcgtca ccaaaaagcg caaactggag tccactgaga
53761 gccgcagcag cttctcacag cacgcacgca ctagcgggcg cgtggccgtg gaggaggtgg
53821 atgaggaggg caagtttgtc cggctgcgca acaagtccaa tgag
```

= 223 b

```
ga ccagtccatg
54361 ggcaattggc agatcaagcg ccagaatgga gatgatccct tgctgactta ccggttccca
54421 ccaaagttca cctgaagcg tgggcaggtg gtgacg
```

= 108 b

```
atctgggctg caggagctgg ggccaccac agcccccta ccgacctggt gtggaaggca
54601 cagaacacct ggggctgcgg gaacagcctg cgtacggctc tcatcaactc cactggggaa
```

= 120 b

```
gaagtggcc atgcgcaagc tggcgcctc agtgactgtg gttgaggacg acgaggatga
55141 ggatggagat gacctgctcc atcaccacca c
```

= 90 b

```
ggctc
55921 ccactgcagc agctcggggg accccgctga gtacaacctg cgtcgcgca ccgtgctgtg
55981 cgggacctgc gggcagcctg ccgacaaggc atctgccagc ggctcaggag ccaggtggg
= 126 b 56041 cggacceatc tcctctggct cttctgcctc cagtgtcacg gtcactcgca gctaccgcag
56101 tgtggggggc agtgggggtg gcagcttcgg ggacaatctg gtcacccgct cctacctct
56161 gggcaactcc agccccgaa ccag
```

La suma de totes les bases anteriors dona 1824, per tant la posició 1824 de la CDS és la 56041 del gen.

```
agc cccagaact
56521 gcagcatcat gtaa
```