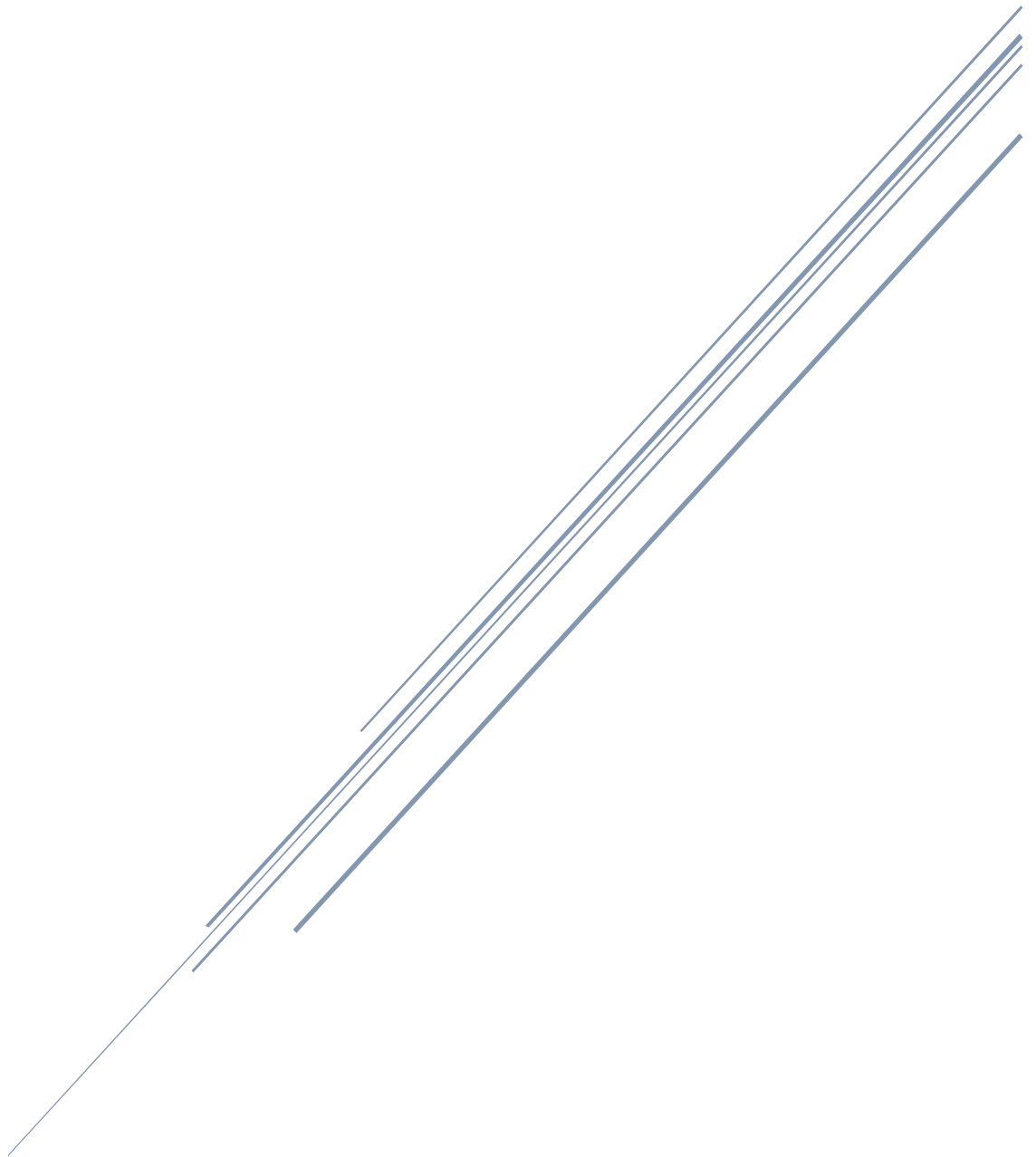


PRÁCTICA 3: VARIACIONES

Sistemas de Información Aplicados a la Bioinformática:
Gestión de Datos Genómicos



MITSS
Sergi Sanz

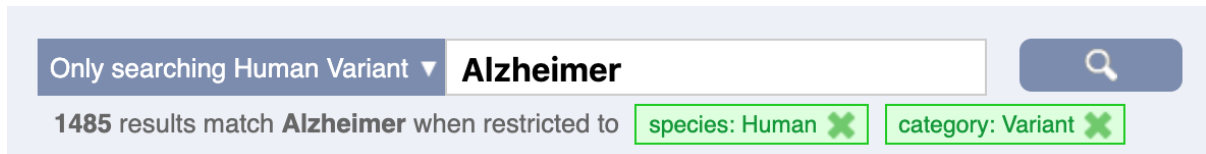
Índices:

EJERCICIO 3 – BUSCAR VARIACIONES SOBRE EL ALZHEIMER	2
3.3.- ¿CUÁNTAS VARIACIONES MUESTRA EN TOTAL CADA BASE DE DATOS?	2
3.4.- SEGÚN CLINVAR ¿CUÁNTAS VARIACIONES SE CONSIDERAN PATOGENICAS PARA LA ENFERMEDAD DE ALZHEIMER?	2
EJERCICIO 4 – SECUENCIA DE REFERENCIA	2
4.1.- ¿QUÉ SECUENCIA DE REFERENCIA USA CADA BASE DE DATOS PARA REPRESENTAR LA LOCALIZACIÓN DE UNA VARIACIÓN EN EL GENOMA HUMANO?	2
EJERCICIO 6 – INFORMACIÓN ASOCIADA AL POLIMORFISMO	2
6.1.- SEGÚN LA INFORMACIÓN DISPONIBLE EN CADA BASE DE DATOS ¿QUÉ FENOTIPOS ESTÁN ASOCIADOS A ESTE POLIMORFISMO?	2
6.2.- ¿CUÁL ES EL SIGNIFICADO CLÍNICO DEL POLIMORFISMO CON RESPECTO AL FENOTIPO CARDIOMIOPATÍA DILATADA?	3
6.3.- ¿QUÉ GEN O GENES ESTÁN AFECTADOS POR ESTA VARIACIÓN?	3
6.4.- ¿QUÉ TÉRMINO UTILIZA CADA UNA DE LAS BASES DE DATOS PARA REFERIRSE A ESTE POLIMORFISMO?	3
6.5.- ¿CUÁL ES EL CAMBIO QUE PRODUCE EL POLIMORFISMO A NIVEL DE NUCLEÓTIDOS? ¿Y A NIVEL DE AMINOÁCIDOS?	3
EJERCICIO 7 – FRECUENCIA ALÉLICA	4
7.1.- ¿CUÁNTOS TIPOS DE FRECUENCIAS ALÉLICAS PROPORCIONA CLINVAR?	4
7.2.- SEGÚN ENSEMBL Y LOS DATOS OBTENIDOS DE 1000 GENOMES PROJECT ¿EN QUÉ PORCENTAJE DE LA POBLACIÓN EUROPEA SE ENCUENTRA LA VARIACIÓN?	4
EJERCICIO 8 – FRECUENCIA ALÉLICA	4
8.1.- ¿CUÁNTOS TIPOS DE VARIACIONES CONSIDERA ENSEMBL?	4
8.2.- ¿CUÁNTOS TIPOS DE VARIACIONES CONSIDERA CLINVAR?	5
EJERCICIO 9 – DBSNP	6
9.3.- IDENTIFICAR CÓMO REPRESENTA DBSNP LOS ALELOS QUE SE PUEDEN DAR PARA EL POLIMORFISMO BUSCADO.	6
9.4.- ¿CUÁNTOS TIPOS DE VARIACIONES CONSIDERA DBSNP?	8

Ejercicio 3 – Buscar variaciones sobre el Alzheimer

3.3.- ¿Cuántas variaciones muestra en total cada base de datos?

En total se muestran 1485 resultados en NCBI:



Only searching Human Variant ▾ **Alzheimer** 🔍

1485 results match **Alzheimer** when restricted to species: Human ✕ category: Variant ✕

Según Ensembl se muestran un total de 354 variaciones:

Search results

Items: 1 to 100 of 354

3.4.-Según ClinVar ¿cuántas variaciones se consideran patogénicas para la enfermedad de Alzheimer?

En este caso, tenemos dos campos, los que sabemos que ya son patógenos que son 84 y los que se etiquetan como probable patógenos que son 26.

Ejercicio 4 – Secuencia de referencia

4.1.- ¿Qué secuencia de referencia usa cada base de datos para representar la localización de una variación en el genoma humano?

La secuencia de referencia utilizada para cada base de datos es la de GRCh37 y GRCh38.

Ejercicio 6 – Información asociada al polimorfismo

6.1.- Según la información disponible en cada base de datos ¿Qué fenotipos están asociados a este polimorfismo?

En la base de datos NCBI encontramos:

- Enfermedad de Alzheimer, tipo 3
- Enfermedad de Alzheimer
- ClinVar: fenotipo no especificado
- ClinVar: fenotipo no especificado
- Miocardiopatía dilatada, dominante
- Enfermedad de Alzheimer familiar de inicio temprano

Mientras que en la base de datos Ensembl encontramos:

- No especificado
- No provisto
- Enfermedad de Alzheimer
- Enfermedad de Alzheimer familiar de inicio temprano
- Miocardiopatía dilatada, dominante
- Enfermedad de Alzheimer, tipo 3
- Demencia frontotemporal
- Enfermedad de Pick

6.2.- ¿Cuál es el significado clínico del polimorfismo con respecto al fenotipo Cardiomiopatía dilatada?

EFO:0000407, EFO:0009142, HP:0001644

6.3.- ¿Qué gen o genes están afectados por esta variación?

Por esta variación se ve afectada el gen ENSG00000080815

6.4.- ¿Qué término utiliza cada una de las bases de datos para referirse a este polimorfismo?

Ensembl utiliza el termino con el que lo hemos buscado que es rs17125721, mientras que ClinVar utiliza VCV000098094.1 .

6.5.- ¿Cuál es el cambio que produce el polimorfismo a nivel de nucleótidos? ¿Y a nivel de aminoácidos?

El HGVS del polimorfismo es el siguiente: NM_000021.4(PSEN1):c.953A>G (p.Glu318Gly).

Por tanto, sabemos que NM_000021.4 es la secuencia genética, mientras que la letra “c” se utiliza para una secuencia de referencia de ADN codificante. Por otra parte, se producirá una sustitución de el nucleótido A en la posición c953 en cambió a una G. Mientras que p.Glu318Gly hace referencia a una proteína, que consiste en una secuencia de aminoácidos (no hay que olvidar que p. es la información complementaria sobre la componente proteica que ha sido codificada por el ARN codificante denotado por la secuencia c).

Ejercicio 7 – Frecuencia alélica

7.1.- ¿Cuántos tipos de frecuencias alélicas proporciona ClinVar?

1000 Genomes Project 0.00559

NHLBI Exome Sequencing Project (ESP) Exome Variant Server 0.01384

Trans-Omics for Precision Medicine (TOPMed) 0.01288

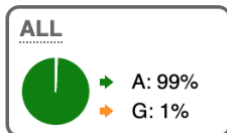
The Genome Aggregation Database (gnomAD) 0.01697

Exome Aggregation Consortium (ExAC) 0.01530

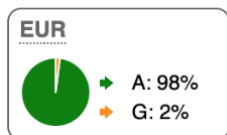
The Genome Aggregation Database (gnomAD), exomes 0.01461

7.2.- Según Ensembl y los datos obtenidos de 1000 Genomes Project ¿En qué porcentaje de la población europea se encuentra la variación?

En este caso, nos encontramos con que:



Lo cual significa que el 99% de la población tiene el alelo de referencia A mientras que solo el 1% de la población tiene el alelo de referencia G.



Mientras que en el caso de Europa podemos apreciar como el 98% de la población tiene el alelo de referencia A mientras que solo el 2% de la población tiene el alelo de referencia G.

Ejercicio 8 – Frecuencia alélica

8.1.- ¿Cuántos tipos de variaciones considera Ensembl?

En este caso encontramos que Ensembl considera:

- Variantes de secuencia
 - SNP
 - Insertion
 - Deletion
 - Indel
 - Substitution
- Variantes estructurales
 - CNV
 - Inversion
 - Translocation

- Clases variantes
 - SNV
 - Substitution
 - Alu_insertion
 - complex_structural_alteration
 - complex_substitution
 - copy_number_gain
 - copy_number_loss
 - copy_number_variation
 - duplication
 - interchromosomal_breakpoint
 - interchromosomal_translocation
 - intrachromosomal_breakpoint
 - intrachromosomal_translocation
 - inversion
 - loss_of_heterozygosity
 - mobile_element_deletion
 - mobile_element_insertion
 - novel_sequence_insertion
 - short_tandem_repeat_variation
 - tandem_duplication
 - translocation
 - deletion
 - indel
 - insertion
 - sequence_alteration
 - probe

8.2.- ¿Cuántos tipos de variaciones considera ClinVar?

En este caso encontramos que ClinVar considera:

- Alu Deletion
- Alu Insertion
- Complex Chromosomal Rearrangement
- Complex Substitution
- Copy Number Gain
- Copy Number Loss
- Copy Number Variation
- Deletion
- Duplication
- HERV Deletion
- Indel

- Insertion
- Interchromosomal Translocation
- Intrachromosomal Translocation
- Inversion
- LINE1 Deletion
- LINE1 Insertion
- Microsatellite
- Mobile Element Deletion
- Mobile Element Insertion
- Monomeric Repeat
- Multiple Nucleotide Polymorphism
- Multiple Nucleotide Variation
- No Alteration
- Novel Sequence Insertion
- Sequence Alteration
- Single Nucleotide Variant
- SVA Deletion
- SVA Insertion
- Tandem Duplication
- Translocation

Ejercicio 9 – dbSNP

9.3.- Identificar cómo representa dbSNP los alelos que se pueden dar para el polimorfismo buscado.

En la base de datos dbSNP obtenemos los siguientes alelos: C>A / C>G / C>T

Allele: A (allele ID: [33128](#))

ClinVar Accession	Disease Names	Clinical Significance
RCV000019715.28	Alzheimer disease, type 1	Pathogenic

Allele: G (allele ID: [33144](#))

ClinVar Accession	Disease Names	Clinical Significance
RCV000019733.27	Alzheimer disease, type 1	Pathogenic
RCV000518713.1	not provided	Pathogenic

Allele: T (allele ID: [33127](#))

ClinVar Accession	Disease Names	Clinical Significance
RCV000019714.31	Alzheimer disease, type 1	Pathogenic
RCV000020308.2	Alzheimer's disease	Pathogenic
RCV000084575.1	not provided	Not-Provided

9.4.- ¿Cuántos tipos de variaciones considera dbSNP?

Las variaciones que considera dbSNP son las siguientes:

Sequence name	Change
APP RefSeqGene	NG_007376.2:g.284345G>T
APP RefSeqGene	NG_007376.2:g.284345G>C
APP RefSeqGene	NG_007376.2:g.284345G>A
APP RefSeqGene	NG_007376.1:g.284037G>T
APP RefSeqGene	NG_007376.1:g.284037G>C
APP RefSeqGene	NG_007376.1:g.284037G>A
GRCh37.p13 chr 21	NC_000021.8:g.27264096C>A
GRCh37.p13 chr 21	NC_000021.8:g.27264096C>G
GRCh37.p13 chr 21	NC_000021.8:g.27264096C>T
GRCh38.p12 chr 21	NC_000021.9:g.25891784C>A
GRCh38.p12 chr 21	NC_000021.9:g.25891784C>G
GRCh38.p12 chr 21	NC_000021.9:g.25891784C>T