



UNIVERSIDAD
NACIONAL
DE LA PLATA

Facultad de Informática

Soporte al proceso de análisis e interpretación de variantes genéticas en Exoma humano

Trabajo de tesis para obtener el grado de Magíster en Ingeniería del
Software

Tesista: Cecilia Moreyra

Director: Dr. Andrés Rodríguez

Co-directora: Dra. Alejandra Garrido

La Plata, Noviembre 2022

Índice

Resumen.....	6
Capítulo 1: Introducción.....	7
1.1 Contexto de la tesis: el diagnóstico genético.....	7
1.2 Objetivos de la tesis	9
1.3 Contribuciones	9
1.4 Organización de la tesis.....	10
Capítulo 2: Marco teórico	11
2.1 Introducción a los conceptos fundamentales de genética	11
2.2 Introducción al Exoma.....	19
2.3 Trabajos relacionados	20
2.4 Herramientas de asistencia al genetista	23
2.5 Introducción al concepto de diseño centrado en el usuario, escenarios y prototipado ..	29
2.6 Introducción al concepto de usabilidad y métodos de evaluación de usabilidad	32
Capítulo 3: Descripción del Problema	34
3.1 Secuenciación del Exoma	34
3.2 Detalle del archivo Excel	36
3.3 Análisis e Interpretación	40
Capítulo 4: Modelado del proceso de toma de decisiones, planteo de escenarios.	43
4.1 Escenario previo al filtrado de las variantes presentadas en el Exoma	44
4.1.1 Registro de los datos de la historia clínica del paciente	44
4.1.2 Priorización de genes a través del Panel Bioinformático	46
4.2 Escenario de filtrado de las variantes presentadas en el Exoma	47
4.2.1 Filtros involucrados	47
4.2.2 Trazabilidad del proceso de filtrado de las variantes presentadas en el Exoma	49
4.3 Escenario posterior al filtrado de las variantes presentadas en el Exoma, generación del Diagnóstico.....	51
4.3.1 Registro de la consulta manual a bases de datos de fenotipos	51
4.3.2 Registro de variables patológicas.....	52
4.3.3 Generación de informes.....	52
4.4 Escenario de retro-alimentación de diagnósticos futuros	54
4.4.1 Relación de la variante con el fenotipo.....	54
4.4.2 Frecuencia poblacional de enfermedades monogénicas en la población argentina.	55
4.5 Retroalimentación del Panel Bioinformático	55

Capítulo 5: Especificación de los requerimientos de la solución	57
5.1 Requerimientos de Visualización del proceso de filtrado del Exoma.	57
Capítulo 6: Solución propuesta	61
6.1 Stock Screener de Investing.com como referencia.....	61
6.2 Prototipo herramienta Exoma Browser	62
6.2.1 Escenario previo al filtrado de variantes presentadas en el Exoma	63
6.2.2 Escenario de filtrado de las variantes presentadas en el Exoma	64
6.2.3 Escenario posterior al filtrado de las variantes presentadas en el Exoma, generación del diagnóstico	66
6.2.4 Escenario de retro-alimentación de diagnósticos futuros	69
6.3 Validación del producto	70
6.3.1 Caso de Validación: Dr. Gabriel Ercoli, médico Genetista.....	70
6.3.2 Caso de Validación: Mg. Matias Butti, Director de la empresa Zoigen & genomIT ...	71
7. Conclusiones.....	73
7.1 Conclusión general	73
7.2 Trabajo Futuro.....	74
Bibliografía.	75
Apéndices	80
Apéndice 1. Informe de resultados.....	80
Apéndice 2. Informe clínico.....	84
Apéndice 3. Minutas de Reunión	87
Apéndice 4. Cuestionario SUS – Gabriel Ercoli.....	103
Apéndice 5. Cuestionario SUS – Matias Butti.....	106
Glosario	109

Índice de Figuras

Figura 1: Abstracción de la gran palabra en un lenguaje de 4 letras (A, C, T, G)	12
Figura 2: Dogma central de la biología molecular, generación de la proteína	12
Figura 3: Dogma central de la biología molecular, transcripción	13
Figura 4: Dogma central de la biología molecular, traducción	13
Figura 5: NGS raw data-BRAF V600E mutación en un melanoma [5]	14
Figura 6: Genoma de Referencia.....	15
Figura 7: Cambios con respecto al Genoma de Referencia	15
Figura 8: Browser de un genoma	17
Figura 9: Bioinformática, campo de acción.....	18
Figura 10: Ciclo de flujo de trabajo generalizado entre la plataforma de gestión de datos LORIS y la plataforma de procesamiento CBRAIN [13].	21
Figura 11: Gen IO, Interfaz de Usuario [15]	24
Figura 12: VCF-Server, Módulos funcionales [17].....	25
Figura 13: VCF Explorer, Carga de archivos [19]	26
Figura 14: VCF Explorer, Filtrado y anotación de variantes [19].....	26
Figura 15: VCF Explorer, Visualización de Variantes [19].....	27
Figura 16: ClinLabGeneticist, Workflow [21].....	28
Figura 17: Prototipo de un editor de clases multitáctil [26]	31
Figura 18: Proceso de secuenciación, análisis e interpretación del exoma	35
Figura 19: Historia clínica del Paciente registrada por el médico genetista	45
Figura 20: Stock Screener, panel principal [38].....	62
Figura 19: Exoma Browser, panel principal [39].	63
Figura 21: Exoma Browser, registro de los datos de la Historia Clínica del Paciente [39].....	63
Figura 22: Exoma Browser. Panel Bio-Informático. Búsqueda por fenotipo [39].	64
Figura 23: Exoma Browser, panel de filtrado de variantes [39].....	65
Figura 24: Exoma Browser, solapas con columnas identificatorias de la variante [39].	65
Figura 25: Exoma Browser, solapas sin columnas identificatorias de la variante [39].	66
Figura 26: Exoma Browser, paginación [39].....	66
Figura 27: Exoma Browser, registro de la consulta manual a bases de datos de fenotipos [39]	67
Figura 28: Exoma Browser, Informe de Resultados [39].....	68
Figura 29: Exoma Browser, Informe Clínico [39].....	69
Figura 30: Exoma Browser, tooltip de variante candidata relacionada con el fenotipo [39].	70

Índice de Tablas

Tabla 1: Descripción de las columnas representativas del Exoma.....	37
Tabla 2: Extracto de una fila del exoma que posee 2 posiciones del transcripto de un gen	39
Tabla 3: Transcripto de posición 1 dentro del extracto	39
Tabla 4: Transcripto de posición 2 dentro del extracto	40
Tabla 5: Planteo de escenarios.....	44
Tabla 6: Columnas de un panel bioinformático	47

Resumen

El avance de las ciencias médicas en especial de la bioinformática, la genética y el diagnóstico genético han traído grandes cambios en los tratamientos de las enfermedades y por ende en nuestra calidad de vida. De esto surge la necesidad de proveer a los médicos de herramientas informáticas cada vez más complejas en apoyo al trabajo con los pacientes y sus familiares.

En esta tesis se plantean los escenarios llevados a cabo, por el médico Genetista, para ejecutar el proceso de análisis e interpretación de variantes genéticas, a través de un Exoma Humano, que llevan a la detección y diagnóstico de enfermedades monogénicas en un paciente.

Para tal fin, se modela el proceso de toma de decisiones y los distintos escenarios, la trazabilidad en la ejecución de los filtros de las variantes presentadas, se muestra cómo se arriba a un diagnóstico, cómo se retroalimenta un panel bioinformático, que cumplirá la función de base de datos propia, y también cómo se retroalimentan los diagnósticos futuros. Se especifican los requerimientos y se propone una solución en base a la construcción de un prototipo que cumple con todo lo requerido. Finalmente, se valida dicha solución con un médico genetista y se arriba a las conclusiones.

Capítulo 1: Introducción

1.1 Contexto de la tesis: el diagnóstico genético

La detección, lectura y cuantificación de material genético en una muestra biológica ha mostrado en los últimos años un significativo impacto en todas las áreas de la salud, sobre todo en las áreas de las enfermedades infecciosas, el cáncer y las enfermedades poco frecuentes. [1]

El desarrollo de nuevas tecnologías, más rápidas y precisas ha transformado al [diagnóstico molecular](#) en una herramienta clave para el equipo clínico, en directo beneficio del paciente en cuanto a la definición de su pronóstico, seguimiento, tratamiento y asesoramiento sobre riesgos familiares [1]. Una de las nuevas tecnologías es la secuenciación del [exoma](#) completo, que consiste en realizar una lectura de las regiones de mayor relevancia de miles de [genes](#) del individuo, en la búsqueda de variantes clínicamente asociadas con su [fenotipo](#) o con riesgos para la salud.

En las últimas décadas ha tomado especial importancia la identificación de evidencia genética y experimental relevante que respalde la relación entre variantes de ADN (cambios con respecto a un genoma “de referencia”) y una o más enfermedades. Es decir, dado un fenotipo particular y un problema conocido, se trata de la identificación de esas variantes que son relevantes para el fenotipo considerado. Este problema es especialmente importante en contextos de diagnóstico clínico [1]. En efecto, las variaciones genéticas no solo causan características físicas diferentes entre humanos, sino que también pueden ser potencial causa de [enfermedades monogénicas](#) (las que obedecen a la presencia de una variación genética sin incidencia de otros factores) o predisponer genéticamente a [enfermedades multigénicas y multifactoriales](#) (cuando se agrupan múltiples variantes en genes distintos y se suma el impacto del medio ambiente).

En este contexto, el diagnóstico genético es el proceso que permite extraer conocimiento de los datos genéticos con el fin de contribuir al diagnóstico de una determinada enfermedad. Consiste en la identificación de variaciones genéticas que causan o predisponen a la enfermedad. Si consideramos que un exoma puede incluir la presencia de varios centenares

de variantes genéticas, el diagnóstico se convierte en un proceso de comparación de datos que es muy complejo y requiere mucho tiempo.

Una forma sistemática de realizar diagnóstico genético es la metodología SILE (Search, Identification, Load, Exploitation) [2] que consta de cuatro pasos:

1. Búsqueda (los investigadores buscan bases de datos biológicas externas y seleccionan aquellas que sirvan para propósitos de análisis);
2. Identificación (para cada base de datos, identifican el contenido adecuado para el análisis);
3. Carga (los archivos de variantes respecto del genoma de referencia, o [archivos VCF](#) y el contenido de las bases de datos biológicas se cargan en una plataforma (repositorio central).
4. Explotación (los investigadores exploran, operan, analizan y extraen conocimiento de datos genéticos extrayendo conclusiones de los mismos).

Este escenario motiva diversos desafíos para el desarrollo de la tesis, como la cantidad y variedad de dimensiones que debe manejar el genetista para obtener el diagnóstico; aún en el caso de prácticas consensuadas como SILE, cada profesional lleva adelante un proceso con criterios e idiosincrasias propias que deben ser respetadas; la necesidad de facilitar diferentes interacciones con la información.

Las herramientas de software que permitan soportar un proceso de diagnóstico genético de estas características deberán facilitar la identificación en el exoma de variantes genéticas respecto al genoma de referencia, el volumen del archivo a analizar por el genetista, la variedad y tipo de los datos que se presentan, las múltiples bases de datos consultadas para el análisis y la interpretación y las tecnologías involucradas requieren el manejo de volumen y velocidad. Al mismo tiempo, la concepción, diseño e implementación de estas herramientas presupone desafíos variados en términos de visualización de grandes volúmenes de datos, personalización de procesos y registro de las decisiones en cada etapa.

1.2 Objetivos de la tesis

Objetivo General

El propósito principal de esta tesis es dar soporte al médico genetista en el proceso de diagnóstico genético, en particular el acceso a la información sobre la composición genética de sus pacientes, la identificación de los posibles escenarios de detección y diagnóstico y el registro de cada uno de los pasos y datos llevados a cabo en el proceso.

Objetivos específicos

- Modelar el proceso de toma de decisiones para el filtrado y selección de variantes de un exoma relevantes llevado a cabo por el médico genetista para alcanzar un informe de diagnóstico de un paciente.
- Proveer herramientas de asistencia, basadas en una visualización eficaz, para la ejecución de ese proceso de análisis e interpretación con mayor rapidez y sensibilidad.
- Acelerar los tiempos de interpretación.
- Aumentar la sensibilidad (en este caso la tasa de detección de las variantes patogénicas causales de la enfermedad o que sean relevantes para su salud).

1.3 Contribuciones

La presente tesis ofrece dos contribuciones. En primer lugar, la especificación de requerimientos para herramientas de soporte al proceso de diagnóstico genético basada en Escenarios de Uso. Se plantean los escenarios de uso del médico genetista durante la ejecución del proceso de análisis e interpretación de variantes genéticas, a través de un exoma humano, que llevan a la detección y diagnóstico de enfermedades monogénicas en un paciente y tiene como principal contribución la ayuda y asistencia en el proceso de toma de decisiones hasta llegar al diagnóstico y seguimiento.

La segunda contribución de esta tesis es la presentación de un prototipo de aplicación denominada “Exoma Browser” que cumple con el objetivo principal que es dar soporte al médico genetista en el proceso de análisis e interpretación de variantes genéticas en el exoma humano, diagnóstico de enfermedades y registro de cada uno de los pasos llevados a cabo en el proceso. Dicho prototipo está basado en la aplicación “Stock Screener” del sitio Investing.com [3] diseñada para el dominio de inversiones financieras que, en términos funcionales, aplica filtros a un gran volumen de datos para facilitar y acelerar criterios complejos de decisión. Stock Screener sirvió como punto de partida para el diseño de la pantalla de la herramienta Exoma Browser, propuesta como solución, con lo cual se planteó el uso de un prototipo ya explorado como herramienta de descubrimiento, comprobación de los requerimientos y evolución del producto final.

1.4 Organización de la tesis

En el capítulo 2 se presenta el marco teórico e introductorio de los conceptos fundamentales de genética.

En el capítulo 3 se realiza una presentación general de la problemática a resolver.

En el capítulo 4 se realiza el modelado de dicha problemática a través de los escenarios llevados a cabo en el proceso de decisión del diagnóstico: paso previo, ejecución del filtrado y posterior del análisis de variantes y cómo se realiza también la retroalimentación de diagnósticos futuros.

En el capítulo 5 se recopilan los requerimientos que surgen en base al relevamiento de los escenarios antes mencionados observando dinámicamente el trabajo y metodología llevados a cabo por el genetista en cada paso.

En el capítulo 6 se describe la herramienta propuesta para cumplir los objetivos y requerimientos.

En el capítulo 7 se presentan la conclusión general y el trabajo futuro.

Capítulo 2: Marco teórico

En este capítulo se introducen los conceptos fundamentales genética y exoma, relacionados al campo de aplicación y trabajos relacionados a los mismos. También se presentan los conceptos de escenarios de uso, prototipos y pruebas de usabilidad relacionados a la Ingeniería de Software.

2.1 Introducción a los conceptos fundamentales de genética

Para poder comprender el proceso de evaluación y análisis de variantes del exoma, es necesario conocer previamente algunos conceptos básicos de genética. Se presenta a continuación una síntesis simplificada que permitirá comprender este trabajo de tesis, pero no se intenta abarcar con estos la complejidad de la temática. Para una mayor profundidad se recomienda consultar bibliografía del área [4]

ADN

La célula humana se divide en 23 pares de [cromosomas](#) heredados de los padres. Cada uno de los cromosomas es una molécula de ADN. La molécula de ADN es una secuencia de bloques de construcción que son los nucleótidos. Es bicatenaria y complementaria.

Abstrayendo el ADN

Los genes son regiones particulares que están dentro del ADN (de la “gran palabra”) y codifican la instrucción para generar un tipo de [proteína](#) particular. El código es en ACTG instrucciones (Figura 1).

```
TCCCGGAAGGTTTGCGGCGCACGCATGCATGTAAGCTGCTGTCGCGCGCATTCTGTGCTGAACCCGTCGTAGHC
TGCTGATCGTACAACuuuGAGTCGTAGCTGCAGTGCAGCTGCATTGCTAGCGTAGCTGCCGCTAGCTTATTATAGCTG
ACTGATCTATTAGCGTATGCTGATTATCGTATGCTTAGCTATGACATTTAGCTAGTGTATATTATGCTGATCTCTA.....
```

Figura 1: Abstracción de la gran palabra en un lenguaje de 4 letras (A, C, T, G)

Variante

Una variante es un valor que un individuo presenta en esa posición diferente al de referencia. Si se presenta una variante dentro de la instrucción podría producirse una proteína anómala, no funcional y por ende células, tejidos y órganos no funcionales.

Dogma central de la biología molecular

El dogma central de la biología molecular establece los pasos para la generación de la proteína es decir la transcripción y traducción (Figuras 2, 3 y 4). Una vez identificado el gen que contiene la instrucción se copia solo el fragmento del mismo en una molécula de ARN intermedia que es más chica y más fácil de manipular. Una vez producida esta molécula sale al citoplasma en donde se produce la traducción.

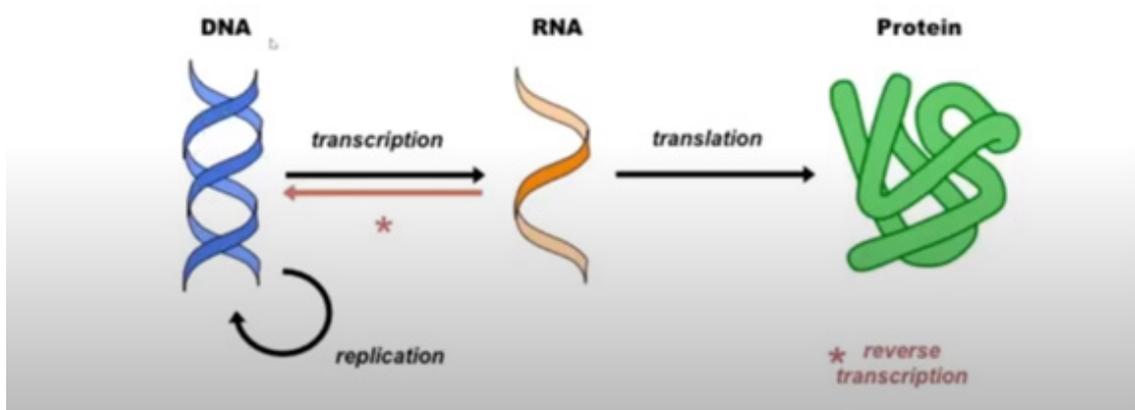


Figura 2: Dogma central de la biología molecular, generación de la proteína

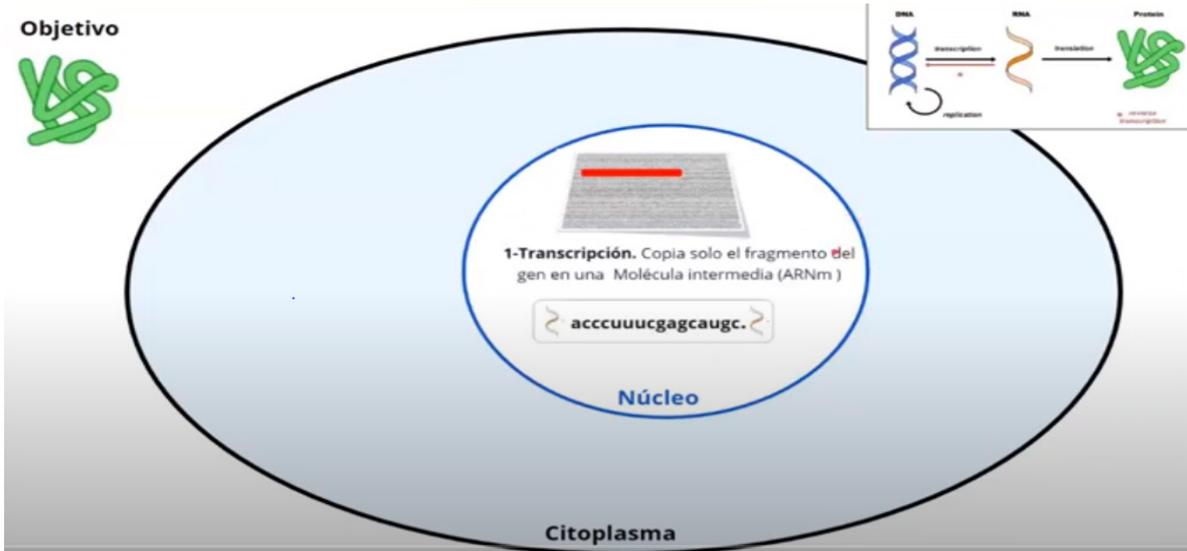


Figura 3: Dogma central de la biología molecular, transcripción

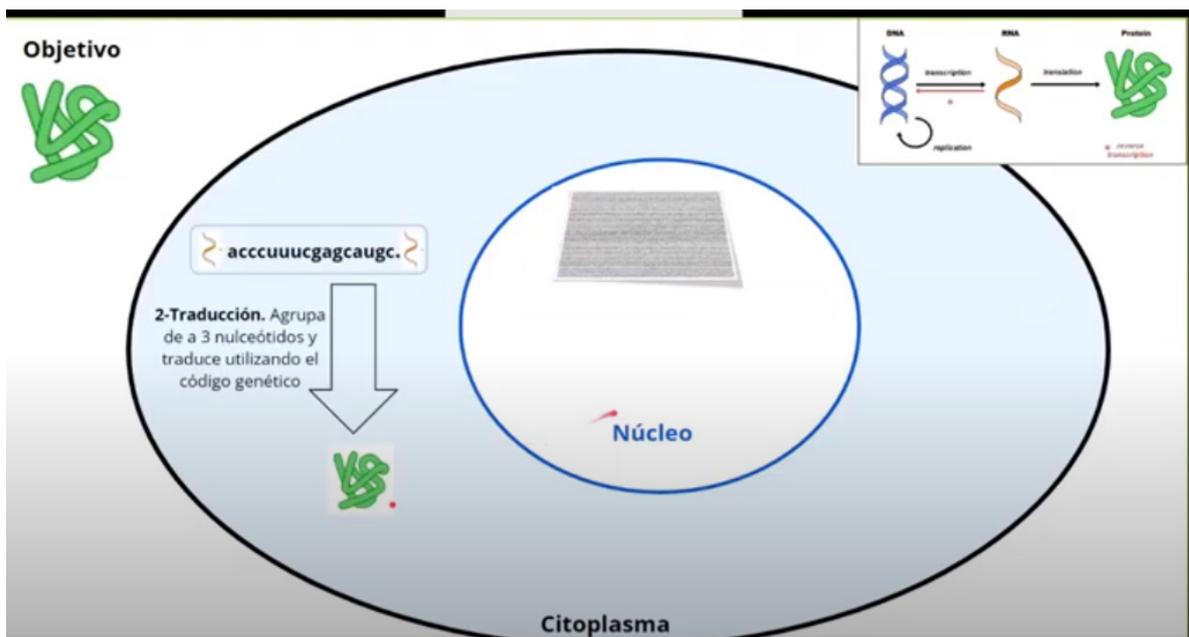


Figura 4: Dogma central de la biología molecular, traducción

El proceso se repite por cada tipo distinto de proteína. El ARN se podrá reutilizar pudiendo haber varias moléculas de ARN dependiendo la cantidad de proteínas que necesite.

En conclusión, el [genoma](#) nuclear de un individuo es el total de su material genético nuclear. Los genomas de individuos con distintas características, tanto clínicas como

moleculares, pueden ser comparados y se puede identificar la asociación de determinadas variantes en los genes con predisposición a ciertas enfermedades. También puede identificarse pacientes que responden a un tratamiento de los que no, para encontrar informática y estadísticamente cuáles son las letras diferentes entre un grupo y otro y a partir de ahí empezar a generar hipótesis. Hay drogas que funcionan solamente en pacientes que tienen determinada variante.

Lectura del ADN

Se realiza utilizando [técnica de NGS \(Next Generation Sequencing\)](#). Con dicha técnica se puede secuenciar todo el genoma humano (Figura 5). Las plataformas de NGS realizan la secuenciación de millones de pequeños fragmentos de ADN en paralelo. El análisis bioinformático se utiliza para unir estos fragmentos mediante el mapeo de las lecturas individuales con el genoma humano de referencia. Cada una de las tres mil millones de bases del genoma humano se secuencian varias veces, lo que proporciona una gran profundidad para obtener datos precisos y una visión de la variación inesperada del ADN. La técnica NGS puede utilizarse para secuenciar genomas enteros o limitarse a áreas específicas de interés, incluidos los 22.000 genes codificantes (un exoma completo) o pequeños números de genes individuales [5].

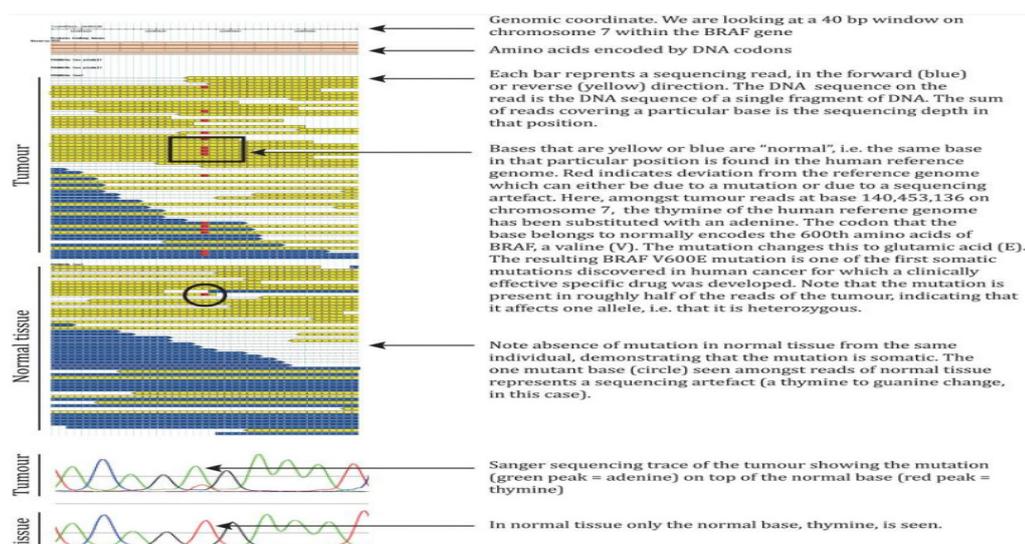


Figura 5: NGS raw data-BRAF V600E mutación en un melanoma [5]

Genoma de referencia

El genoma de referencia (Figura 6) nos dice para cada posición del genoma cuál es la letra más representativa.

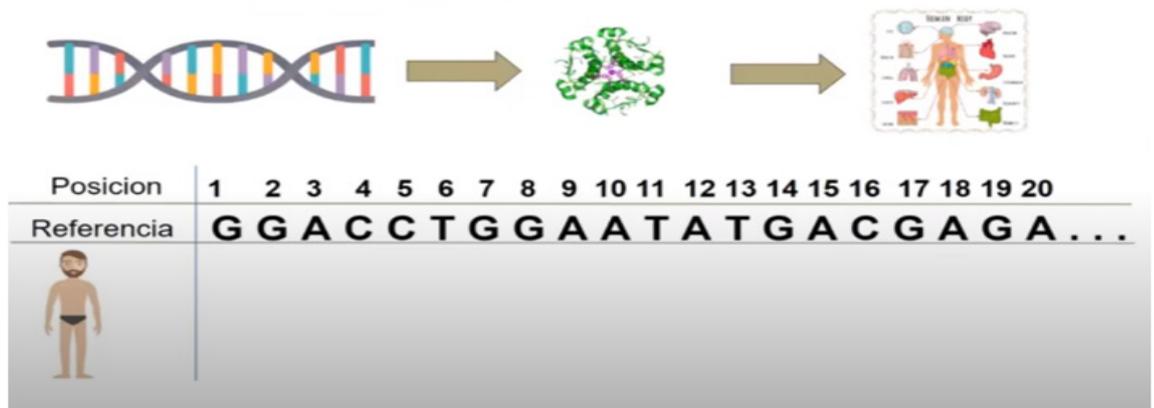


Figura 6: Genoma de Referencia

Se toma la muestra y se manda a secuenciar. Entonces por cada posición se pueden encontrar variantes, se evalúa qué produce esa variante respecto al genoma de referencia. No necesariamente está asociada a una patología.

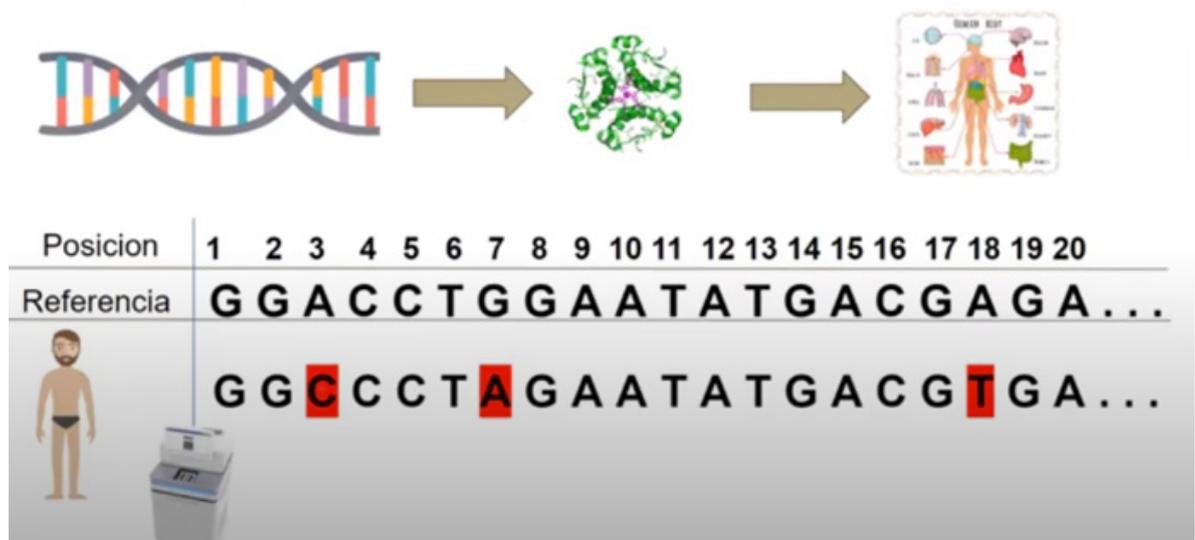


Figura 7: Cambios con respecto al Genoma de Referencia

Tal como se muestra en la figura 6 las posiciones 3, 7 y 18 tienen variantes (Figura 7). Existen bases de datos que categorizan cada una de las variantes en benignas, neutrales o patogénicas. Si tengo una variante en una región que no está dentro de un gen probablemente no me traiga ningún problema porque no será utilizada para generar proteínas.

Fenotipo

Un fenotipo es la característica clínica, conductual y de comportamiento del paciente, asociada con una patología. Por ejemplo, si el paciente presenta brazos, piernas cortas y cabeza grande son fenotipos asociados a la acondroplasia, otro ejemplo es un paciente que presenta marcha inestable, debilidad generalizada y dificultad para hablar son fenotipos relacionados a la ataxia.

Enfermedades Monogénicas y multigénicas

Las enfermedades monogénicas son provocadas por la variante de 1 solo gen y sin factores ambientales que puedan evitar la aparición de la enfermedad. Las multigénicas son aquellas que agrupan múltiples variantes en genes distintos y se suma el impacto del medio ambiente.

Las enfermedades monogénicas pueden ser dominantes o recesivas, si son dominantes se prioriza el hallazgo de una variante patogénica en el mismo gen (en una de las dos copias, la materna o la paterna), si son recesivas se prioriza el hallazgo de dos variantes patogénicas en el mismo gen (en las dos copias, la materna y la paterna). La relación de las enfermedades monogénicas con respecto a los fenotipos se encuentra en una Ontología que se llama HPO (Human Phenotype Ontology) [6]. Las patologías de OMIM [7] se anotan con los términos HPO que le corresponden. Entonces para cada patología se pueden asociar los fenotipos que están descritos para esa patología. Existen herramientas, tales como Phenomizer [8], que permiten ingresar los fenotipos para determinar las patologías asociadas a un fenotipo y por ende los genes asociados a un fenotipo.

En conclusión, los fenotipos son identificados y luego buscados en las bases de datos que permiten encontrar genes que estén relacionados con patologías que están relacionadas con esos fenotipos.

Browser de un Genoma

El browser de un genoma (Figura 8) soporta la exploración y análisis de los diferentes tipos de datos moleculares en el contexto del genoma completo incluyendo la anotación de genes.

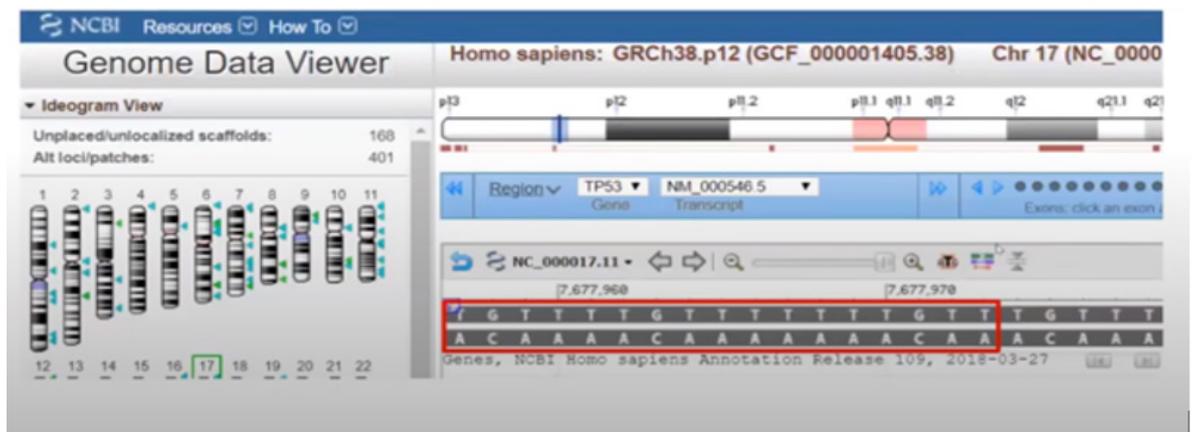


Figura 8: Browser de un genoma

Genómica y genética

La genómica se refiere al estudio del genoma completo de un organismo, mientras que la genética se refiere al estudio de un gen en concreto [9]. Se encarga de trabajar con las letras ACTG. Estudia las secuencias génicas.

Bioinformática

La bioinformática (Figura 9) es una sub-disciplina de la biología y las ciencias computacionales que se encarga de adquirir, almacenar, analizar y diseminar la información biológica, en gran parte correspondiente a las secuencias de ADN y aminoácidos [10]. Se basa en la propiedad Genetic Linkage que dice que las letras se transmiten de a bloques y existen base de datos de bloques. Si hay una letra que no se puede leer con los algoritmos interactuando con las bases de datos de bioinformática es posible inferirla. También se encarga de diseñar la base de datos genómica del paciente para encontrar asociaciones entre determinadas variantes genéticas y determinadas características que interesen estudiar, en base a esto se realizan la búsqueda/inferencia de los significados de una variante para saber si es benigna, patogénica o neutral.

La bioinformática también se encarga del armado de [paneles bioinformáticos](#) donde se comparan 2 secuencias para saber si pertenecen o no a la misma región del genoma. Las secuencias son bastante largas, los alineamientos de secuencias se realizan a través de algoritmos que permiten comparar las características biológicas. Los genetistas se valen de las bases de datos bioinformáticas que permiten comparar por ejemplo la secuencia de un gen humano con la secuencia de un gen de otra especie de la cual ya se conoce la función. Partiendo de ahí se puede idear una hipótesis.

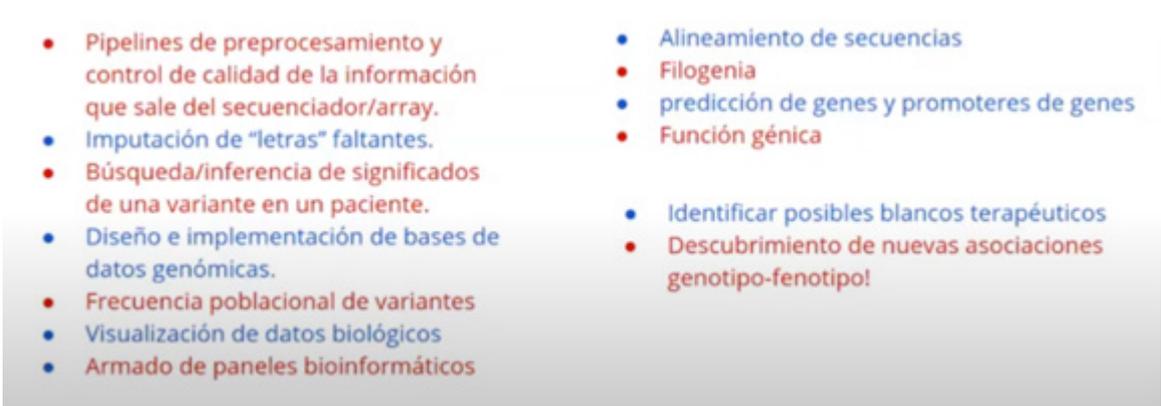
- 
- Pipelines de preprocesamiento y control de calidad de la información que sale del secuenciador/array.
 - Imputación de "letras" faltantes.
 - Búsqueda/inferencia de significados de una variante en un paciente.
 - Diseño e implementación de bases de datos genómicas.
 - Frecuencia poblacional de variantes
 - Visualización de datos biológicos
 - Armado de paneles bioinformáticos
 - Alineamiento de secuencias
 - Filogenia
 - predicción de genes y promotores de genes
 - Función génica
 - Identificar posibles blancos terapéuticos
 - Descubrimiento de nuevas asociaciones genotipo-fenotipo!

Figura 9: Bioinformática, campo de acción

2.2 Introducción al Exoma

Cuando la cantidad de genes que podrían explicar el diagnóstico presuntivo son múltiples, o bien son pocos, pero de gran longitud, está indicado el pedido de un exoma. El exoma es la parte del genoma formado por los exones, es decir, las partes codificantes de los genes que formarán parte del ARN mensajero maduro y, al ser éste traducido por la maquinaria celular, darán lugar a las proteínas, es la parte funcional más importante del genoma y la que contribuye en mayor medida al fenotipo final de un organismo. Aproximadamente, corresponde al 1.5% del genoma humano, pero en él se encuentran entre 30% y 50% de las alteraciones genéticas que causan patologías monogénicas (existe variabilidad según el tipo de órgano afectado o patología se trate) [11].

El exoma trata solamente enfermedades monogénicas. Si una fila aparece es porque hay una variante. Se vuelca en los archivos Archivo VCF que contienen las variantes respecto del genoma de referencia. Cada renglón es una variante. Este archivo contiene las divergencias con respecto al mapa de genes ideal. Luego ese archivo VCF es transformado al formato Excel [12].

Escenarios donde se indica el Exoma

El médico genetista u otro especialista realizan un diagnóstico presuntivo luego de entrevistar, revisar y realizar estudios complementarios en un paciente con una enfermedad que podría ser de causa genética. Posteriormente se solicitan estudios genéticos para confirmar esta sospecha. Cuando la sospecha es muy clara y el síndrome sospechado es provocado por uno o pocos genes, se piden paneles genéticos o estudios de un único gen o variante, que solo secuencien esa porción del ADN.

El objetivo de estos estudios es encontrar la causa genética que explique la enfermedad y de esta forma tomar medidas terapéuticas y/o preventivas, acordes al paciente y su familia. Pueden presentarse dos escenarios:

1. Enfermedades monogénicas. Cuando existen unos pocos genes que alterados producen la enfermedad no se pide el exoma, se piden los paneles que me digan los poquitos genes. Cuando se superan los 20 o 30 genes se pide un exoma.
2. Enfermedades multigénicas (multifactoriales donde si afectan los factores ambientales).

Bajo estos escenarios el genetista parte de un diagnóstico presuntivo, este diagnóstico que se presupone antes de pedir los estudios que lo confirman y para llegar al diagnóstico definitivo indica el pedido del exoma.

2.3 Trabajos relacionados

Los trabajos que se presentan posteriormente fueron encontrados en distintas publicaciones de bioinformática realizadas con el objetivo de compartir, explorar, desarrollar, implementar, integrar y analizar datos bioinformáticos de un paciente en equipos clínicos multidisciplinares. Se encontraron solo dos relacionados al ámbito de la visualización de conjuntos de múltiples tipos de datos genómicos y genéticos de un individuo de gran volumen.

Integración de [datos ómicos](#) y datos fenotípicos en un framework multimodal, extensible y unificado

El análisis de los datos "ómicos" suele ser un proceso largo y segmentado, que abarca múltiples etapas, desde la recolección inicial de datos hasta el procesamiento, el control de calidad y la visualización [13]. La naturaleza multimodal de los recientes análisis genómicos hace que este proceso sea difícil de automatizar y estandarizar; consecuentemente, los usuarios suelen recurrir a intervenciones manuales que comprometen la fiabilidad y reproducibilidad de los datos. Esto, a su vez, puede producir múltiples versiones de los conjuntos de datos en los sistemas de almacenamiento. Como resultado, los científicos pueden perder mucho tiempo y recursos intentando ejecutar y supervisar sus flujos de trabajo analíticos y encontrar dificultades para compartir los datos versionados. En 2015, el centro Ludmer de Neuroinformática y Salud Mental de la Universidad McGill formó un grupo de trabajo de genética y epigenética. Los objetivos de este grupo de trabajo fueron:

1. Diseñar un proceso automatizado para los datos epi-genéticos que consolide conjuntos de datos heterogéneos en la plataforma de datos de código abierto LORIS.
2. Agilizar el análisis de los datos.
3. Integrar los resultados con la información de procedencia.
4. Facilitar el intercambio estructurado y versionado de pipelines para una reproducibilidad optimizada utilizando entornos de computación de alto rendimiento (HPC, high-performance computing) a través del portal de procesamiento CBRAIN.

Para cumplir dichos objetivos el grupo de trabajo desarrolló el “Framework omics” (Figura 10) que como resultante fue muy beneficioso ya que tiene la habilidad de:

1. Integrar múltiples tipos de conjuntos de datos biológicos y multimodales (de imagen, clínicos, demográficos y de comportamiento).
2. Automatizar el proceso de lanzamiento de pipelines de análisis en plataformas HPC.
3. Eliminar las barreras bioinformáticas inherentes a este proceso.
4. Garantizar la estandarización y la puesta en común de los pipelines de procesamiento para mejorar la coherencia computacional.
5. Almacenar los resultados en una interfaz web consultable.
6. Ofrecer herramientas de visualización para ver mejor los datos.
7. Proporcionar los mecanismos para garantizar la usabilidad y la reproducibilidad.

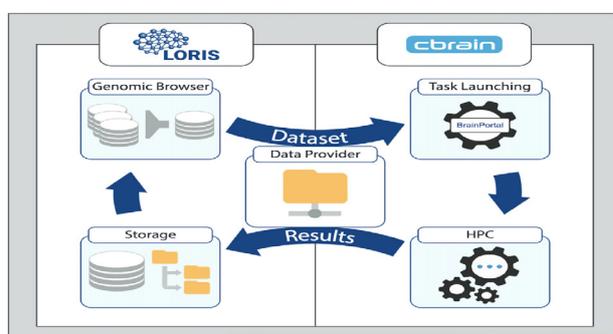


Figura 10: Ciclo de flujo de trabajo generalizado entre la plataforma de gestión de datos LORIS y la plataforma de procesamiento CBRAIN [13].

Al igual que el Framework “omics”, la solución Exoma Browser propuesta en esta tesis, integra múltiples conjuntos de datos biológicos (clínicos, demográficos y de comportamiento) y almacena los datos en una interfaz web consultable que ofrece herramientas de visualización para ver mejor los datos, aunque no utiliza entornos de computación de alto rendimiento (HPC).

Sapientia: acelerar el diagnóstico de las enfermedades poco frecuentes

Congenica (Cambridge, Reino Unido) es un líder mundial en el desarrollo de tecnologías de descubrimiento y diagnóstico basadas en el genoma. La plataforma tecnológica Sapientia de Congenica [14] permite que el análisis de la secuencia del genoma completo sea fácilmente interpretado y presentado dentro de un informe de diagnóstico clínicamente procesable. Dicha plataforma fue desarrollada por científicos que trabajan en el Proyecto del Genoma Humano y en el proyecto “Developing Developmental Diseases”. La plataforma de análisis genómico clínico Sapientia, permite el procesamiento de conjuntos de datos secuencia de ADN humano con su correspondiente fenotipo clínico para proporcionar un diagnóstico molecular definitivo y una interpretación para el tratamiento, la estratificación de los pacientes y mejorar los descubrimientos terapéuticos.

Los médicos pueden ver sólo uno o dos casos de una enfermedad poco frecuente en su carrera, y no todos los pacientes con una enfermedad concreta mostrarán los mismos síntomas. Con Sapientia pueden tener acceso a una base de conocimientos de referencia que contiene otros pacientes con síntomas similares con los que comparar al paciente aceleraría enormemente el diagnóstico para acelerar enormemente los tiempos de diagnóstico. Sapientia está construida en torno a herramientas y nomenclatura estandarizadas con lo cual permite facilitar la búsqueda de datos y las comparaciones. Los conocimientos que están surgiendo del estudio de las enfermedades poco frecuentes están haciendo avanzar significativamente la comprensión de las causas y tienen el potencial de proporcionar información sobre importantes vías controladas por los genes implicados. Algunas características de Sapientia son:

- Permitir la creación de paneles virtuales de genes, listas curadas de genes para su inclusión o exclusión y priorización de listas de genes para conjuntos de datos más grandes, como exomas o genomas.

- Permitir acceder a anotaciones de variantes curadas manualmente por expertos clínicos.
- Los usuarios también pueden integrar sus propios datos heredados para mejorar sus capacidades de diagnóstico.
- El fenotipo asociado puede describirse con un selector de términos del Proyecto de Ontología del Fenotipo Humano incorporado.
- Acelerar la interpretación clínica de las variantes mediante la integración de herramientas online como OMIM, PubMed y herramientas de búsqueda con las principales bases de datos de variantes, en un único entorno de trabajo.

Al igual que Sapiencia, la solución Exoma Browser propuesta en esta tesis, muestra un panel virtual de genes para su análisis y posterior diagnóstico y el médico genetista puede integrar sus propios datos para mejorar las capacidades de diagnóstico, aunque o posee integración con herramientas online como OMIM u otras.

2.4 Herramientas de asistencia al genetista

Se presentan a continuación herramientas disponibles para asistir a los médicos genetistas en el proceso de análisis e interpretación de variantes genéticas en el exoma humano.

Dichas herramientas fueron seleccionadas de sitios de publicación de artículos académicos sobre genética molecular y medicina genómica, explorando aquellas útiles para la gestión y evaluación de variantes genéticas y el diagnóstico de enfermedades

Se han elegido para la presentación GENIO, VCF-Server, VCF-Explorer y ClinLabGeneticist por ser las más utilizadas en los hospitales y/o laboratorios clínicos.

Herramienta GENIO

La herramienta paga GenIO [15] (Figura 11) ha sido diseñada para ayudar a los investigadores clínicos de genómica y a médicos genetistas en el proceso de diagnóstico de enfermedades de causa genética. La herramienta identifica las variantes más probables que

causan una enfermedad utilizando información genómica y clínica. Las variantes identificadas en un genoma completo y la de los estudios de secuenciación de exoma completo son anotados, clasificados y filtrados por significancia clínica. Se identifican los genes candidatos asociados con los síntomas del paciente, las sospechas de enfermedad y con hallazgos complementarios para obtener las variantes de genes candidatas asociadas a enfermedades recesivas o dominantes. Adicionalmente, siguiendo las normas y lineamientos del “American College of Medical Genetics and Genomics” y de la Association of Molecular Pathology”, son identificadas todas las variantes potencialmente patogénicas que pueden contribuir a la enfermedad y a los hallazgos secundarios.

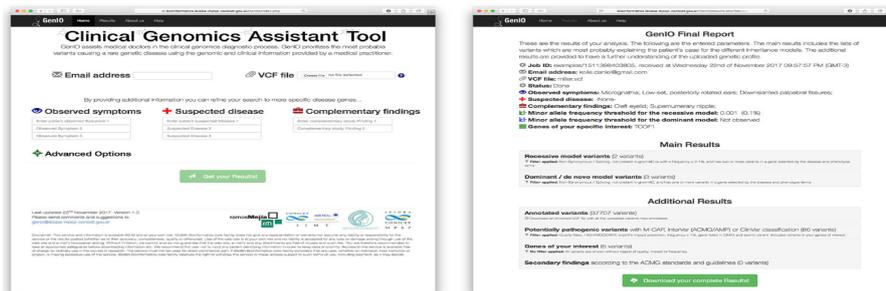


Figura 11: Gen IO, Interfaz de Usuario [15]

GenIO es una herramienta que integra la información clínica del paciente, los síntomas, la enfermedad sospechada y los hallazgos complementarios. Luego en su pipeline de procesamiento consulta a las bases de datos de variantes genéticas generando como salida un reporte. Esta herramienta para asistir a los médicos genetistas se encuentra libremente disponible en: <https://bioinformatics.ibioba-mpsp-conicet.gov.ar/GenIO/> [16].

Herramienta VCF-Server

El VCF-Server [17] fue desarrollado como una herramienta de visualización web para apoyar el análisis interactivo de datos de variantes genéticas. Permite a los investigadores y médicos genetistas gestionar, anotar, filtrar, consultar y exportar variantes de forma rápida y eficaz. Una vez que se carga el archivo VCF, VCF-Server permite a los usuarios anotar el VCF con bases de datos de uso común o anotaciones de variantes definidas por el usuario.

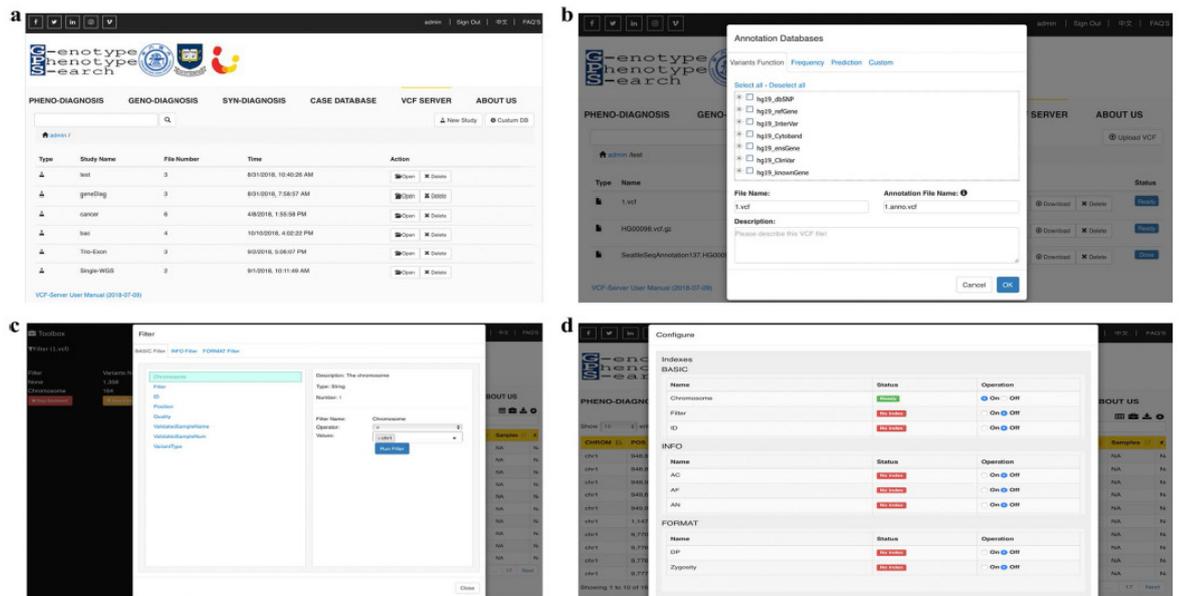


FIGURE 3 Screenshots of different functional modules on VCF-Server. (a) VCF file management online. (b) Annotating VCF with commonly used annotation databases. (c) Variant filtering and visualization. (d) VCF index management

Figura 12: VCF-Server, Módulos funcionales [17]

La información de variantes en el VCF se muestra visualmente a través de una interfaz gráfica interactiva (Figura 12). Los usuarios pueden filtrar las variantes con reglas de filtrado flexible y las variantes priorizadas se pueden exportar localmente para su posterior análisis. Esta herramienta para asistir a los médicos genetistas se encuentra libremente disponible en: <https://www.diseasegps.org/VCF-Server?lan=eng> [18].

Herramienta VCF-Explorer

Es una herramienta de software Open Source diseñada para realizar análisis de grandes archivos VCF. La interfaz gráfica de usuario (Figuras 13, 14 y 15) facilita el filtrado de variantes con respecto a las anotaciones a nivel de muestra o de variante [19].

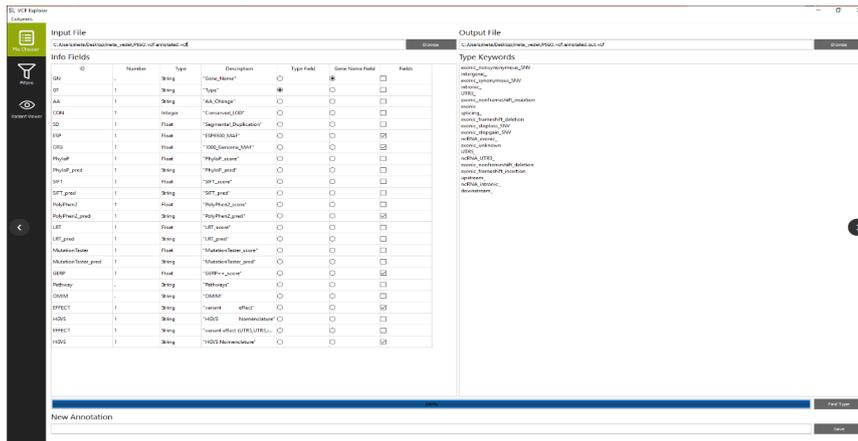


Figura 13: VCF Explorer, Carga de archivos [19]

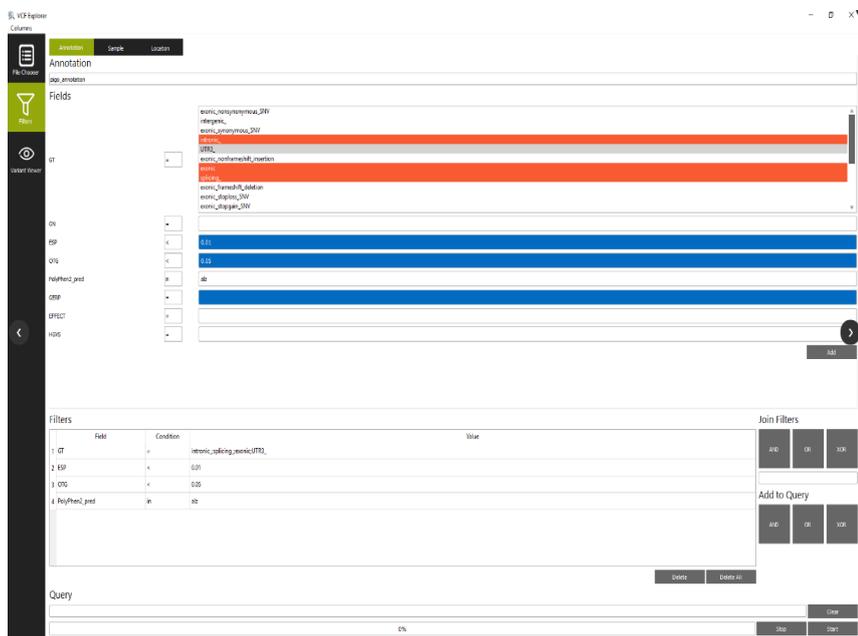


Figura 14: VCF Explorer, Filtrado y anotación de variantes [19]

Variants

Chr	POS	ID	REF	ALT	QUAL	FILTER	GN	GT	AA	CON	SD	ESP	OTG	PloidyP
5	61745748	-	TT	T	0	-	IPD11	intronic_	-	529	0	0	0	0
5	64072617	-	A	AAATTATGAC	0	-	CVIC27	intronic_	-	657	0	0	0.59	0
5	79257859	rs405482	G	A	0	-	TRES4	intronic_	-	471	0	0	0.46	0
5	90052268	-	C	CC	0	-	GPR99	intronic_	-	397	0	0	0	0
5	96116124	-	T	TT	0	-	ERAF1	intronic_	-	534	0	0	0	0
5	96332990	-	CC	C	0	-	LNP1P	intronic_	-	662	0	0	0	0
5	100203463	-	TT	T	0	-	FER	intronic_	-	0	0	0	0	0
5	110794822	rs2116824	A	G	0	-	CANM4	intronic_	-	0	0	0.699708	0.64	0
5	137454338	rs6876375	C	G	0	-	NMES	intronic_	-	0	0	0	0.96	0
5	140992259	rs200166308	G	A	0	-	ADNREP37MM_D01...	splicing	-	371	0	0.004486	0.0014	0
5	173511992	rs62389230	C	T	0	-	FAM1538	intronic_	-	0	0.99	0	0.5	0
5	176729746	-	OCC	C	0	-	RAB24	intronic_	-	601	0	0	0	0
5	176729747	-	CC	C	0	-	RAB24	intronic_	-	601	0	0	0	0
6	7368977	rs1324502	A	G	0	-	CAGE1	intronic_	-	0	0	0.356423	0.36	0
6	11192375	rs1465131	C	T	0	-	NEDD9	intronic_	-	401	0	0.049008	0.9	0
6	24174955	rs9350755	G	C	0	-	DCDC2	UTR3	-	0	0	0.702906	0.76	0
6	3040644	-	TT	T	0	-	MARK14	intronic_	-	627	0	0	0	0
6	36979745	rs1724005	G	A	0	-	FOD2	intronic_	-	0	0	0.29694	0.38	0
6	42942779	rs9986447	A	G	0	-	PEX3	intronic_	-	355	0	0.376672	0.3	0
6	43416840	rs201244769	T	C	0	-	ABCC10	intronic_	-	0	0	0.003229	0	0
6	44102000	rs11967633	G	A	0	-	TMM63B	intronic_	-	0	0	0.126895	0.24	0
6	44132421	-	TCTT	T	0	-	TMM63B	intronic_	-	0	0	0	0	0
6	46995901	rs76753338	T	C	0	-	GPR110	intronic_	-	313	0	0.00231	0.03	0
6	52274821	-	CAC	C	0	-	TRAM2	intronic_	-	472	0	0	0	0

Samples

Sample	Genotype	Genotype Score	Coverage
454	0/0	9	3
455	1/1	3	1
456	0/1	99	6
457	0/1	35	2

Number of Variant Positions: 584

C:\Asesmete\Desktop\met_vedek\PI902_vcf.annotated.out.vcf

Figura 15: VCF Explorer, Visualización de Variantes [19]

VCF-Explorer procesa el archivo, línea por línea, para cada consulta. Este enfoque elimina el requisito de un gran uso de memoria o el manejo de una estructura de base de datos. Su versión 1.0 como aplicación de escritorio, se encuentra libremente disponible: <http://vcfexplorer.sourceforge.net/> [20].

Herramienta ClinLabGeneticist

La herramienta ClinLabGeneticist [21] (Figura 16) implementa un workflow para la gestión y evaluación de variantes en pruebas genéticas y diagnóstico de enfermedades. Establece un data source de anotaciones de variantes para la identificación de variantes patogénicas. Despliega un panel de control para ayudar al proceso jerárquico de revisión, de varios pasos, que conduce a las decisiones clínicas finales sobre la evaluación de variantes genéticas y posee una base de datos central para archivar todos los datos de las pruebas genéticas, las notas y los comentarios a lo largo del proceso de revisión, los datos de validación de variantes mediante secuenciación, así como los informes clínicos finales para futuras consultas.

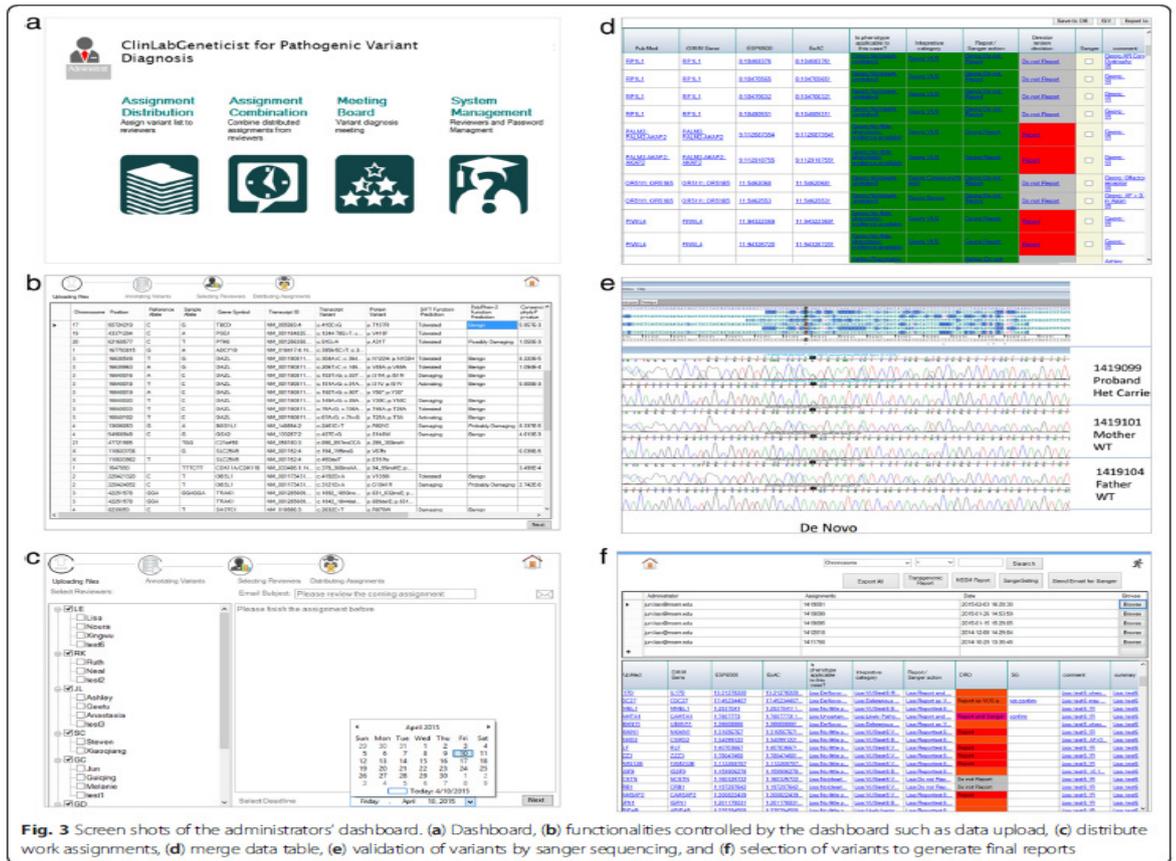


Figura 16: ClinLabGeneticist, Workflow [21]

Todo el flujo de trabajo, incluyendo la entrada de datos, la distribución de las asignaciones de trabajo, la evaluación y revisión de variantes, la selección de variantes para la validación, la generación de informes y las comunicaciones entre el personal, está integrado en una única herramienta de gestión. Está disponible gratuitamente para el ámbito académico en <http://rongchenlab.org/software/clinlabgeneticist> [22].

Conclusiones acerca de las herramientas relacionadas

Las principales falencias detectadas en las herramientas analizadas son las siguientes:

En primer lugar, obligan al usuario al proceso de anotación de variantes y no contemplan el registro del proceso completo (interacción con el paciente, generación de un diagnóstico presuntivo, generación de resultados). Además, algunas de ellas trabajan con archivos sin datos unificados que permitan retroalimentar diagnósticos futuros.

Los reportes generados por dichas herramientas carecen de una trazabilidad que permita visualizar las razones para las conclusiones inferidas sin permitir consultar un historial de interacciones. A su vez carecen de una visualización representativa que funcione como aporte significativo al profesional consultante. Todas carecen de panel bioinformático propio.

2.5 Introducción al concepto de diseño centrado en el usuario, escenarios y prototipado

Los escenarios de uso capturan explícitamente lo que experimentan los usuarios a medida que avanzan hacia el logro de una solución ideal y ayudan a refinar cualquier requisito en el futuro. Los escenarios muestran caminos tomados para completar tareas, procesos de decisión, puntos de fricción en la ejecución diaria y su experiencia. Debe observarse cómo son llevadas a cabo las soluciones en orden de poder realizar una especificación de los requerimientos y situaciones de uso y actividades de un usuario [23].

La ingeniería de requerimientos y el diseño basados en escenarios de uso son una práctica reconocida en la Ingeniería de Software. Los escenarios de uso se pueden utilizar para evaluar cómo actúan los usuarios en un sistema o entorno determinado, visualizar aspectos y dar solución observando sus contextos de uso, actividades, necesidades, experiencias, comportamientos y objetivos. Dentro del diseño basado en escenarios hay factores que influyen: antecedentes, motivaciones, tareas, contexto de uso, entorno, retos, los escenarios son concisos y representan una instantánea de la experiencia del usuario [24].

Los escenarios son historias sobre las personas y sus actividades [25]. Por ejemplo, un contable desea abrir una carpeta en el escritorio del sistema para acceder a una nota sobre presupuestos. Sin embargo, la carpeta está tapada por una hoja de cálculo de presupuestos que el contable desea consultar mientras lee la nota. La hoja de cálculo es tan grande que casi llena la pantalla. El contable hace una pausa de varios segundos, cambia el tamaño de la hoja de cálculo, la aparta parcialmente de la pantalla, abre la carpeta, abre la nota, cambia el tamaño y la posición de la nota, y sigue trabajando. Este es escenario de trabajo específica la gestión de ventanas y la funcionalidad de cambio de aplicaciones de forma vívida y precisa: la

gente necesita coordinar las fuentes de información para comparar, copiar e integrar los datos de múltiples aplicaciones; las pantallas inevitablemente se desordenan; la gente necesita encontrar y reordenar las ventanas en la pantalla.

Los escenarios ponen de relieve los objetivos sugeridos por la apariencia y el comportamiento del sistema, lo que las personas intentan hacer con el sistema, los procedimientos que se adoptan, los que no se adoptan, los que se llevan a cabo con éxito o de forma errónea y las interpretaciones que las personas hacen de lo que les ocurre. Los escenarios tienen elementos característicos. Mencionan o presuponen una configuración. Los escenarios incluyen agentes o actores, cada uno de los cuales suele tener metas u objetivos. Todo escenario incluye al menos un agente y al menos un objetivo. Cuando hay más de un agente o un objetivo implicado, pueden tener un protagonismo diferente en el escenario.

Los escenarios tienen una trama, incluyen secuencias de acciones y acontecimientos, cosas que hacen los agentes, cosas que les ocurren, cambios en las circunstancias del escenario, etc. Las acciones y eventos particulares pueden facilitar, obstruir o ser irrelevantes para determinados objetivos. Representar el uso de un sistema o una aplicación a través de escenarios con un conjunto de interacciones hace que ese uso sea explícito y, al hacerlo, orienta el diseño y el análisis hacia una visión más amplia de las soluciones. Pueden ayudar a los diseñadores y analistas a centrar la atención en los supuestos sobre las personas y sus tareas que están implícitos en los sistemas y aplicaciones.

Las representaciones de escenarios pueden elaborarse como prototipos, mediante el uso de herramientas de storyboard, vídeo y prototipado rápido. Son los contextos mínimos para desarrollar una lógica de diseño orientada al uso: una determinada decisión de diseño puede evaluarse y documentarse en función de sus consecuencias específicas dentro de determinados escenarios. Así, los escenarios pueden proporcionar un marco para la interacción persona-ordenador basada en el diseño.

En la Ingeniería de Software el prototipado se considera esencial para los procesos de diseño centrados en el usuario. Los prototipos hacen que las ideas abstractas cobren vida, y

son una forma de involucrar al usuario final en la evaluación de las ideas de diseño. Ayudan a probar los requerimientos con el usuario, a fomentar la reflexión sobre el diseño y la elección de alternativas, y a realizar evaluaciones de usabilidad. En el contexto del diseño iterativo, los prototipos se consideran en general representaciones o implementaciones parciales y gradualmente refinadas del sistema interactivo que se va a producir. La visión de los prototipos orientada al producto debe ser complementada con una segunda perspectiva que pone más énfasis en el hecho de que los artefactos, las personas y las prácticas co-evolucionan [26]. A modo de ejemplo se presenta el prototipo de un tablero interactivo multi-táctil utilizado en reuniones para modelar UML. El prototipo está diseñado para permitir la interacción simultánea en un diagrama compartido en un grupo de trabajo. Los miembros del equipo pueden trabajar en copias temporales de los elementos del diagrama en su espacio personal. Las modificaciones son inmediatamente visibles en el espacio del grupo. La esa "propiedad" temporal se indica con marcos de color alrededor de los elementos seleccionados. En este ejemplo se pide a los diseñadores que utilicen una nueva tecnología (multitouch tabletops) para una herramienta de modelado colaborativo como se muestra en la Figura 17.

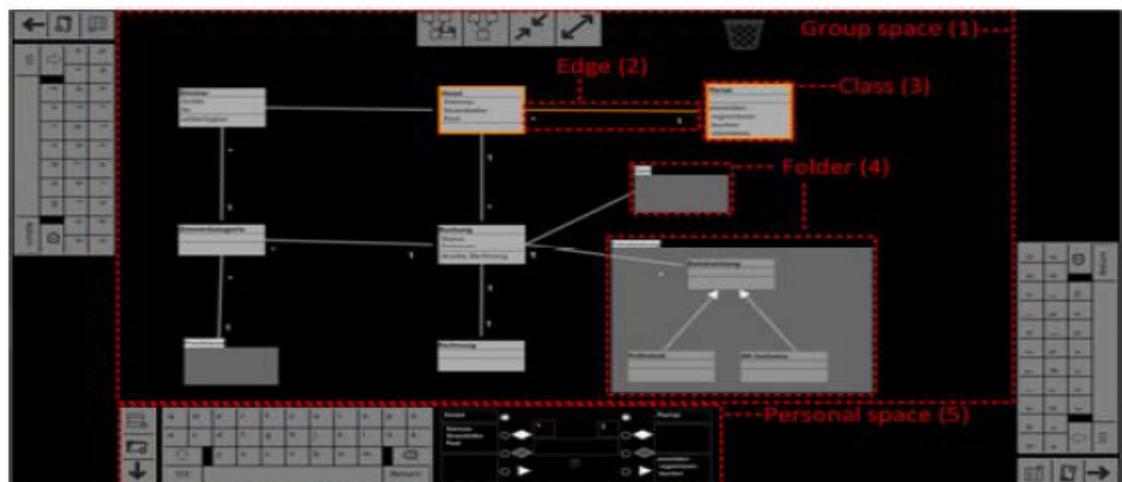


Figura 17: Prototipo de un editor de clases multitáctil [26]

Para realizar el diseño de la aplicación se seguirá la metodología de diseño centrado en el usuario. El objetivo del DCU es aumentar la usabilidad (el grado en que un usuario encuentra fácil navegar por una interfaz). El término "usabilidad" también puede referirse a los métodos que pueden utilizarse para mejorar el diseño de la interfaz de una aplicación [40]. Para cumplir dicho objetivo se realizó el modelado a través escenarios de uso y a través de prototipos como herramienta que muestra el resultado del modelo.

2.6 Introducción al concepto de usabilidad y métodos de evaluación de usabilidad

A efectos de asegurar que la herramienta tenga un nivel alto de usabilidad y que el genetista pueda realizar su tarea de manera segura, efectiva y eficiente y así cumplir con el objetivo de hacer más ágil el proceso de diagnóstico se definen las distintas pruebas de usabilidad que pueden realizarse tales como revisiones de expertos y pruebas de aceptación.

El creciente interés por la interacción persona-ordenador se debe a el deseo de los diseñadores de interfaz de usuario de mejorar la experiencia de los usuarios. Esta estrecha interacción con la comunidad de usuarios conduce a una serie de tareas de referencia que constituyen la base de los objetivos y las medidas de usabilidad. Para cada tipo de usuario y cada tarea, los objetivos medibles precisos guían al diseñador a través del proceso de pruebas de usabilidad. La norma ISO 9241 sobre Ergonomía de la interacción entre el ser humano y el sistema (ISO, 2013) se centra en objetivos deseables: eficacia, eficiencia y satisfacción [27]. Las medidas de usabilidad, revisiones de expertos y pruebas de aceptación, que se presentan a continuación se centran en los dos últimos objetivos y conducen más directamente a la evaluación práctica.

Revisiones de expertos

Las revisiones de expertos dependen de la disponibilidad de expertos (cuya experiencia puede ser en usabilidad o en interfaz de usuario) como consultores que recorren las funciones clave de la interfaz con un enfoque disciplinado. Las revisiones de expertos pueden realizarse al principio o al final de la fase de diseño. El resultado puede ser un informe formal con problemas identificados o recomendaciones de cambios. Los revisores pueden anotar posibles problemas para discutirlos con los diseñadores, pero el desarrollo de las soluciones debe dejarse en manos de éstos [27].

Las revisiones de expertos suelen durar entre medio día y una semana, aunque puede ser necesario un largo periodo de formación para explicar el ámbito de la tarea o los

procedimientos operativos. Hay una gran variedad de métodos de revisión de expertos entre los que elegir [27].

Pruebas de aceptación

Las pruebas de aceptación son un que garantizan la calidad del producto y determinan hasta qué punto una aplicación cumple con la aprobación de los usuarios finales. Pueden establecerse criterios medibles para la interfaz de usuario como por ejemplo los siguientes [27]:

- Tiempo de aprendizaje de funciones específicas por parte de los usuarios.
- Velocidad de ejecución de la tarea.
- Tasa de errores de los usuarios.
- Retención de comandos por parte de los usuarios a lo largo del tiempo.
- Satisfacción subjetiva del usuario.

Otros criterios, como la satisfacción subjetiva del usuario, la comprensibilidad de los resultados, el tiempo de respuesta del sistema, la documentación o el atractivo gráfico, también pueden considerarse en las pruebas de aceptación de productos. El objetivo central de las pruebas de aceptación no es detectar fallos, sino verificar el cumplimiento de los requisitos.

El objetivo de las revisiones por parte de expertos y de las pruebas de aceptación es forzar la mayor parte posible del desarrollo evolutivo en la fase previa al lanzamiento, cuando el cambio es relativamente fácil y menos costoso de llevar a cabo [27].

Capítulo 3: Descripción del Problema

El presente capítulo describe la problemática actual del proceso de análisis e interpretación del exoma llevado a cabo por el genetista. Luego de obtener el archivo Excel con las variantes de la muestra con respecto al genoma de referencia el, genetista comienza la labor de análisis e interpretación. Debido a que el archivo Excel cuenta con 4 solapas, cada una con una tabla de 29 columnas y además cada tabla puede tener hasta 50 mil filas, el proceso de filtrado se hace engorroso y difícil de ejecutar. No están centralizados los exomas de todos los pacientes y no hay trazabilidad entre la historia clínica del paciente, su exoma y los informes de resultados.

3.1 Secuenciación del Exoma

El proceso de secuenciación que lleva al reporte de las variantes del exoma implicadas en posibles patologías tiene varias etapas (Figura 18). Se inicia con una toma de muestra de sangre, a partir de la cual se extrae el ADN y se envía a secuenciar el exoma utilizando técnica de NGS (Next Generation Sequencing). El resultado de la secuenciación es una serie de archivos de texto plano y archivos de QA utilizados para descartar información por debajo del umbral de calidad esperado.

Los archivos planos son pre-procesados por un pipeline bioinformático y transformados a un archivo estándar que contiene las variantes de la muestra respecto al genoma de referencia. Este archivo será procesado por el equipo de bioinformática para generar una planilla Excel que presentará las variantes halladas en los genes relacionados con el diagnóstico presuntivo, las variantes patogénicas y probablemente patogénicas en todos los genes, las variantes en los 73 genes de ACMG - American College of Medical Genetics and Genomics [28] y todas las variantes de todos los genes. La información sobre los 73 genes de ACMG se ofrece de forma optativa a los pacientes, ya que, aunque las variantes no estén relacionadas con el diagnóstico presuntivo tienen una accionabilidad definida y aceptada por la comunidad médica.

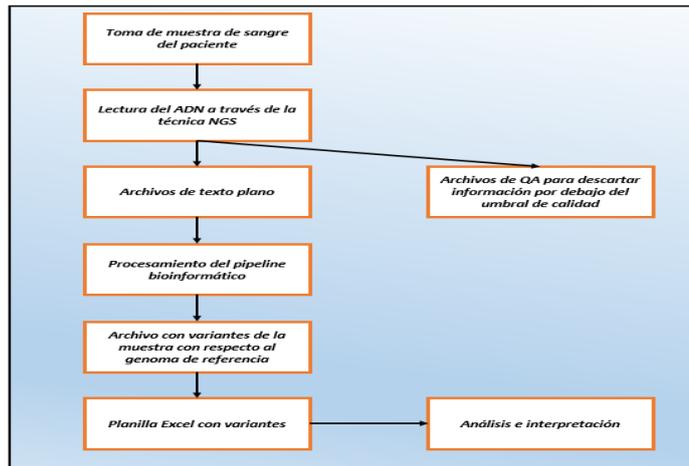


Figura 18: Proceso de secuenciación, análisis e interpretación del exoma

En todos los casos las variantes presentan:

Campos Descriptivos y en qué [cigosis](#) se encuentra la variante:

- Ubicación genómica.
- Gen dentro del cual está presente la variante.

Variante en [heterocigosis](#) u [homocigosis](#) (si las dos copias del gen, paterna y materna, tienen la variante, o sólo una de ellas).

Campos Anotados:

- Frecuencia de esa variante en bases de datos de personas sanas (si muchos individuos sanos la tienen, probablemente no sea causa de enfermedad, y viceversa).
- Predicción de la patogenicidad según algoritmos bioinformáticos (consideran comparación con otras especies y tipo de efecto que tiene la variante sobre la proteína codificada por el gen).
- Significancia clínica: Pathogenic, Likely Pathogenic, VUS (Variant of uncertain significance), Benign, Likely Benign. Existen bases de datos con este registro. Ejemplo Clinvar (Base de datos de afirmaciones sobre la importancia clínica y la relación fenotípica de la variación humana [29]) o HGMD (Human Gene Mutation Database [30])

- Alteración biológica que produce: interrupción de la síntesis de la proteína, introducción de otro aminoácido, perder o ganar algún fragmento de gen, modificar su procesamiento, entre otras.
- Localización de la variante dentro del gen (algunas regiones del gen pueden ser más “afectables” por variantes genéticas y provocar patología, y otras menos), presencia o ausencia de otras variantes de similar efecto en las cercanías de la variante.

3.2 Detalle del archivo Excel

El archivo Excel que contiene las variantes de la muestra respecto al genoma de referencia y que se presenta como caso de estudio fue organizado de la siguiente manera:

Solapas

- Todo: Contiene todas las variantes presentes en los genes capturados por el experimento del exoma.
- Variantes genes fenotipo: Filtra los datos de la solapa Todo por genes asociados con el diagnóstico presuntivo (fenotipos patológicos), seleccionados manualmente tras la valoración del caso por parte del genetista.
- Variantes Pathogenic y L. Pathogenic: Contiene las variantes patogénicas y probablemente patogénicas en todos los genes capturados por el experimento.
- Variantes genes Incidental Findings: Contiene las variantes en los 73 genes de ACMG. Son genes que, en caso de encontrar variantes en ellos, aunque no estén relacionadas con la patología en estudio se pueden informar, si el paciente así lo desea. Son hallazgos más allá de los genes del fenotipo específico. Presenta alteraciones que podrían tener un impacto accionable para el paciente. En esta solapa sólo se informan las variantes patogénicas claras, y no las que sean dudosas mientras que en las otras solapas puede a veces haber alguna dudosa.

Columnas

El exoma cuenta con 29 columnas, 12 representativas de los metadatos identificatorios de la variante, 13 que representan [impacto funcional](#), [frecuencia poblacional](#), [predictores bioinformáticos](#) y [significancia clínica](#) y 4 de llenado manual por el médico genetista. A continuación, en la Tabla 1 se detallan cada una de ellas.

Denominación de la columna	Significado	
CHROM	Cromosoma dentro del cual está presente la variante	Metadatos identificatorios de la variante
POS	Ubicación genómica	
REF	Alelo (valor) de referencia	
ALT	Alelos diferentes al de referencia (en general 1 pero puede haber más)	
GEN[0].GT	Cigosis. Variante en heterocigosis u homocigosis (si las dos copias del gen, paterna y materna, tienen la variante, o sólo una de ellas)	
GEN[0].AD	Profundidad de lectura	
GEN[0].DP	Reads (fragmentos) leídos en esa posición se distinguen en cuántos eran de un alelo y cuántos del otro.	
ID	Identificador de las variante	
ANN[*].GENE	Gen en el que está presente la variante. Si bien no es común, podría ocurrir que la variante esté presente en más de un gen en los casos donde más de un gen tienen regiones superpuestas	
ANN[*].HGVS_C	Variante en ADN	
ANN[*].HGVS_P	Variante en proteína	
ANN[*].FEATUREID	Transcriptos. Cada gen tiene distintos bloques (exones). Un transcripto es una forma particular de combinar esos bloques. Serán todos los transcriptos que incluyan el exon (bloque) que contiene esa posición	
ANN[*].EFFECT	Efecto funcional de la variante	
ANN[*].IMPACT	Impacto funcional en la variante (high, moderate, low)	
dbNSFP_ExAC_AF	Frecuencia poblacional de la variante en bases de datos de personas sanas	
AN_Adj		
AC_Adj		
ExAC_AF		
dbNSFP_Polyphen2_HDIV_pred	Predictores bioinformáticos de patogenicidad. A través de ellos podemos saber si es probable que esta variante/cambio va a generar una patología o no	
dbNSFP_Polyphen2_HVAR_pred		
dbNSFP_SIFT_pred		
dbNSFP_MutationTaster_pred		
dbNSFP_MutationAssessor_pred		
CLNSIG	Significancia/información clínica de la variante	Datos relacionados al diagnóstico
CLNSIGCONF		
Terminos HPO - Enfermedad	Tipificación del fenotipo en términos HPO y/o denominación del fenotipo	
Patrón de Herencia	Si se sospecha un patrón recesivo se priorizará el hallazgo de dos variantes patogénicas en el mismo gen, o bien una sola en estado homocigota (en las dos copias del gen, la materna y paterna). Si se sospecha un patrón dominante, se espera encontrar un resultado heterocigota	
Relación con el Fenotipo	Si los genes guardan relación con el fenotipo o no	
Es Variante Candidata	Si una variante es candidata a ser patológica o no	

Tabla 1: Descripción de las columnas representativas del Exoma

Representación de los transcritos

Una variante puede nombrarse distinto según el [transcripto](#) del gen al cual nos estemos refiriendo con efectos distintos sobre la proteína y la función resultante. Es por esto que en las columnas ANN[*].GENE, ANN[*].HGVS_C, ANN[*].HGVS_P, ANN[*].FEATUREID, ANN[*].EFFECT, ANN[*].IMPACT, dbNSFP_Polyphen2_HDIV_pred, dbNSFP_Polyphen2_HVAR_pred, dbNSFP_SIFT_pred, dbNSFP_MutationTaster_pred y dbNSFP_MutationAssessor_pred se presenta como contenido una cadena de caracteres alfanuméricos separados con comas en representación de las distintas posiciones a las que refiere, cada una respectiva a la otra, a un determinado transcripto de interés. Por ejemplo, en un extracto de exoma indicamos los transcritos de posición 1 y 2 respectivamente según se muestra en las tablas 2, 3 y 4 respectivamente:

AN N[*] .G ENE	AN N[*] .H GV S_C	AN N[*] .H GV S_P	AN N[*] .FE AT UR EID	AN N[*] .EF FEC T	AN N[*] .I MP ACT	db NSF P_E xAC _AF	AN _Ad j	AC_ Adj	ExA C_A F	db NSF P_P oly phe n2_ HDI V_p red	db NSF P_P oly phe n2_ HV AR_ pre d	db NSF P_S IFT _pr ed	db NSF P_ Mu tati onT ast er_ pre d	dbNSFP _Mutat ionAsse ssor_pr ed
SA MD 11, SA MD 11	n.- 984 A> G,n .- 104 7A> G	p.GI y48 0GI y,p. Gly 387 Gly, p.GI y30 6GI	ENS T00 000 466 827 ,EN ST0 000 046 494	ups tre am _ge ne_ vari ant, ups tre am	MO DIFI ER, MO DIFI ER, MO DIFI ER		584 38	542 05	0.9 275 642 56	B B	T	P	N	

		y	8	_ge ne_ vari ant										
--	--	---	---	---------------------------	--	--	--	--	--	--	--	--	--	--

Tabla 2: Extracto de una fila del exoma que posee 2 posiciones del transcripto de un gen

AN N[*]].G ENE	AN N[*]].H GV S_C	AN N[*]].H GV S_P	AN N[*]].FE AT UR EID	AN N[*]].EF FEC T	AN N[*]].I MP AC T	db NSF P_E xAC _AF	AN _Ad j	AC_ Adj	ExA C_A F	db NSF P_P oly phe n2_ HDI V_p red	db NSF P_P oly phe n2_ HV AR_ pre d	db NSF P_S IFT _pr ed	db NSF P_ Mu tati onT ast er_ pre d	dbNSFP _Mutat ionAsse ssor_pr ed
SA MD 11	n.- 984 A> G	p.Gl y48 OGl y	ENS T00 000 466 827	ups trea m_g ene _va rian t	MO DIF IER		584 38	542 05	0.9 275 642 56	B B T P N	B B T P N	T P N	P N	N

Tabla 3: Transcripto de posición 1 dentro del extracto

AN N[*]].G ENE	AN N[*]].H GV S_C	AN N[*]].H GV S_P	AN N[*]].FE AT UR EID	AN N[*]].EF FEC T	AN N[*]].I MP AC T	db NSF P_E xAC _AF	AN _Ad j	AC_ Adj	ExA C_A F	db NSF P_P oly phe n2_ HDI V_p red	db NSF P_P oly phe n2_ HV AR_ pre	db NSF P_S IFT _pr ed	db NSF P_ Mu tati onT ast er_ pre	dbNSFP _Mutat ionAsse ssor_pr ed
--------------------------	--------------------------------	--------------------------------	---------------------------------------	--------------------------------	------------------------------------	--------------------------------	----------------	------------	-----------------	--	---	--------------------------------------	---	--

												d		d	
SA	,n.-	p.GI	ENS	ups	MO										
MD	104	y30	T00	trea	DIF										
11	7A>	6GI	000	m_g	IER										
	G	y	464	ene											
			948	_va											
				rian											
				t,up											
				stre											
				am											
				_ge											
				ne_											
				vari											
				ant											

Tabla 4: Transcrito de posición 2 dentro del extracto

3.3 Análisis e Interpretación

La planilla Excel es interpretada por un grupo interdisciplinario compuesto por profesionales de bioinformática y de genética clínica. El objetivo de la interpretación es encontrar variantes relacionadas con el fenotipo que presenta el paciente, u otras que sean relevantes para su salud.

La primera búsqueda se realiza en bases de datos como Clinvar[19] o HGMD[20] con el objetivo de evaluar si el paciente presenta alguna de las variantes registradas como patogénica para la patología en sospecha. Si se encontrara, se evalúan si hay variantes en genes de los 73 de “incidental findings” que deben ser informados y allí termina la interpretación. De todas formas, se miran las otras variantes en los genes relacionados con el fenotipo y las otras descritas como patogénicas, pero es poco probable que una persona tenga dos patologías genéticas al mismo tiempo. En los casos más complejos, las variantes se analizan en diferentes ejes (descritos en la sección de “ejes de información”) que deberán ser valorados en su conjunto para tomar una decisión.

Finalmente, se complementa la información del Excel con consultas a diferentes bases de datos bioinformáticas.

Ejes de información utilizada durante el proceso de interpretación

Para la interpretación se integra información conocida sobre las variantes e información relevada sobre el paciente y su historia familiar. Realizando un proceso de abstracción, podemos definir los siguientes ejes de información. Analizada cada variante integralmente en todos los ejes, podrá ser filtrada y caracterizada en el contexto del paciente en estudio.

- **Relación de la variante con los fenotipos del paciente.**

Los fenotipos representan las características clínicas, conductuales, de comportamiento del paciente. Por ejemplo, discapacidad intelectual. [HPO](#) (Human Phenotype Ontology) es una ontología estándar para describir fenotipos. Existen a su vez sistemas como Phenomizer que permiten priorizar genes (e indirectamente variantes en esos genes) de acuerdo con el fenotipo que el médico observa.

- **Frecuencia poblacional de la variante.**

El filtro que se hace por frecuencia en población sana considera la frecuencia conocida que tiene la enfermedad; si buscamos por ejemplo una enfermedad de 1/10.000 personas, ninguna variante patogénica en ese gen podría tener una frecuencia que sea mucho mayor que esa (por lo tanto, las descartamos como causales de la enfermedad). Cuando la frecuencia poblacional de una variante excede el 5% de la población, es posible descartarla como causal de una patología monogénica, precisamente porque excede un umbral de la frecuencia de las enfermedades genéticas conocidas en cualquier población (hay algunas excepciones muy precisas y que se analizan puntualmente, como las hemoglobinopatías en etnias africanas).

- **Consecuencia biológica por predicción bioinformática.**

Se utilizan herramientas como SIFT [31], Polyphen-2 [32], CADD [33], Mutation Tester [34] que predigan la consecuencia biológica (en la proteína) de la mutación. También se tiene en cuenta la ubicación de la variante en el gen para determinar cuánto puede

afectar una variante allí (Uniprot [35], String [36]). La interpretación finalmente considera la información en todos los ejes para su clasificación.

Capítulo 4: Modelado del proceso de toma de decisiones, planteo de escenarios.

En este capítulo se presenta el modelado del proceso de toma de decisiones, llevado a cabo por el genetista, para arribar al diagnóstico final. Para modelar el proceso se realizaron diversas entrevistas al Dr. Sebastián Menazzi, médico de planta de la División Genética del Hospital de Clínicas y Director Médico de la empresa Zoigen & genomIT y al Mg. Matias Butti, especialista en ciencias de la computación y en bioinformática y Director de la empresa Zoigen & genomIT en las que fueron transmitiendo su trabajo, caminos tomados, puntos de ejecución y su experiencia y, a través de la observación y evaluación de las actividades, necesidades, características y objetivos del trabajo realizado por el para arribar al diagnóstico de enfermedades monogénicas. Se realizaron un total de 6 entrevistas, 5 en forma remota y una en forma presencial. A continuación, se hace un resumen de los temas tratados y el material entregado en cada una de ellas:

1. Presentación de conceptos fundamentales de genética, exoma su definición y sus solapas y columnas, escenarios donde se indica. Proceso de secuenciación del exoma. Proceso de análisis e interpretación del exoma. Posible visualización de los datos. Se entregó un exoma en formato Excel.
2. Se presentaron los datos relativos al registro de la historia clínica del paciente. El médico genetista entregó un ejemplo del mismo. Se introdujo el concepto de priorización de genes a través del panel bioinformático, se presentó un ejemplo del mismo. Se presentaron los filtros del exoma, explicando cada uno de ellos.
3. Presentación de los filtros del exoma (continuación). Definición de los pasos del proceso de filtrado del exoma, orden de ejecución, conclusiones en cada paso y conclusiones post-filtrado acerca del diagnóstico final.
4. Generación de informes post-diagnóstico. Se presentaron dos modelos de informes de ejemplo. Se introdujo el proceso de retro-alimentación de diagnósticos futuros. Se presentó la especificación de los requerimientos de visualización la cual fue aprobada por los participantes.
5. Consultas acerca del proceso de filtrado y posterior diagnóstico.
6. Presentación del prototipo finalizado el cuál fue aprobado por el médico genetista.

Se plantean los escenarios de uso durante la ejecución del proceso de análisis e interpretación de variantes genéticas en un exoma a través de la observación del proceso completo de toma de decisiones hasta llegar al diagnóstico y seguimiento. En total se identificaron 4 escenarios de uso que se resumen en la Tabla 5.

Escenario	Descripción
Escenario previo al filtrado de las variantes presentadas en el Exoma	Se realiza el registro de los datos del paciente, el motivo de la consulta y el conjunto de términos HPO, también se realiza una primer priorización de genes en el panel bioinformático.
Escenario de filtrado de las variantes presentadas en el Exoma	Se realiza el filtrado de las variantes del exoma aplicando los filtros de significancia clínica, frecuencia poblacional, predictores bioinformáticos, patrón de herencia e impacto funcional de la variante.
Escenario posterior al filtrado de las variantes presentadas en el Exoma, generación del Diagnóstico	Se registra la consulta manual a las bases de datos de fenotipos, las variantes que son patológicas. Se generan los informes clínico y de resultados.
Escenario de retro-alimentación de diagnósticos futuros	Se registra la frecuencia poblacional de las enfermedades en la población argentina y la relación de la variante con el fenotipo. Se retroalimenta el panel informático.

Tabla 5: Planteo de escenarios

Los escenarios se describen a continuación.

4.1 Escenario previo al filtrado de las variantes presentadas en el Exoma

4.1.1 Registro de los datos de la historia clínica del paciente

El proceso se inicia con la consulta del paciente. Se realiza un registro en la historia clínica donde se vuelcan las causas de la consulta del paciente. Actualmente, el genetista realiza dicho registro en papel como se muestra en la Figura 19.

4.1.2 Priorización de genes a través del Panel Bioinformático

Luego del registro de los datos de la historia clínica del paciente el médico genetista realiza una evaluación clínica del paciente (motivo de la consulta, examen físico). A partir de dicha evaluación clínica surgen una enfermedad/patología sospechada y/o un conjunto de términos HPO.

Con estos datos se realiza una búsqueda en el panel bioinformático para obtener un conjunto de genes candidatos que se espera coincida con los existentes en la solapa “Genes Fenotipo”. Las columnas que presentará dicho panel luego de la búsqueda se definen en la tabla 6.

Si no se encuentra nada en la solapa “Genes Fenotipo” entonces puede ser que la priorización inicial esté incompleta y, a partir de allí, se analizan la solapa “Variantes Pathogenic y L. Pathogenic” y/o “Todo” para la búsqueda del gen que tal vez a priori no llegó a estar en el panel bioinformático (porque o no cumple muchos términos HPO o porque se pensó en una patología diferente). Si se encuentra en estas últimas pestañas el gen que coincide con los términos HPO/enfermedad del paciente, entonces la variante pasa a ser una variante candidata y a tener relación con el fenotipo y eventualmente se retroalimentará el panel bioinformático para un próximo paciente que tenga un fenotipo o un término HPO parecido. Si se elige obtener resultados ingresando la enfermedad, entonces se realizará una búsqueda en la base de datos de OMIM (Ontología de enfermedades Mendelianas) por dicho término. Si se elige obtener resultados ingresando los términos HPO entonces se realizará una búsqueda, de dichos términos, en la base de datos Genecards [37] y para completar el resto de los datos de la grilla se busca en OMIM a través del GEN devuelto por Genecards.

Denominación de la columna	Comentarios
Ubicación genómica	Ubicación del gen
Enfermedad	Denominación del fenotipo
Patrón de Herencia	Cigosis. Variante en heterocigosis u homocigosis (si las dos copias del gen, paterna y materna, tienen la variante, o sólo una de ellas)
Número OMIM	Cada entrada de OMIM recibe un número único de seis dígitos
GEN	Gen en el que está presente la variante. Si bien no es común, podría ocurrir que la variante esté presente en más de un gen en los casos donde más de un gen tienen regiones superpuestas
Terminos HPO	Tipificación del fenotipo en términos HPO

Tabla 6: Columnas de un panel bioinformático

4.2 Escenario de filtrado de las variantes presentadas en el Exoma

4.2.1 Filtros involucrados

El proceso de filtrado de las variantes del exoma presenta varios caminos posibles comenzando por el primer filtro que es la frecuencia poblacional. A continuación, se presentan los filtros mediante los cuales se arribará a un diagnóstico:

- **Filtro significancia clínica de la variante**

Este filtro consta de ver si ya está la variante descrita como patogénica o no en bases de datos que incluyen significancia, como ClinVar. Se podría eliminar las que figuran como benignas y evaluar las que son de significado incierto o de interpretaciones conflictivas de patogenicidad. También se podría mirar en detalle aquellas no descritas previamente, que cumplan los criterios siguientes que sugieren patogenicidad.

- **Filtro frecuencia poblacional**

Si la frecuencia de esa variante en bases de datos de personas sanas es muy alta entonces se descarta como posible causa de la enfermedad genética. Generalmente la columna más usada es ExAC_AF por ser la más completa. Se quitan todas las variantes que estén presentes en más de 0.05 de la población. Cuanto más rara sea la enfermedad que tiene el paciente, se puede ser más restrictivo con este valor (sólo

tomar variantes presentes en menos de 0.01 de la población, por ejemplo), ya que no se espera que muchas personas porten una variante patogénica para una enfermedad así de rara. También podría considerarse que alguien sano puede ser portador de una variante patogénica (para una enfermedad recesiva, por ejemplo, o una dominante con penetrancia incompleta, que no exprese necesariamente en todas las personas que la porten), en este caso no se seleccionarían solo variantes ausentes en ExAC_AF. Por otro lado, también puede darse que una variante esté en baja frecuencia o ausente en estas bases pero que sea benigna, ya que hay polimorfismos (cambios genéticos benignos) de baja frecuencia (especialmente si tenemos en cuenta que poblaciones como la latinoamericana no están muy representadas en estas bases).

- **Filtro predictores bioinformáticos de patogenicidad**

Este filtro consta de ver si la variante generará una patología o no. Los más destacados son los que figuran como Damaging (D) o Pathogenic (P). Estos filtros se basan fundamentalmente en análisis in silico de la variante: si por la posición en la proteína – zonas más o menos relevantes para su función- o qué tan conservada está esa ubicación a nivel evolutivo (si en distintas especies es siempre ese aminoácido o puede variar) se puede estimar si la variante generaría alteraciones en el fenotipo.

- **Filtro patrón de herencia (cigosis):**

Este filtro consta de evaluar especialmente las variantes que están en heterocigosis si pensamos que se trata de una patología de origen dominante, o aquellas en homocigosis si pensamos que se trata de una patología de origen recesivo (especialmente si los padres del paciente son parientes entre sí). O podemos no quitar ninguna si se tratara de una heterocigosis compuesta, con variantes patogénicas diferentes cada una en una copia del mismo gen (si pudiera ser también una enfermedad autosómico recesiva). Si su valor es 0/1 significa que es un patrón heterocigota, 1/1 homocigota, 1/2 heterocigota compuesto.

- **Filtro Impacto funcional de la variante**

Este filtro consta de seleccionar las variantes que tienen impacto alto de la columna ANN[*].IMPACT. Esto depende del tipo de variante, se cree que generalmente las sinónimas (cuando no cambia el aminoácido resultante) van a tender a ser benignas, las nonsense o frameshift (alteración grave de la función de la proteína, como

interrupción de su síntesis) tenderán a ser patogénicas y las missense (cambio de un aminoácido por otro) pueden ser patogénicas o benignas. Hay excepciones, pero es una idea general de lo que nos sugiere el tipo de variante y su efecto predicho sobre la proteína. En este punto, también es usada otra información complementaria, pero con un análisis más bien manual: si es missense y no está descrita, pero en ClinVar encontramos que otras missense cercanas en el gen están descritas como patogénicas podemos inferir que quizá esta también lo sea. Lo mismo a la inversa, si muchas missense cercanas son benignas es menos probable que esta variante sea patogénica.

4.2.2 Trazabilidad del proceso de filtrado de las variantes presentadas en el Exoma

Debido a que el proceso de filtrado es repetitivo se procedió a definir un orden específico en cuanto la selección de los filtros y/o solapas en pro a arribar al diagnóstico teniendo en cuenta que la aplicación del filtro “Patrón de Herencia” es independiente del resto de los filtros. A continuación, en la tabla 5, se presenta la tabla de trazabilidad de ejecución y orden habitual de los filtros:

1) Análisis de la solapa GenesFenotipo			
Utilidad vs. otras solapas: ya al tener precargados los genes de interés podemos concentrarnos en variantes que quizás no estén todavía definidas como patogénicas (solapa pathogenic)	Orden habitual de filtros:	Columna frecuencia poblacional (ExAC_AF) valores <0,05 Columna de significancia clínica en Clinvar (CLNSIG) patogénica, likely patogénica o VUS-significado incierto-no descripta Columna de impacto (ANN[*].IMPACT) high, moderate o modifier Columna de predictores (dbNSFP_Polyphen2_HDIV_pred) damaging/pathogenic Si hasta ahí no se encontraron sospechosas, se vuelven a mirar en otro orden, basado en el impacto de la proteína y obviando clinvar:	Columna de impacto (ANN[*].IMPACT) high, moderate o modifier Columna frecuencia poblacional (ExAC_AF) valores <0,05 Columna de predictores (dbNSFP_Polyphen2_HDIV_pred) damaging/pathogenic
	Otra opción es concentrarse en Clinvar:	Columna de significancia clínica en Clinvar (CLNSIG) patogénica, likely patogénica o VUS-significado incierto-no descripta Columna frecuencia poblacional (ExAC_AF) valores <0,05 Columna de predictores (dbNSFP_Polyphen2_HDIV_pred) damaging/pathogenic	

2) Análisis de la solapa Pathogenic			
Utilidad vs. otras solapas: ya al tener precargadas las variantes conocidas como patogénicas, podemos concentrarnos en genes que quizás no estén todavía tan definidos como asociados con el fenotipo del paciente, o que sean diagnósticos diferenciales que quizás no consideramos inicialmente en la consulta.	Orden habitual de filtros:	Columna frecuencia poblacional (ExAC_AF) valores <0,05 Columna de impacto (ANN[*].IMPACT) high, moderate o modifier Columna de predictores (dbNSFP_Polyphen2_HDIV_pred) damaging/pathogenic	
		Columna de genes: ir viendo manualmente qué genes están relacionados con el fenotipo del paciente o no (*)	

3) Análisis de la solapa incidental			
Utilidad vs. otras solapas: esta solapa es especial, porque no guarda relación con el fenotipo / motivo de consulta del paciente sino que son genes que podrían tener impacto de tratamiento (accionables) si se encontrara accidentalmente alguna alteración. Sólo se van a informar las variantes patogénicas/likely claras.	Orden habitual de filtros:		
		Columna de significancia clínica en Clinvar (CLNSIG) patogénica o likely patogénica (SOLO ESAS 2)	
		Columna frecuencia poblacional (ExAC_AF) valores <0,05	
		Columna de impacto (ANN[*].IMPACT) high, moderate o modifier	
		Columna de predictores (dbNSFP_Polyphen2_HDIV_pred) damaging/pathogenic	
	También acá se suele ver en forma independiente de clinvar:		
	Columna frecuencia poblacional (ExAC_AF) valores <0,05		
	Columna de impacto (ANN[*].IMPACT) high, moderate o modifier		
	Columna de predictores (dbNSFP_Polyphen2_HDIV_pred) damaging/pathogenic		

4) Análisis de la solapa Todos			
Utilidad vs. otras solapas: Se usa cuando las otras pestañas no nos dieron información suficiente para el diagnóstico.	Orden habitual de filtros:		
		Columna frecuencia poblacional (ExAC_AF) valores <0,05	
		Columna de impacto (ANN[*].IMPACT) high, moderate o modifier	
		Columna de predictores (dbNSFP_Polyphen2_HDIV_pred) damaging/pathogenic	
		Columna de significancia clínica en Clinvar (CLNSIG) patogénica, likely patogénica o VUS-significado incierto-no descripta	
		Columna de genes: ir viendo manualmente qué genes están relacionados con el fenotipo del paciente o no (se realiza una consulta manual de los genes relacionados al fenotipo o conjunto de fenotipos del paciente, se consultan las bases de OMIM[29] y Phenomizer[22], entre otras.)	

Tabla 6: Trazabilidad del proceso de filtrado de las variantes presentadas en el Exoma

4.3 Escenario posterior al filtrado de las variantes presentadas en el Exoma, generación del Diagnóstico

4.3.1 Registro de la consulta manual a bases de datos de fenotipos

Para completar el proceso de filtrado se realiza una búsqueda manual de los genes relacionados al fenotipo o conjunto de fenotipos del paciente consultando las bases de OMIM y Phenomizer, entre otras. La búsqueda puede realizarse de dos maneras:

- buscar los síntomas del paciente, describiendo cada uno como un "fenotipo", para llegar al cuadro clínico general, o conjunto de fenotipos
- buscar si los genes en los cuales se hallaron variantes se corresponden con ese conjunto de fenotipos.

La forma de tipificar un fenotipo es con los términos HPO. Luego de ubicados esos términos se correlacionan con los genes que pueden ser los causales y se deja registro en la columna "Términos HPO" conjuntamente con el "Patrón de Herencia" y la constancia de "Relación con el fenotipo" (si/no). A modo de ejemplo si una variante es benigna, pero tiene mucha relación con el fenotipo se registra como relacionada al fenotipo, en este caso no será candidata.

4.3.2 Registro de variables patológicas

Una vez encontrada una variante candidata en un gen, relacionada con el fenotipo del paciente y coincidente con el patrón de herencia sugerido por su genealogía se deja registro de que dicha variante es posiblemente patológica.

Finalmente, se hace un "curado" o análisis manual detallado de las variantes candidatas que se obtuvieron del proceso, hasta que, idealmente, se identifica la variante causal o de relevancia para el paciente (o el conjunto de estas). Dicha identificación se realizará a través del registro en la columna "Variante Candidata" (si/no). Las variantes candidatas son de baja frecuencia, por el tipo de cambio en la proteína parecen generar algún daño o tal vez se encuentren en otras bases de datos que otros investigadores las hayan encontrado en pacientes con la misma patología.

4.3.3 Generación de informes

Luego del registro de las variables patológicas se generan dos informes, el Informe de Resultados que se entrega al paciente y el Informe Clínico. En el Informe de Resultados quedará registrado:

- Indicación clínica (causa por la cual fue solicitado el estudio)
- Los resultados, es decir se informa la variante de secuencia asociada con el fenotipo que genera la patología en el paciente y los motivos que justifican por qué esta es la variante patogénica.
- Diagnóstico molecular.
- Metodología utilizada para la detección
- Informe clínico que indica qué cuidados deberán ser hechos a futuro, los signos de la patología, riesgos futuros.
- Interpretación clínico-biológica que describe las evidencias de que la variante es la causa de la enfermedad del paciente, la justificación de por qué la variante es patológica.
- Descripción de los genes hallados intermedios, previos a arribar al resultado final si es que los hubo.
- Información suplementaria.
- Referencias/ontologías consultadas.
- Notas.

A modo de ejemplo, se proporciona un documento descriptivo del Informe de Resultados Anexo I - Informe de resultados estudio de Exoma por secuenciación masiva análisis de hallazgos clínicamente relevantes, Sebastian Menazzi (2020).

En el Informe Clínico quedará registrado:

- Motivo por el cual se realizó el estudio genético. Indicación clínica.
- Resumen de los antecedentes personales y familiares del paciente
- Resultados (variante de secuencia asociada con el fenotipo).
- Diagnóstico molecular. Frecuencia del trastorno, manifestaciones clínicas que tiene. Características a nivel molecular. Motivos, si es que los hubiere, de posibles sospechas del paciente acerca de una patología que no terminó siendo la detectada.
- Controles recomendados. Prevención. Medidas de reducción del riesgo de enfermedades futuras.
- Asesoramiento genético a familiares del paciente, medidas preventivas al respecto.
- Resultados relacionados a la solapa "Incidental Findings".

A modo de ejemplo, se proporciona un documento descriptivo del Informe de Resultados - Anexo II – Informe de asesoramiento genético, estudio de Exoma por secuenciación masiva y estudios complementarios

Por último, de ser necesario, se consultan las ontologías/repositorios de enfermedades monogénicas para postular qué genes pueden ser causales del fenotipo que se observa en relación a la patología diagnosticada para buscar evidencias y publicaciones relacionadas a dicha enfermedad.

4.4 Escenario de retro-alimentación de diagnósticos futuros

El proceso de retro-alimentación de diagnósticos futuros puede ocurrir de dos maneras:

- Detectando enfermedades monogénicas frecuentes o altamente frecuentes en la población Argentina.
- Detectando la relación de la variante con el fenotipo en diagnósticos previos.

4.4.1 Relación de la variante con el fenotipo

Puede darse el caso de encontrar una variante candidata en un paciente con un fenotipo determinado y no tener disponible los datos necesarios para calificarla de patogénica porque no estaba descrita en ninguna base. Si se encontrara algún paciente más con la misma característica entonces debe registrarse esa información a fines de retro-alimentar el proceso. De lo expuesto anteriormente se concluye que, al momento de analizar nuevos exomas, deben generarse alertas en variantes similares acerca de si tienen relación con el fenotipo o no.

4.4.2 Frecuencia poblacional de enfermedades monogénicas en la población argentina

Basados en el conocimiento de que la columna de frecuencia poblacional es representativa solo de países desarrollados sabemos que los países latinoamericanos se encuentran sub-representados, puede ocurrir que en dicha columna haya baja frecuencia lo que induce a pensar que la variante no es patológica, pero, luego de diagnosticar la variante como patológica en tal vez más de 500 exomas, entonces, se puede deducir que es mucho más frecuente en nuestra población.

De lo expuesto anteriormente se concluye que, al momento de analizar nuevos exomas, las variantes patológicas curadas anteriormente deben generar alertas a efectos de ser identificadas como frecuentes o altamente frecuentes en la población argentina, con lo cual se debe calcular en qué proporción de pacientes esta variante apareció.

4.5 Retroalimentación del Panel Bioinformático

Aquellos genes que no fueron encontrados en la búsqueda inicial del panel bioinformático y que fueron registrados por el genetista como “Relación con el Fenotipo” con un “Si”, los términos HPO con lo cual es una “Variante Candidata” entonces deberán retroalimentar la base de datos del panel bioinformático. El fenotipado del paciente, en el proceso de retroalimentación de diagnósticos futuros, cobra importancia debido a varios motivos. Los más importantes se exponen a continuación:

- El experimento de exoma secuencía miles de genes al mismo tiempo, y por el momento el conocimiento sobre el significado de cada variante hallada (o posible) en cada gen es incompleto. Esto significa que hay todavía demasiadas incertezas como para hacer un análisis totalmente libre de sesgos, evaluando e informando en forma automatizada todos los resultados de cada variante (definiendo si son causa clara del problema médico o no).

- Con relación con el sesgo de conocimiento, a veces a nivel biológico se dan situaciones complejas: una única variante en un gen aparentemente recesivo alcanza para dar con una enfermedad, o se combina con otro gen en un modelo de tipo "digénico", o una variante que no aparenta ser demasiado grave en realidad lo es.

Todo esto hace que uno de los factores clave para definir un diagnóstico es si concuerda o no lo que se ve con lo que tiene el paciente y, por lo tanto, preseleccionar de acuerdo con el fenotipo cuáles serán los genes de interés más importantes a analizar resulta clave.

Capítulo 5: Especificación de los requerimientos de la solución

En este capítulo se describirán los requerimientos funcionales que debe satisfacer la herramienta a partir de los escenarios de trabajo relevados en el capítulo 4.

5.1 Requerimientos de Visualización del proceso de filtrado del Exoma.

La herramienta a desarrollar debe contar con los datos generados en el VCF ya disponibles y anotados en una base de datos. Luego de que el genetista entreviste al paciente deberá permitir registrar los datos de la historia clínica del mismo. Las variantes y anotaciones deberán ser desplegadas en forma tabular en el espacio central del browser. Cada columna deberá ser ocultada o desplegada según la relevancia en el filtrado permitiendo al genetista focalizarse en la información más importante. La totalidad de los registros deberá navegarse a través de la paginación. El panel de filtrado deberá mostrar los filtros que se aplican al exoma, cada uno de los filtros se aplica independientemente pudiendo también ser ejecutados conjuntamente, la grilla que presenta los resultados deberá filtrarse dinámicamente.

Opción Historia Clínica

La herramienta deberá permitir, en el proceso previo de filtrado de las variantes, registrar los datos de la historia clínica del paciente según lo definido en el punto “4.1.1 Registro de datos de la historia clínica del paciente”.

Opción Panel Bioinformático

Para la definición del panel bioinformático la herramienta deberá disponibilizar una opción que permitirá ingresar, en un campo de texto, el fenotipo (tanto por término HPO como por enfermedad) y como resultado se mostrará una grilla que muestra las columnas resultado según lo definido en la Tabla 6 del punto “4.1.2 Priorización de genes a través del Panel Bioinformático”.

Opción Filtrado de Variantes

Esta será la opción principal. La herramienta habilitará, en primer lugar, una bandeja de entrada de nuevos exomas por paciente. Una vez seleccionado el paciente al que se desea diagnosticar se visualizarán las 4 solapas que contienen las variantes de la muestra respecto del genoma de referencia: solapa “Todos”, solapa “Variantes genes fenotipo”, solapa “Variantes Pathogenic y L. Pathogenic” y solapa “Variantes genes Incidental Findings” según lo definido en el apartado “Solapas” del punto 3.2. Cada una de las solapas mostrará las columnas definidas en el apartado “Columnas” del punto “3.2 Detalle del archivo Excel” con el contenido del exoma en formato grilla siendo las últimas 4 columnas editables para permitir el llenado manual por parte del genetista, en el escenario posterior al filtrado, de los siguientes campos:

- “Términos HPO”, “Patrón de Herencia”, “Relación con el Fenotipo” según lo definido en el punto “4.3.1 Registro de la consulta manual a bases de datos de fenotipos”.
- “Es variante candidata” según lo definido en el punto “4.3.2 Registro de variables patológicas”.

Las columnas dbNSFP_ExAC_AF y ExAC_AF serán mostradas con un valor con 4 decimales y las columnas indicadas como identificatorias de la variante definidas en el apartado “Columnas” del punto “3.2 Detalle del archivo Excel” aparecerán ocultas en primera instancia y se disponibilizará una opción que permita desplegarlas, en caso de ser necesario. La

grilla que muestra la cantidad total de filas del exoma deberá ser navegable a través de una opción de paginación.

El panel de filtros de las variantes será aplicable indistintamente en cualquiera de las solapas. Dichos filtros serán ejecutados, según lo definido en el punto “4.2.2 Trazabilidad del proceso de filtrado de las variantes presentadas en el Exoma” debiendo las grillas filtrar sus datos dinámicamente en el mismo momento de ejecución del filtro y se contará con la posibilidad de guardar los filtros ejecutados para un próximo análisis sobre el mismo exoma.

Se define la necesidad de contar con un control de rango para seleccionar la posición del transcripto que desea visualizarse del exoma adaptando así el contenido de las columnas a mostrar la posición deseada. Dicho control será agregado en el panel de filtros de variantes debiendo el sistema comportarse, en la visualización del contenido del exoma, según lo definido en el apartado “Representación de los transcriptos” del punto “3.2 Detalle del archivo Excel”. El nivel de profundidad de la navegación de los transcriptos será fijado arbitrariamente en 10 pero es N.

La frecuencia poblacional un control de rango de entre 0 y 0.05. El patrón de herencia será un desplegable con la siguiente lista de opciones seleccionables: 0/1 (Heterocigota), 0/1 (Heterocigota) y 0/1 (Heterocigota). Los Predictores Bioinformáticos serán un desplegable con la siguiente lista de opciones seleccionables: A (Disease Causing Automatic), B (Benign), D (Probably Damaging). La significancia clínica será un desplegable con la siguiente lista de opciones seleccionables: Benign, Likely Benign, Conflicting Interpretations Of Pathogenicity, Uncertain Significance, Likely Pathogenic, Pathogenic, Drug Response, Association, Protective, Risk Factor, Not Provided, Other. El impacto funcional de la variante será un desplegable con la siguiente lista de opciones seleccionables: High, Moderate, Modifier, Low. Para facilitar el proceso de retro-alimentación de diagnósticos futuros, toda vez que aparezcan nuevos exomas:

- Se mostrará una alerta acerca de la relación de variantes similares con el fenotipo, según lo definido en el punto “4.4.1 Relación de la variante con el fenotipo”

resaltando la fila de la grilla que representa la variante donde se detectó la patología con otro color. El contenido de la celda “Relación con el Fenotipo” deberá estar en “SI”.

- Se presentará una alerta de las variantes curadas anteriormente identificándolas como frecuentes o altamente frecuentes indicando en qué proporción de pacientes apareció. Dicha alerta se verá reflejada en la fila de la grilla que representa la variante donde se detectó la patología (columna “Es variante candidata” en “SI”) disponibilizada a través de una ayuda contextual de tipo tooltip con el dato, según lo definido en el punto “4.4.2 Frecuencia Poblacional de enfermedades monogénicas en la población argentina”.

Opción Informe de Resultados, Opción Informe Clínico

Se disponibilizarán dos opciones: una de creación del “Informe de Resultados” y la otra de creación del “Informe Clínico”. Ambas opciones comenzarán a través del llenado de un formulario que luego de ser completado habilite la generación de los mismos en archivo de formato pdf. Los campos de ingreso del formulario, para cada uno de los informes, serán los definidos en el apartado “Generación de informes” del punto 4.3, para cada uno de los informes respectivamente.

Capítulo 6: Solución propuesta

En este capítulo se avanza con las etapas de ideación y prototipado de la herramienta Exoma Browser como solución propuesta para proveer al médico genetista soporte para el proceso de análisis e interpretación de variantes genéticas en el exoma humano y así alcanzar el diagnóstico del paciente. Dicho prototipado utiliza como punto de partida el prototipo de Stock Screener del sitio Investing.com.

6.1 Stock Screener de Investing.com como referencia

Un Stock Screener es una manera de buscar el éxito en el mercado de valores averiguando cuáles son las mejores acciones en las que invertir. Hay miles de acciones cotizando en los mercados entre las que los inversores deben elegir filtrando una lista de acciones potenciales hasta llegar a las que se ajustan criterios propios, ya sea de rentabilidad, riesgo o tamaño [38].

Investing.com Search the website... Sign In / Free Sign Up

BYBIT Obtené un bono de \$50 en BTC. Además, ¡disfrutá del Spot trading sin cargo! [Regístrate ahora](#)

Stock Screener

United States | All Exchanges | All Sectors | All Industries | All Equity Types | [Advanced Screener](#)

Add Criteria

- Popular
- Ratios
- Price
- Volume & Volatility
- Fundamental
- Advanced Metrics
- Dividends
- Technical Indicators

Edit Criteria (Matches: 3041) Save | Clear All

- 1-Year Change: -100.00 | 372.90K
- Market Cap: 1.00 | 2.28T
- Dividend Yield (%): 0.01 | 2.51K

Screener Results (3041) Download Results | Add to Watchlist

Overview	Fundamental	Technical	Performance	Candlestick Patterns		
Name	Symbol	Exchange	Last	Chg. %	Market Cap	Vol.
Apple	AAPL	NASDAQ	141.47	-5.59%	2.28T	78.90M
Microsoft	MSFT	NASDAQ	234.96	-2.53%	1.75T	13.87M
Broadcom Mandatory	AVGOP	NASDAQ	1420.14	-1.65%	732.32B	16.81K
UnitedHealth	UNH	NYSE	504.12	-1.91%	472.21B	1.30M
J&J	JNJ	NYSE	164.27	-1.26%	432.21B	3.52M
Visa A	V	NYSE	179.29	0.06%	371.15B	5.45M
Exxon Mobil	XOM	NYSE	88.55	-0.35%	370.04B	10.00M
Taiwan Semiconductor	TSM	NYSE	68.69	-4.77%	356.51B	9.46M
Tencent Holdings	TCTZF	OTC Markets	34.305	-3.59%	324.90B	15.86K
Roche Holding Participation	RHHVF	OTC Markets	323.38	-2.3%	315.95B	54.19K

	Bank of America Corp Pe ADR	BAC_pe	NYSE	19.34	1.10	249.65B	18.17K
	Bank of America Corp Pg ADR	BML_pg	NYSE	19.21	-0.70	249.65B	7.07K
	Bank of America Corp Pj ADR	BML_pj	NYSE	19.69	1.27	249.65B	4.75K
	Bank of America Corp PI ADR	BML_pl	NYSE	20.05	1.83	249.65B	26.21K
	Bank of America Corp PI Pref	BAC_pl	NYSE	1.20K	0.91	249.65B	6.71K
	Pfizer	PFE	NYSE	44.43	0.77	249.36B	16.46M
	Samsung Electronics Co	SSNLF	OTC Markets	40.60	0.00	246.78B	0.00
	Coca-Cola	KO	NYSE	56.98	1.06	246.42B	21.45M
	PepsiCo	PEP	NASDAQ	168.67	1.60	232.75B	5.27M
	Novo Nordisk ADR	NVO	NYSE	99.36	3.61	226.97B	1.38M
	Merck&Co	MRK	NYSE	86.78	1.10	219.84B	6.83M
	Costco	COST	NASDAQ	488.29	2.09	216.15B	2.41M
	Alibaba ADR	BABA	NYSE	80.99	4.01	214.42B	20.58M
	Alibaba	BABAF	OTC Markets	9.93	1.85	214.42B	11.11K
	Industrial Commercial Bank of ...	IDCBY	OTC Markets	9.32	-0.43	214.36B	213.37K

1 2 3 4 5 6 7 8 9 10

Next →

Figura 20: Stock Screener, panel principal [38].

La herramienta Stock Screener de Investing.com (Figura 20) presenta como características un panel de filtros con posibilidad de guardarlos para próximas operaciones, el manejo de solapas que muestran distintas vistas de las empresas que poseen acciones, la posibilidad de bajar los resultados a un archivo, el filtrado automático de los resultados ni bien se aplican los filtros y la opción de paginación por tratarse de un gran volumen de datos que requieren análisis y criterios complejos de decisión. Es por eso que fue tomado como modelo de base ya que satisface requerimientos similares a los de Exoma Browser.

6.2 Prototipo herramienta Exoma Browser

Con el objetivo de cumplir los requerimientos del proceso de análisis e interpretación de variantes del exoma, llevado a cabo por el médico genetista, se presenta la herramienta Exoma Browser (disponible online) [39] que permite la ejecución de los escenarios hasta arribar al diagnóstico final y también permite registrar los datos que permiten retro-alimentar diagnósticos futuros. Dicha herramienta fue construida en base a una Arquitectura Web y con las siguientes tecnologías: HTML, Javascript, CSS, Bootstrap y JQuery.

El Exoma Browser (Figura 19) acelera los tiempos de evaluación diagnóstica, concentra los datos del paciente y su exoma y aumenta la sensibilidad del proceso de detección de variantes patogénicas facilitando también el registro de los términos HPO asociados, su patrón

de herencia, si guarda relación con el fenotipo o no y finalmente si es una variante candidata a ser patogénica a través de los campos editables de la grilla.

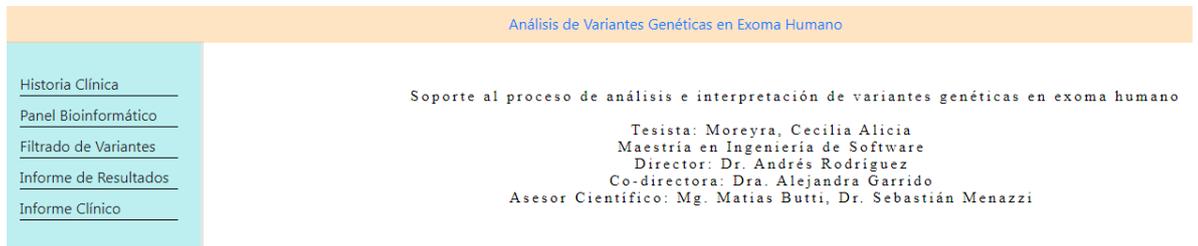


Figura 19: Exoma Browser, panel principal [39].

6.2.1 Escenario previo al filtrado de variantes presentadas en el Exoma

En el escenario previo al filtrado de las variantes se permitirá cargar, a través de un formulario accesible desde la opción Historia Clínica, los datos relacionados al paciente asociados a un Número de Historia Clínica (Figura 21).

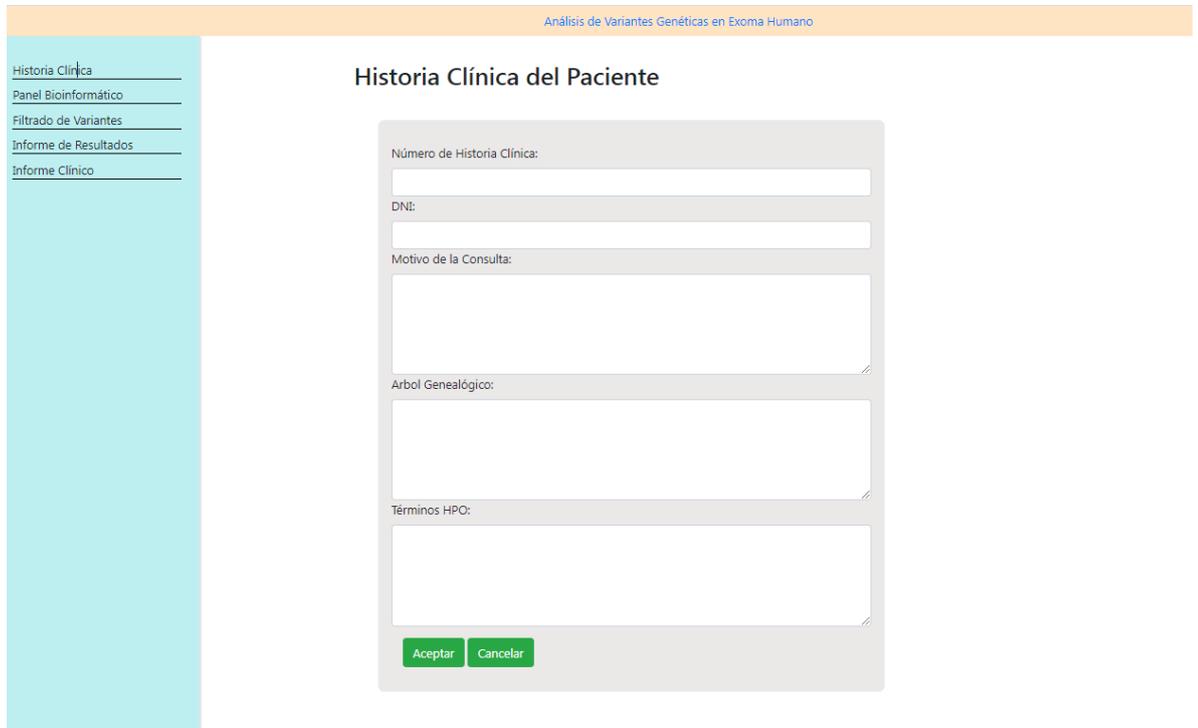


Figura 21: Exoma Browser, registro de los datos de la Historia Clínica del Paciente [39].

A efectos de que el genetista pueda realizar la priorización de los genes la herramienta Exoma Browser presenta el panel bioinformático que permite realizar la búsqueda por fenotipo, ya sea por término HPO como por enfermedad en búsqueda del gen que determina si la variante es patológica o no lo es (Figura 22).

Ubicación Genómica	Enfermedad	Patrón de Herencia	Nro OMIM	Gen	Términos HPO
2q23.3	Autosomal recessive inheritance	AD	616055	CACNA1A	HP:0000006
2q24.3	Autosomal recessive inheritance	AD	616056	KCNK1	HP:0000007
19p13.13	Abnormal paranasal sinus morphology	AD	616057	EA7	HP:0000245
19q13	Abnormal nasal mucosa morphology	AD	616058	EA4	HP:0410042
12p13.32	Diabetes mellitus	AD	616056	EA3	HP:0000819

Figura 22: Exoma Browser. Panel Bio-Informático. Búsqueda por fenotipo [39].

6.2.2 Escenario de filtrado de las variantes presentadas en el Exoma

En el escenario principal de filtrado de las variantes el genetista selecciona, de una bandeja de entrada de nuevos exomas, con qué paciente querrá trabajar. Luego de esto es presentado un panel de selección de los filtros involucrados en el proceso de obtención del diagnóstico: Nro de Transcripto, Frecuencia Poblacional, Patrón de Herencia, Predictores Bioinformáticos, Significancia Clínica, Impacto Funcional de la Variante (Figura 23).

- Historia Clínica
- Panel Bioinformático
- Filtrado de Variantes
- Informe de Resultados
- Informe Clínico

Análisis de variantes de exoma

Exoma - Historia Clínica Nro: 392845
 Exoma - Historia Clínica Nro: 765839
 Exoma - Historia Clínica Nro: 294776

Guardar Filtros | Borrar Filtros | Bajar Exoma

Nro. de Transcrito

Patrón de Herencia

Significancia clínica

Frecuencia Poblacional

Predictores Bioinformáticos

Impacto funcional de la variante

Figura 23: Exoma Browser, panel de filtrado de variantes [39].

A medida que los filtros van siendo seleccionados, dinámicamente se filtran los datos presentados en las grillas. De ser necesario el genetista podrá guardar la selección de los filtros para continuar el proceso de diagnóstico en otro momento a través de la opción “Guardar Filtros” que brinda la herramienta. Las solapas están listas para ser seleccionadas y despliegan una grilla con las columnas y sus correspondientes datos. Se presenta una opción para ocultar los datos de identificación de la variante a través del botón “Ver identificación” (Figuras 24 y 25).

Análisis de Variantes Genéticas en Exoma Humano

- Historia Clínica
- Panel Bioinformático
- Filtrado de Variantes
- Informe de Resultados
- Informe Clínico

Genes Fenotipo Pathogenic Incidental Todos

Ver Identificación

CHROM	POS	REF	ALT	GEN[0].GT	GEN[0].AD	GEN[0].DP	ID	ANN[*].GENE	ANN[*].HGVS_C	ANN[*].HGVS_P	ANN[*].FEATUREID	ANN[*].EFFECT	ANN[*].IMPACT	GENSFP_ExAC
chr12	19873407A	G	T	1/1	0 16	16	rs4369465							
chr12	20724606G	A	G	0/1	33 19	52	rs7966459							
chr12	20766559C	T	T	1/1	0 117	117	SNP12-20657826/rs10743380							
chr12	20766685G	C	T	1/1	0 25	25	rs7308381							
chr12	20852546C	T	D	1/1	40 23	63	SNP12-20743813/rs34243130	SLCO1C1	c.*2637C>T		ENST00000534996	downstream_gene_variant	MODIFIER	
chr12	20864356C	T	D	1/1	37 29	66	rs7963279	SLCO1C1	c.*250>T		ENST00000539415	3_prime_UTR_variant	MODIFIER	
chr12	20903757T	C	T	1/1	0 45	45	rs10841611	SLCO1C1	c.1916>31T>C		ENST00000265509	intron_variant	MODIFIER	
chr12	20905250C	T	T	1/1	0 52	52	rs8487138	SLCO1C1	c.*1614C>T		ENST00000539415	3_prime_UTR_variant	MODIFIER	
chr12	21011480T	G	D	1/1	107 82	189	rs4149117.261184	SLCO1B3	c.334T>G	p.Ser112Ala	ENST00000381545	missense_variant	MODERATE	
chr12	21013946C	T	D	1/1	86 33	89	rs1195.rs3764009	SLCO1B3	p.302>3C>T		ENST00000545880	splice_region_variant&intron_variant	LOW	
chr12	21015760G	A	D	1/1	83 85	118	rs1196.rs7311358	SLCO1B3	c.171D>A	p.Met57Ile	ENST00000544370	missense_variant	MODERATE	
chr12	21038411A	G	D	1/1	51 59	110	rs1183.rs3053998	SLCO1B7	c.*359>24996A>G		ENST00000544957	intron_variant	MODIFIER	
chr12	21054366G	A	D	1/1	38 36	73	rs3784006.440268	SLCO1B7	c.*359>4286A>A		ENST00000544957	intron_variant	MODIFIER	
chr12	21054918T	C	D	1/1	5 18	23	rs2062516	SLCO1B3	c.1865>517T>C		ENST00000381545	intron_variant	MODIFIER	
chr12	21175891A	G	D	1/1	17 6	23	rs10841725							
chr12	21175703G	A	D	1/1	16 11	27	rs10841726							
chr12	21175719A	G	D	1/1	20 13	33	rs10841727							
chr12	21176135T	C	D	1/1	77 58	135	SNP12-21067402/rs1546308							
chr12	21331599T	C	D	1/1	56 7	126	rs4149057.259884							
chr12	21331840G	A	D	1/1	33 49	82	rs374113543							
chr12	21472254G	A	D	1/1	33 33	66	rs10505872	SLCO1A2	c.61>397C>T		ENST00000544290	intron_variant	MODIFIER	
chr12	21580766C	A	D	1/1	15 12	27	rs2101695	PYROXD1	c.64>49C>A		ENST00000544870	intron_variant	MODIFIER	
chr12	21593404G	C	1/1	0 29	29	rs4707023	PYROXD1	c.*49>229C>C			ENST00000535982	intron_variant	MODIFIER	
chr12	21596204C	T	D	1/1	2 4	6	rs1963817	PYROXD1	c.146>17C>T		ENST00000544870	intron_variant	MODIFIER	
chr12	21608011G	T	D	1/1	28 22	50	rs2192182	PYROXD1	c.415>55G>T		ENST00000544870	intron_variant	MODIFIER	
chr12	21608043A	G	D	1/1	30 24	54	rs2192183	PYROXD1	c.415>23A>G		ENST00000544870	intron_variant	MODIFIER	
chr12	21609761C	CAACTD	1/1	38 21	59	rs34335650.rs60594013	PYROXD1	c.*76>73A.*76>735resCTAA			ENST00000375266	intron_variant	MODIFIER	

Figura 24: Exoma Browser, solapas con columnas identificatorias de la variante [39].

Análisis de Variantes Genéticas en Exoma Humano

Historia Clínica

Panel Bioinformático

Filtrado de Variantes

Informe de Resultados

Informe Clínico

Genes Fenotipo	Pathogenic	Incidental	Todos											
Ver Identificación														
ANN ¹ LEFFECT	ANN ¹ IMPACT	dbNSFP_ExAC_AF	AN_Adj	AC_Adj	ExAC_AF	dbNSFP_Polyphen2_HDIV_pred	dbNSFP_Polyphen2_HVAR_pred	dbNSFP_SIFT_pred	dbNSFP_MutationTaster_pred	dbNSFP_MutationAssessor_pred	CLNSIG	CLNSIGCONF	Términos HPO - Fenotipo	Patrón de Herencia
				121388	112867	0.5298								
				112484	109560	0.9482								
downstream_gene_variant	MODIFIER			120274	7577	0.6630								
3_prime_UTR_variant	MODIFIER			121338	13191	0.2735								
intron_variant	MODIFIER			121286	95464	0.5397								
3_prime_UTR_variant	MODIFIER			119576	63549	0.5359								
missense_variant	MODERATE			121358	97827	0.8945							Benign	
splice_region_variant&intron_variant	LOW			118852	95861	0.8066							Benign	
missense_variant	MODERATE			121360	97617	0.8044							Benign	
intron_variant	MODIFIER			121296	97604	0.8047							Benign	
intron_variant	MODIFIER			121376	99311	0.8182							Benign	
intron_variant	MODIFIER			10526	7850	0.7458								
				119570	113543	0.9498								
				121322	83821	0.5280							Benign	
				121220	2	0.0000								
intron_variant	MODIFIER													
intron_variant	MODIFIER			86574	43597	0.5036								
intron_variant	MODIFIER			120750	120193	0.9954								
intron_variant	MODIFIER			97860	45884	0.4696								
intron_variant	MODIFIER													
intron_variant	MODIFIER			113462	55595	0.4900								
intron_variant	MODIFIER			10058	4591	0.4565								
intron_variant	MODIFIER			114890	55781	0.4851								
intron_variant	MODIFIER			121346	58302	0.4605								
downstream_gene_variant	MODIFIER			77304	28726	0.3716								
downstream_gene_variant	MODIFIER			120780	51868	0.4280								
downstream_gene_variant	MODIFIER				336	NaN								
intron_variant	MODIFIER			120476	51553	0.4279								
intron_variant	MODIFIER													
3_prime_UTR_variant	MODIFIER			116852	91346	0.7817								
intron_variant	MODIFIER													
intron_variant	MODIFIER			34852	15284	0.4385								
													Likely_benign	
				119790	91773	0.7861							Benign	
				116050	88861	0.7557							Benign	
				120054	92211	0.7881							Benign	
				121340	98814	0.7896							Benign	
				121384	93291	0.7886							Benign	
				121156	120027	0.9507								
intron_variant	MODIFIER			121302	118940	0.9805							Benign	
				121056	120998	0.9995							Benign/Likely_benign	
				121300	121175	0.9990							Benign/Likely_benign	
				121278	120905	0.9969							Benign/Likely_benign	
				121300	121175	0.9990							Benign/Likely_benign	

Figura 25: Exoma Browser, solapas sin columnas identificatorias de la variante [39].

Se podrá paginar para navegar la totalidad de los registros (Figura 26).

Análisis de Variantes Genéticas en Exoma Humano

Historia Clínica

Panel Bioinformático

Filtrado de Variantes

Informe de Resultados

Informe Clínico

intron_variant	MODIFIER			121220	2	0.0000								
intron_variant	MODIFIER			86574	43597	0.5036								
intron_variant	MODIFIER			120750	120193	0.9954								
intron_variant	MODIFIER			97860	45884	0.4696								
intron_variant	MODIFIER													
intron_variant	MODIFIER			113462	55595	0.4900								
intron_variant	MODIFIER			10058	4591	0.4565								
intron_variant	MODIFIER			114890	55781	0.4851								
intron_variant	MODIFIER			121346	58302	0.4605								
downstream_gene_variant	MODIFIER			77304	28726	0.3716								
downstream_gene_variant	MODIFIER			120780	51868	0.4280								
downstream_gene_variant	MODIFIER				336	NaN								
intron_variant	MODIFIER			120476	51553	0.4279								
intron_variant	MODIFIER													
3_prime_UTR_variant	MODIFIER			116852	91346	0.7817								
intron_variant	MODIFIER													
intron_variant	MODIFIER			34852	15284	0.4385								
													Likely_benign	
				119790	91773	0.7861							Benign	
				116050	88861	0.7557							Benign	
				120054	92211	0.7881							Benign	
				121340	98814	0.7896							Benign	
				121384	93291	0.7886							Benign	
				121156	120027	0.9507								
intron_variant	MODIFIER			121302	118940	0.9805							Benign	
				121056	120998	0.9995							Benign/Likely_benign	
				121300	121175	0.9990							Benign/Likely_benign	
				121278	120905	0.9969							Benign/Likely_benign	
				121300	121175	0.9990							Benign/Likely_benign	

« 1 2 3 4 5 6 7 8 9 10 »

Figura 26: Exoma Browser, paginación [39].

6.2.3 Escenario posterior al filtrado de las variantes presentadas en el Exoma, generación del diagnóstico

En el escenario posterior al filtrado de las variantes el genetista registrará los “Términos HPO”, el “Patrón de Herencia” y la “Relación con el Fenotipo” (si/no) luego de buscar manualmente los genes relacionados al fenotipo o conjunto de fenotipos. Estos datos serán ingresados manualmente en las columnas destinadas a tal efecto, denominadas con el mismo nombre. Si una variante se encuentra relacionada con el fenotipo, opción “si”, la fila correspondiente se resaltará con otro color (Figura 26).

También se registrará si una variante es posiblemente patológica en la columna “Variante Candidata” (si/no). Si una variante es candidata a ser patológica, opción “si”, la fila correspondiente se resaltará con otro color (Figura 27).

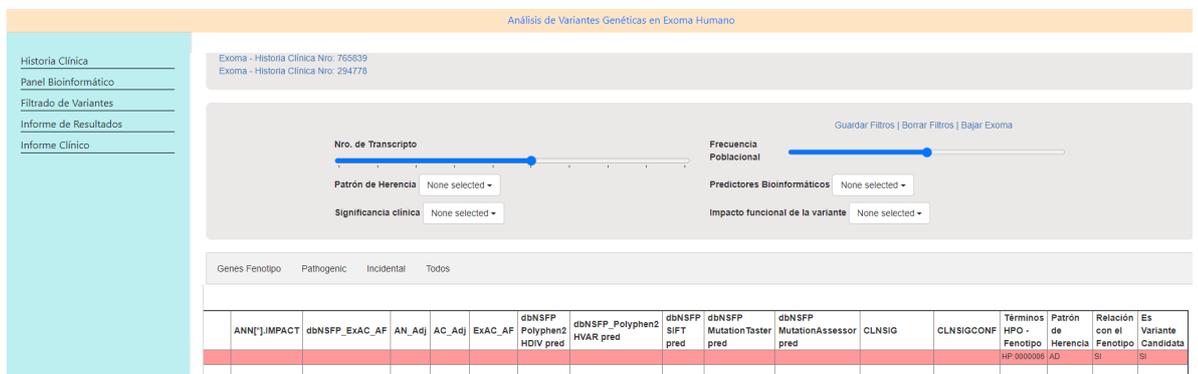


Figura 27: Exoma Browser, registro de la consulta manual a bases de datos de fenotipos [39]

Luego del registro de las variantes patológicas el genetista procederá a la generación de informes cargando los datos en los formularios y ejecutando la opción de impresión (Figuras 28 y 29).

Análisis de Variantes Genéticas en Exoma Humano

Historia Clínica

Panel Bioinformático

Filtrado de Variantes

Informe de Resultados

Informe Clínico

Informe de Resultados

Indicación clínica

Resultados

Diagnóstico molecular

Metodología utilizada para la detección

Informe clínico

Interpretación clínica/Bioinformática

Descripción de genes intermedios

Información suplementaria

Referencias/Ontologías consultadas

Notas

Aceptar
Cancelar

La operación se ha realizado con éxito. [Descargar Informe](#)

Figura 28: Exoma Browser, Informe de Resultados [39].

Análisis de Variantes Genéticas en Exoma Humano

Historia Clínica

Panel Bioinformático

Filtrado de Variantes

Informe de Resultados

Informe Clínico

Informe Clínico

Motivo por el cual se realizó el estudio genético

Resumen de los antecedentes personales y familiares del paciente

Resultados (variante de secuencia asociada con el fenotipo)

Diagnóstico molecular

Controles recomendados

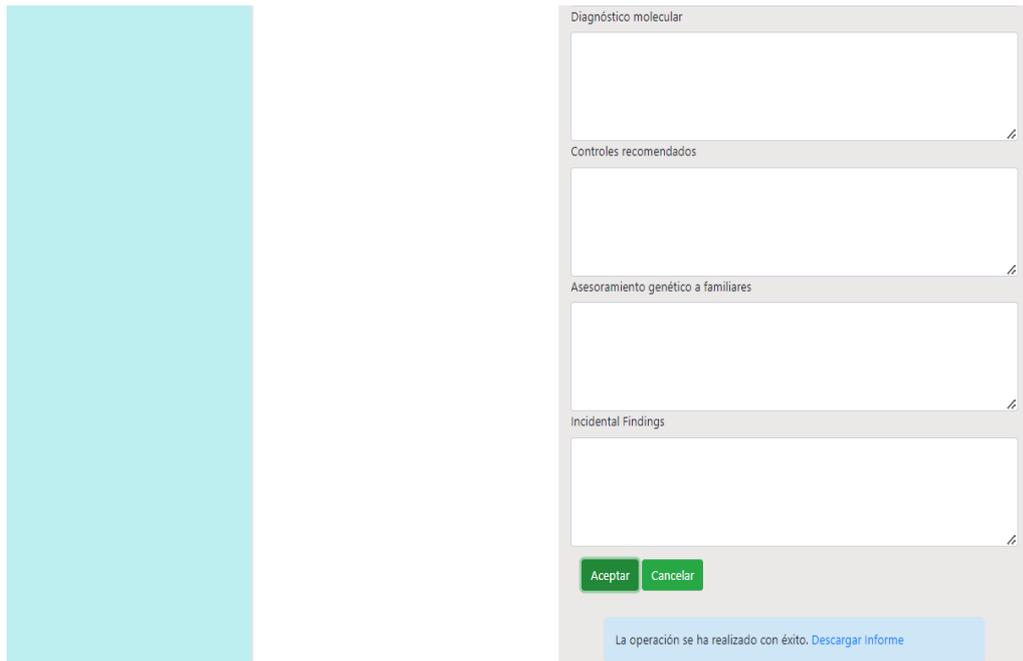


Figura 29: Exoma Browser, Informe Clínico [39].

6.2.4 Escenario de retro-alimentación de diagnósticos futuros

En futuros exomas si se encontrara que hay una variante similar a otra analizada anteriormente e identificada como relacionada el fenotipo entonces la fila aparecerá resaltada con otro color.

Con respecto a la frecuencia poblacional de enfermedades monogénicas se muestra, a través de una alerta del estilo tooltip, en qué cantidad de pacientes fue encontrada esta variante que anteriormente se diagnosticó como patogénica (Figura 30). La fila de dicha variante se presenta resaltada en la grilla con un color diferente.

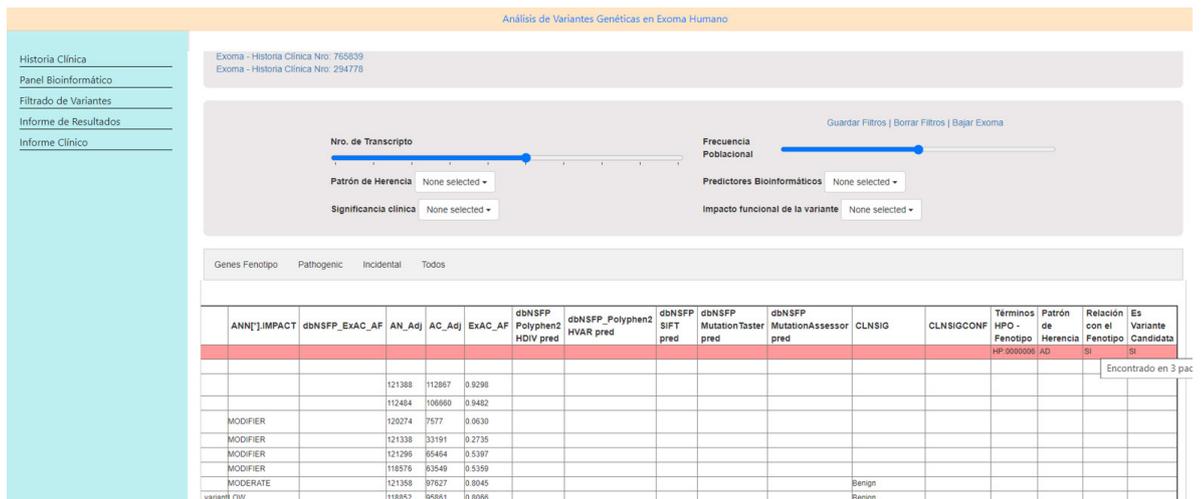


Figura 30: Exoma Browser, tooltip de variante candidata relacionada con el fenotipo [39].

6.3 Validación del producto

En este capítulo se presentan dos casos de validación de forma de mostrar que la herramienta propuesta es válida para ser usada por otros profesionales dentro de la misma especialidad. Según los conceptos introducidos en la sección [2.6 Introducción al concepto de usabilidad y métodos de evaluación de usabilidad](#), se realizó una revisión por expertos a través de entrevistas realizadas a dos profesionales de la salud que desempeñan actualmente su cargo en áreas relacionadas a la genética y la bioinformática.

En dichas entrevistas se mostró la herramienta, se realizaron pruebas de usabilidad y se presentó un cuestionario SUS (System Usability Scale) [41] para ser respondido con su nivel de satisfacción y con la finalidad de documentar todos los criterios medibles tales como aprendizaje, performance, tasa de errores y otros con el objetivo de evolucionar el desarrollo propuesto.

6.3.1 Caso de Validación: Dr. Gabriel Ercoli, médico Genetista

El Dr. Gabriel Ercoli desempeña actualmente su función como Director Médico de “Genesis, Centro Integral de Genética”. Se lo convocó a una entrevista en la cual se mostró la herramienta y sus principales módulos en funcionamiento y luego el mismo ejecutó la prueba.

En su opinión la misma: “tiene sentido, cierra, va por el camino adecuado”.

No encuentra inconsistencias. Es útil con respecto a la retro-alimentación del panel informático que se lleva a cabo luego de todo el proceso de filtrado ya que “es muy importante tener tu propia base”. La terminología es la que usa en su trabajo diario, es fácil de entender y le es familiar.

Las opciones que presenta son correctas, no agregaría ni restaría ninguna, el proceso está “completamente contemplado del principio al final”.

Con respecto a las bases de datos genéticos consultadas durante todo el proceso (tales como HPO [6], OMIM [7], Genecards [37], Phenomizer [8] y otras) cree que no es necesario integrarlas a la herramienta debido al bajo volumen de Exomas que se analizan por mes. Le vería sentido si el volumen de trabajo lo requiere, es decir si se tuvieran 100 Exomas por mes entonces significaría un aumento significativo en los tiempos de evaluación diagnóstica, pero por ahora no lo ve imprescindible.

Como salvedad observó que según él la herramienta es para biólogos moleculares, bioinformáticos, bioquímicos o laboratorios de genómica ya que del lado de un genetista clínico él ya recibe el informe de resultados elaborado y no realiza todo el análisis.

Sus respuestas al cuestionario se encuentran en el [Apéndice 4. Cuestionario SUS – Gabriel Ercoli](#).

6.3.2 Caso de Validación: Mg. Matias Butti, Director de la empresa Zoigen & genomIT

El Mg. Matias Butti observó que, por cada exoma que nos derivan médicos de diferentes especialidades necesitamos, luego de correr el proceso bioinformático que identifica las variantes respecto al genoma de referencia, interpretar cada una de las variantes y generar el reporte. Una herramienta de este estilo la usaríamos para cada exoma que procesamos.

Con respecto a su usabilidad el Mg. Matias Butti la considera fácil de usar y cree que mejora notablemente el uso de un Excel para la interpretación de variantes. La automatización en la generación del reporte también es un plus. Es una buena primera aproximación a la herramienta que dispondría en su equipo para la interpretación de variantes de exoma en el contexto de enfermedades raras y de cáncer hereditario.

Los términos utilizados demuestran un proceso de abstracción adecuado para su mejor comprensión. Mencionó que, de todas maneras, agregaría tooltips (o algún otro elemento más avanzado de HCI) sobre cada una de las siglas para explicar de qué se tratan.

Cree que no necesitaría soporte de un técnico para utilizar la herramienta: “Un buen manual de usuario y videos de uso alcanzarían para que yo o cualquier otro especialista en el tema pueda comprender su uso”. El sistema es intuitivo. Un usuario no experto debería primero capacitarse en el tema, pero esto es esperable como consecuencia de su complejidad.

Agregaría una opción para trabajar más en los datos de la genealogía. Al no ser estructurada podría ser difícil aprovechar dicha información para automatizar un sistema de

soporte a la toma de decisiones. Por ejemplo, al cargar un síntoma en el caso índice, poder indicar también quienes son los familiares que lo tienen.

El Mg. Matias Butti cree que sería conveniente integrar las consultas a bases de datos genéticos realizadas por fuera de la herramienta” Probablemente sea conveniente bajar las bases de datos e importarlas a tu plataforma en vez de hacer una integración online (por una cuestión de performance). Es la estrategia que solemos seguir”.

Sus respuestas al cuestionario se encuentran en el [Apéndice 5. Cuestionario SUS – Matias Butti](#)

7. Conclusiones

7.1 Conclusión general

El objetivo de este trabajo fue proveer al médico genetista de una herramienta que permita facilitar el proceso de análisis e interpretación de variantes genéticas en un estudio de exoma humano. Durante el transcurso de las sucesivas entrevistas el genetista fue mostrando su trabajo a través de distintos escenarios, especificaciones, actividades, ejecución dinámica del filtrado de variantes a través de un Excel, el rol de cada filtro, la navegación por las distintas solapas, representación de los datos manejados y de las bases de datos consultadas en el proceso, datos recopilados del paciente previo a la ejecución del filtrado y la conclusión final del diagnóstico plasmada en dos informes.

Los escenarios fueron analizados cuidadosa y detalladamente a través de pequeñas entrevistas que revelaron necesidades, un objetivo que es arribar a un diagnóstico detectando las variantes candidatas a ser patogénicas y también una visión a futuro: detectar aquellas enfermedades que son frecuentes o altamente frecuentes en la población argentina y retroalimentar un panel bioinformático con datos recolectados en el proceso de diagnóstico de enfermedades en dicha población argentina. La relación entre los distintos escenarios fue representada desde la perspectiva de su trabajo diario y de su experiencia.

La suma de todas estas tareas condujo al modelado del proceso de toma de decisiones, la especificación de los requerimientos necesarios para la herramienta y al diseño de una interfaz de usuario capaz de cumplir con el comportamiento necesario, en interacción con el usuario, para llevar a cabo la tarea.

Por último, se diseñó la solución Exoma Browser que, en base al modelado del proceso de toma de decisiones llevado a cabo por el genetista, provee asistencia para arribar al diagnóstico final con mayor rapidez acelerando los tiempos de interpretación, permitiendo la retro-alimentación de diagnósticos futuros y habilitando un panel bio-informático propio.

Esta tesis aporta a la comunidad médica, dedicada al área genética, elementos para poder realizar un diagnóstico integral en base al análisis del Exoma Humano y generar una base de datos genéticos propia con el registro del trabajo realizado, pudiendo de esta manera mejorar los tiempos de evaluación diagnóstica, tener la información centralizada y contar con información para diagnósticos futuros. La herramienta ha sido validada con dos profesionales diferentes, el primero con amplia experiencia dentro de la genética clínica y el segundo con amplia experiencia en el campo de la bioinformática. Ambos se han mostrado satisfechos en cuanto a la usabilidad y aportes de la herramienta lo cual demuestra que cumple con los objetivos pautados.

7.2 Trabajo Futuro

Como principal trabajo futuro se sugiere que, en caso de ser necesario porque el volumen de trabajo lo justifique, sea integrada la consulta manual a las bases de datos de datos genéticas tales como OMIM, Phenomizer, entre otras.

Para tal fin, se podrían construir aplicaciones de orquestación utilizando las apis que cada una de las bases publica para su consulta. De esta manera se unificaría, en una sola herramienta, todo el trabajo realizado por el genetista.

Podría también construirse un sistema de recomendación basado en la lectura previa de las mencionadas apis.

Podría también a futuro ser desarrollado un módulo de contribución e integración a las bases de datos de estadísticas de enfermedades detectadas en la población argentina, a efectos de tomar decisiones de prevención en las patologías más frecuentes.

Para concluir, el diseño de la interfaz de usuario puede utilizarse como punto de partida para un proceso de desarrollo e implementación de la herramienta Exoma Browser.

Bibliografía.

[1] Óscar Pastor, Ana Palacio León, José Fabián Román Reyes, Alberto Simón García and Juan Carlos Rodenas Casamayor (2020). Using conceptual modeling to improve genome data management.

Briefings in Bioinformatics, Oxford. doi: 10.1093/bib/bbaa100

[2] Carlos Iñiguez-Jarrín, José Ignacio Panach, Oscar Pastor López (2018). Defining Interaction Design Patterns to Extract Knowledge from Big Data. Springer International Publishing AG, part of Springer Nature 2018. https://doi.org/10.1007/978-3-319-91563-0_30.

[3] Investing.com. (2022). <https://www.investing.com/stock-screener/>

[4] David Clark, Nanette Pazdernik, Michelle McGehee (2018). Molecular Biology, 3rd Edition.

[5] National Library of Medicine. National Center for Biotechnology Information. (2013, August) What is next generation sequencing?. doi: 10.1136/archdischild-2013-304340

[6] Human Phenotype Ontology. (last release: 2022, April). <https://hpo.jax.org/app/>

[7] Department of Genetic Medicine Johns Hopkins University. (Last release: 2022, August). An Online Catalog of Human Genes and Genetic Disorders. <https://www.omim.org/>

[8] Köhler et al. (2009). Clinical diagnostics in human genetics with semantic similarity searches in ontologies. Am J Hum Genet. Vol. 85 (4) pp. 457-64.

Köhler et al. (2017). The Human Phenotype Ontology. Nucleic Acids Research. doi: <https://doi.org/10.1093/nar/gkw1039>.

<http://compbio.charite.de/phenomizer/>

[9] National Human Genome Research Institute (Last Release: 2022, July). Glosario parlante de términos genómicos y genéticos. <https://www.genome.gov/es/genetics-glossary/Genomica>.

[10] National Human Genome Research Institute (Last Release: 2022, July). Glosario parlante de términos genómicos y genéticos. <https://www.genome.gov/es/genetics-glossary/Bioinformatica>

[11] Ng PC, Levy S, Huang J, Stockwell TB, Walenz BP, et al. (2008). Genetic Variation in an Individual Human Exome. *PLoS Genetics* 4(8): e1000160. doi:10.1371/journal.pgen.1000160

[12] Sebastian Menazzi, Matias Butti. (2020) Ejemplo Exoma. https://docs.google.com/spreadsheets/d/1qN_GQj3_WgMkX_EWv5iF7ZduGcu5Awlr/edit#gid=1764345462

[13] Samir Das, Xavier Lecours Boucher, Christine Rogers, Carolina Makowski, François Chouinard-Decorte, Kathleen Oros Klein, Natacha Beck, Pierre Rioux, Shawn T. Brown, Zia Mohaddes, Cole Zweber, Victoria Foing, Marie Forest, Kieran J. O'Donnell, Joanne Clark, Michael J. Meaney, Celia M. T. Greenwood, Alan C. Evans (2018, December). McGill Centre for Integrative Neuroscience, Montreal Neurological Institute, McGill, University, Douglas Hospital Research Centre, McGill University, Ludmer Centre for Neuroinformatics & Mental Health, Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, QC, Canada. *Integration of "omics" Data and Phenotypic Data Within a Unified Extensible Multimodal Framework*.
doi: 10.3389/fninf.2018.00091

[14] Congenica Ltd, Wellcome Genome Campus, Hinxton, Cambridge, UK (2016). Sapiientia: accelerating rare disease diagnosis and treatment.

[15] Daniel Koile, Marta Cordoba, Maximiliano de Sousa Serro, Marcelo Andres Kauffman, Patricio Yankilevich (2018). GenIO: a phenotype-genotype analysis web server for clinical genomics of rare diseases. Koile et al. *BMC Bioinformatics* (2018). Doi: 10.1186/s12859-018-2027-3.

[16] Ramos Mejía RM, Conicet, Universidad Austral Ciencias biomédicas, Sociedad Max Planck para la Promoción de las Ciencias (2017, November). <https://bioinformatics.ibioba-mpsp-conicet.gov.ar/GenIO/>

[17] Jianping Jiang¹, Jianlei Gu, Tingting Zhao, Hui Lu. (2019). VCF-Server: A web-based visualization tool for high-throughput variant data mining and management. *Molecular Genetics and Genomics Medicine*, Wiley.

[18] <https://www.diseasegps.org/VCF-Server?lan=eng>

[19] Mete Akgün, Hüseyin Demirci. (2017). VCF-Explorer: filtering and analyzing whole genome VCF files. Published by Oxford University Press.

[20] Sourforge Headquarters (Last Release: 2016, July). <https://sourceforge.net/projects/vcfexplorer/>.

[21] Jinlian Wang, Jun Liao, Jinglan Zhang, Wei-Yi Cheng, Jörg Hakenberg, Meng Ma, Bryn D. Webb, Rajasekar Ramasamudram-chakravarthi, Lisa Karger, Lakshmi Mehta, Ruth Kornreich, George A. Diaz, Shuyu Li, Lisa Edelmann, Rong Chen (2015). ClinLabGeneticist: a tool for clinical management of genetic variants from whole exome sequencing in clinical genetic laboratories. Wang et al. *Genome Medicine* (2015) 7:77. Doi: 10.1186/s13073-015-0207-6

[22] Rong Chen Laboratory (Last release: 2015). Tool to assist geneticist for diagnosis of pathogenic variants from WES data <http://rongchenlab.org/software/clinlabgeneticist>.

[23] Interaction Design Foundation. User Scenarios. <https://www.interaction-design.org/literature/topics/user-scenarios>

[24] Alistair Sutcliffe (2003). Scenario-based Requirements Engineering. University of Manchester Institute of Science & Technology (UMIST) PO Box 88, Manchester, M60 1QD, UK

[25] John M. Carroll, Massachusetts Institute of Technology (2000). Making Use Scenario-Based Design of Human-Computer Interactions. The MIT Press Cambridge, Massachusetts London, England.

[26] Björn Pullwer, Anke Dittmar. University of Rostock, Department of Computer Science, Rostock. (2015, July). Prototypes as Tools for Discovery: A Case Study on Multi-Touch UML Modeling.

Doi: <http://dx.doi.org/10.1145/2788412.2788422>

[27] Ben Shneiderman, Catherine Plaisant, Maxine Cohen, Steven Jacobs, Niklas Elmqvist (2018). Designing the User Interface Strategies for Effective Human-Computer Interaction. Six edition. Ed. Pearson.

[28] Sue Richards, Nazneen Aziz, Sherri Bale, David Bick, Soma Das, Julie Gastier-Foster (2015). Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 405–424. doi:10.1038/gim.2015.30.

[29] National Library of Medicine. National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov/clinvar/>

[30] D.N. Cooper, E.V. Ball, P.D. Stenson, A.D. Phillips, K. Evans, S. Heywood, M.J. Hayden, M.M. Chapman, M.E. Mort, L. Azevedo, D.S. Millar. The Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff. <http://www.hgmd.cf.ac.uk/>

[31] Pauline Ng. Server support by Bioinformatics Institute in Singapore (Last release: 2017, March) Sorting Intolerant from Tolerant. <https://sift.bii.a-star.edu.sg/>

[32] PolyPhen-2 prediction of functional effects of human nsSNPs. <http://genetics.bwh.harvard.edu/pph2/>

[33] University of Washington, Hudson-Alpha Institute for Biotechnology and Berlin Institute of Health (2013-2021). Combined Annotation Dependent Depletion. <https://cadd.gs.washington.edu/snv>

[34] Berlin Institute of Health (2021). Mutation Taster. <http://www.mutationtaster.org>

[35] European Molecular Biology Laboratory of European Bioinformatics Institute. (Last Release: 2022, March). Protein sequence and functional information. <https://www.uniprot.org/>

[36] SIB - Swiss Institute of Bioinformatics, CPR - Novo Nordisk Foundation Center Protein Research, EMBL - European Molecular Biology Laboratory (Last release: 2022) . Protein-Protein Interaction Networks Functional Enrichment Analysis. String Database. <https://string-db.org/>

[37] Weizmann Institute of Science of Israel, LifeMap Sciences of USA. (Last release: 2017). The GeneCards human gene database. <https://www.genecards.org/cgi-bin/carddisp.pl?gene=CFTR>

[38] ZonaValue.com. Global Financial Marketplace, Valencia. (2022, March) <https://zonavalue.com/blog/estrategias-de-inversion/que-es-un-stock-screener-y-cuales-son-sus-ventajas>

[39] Cecilia Moreyra. Universidad Nacional de La Plata (Last release: 2022, February). Prototipo de la aplicación Exoma Browser. <https://exoma-browser.000webhostapp.com/>

[40] Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications, 37(4), 445-456.

[41] <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>

Apéndices

Apéndice 1. Informe de resultados



INFORME DE RESULTADOS ESTUDIO DE EXOMA POR SECUENCIACIÓN MASIVA ANÁLISIS DE HALLAZGOS CLÍNICAMENTE RELEVANTES

Paciente: ----
Fecha de nacimiento: ----
DNI: ----

Recepción de la muestra: 15/01/2020
Fecha de informe: 18/05/2020

ESTUDIO SOLICITADO

Exoma clínico de paciente afectado con informe de hallazgos relevantes relacionados con fenotipo.

INDICACIÓN CLÍNICA:

Paciente de 40 años evaluado por antecedente personal de poiquiloderma congénito y acortamiento muscular, sin antecedentes familiares de relevancia. Diagnóstico presuntivo de síndrome de Rothmund Thomson.

RESULTADOS

Se evaluaron inicialmente genes relacionados con síndrome de Rothmund Thomson, incluyendo *AGPAT2*, *ALDH18A1*, *B3GALT6*, *B4GALT7*, *BANF1*, *BLM*, *BSCL2*, *ERCC2*, *ERCC3*, *ERCC4*, *ERCC5*, *ERCC6*, *ERCC8*, *FBN1*, *LMNA*, *PDGFRB*, *POLD1*, *PYCR1*, *RECQL4*, *SLC25A24*, *WRN* y *ZMPSTE24*. El porcentaje de cobertura > 20X de estos genes se consideró adecuado para proceder con el análisis de variantes de secuencia. Este análisis no arrojó ningún resultado potencialmente asociado con el fenotipo o de significancia incierta. A continuación se procedió con la ampliación del análisis bioinformático a 42 genes relacionados con diagnósticos diferenciales de este cuadro.

Se identificó la presencia de la siguiente variante de secuencia asociada con el fenotipo de la paciente:

Gen	Cambio nucleotídico*	Proteína	Efecto	Cigotidad	Cobertura
<i>FAM111B</i>	c.1882A>C	p.Ser628Arg	Probablemente patogénica	Heterocigosis	[90,69](159X)

*Se ha empleado la nomenclatura recomendada por la Human Variation Society (HGVS v2.0) [1].

INTERPRETACIÓN CLÍNICO-BIOLÓGICA

La variante c.1882A>C (Figura 1) en el transcrito NM_198947.3 del gen *FAM111B* (OMIM *615584) [2] no ha sido descrita previamente en bases de datos poblacionales, incluyendo GnomAD o ExAC, o en bases de datos de variantes como ClinVar [3]. La variante es una sustitución de una adenina por una citosina con predicción de cambio del aminoácido serina en la posición 628 del gen por una arginina. El mismo codón ha sido reportado previamente afectado por una variante confirmada como patogénica, y se han descrito otras variantes *missense* en codones circundantes que también se asocian con la patología [4, 5, 6]. En investigaciones posteriores se reportó como patogénica una variante en la que la serina es reemplazada por arginina, mismo aminoácido que la variante presente en el paciente, lo que refuerza la predicción de su patogenicidad [7]. El dominio de la proteína en cuestión, en el cual se identifica una cadena de serinas y cisteínas, podría ser considerado por lo tanto un *hotspot* para variantes de tipo *missense*.

Con la información disponible hasta el momento, clasificamos esta variante como **probablemente patogénica**.

Dado que el patrón de herencia del síndrome relacionado con variantes patogénicas en *FAM111B* es autosómico dominante, el hallazgo de una variante en heterocigosis confirmaría el diagnóstico de esta patología. Se sugiere asesoramiento genético.

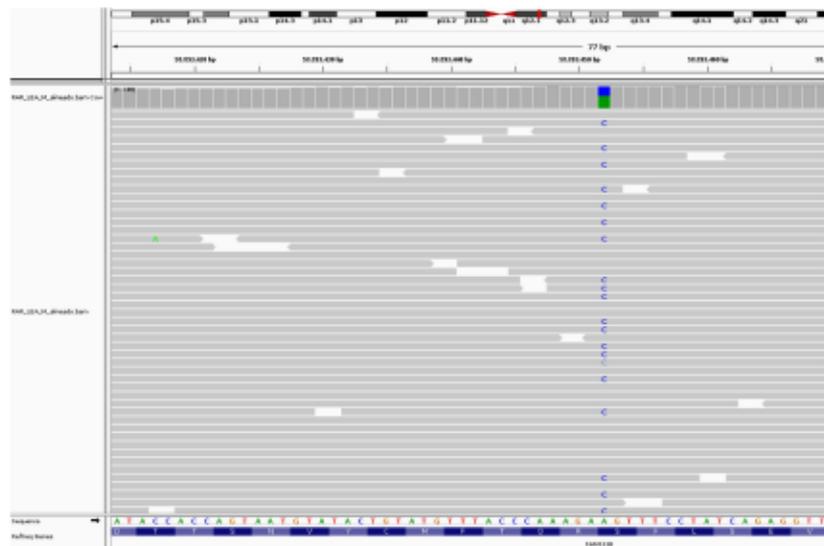


Figura 1: Variante *missense* c.1882A>C en el gen *FAM111B*.

METODOLOGÍA (estudio de Secuenciación del Exoma completo)

1. Cuantificación del ADN a partir de las muestras remitidas.
2. Preparación de una librería de fragmentos del genoma.

3. Selección de las regiones objeto de estudio usando el método de captura SureSelectXT Human All Exon V7 (Agilent Technologies). Estas regiones incluyen exones y regiones intrónicas adyacentes de los genes objeto de estudio.

4. Amplificación clonal y secuenciación de las regiones seleccionadas en la plataforma Illumina HiSeq siguiendo la estrategia de paired-ends.

5. Análisis bioinformático. Se alinean las secuencias de ADN obtenidas contra el genoma de referencia humano (hg19) y se lleva a cabo la búsqueda de variantes. Este análisis considera variantes aquellas alteraciones que poseen una cobertura total $\geq 20X$. Luego, las variantes son anotadas contra diversas bases de datos como ClinVar, dbSNP, ExAC, entre otras.

6. Teniendo en cuenta la información disponible del paciente, el algoritmo diagnóstico utilizado en este estudio consistió en:

- Selección de genes descritos en las bases de datos OMIM y GTR como genes asociados a enfermedad.
- Análisis de las variantes descritas en ClinVar y variantes con efecto deletéreo (frameshift, stop codon, nonsense, essential splicing, etc).
- Análisis de las variantes de significado incierto (inserciones o deleciones in frame, missense, sinónimas o variantes en regiones de splicing +/-8nt) en los genes evaluados.

Limitaciones de la técnica:

- La cobertura global para esta muestra depende de las características de las regiones analizadas, pudiendo existir regiones con una cobertura menor a 10X. Las regiones especialmente sensibles a la disminución de cobertura son aquellas con elevada complejidad (alto contenido en GC, regiones repetitivas, etc.).
- Mutaciones por inserción de más de 4 nucleótidos o por deleción de más de 11 nucleótidos no pueden ser detectadas mediante la metodología empleada.
- La tecnología NGS basada en la captura/enriquecimiento mediante hibridación de sondas, no permite distinguir entre regiones de alta homología (genes homólogos, pseudogenes, familias de genes, etc.) pudiendo dar lugar a falsos negativos.
- Este estudio se basa en la secuenciación de la región codificante y regiones intrónicas adyacentes del gen. Grandes deleciones o duplicaciones y reordenamientos, o mutaciones que afecten a regiones intrónicas profundas del gen, no son detectadas mediante la metodología utilizada.
- Debido a limitaciones técnicas las regiones de homopolímeros (repeticiones de 8 o más pares de bases) han sido excluidas del estudio.
- Esta metodología no permite la detección de mosaicismos en baja proporción.
- El estudio mediante este diseño de exoma no incluye los genes localizados en el genoma mitocondrial.

INFORMACIÓN SUPLEMENTARIA

El análisis específico de variantes patogénicas presentes en aquellos genes definidos por la ACMG como accionables (59 genes) en los individuos incluidos en el estudio, se encuentra disponible previa petición por parte del profesional tratante y previo envío de los consentimientos informados de los individuos incluidos en el estudio.

REFERENCIAS

1. Den Dunnen and Antonarakis (2001). Hum Genet 109: 121 -124.
2. Online Mendelian Inheritance in Man. OMIM #166200. (omim.org/)
3. NCBI ClinVar database (ncbi.nlm.nih.gov/clinvar/)
4. Mercier S et al. Am J Hum Genet. 2013 Dec 5;93(6):1100-7.
5. Mercier S et al. Orphanet J Rare Dis. 2015 Oct 15;10:135.
6. ncbi.nlm.nih.gov/clinvar/variation/120219
7. Goussot R et al. JAAD Case Rep. 2017 Mar 19;3(2):143-150.

Notas:

- Se asume la existencia del consentimiento informado para proceder a la realización del presente estudio genético.
- Este test fue desarrollado y su performance determinada por genomIT SA (genomit.com.ar).
- Los datos obtenidos en este estudio son confidenciales y deben ser manejados sobre la base de estrictos principios de privacidad.
- La interpretación biológica y clínica realizada en el informe se ha hecho en base a la bibliografía disponible hasta la fecha de emisión del informe.
- Conservación de datos: los datos, en caso que el paciente así lo autorice en el consentimiento informado, serán anonimizados para luego ser utilizados con fines de investigación.
- El resultado del estudio debe ser interpretado por el médico especialista dentro del contexto clínico del paciente.



Dr. Sebastián Menazzi
Médico genetista
MN 135.855
Director Médico – genomIT

Apéndice 2. Informe clínico



INFORME DE ASESORAMIENTO GENÉTICO ESTUDIO DE EXOMA POR SECUENCIACIÓN MASIVA Y ESTUDIOS COMPLEMENTARIOS

Paciente: ----	Fecha de nacimiento: ----
DNI: ----	Fecha de informe: 18/05/2020

ESTUDIO SOLICITADO

Exoma clínico de paciente afectado con informe de hallazgos relevantes relacionados con fenotipo

INDICACIÓN CLÍNICA:

Paciente evaluado por poiquiloderma congénito y acortamiento muscular.

Diagnóstico presuntivo de síndrome de Rothmund Thomson.

ANTECEDENTES PERSONALES Y FAMILIARES

Paciente de 40 años evaluado por antecedente personal de poiquiloderma congénito, diagnosticado a los 5 años, con acortamiento muscular detectado en la adolescencia. A nivel cutáneo presenta rash e hipersensibilidad cutánea, tiene escasa sudoración, cejas ralas, limitación a extensión completa de codos y dedos. Adicionalmente presentó estrabismo en la infancia, que requirió cirugía a los 12 años, y actualmente tiene hipermetropía, hipotiroidismo, hipoparatiroidismo, malabsorción intestinal, varicocele, osteopenia y edemas en miembros inferiores con asimetría (suele tener mal retorno linfático). No se constatan antecedentes familiares de relevancia. Desarrollo neuromadurativo normal.

RESULTADOS

Se identificó la presencia de la siguiente variante asociada con un diagnóstico diferencial del presuntivo, que es compatible con el fenotipo del paciente, en el gen *FAM111B*: c.1882A>C. Se trata de una variante novel y fue definida como probablemente patogénica.

ASESORAMIENTO GENÉTICO

Las variantes patogénicas y probablemente patogénicas en *FAM111B* se han asociado con un cuadro denominado Poiquiloderma Hereditario Fibrosante con Contractura de Tendones, Miopatía y Fibrosis Pulmonar (POIKTMP). Se trata de un trastorno muy poco frecuente, con no más de 26 individuos identificados afectados a nivel mundial, y se caracteriza por aparición temprana de poiquiloderma (dermatosis que cursa con placas de piel atrófica, eritema con telangiectasias y áreas de hiperpigmentación que alternan con hipopigmentación), especialmente en mejillas, acompañado por hipotricosis, hipohidrosis y linfedema leve de las extremidades. Adicionalmente se observa contracturas musculares múltiples (por alteraciones tendinosas, especialmente en miembros inferiores), miopatía con debilidad muscular progresiva, restricción pulmonar (debida o no a fibrosis pulmonar), insuficiencia del páncreas exocrino y ocasionalmente anomalías hepáticas, hematológicas, oculares o tiroideas. Se ha descrito variabilidad clínica intrafamiliar, con personas afectadas sólo por el fenotipo cutáneo sin las contracturas musculares, y si bien las manifestaciones cutáneas suelen mejorar con el tiempo, las extracutáneas pueden acentuarse. La penetrancia del cuadro sería de 100%. Dado que el síndrome de Rothmund Thompson representa el principal diagnóstico diferencial del cuadro, es frecuente que este sea la primera sospecha durante la infancia de los pacientes.

Si bien el hipoparatiroidismo no se ha descrito asociado específicamente con este síndrome, y representaría un hallazgo novel en el paciente, en el síndrome relacionado con variantes en *FAM111A* (cercano a *FAM111B*) se han detectado alteraciones en el metabolismo fosfocálcico, lo que podría sugerir una interacción entre ambos genes.

Esta entidad por el momento no cuenta con tratamientos curativos o dirigidos a la anomalía molecular. Sin embargo, se recomiendan diversas evaluaciones periódicas:

- Dermatología
- Neurología (idealmente un servicio con experiencia en patología neuromuscular), con eventual derivación a ortopedia o indicación de tratamientos de fisioterapia o kinesiología
- Neumonología (control regular de la función pulmonar)
- Laboratorios de función hepática y parámetros hematológicos
- Gastroenterología y hepatología, en caso de constatarse anomalías hepáticas o malabsorción
- Evaluación oftalmológica

Asimismo, se sugiere adoptar medidas de prevención para factores de riesgo conocidos y evaluar terapias dirigidas al manejo sintomático:

- Evitación de la exposición excesiva a luz solar o fuentes importantes de calor (incluyendo control rápido de procesos febriles)
- Drenaje linfático manual o compresiones para control de linfedema
- Tratamientos tópicos o intervenciones cosméticas para lesiones cutáneas (según decisión del paciente y su médico tratante)
- Terapia física y ejercicio para reducir el impacto de las contracturas
- Suplementación con calcio y vitamina D para evitar osteopenia secundaria a debilidad muscular
- Ventilación no invasiva (de ser necesario)
- Terapia de reemplazo con enzimas pancreáticas (en casos de insuficiencia) o ácido ursodesoxicólico (ante colestasis)
- Corrección quirúrgica de cataratas

El patrón de herencia descrito para la patología es autosómico dominante, lo que implica que al menos un alelo (copia del gen) se encuentra alterado. En el análisis de la genealogía del paciente no se identifican casos similares al suyo, por lo que se interpreta un fenómeno de aparición *de novo* de la variante en él. Esto implica bajo riesgo de la enfermedad en otros miembros de su familia, a excepción de su eventual descendencia biológica.

En casos de parejas en las que un individuo sea afectado y deseen buscar embarazo el riesgo de recurrencia en su descendencia se estima en 50% por embarazo, en forma independiente del sexo. Existen en la actualidad técnicas en medicina reproductiva capaces de reducir el riesgo de transmisión de una variante patogénica para una patología autosómico dominante a la descendencia en los casos de alto riesgo (un padre afectado por la enfermedad), incluyendo el uso de gametas donadas, el diagnóstico prenatal y el diagnóstico genético preimplantatorio para enfermedades monogénicas (PGT-M), que deben ser discutidas con el médico tratante en especial en casos con alto riesgo que se encuentren en edad reproductiva.

Estas recomendaciones surgen de la evidencia médica actual, los datos referidos por el consultante y los estudios realizados durante el asesoramiento, por lo tanto pueden modificarse ante la aparición de nueva evidencia médica o información familiar. Sugerimos continuar en contacto periódico de actualización para eventual cambio de recomendaciones.

El laboratorio permanece a disposición de la consultante, su familia y el médico tratante para el asesoramiento de aquellos que lo requieran o futuras consultas de actualización.



Dr. Sebastián Menazzi
Médico genetista
MN 135.855
Director Médico – genomIT

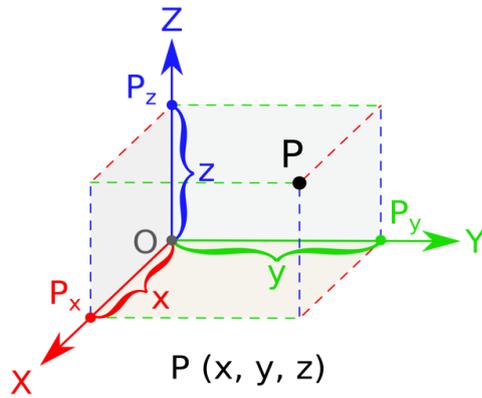
Apéndice 3. Minutas de Reunión

Fecha	06/09/2020
Modalidad	Virtual
Objetivo	Presentación del dominio. Exoma su definición y sus solapas y columnas. Escenarios donde se indica. Proceso de secuenciación del exoma. Proceso de análisis e interpretación del exoma. Posible visualización de los datos.
Asistentes	Sebastián Menazzi, Cecilia Moreyra, Matias Butti, Andres Rodriguez, Alejandra Garrido.
Duración	2 horas

Se presentó un ejemplo de Exoma [9], sus solapas, campos y escenarios donde se indica. También, términos y conceptos relacionados al dominio a saber:

- Matias Butti: Definición de exoma.
- Matias Butti: Escenarios en los que se indica el exoma:
Escenario 1: Enfermedades monogénicas en las que la variante que la explique puede estar en uno de varios genes (más de 20-30).
Escenario 2: 59 Incidental findings.
Escenario 3: Todas las variantes.
- Matias Butti: Solapas del exoma y su finalidad:
Variantes genes fenotipo, Variantes Pathogenic y Likely Pathogenic, Incidental Findings, Todo.
- Matias Butti: Columnas del exoma y su significado. Campos descriptivos y campos anotados. Detalle de las columnas en el archivo Excel.
- Matias Butti: Proceso de secuenciación del exoma, su definición.

- Matias Butti: Proceso de análisis e interpretación del exoma, su definición y la forma en la que se lleva a cabo.
- Matias Butti: Posible Visualización de los Datos:



Si Z es frecuencia poblacional (de mayor a menor), Y es concordancia con fenotipo (menor a mayor) y X es patogenicidad según predictores, el punto P indicaría una variante que probablemente sea la causal del problema del paciente (candidata)

Se podría así visualizar rápidamente qué variante o grupo de variantes tengo que ver en mayor detalle, en lugar de recurrir a filtros del excel

- Matias Butti: Ejes de información utilizada sobre el proceso de interpretación.

Fecha	10/09/2020
Modalidad	Virtual
Objetivo	Registro de la historia clínica del paciente. Priorización de genes a través del panel bioinformático. Presentación de los filtros del exoma.
Asistentes	Matias Butti, Sebastián Menazzi, Cecilia Moreyra.
Duración	2 horas

Etapa previa al filtrado de variantes:

- Sebastián Menazzi:

Registro de la historia clínica del paciente, campos que se registran.

Priorización de genes a través del panel bioinformático. Búsqueda de los términos HPO relacionados a la patología del paciente en el panel bioinformático para obtener el conjunto de genes candidatos. Búsqueda por enfermedad en el panel bioinformático para obtener el conjunto de genes candidatos.

Etapa de filtrado de variantes. Presentación de los filtros y sus finalidades.

- Sebastián Menazzi: El transcripto es un ID, está dentro de las columnas que como finalidad tienen la “identificación de la variante”. Los otros campos que están anotados tienen la “información biológica de la variante”.
- Sebastián Menazzi: Una variante puede nombrarse distinto según el transcripto del gen al cual nos estemos refiriendo con efectos distintos sobre la proteína y la función resultante.

Es por esto que en las columnas **ANN[*].GENE, ANN[*].HGVS_C, ANN[*].HGVS_P, ANN[*].FEATUREID, ANN[*].EFFECT, ANN[*].IMPACT, dbNSFP_Polyphen2_HDIV_pred, dbNSFP_Polyphen2_HVAR_pred, dbNSFP_SIFT_pred, dbNSFP_MutationTaster_pred y dbNSFP_MutationAssessor_pred** se presenta como contenido una cadena de caracteres alfanuméricos separados con comas en representación de las distintas

posiciones que refieren, cada una respectiva a la otra, a un determinado transcripto de interés.

- Sebastián Menazzi: Columna ID: Lo que unifica transcriptos y variantes, es para mutaciones que ya son conocidas.

Sebastián Menazzi: Hay dos maneras de identificar una variante, una es con el TRS, sinó con el transcripto y la notación HGVS.

- Sebastián Menazzi: La columna ANN[*].HGVS_C puede sintetizarse tomando solo el primer valor (previo a la primer coma) ejemplo:

n.*945C>T,n.*945C>T,n.*945C>T,n.*986C>T,n.*993C>T,n.41535G>A,n.666G>A,n.659G>A,n.497G>A quedaría como **n.*945C>T**

Entendiéndose que el resto de los valores serían sinónimos pero referidos a otros transcriptos. El resto de los datos debe quedar igualmente visible.

- Sebastián Menazzi: La columna ANN[*].FEATUREID C puede sintetizarse tomando solo el primer valor (previo a la primer coma) ejemplo:

ENST00000438504,ENST00000541675,ENST00000423562,ENST00000488147,ENST00000538476,ENST00000450305,ENST00000456328,ENST00000515242,ENST00000518655 quedaría como **ENST00000438504.**

El resto de los datos debe quedar igualmente visible.

- Sebastián Menazzi: Ambas simplificaciones pueden leerse como: “La variante n.*945C>T se encuentra en el transcripto ENST00000438504.

- Sebastián Menazzi: Hay variantes que están en los transcriptos. Los transcriptos provienen de genes.

- Sebastián Menazzi: Los transcriptos se consultan de dos bases de datos diferentes:

La europea:

https://www.ensembl.org/Homo_sapiens/Transcript/ldhistory?t=ENST00000438504

La americana: https://www.ncbi.nlm.nih.gov/nuccore/NM_004006.2

(Ref Seq) los transcritos empiezan con NM

- Sebastián Menazzi: Columnas ANN[*].HGVS_C y ANN[*].HGVS_P: Toda variante en proteína tiene asociada una variante en un gen pero no necesariamente es así a la inversa. Todas las que tienen variante p por definición tienen una variante c pero puede haber algunas que generalmente se describen como n que no tengan un correspondiente en p.
- Sebastián Menazzi: Notación HGVS. Links para accederla: <https://varnomen.hgvs.org/> o <https://varnomen.hgvs.org/recommendations/DNA/variant/substitution/> más puntualmente
- Sebastián Menazzi: La columna ANN[*].HGVS_P puede ser utilizada solo a efectos informativos ya que los predictores bioinformáticos y la frecuencia poblacional están referidas al c no al p.
- Sebastián Menazzi: La columna ANN[*].EFFECT se puede recortar la cadena “_gene_variant” las variantes que comienzan con “splice_region_variant&intron” pueden reemplazarse como “splice”.
- Sebastián Menazzi: La columna ANN[*].IMPACT puede ser utilizada solo a efectos informativos.
- Sebastián Menazzi: Los predictores bioinformáticos de patogenicidad (columnas finalizadas en “_pred”) dentro de las celdas que no tienen una letra pueden tomarse como vacías. Las celdas que poseen una letra puede estar acompañada de otros caracteres, pero puede ser leídas solo como esa letra.
- Sebastián Menazzi: La columna CLNSIG: es necesario realizar una equivalencia a los campos que describen la significancia clínica de la variante:
 - Malign / Likely malign
 - Benign / likely Benign

- VUS

- Sebastián Menazzi: Columna CLNSIGCONF desestimarla.
- Cecilia Moreyra: Matias: Necesitaría que me digas si a nivel filas del exoma existe algún criterio de simplificación. Tal como veníamos hablando con las columnas.
- Matias Butti: Cada fila representa una variante genética diferente y no un metadato o anotado de otra variante, por lo tanto entiendo que no se puede simplificar o juntar con otras filas (excepto que se agrupen todas las filas que pertenecen a un mismo gen, teniendo en cuenta que el fenotipo que se busca es a nivel gen y no a nivel variante). Lo que ocurre a nivel de filas son filtros por distintos criterios. Por ejemplo, para buscar variantes (filas) que puedan explicar el cuadro sintomático de una enfermedad rara (menos de 1 en 2000) nos quedamos con solo las que tengan una frecuencia poblacional baja y luego las que sean pathogenic o likely pathogenic.
Si con simplificación te referís a agrupación para descartar el nivel más atómico, yo no lo veo posible porque la unidad de análisis es la variante.
Lo que sí se podría pensar, llevándolo a un modelo BI, son dimensiones jerárquicas sobre las que podés hacer drill down y roll up.
Aunque, mirando los campos la única dimensión que veo es estructura con 3 niveles: variante, gen, cromosoma. Luego algunos campos numéricos (serían las métricas en un modelo BI) se podrían recalcular según el nivel donde te pares y para otros campos no numéricos se podría mostrar estadística descriptiva de la agrupación. Por ejemplo, en el gen cuantas variantes patogénicas hay o la calidad promedio del gen.

Fecha	24/05/2021
Modalidad	Virtual
Objetivo	Presentación de los filtros del exoma (continuación). Definición de los pasos del proceso de filtrado del exoma. Conclusiones post-filtrado.
Asistentes	Sebastián Menazzi, Cecilia Moreyra.
Duración	1 hora y media

- Sebastián Menazzi: Columnas identificatorias y de metadatos:

CHROM, POS, REF, ALT, GEN[0].GT, GEN[0].AD, GEN[0].DP: son los datos específicos del paciente

Metadatos:

- **ID:** tiene un código inequívoco.
- **ANN[*].GENE:** Nombre del GEN, puede haber diferentes o repetidos, se observa el transcritto.
- **ANN[*].HGVS_C y ANN[*].HGVS_P:** Cambio a nivel del ADN copy
- **ANN[*].FEATUREID:** Transcritto, hay transcrittos diferentes que pertenecen al mismo gen o a veces puede haber transcrittos diferentes para genes distintos.
- **ANN[*].EFFECT y ANN[*].IMPACT:** son predictores de efecto (ver valores de cada uno). Muchas veces se suele desprender el diagnóstico de la enfermedad de la evaluación de dichos impactos. Puede haber una correlación con el diagnóstico.
- **Frecuencia poblacional:** Nota: en el ejemplo que nos proporcionó Matias en primera instancia las frecuencias están de la columna O hasta la R. En el ejemplo visto en esta reunión las frecuencias están desde la O hasta la Y. A partir de aquí se produce un desfasaje. La columna más utilizada es la

ExAC_AF por ser la más completa. Se filtran los cambios producidos con más frecuencia en la población

- **Predictores bioinformáticos:** nos dicen si es probable que esta variante/cambio va a generar una patología o no. Los valores que toman varían según el proveedor.
- **CLNSIG y CLNSIGCONF:** Columnas que nos proveen información clínica. Nota: Mismo caso que con la frecuencia poblacional. Difieren en cantidad entre el ejemplo provisto inicialmente con el visto en la reunión. Se pueden seleccionar algunos síndromes o cuadros genéticos. Pueden proporcionar información útil. CLNSIG es la de ClinVar, es la más utilizada entonces podemos ver si esa variante ya está reportada como patológica. Proveen información suplementaria.
- Sebastián Menazzi: Proceso de filtrado del exoma.
 - **Posible primer camino:** solapa Genes del Fenotipo
 - Primer filtro: ir a las columnas de información clínica, ejemplo **CLNSIG (Clinvar)** y ver si ya está la variante descrita como patogénica o no. Se podría eliminar las que figuran como benignas y evaluar las que son de significado incierto o de interpretaciones conflictivas de patogenicidad.
 - Segundo filtro importante sería la frecuencia poblacional en ExAC_AF y quitamos todas las variantes que estén presentes en más de 0.005 de la población. Las variantes surgidas de dichos filtros son revisadas una por una.
 - Tercer filtro importante: Análisis de los predictores bioinformáticos siendo los más destacados los que figuran como Damaging (D) o Pathogenic (P). A partir de aquí se analizan el resto de los datos.
 - Cuarto filtro: El patrón de herencia (columna GEN[0].GT). Podría por ejemplo quitarse las variantes que están en heterocigosis si

pensamos que se trata de una patología de origen dominante o las homocigosis si pensamos que se trata de una patología de origen recesivo. O podemos no quitar ninguna si se tratara de una heterocigosis compuesta.

- Teniendo en cuenta que Clinvar puede no proveer datos exactos se pueden ejecutar los siguientes filtros:
 - Primer filtro: sería la frecuencia poblacional en ExAC_AF y quitamos todas las variantes que estén presentes en más de 0.005 de la población.
 - Segundo filtro: columna **ANN[*].IMPACT** seleccionamos las que tienen un impacto alto quitando las de bajo impacto.
 - Tercer filtro: Predictores bioinformáticos si hay varios predictores que hablan de Damaging entonces posiblemente haya un efecto grave.
 - Cuarto filtro: Ver descripción en Clinvar si está como patogénica.

- **Posible segundo camino:** Solapa de Variantes Patogénicas

- **Posible tercer camino:** Solapa Incidental Findings:

Son hallazgos más allá de los genes del fenotipo específico. Se revisa porque podría haber alguna alteración que tenga un impacto accionable para el paciente.

Para todo el proceso tener en cuenta:

- El orden de los caminos es indistinto, la frecuencia poblacional es muy importante.
- Las solapas de variantes patogénicas e Incidental Findings se filtran bajo los mismos criterios, la solapa Patogénicas en realidad nos

sirve por si se nos escapó algún gen candidato del análisis inicial, y en cuanto a Incidental Findings lo único que cambia es que sólo se informan las variantes patogénicas claras, y no las que sean dudosas (mientras que en las otras solapas podemos a veces informar una dudosa).

Etapas posteriores al filtrado de las variantes.

- **Conclusiones**
 - Finalmente, se marcan con color las variantes sospechosas y se vuelve a revisar todo para re-confirmar la sospecha. De ser necesario, se consultan las ontologías/repositorios de enfermedades monogénicas para postular qué genes pueden ser causales del fenotipo que observamos (los candidatos que pondremos en la solapa correspondiente), o sea al principio del análisis. De todas formas OMIM puede ser consultada varias veces a lo largo del proceso para ver características del gen que estemos revisando, al igual que Genereviews, Clinvar, Uniprot, Genecards, LOVD, entre otras.

Fecha	20/11/2021
Modalidad	Virtual
Objetivo	Generación de informes post-diagnóstico. Retro-alimentación de diagnósticos futuros. Presentación de la especificación de los requerimientos de visualización.
Asistentes	Sebastián Menazzi, Cecilia Moreyra.
Duración	1 hora y media

- Sebastián Menazzi: Luego del registro de las variables patológicas se generan dos informes, el Informe de Resultados que se entrega al paciente y el Informe Clínico. Se presentan ejemplos de cada uno de ellos y se definen los campos que quedan registrados en los mismos.
- Sebastián Menazzi:
 - Definición de la frecuencia poblacional de enfermedades monogénicas en la población argentina.
 - Definición de la relación de la variante con el fenotipo.
 - Proceso de Retroalimentación del panel bio-informático.
- Cecilia Moreyra: Se presenta la especificación de los requerimientos de visualización para todas las etapas del proceso. Dicha especificación es aceptada por Sebastián Menazzi.

Fecha	21/09/2021
Modalidad	Virtual
Objetivo	Consultas acerca del proceso de filtrado y posterior diagnóstico.
Asistentes	Matias Butti, Sebastián Menazzi, Cecilia Moreyra.
Duración	1 hora y media

- Cecilia Moreyra: Trazabilidad: Hay un punto donde es necesario consultar manualmente la base de OMIM para ver qué genes están relacionados con el fenotipo del paciente en la solapa Pathogenic. Esto, podríamos automatizarlo?

Matías Butti: El pipeline no hace interpretación fenotípica, solo toma los datos crudos. Existe una API para poder consultar OMIM.

Sebastián Menazzi: La base de OMIM provee información que debería ser procesada y seleccionada por el genetista para su registro en la aplicación. Dicha información también se encuentra en Phenomizer conjuntamente con los términos HPO.

Matias Butti: El último paso sería la revisión manual de papers.

- Cecilia Moreyra: Retroalimentación: Puede existir una posible retroalimentación del proceso luego de ser finalizado que implica consultas a BD para definir criterios útiles que luego sirvan para futuras conclusiones. Esto, podríamos automatizarlo?

Matias Butti: si, puede ser automatizado.

Sebastián Menazzi: Se buscaría en nuestra propia base de datos en los casos donde se encuentren variantes específicas y existan casos anteriores de pacientes con

la misma mutación y no haya podido confirmarse la coincidencia con el fenotipo y después de X cantidad de casos confirmados se llega finalmente al diagnóstico.

Basados en el concepto de que la columna de frecuencia poblacional es representativa solo de países desarrollados, los países latinoamericanos se encuentran sub-representados, entonces en la frecuencia poblacional puede ocurrir que haya baja frecuencia lo que induce a pensar que no es patológico pero luego de encontrar la variante patológica en tal vez más de 500 exomas entonces se deduce que es mucho más frecuente acá que allá.

Variantes patológicas curadas anteriormente deberían generar alertas de aparecer en exomas posteriores. Variantes encontradas con frecuencia en la Argentina deben ser indicadas como frecuentes en nuestra población.

- Cecilia Moreyra: Viendo el patrón de comportamiento en el filtrado se observa que son aplicados exactamente los mismos filtros a través de las distintas solapas, solo que con distinto orden. El resultado final se presume sería el mismo? por qué no generar el Excel sin las filas que no aportan información?

Matias Butti: Al ser las solapas sub-conjuntos de genes, los resultados arrojados por el filtrado en cada una de ellas se comportan como un diagrama de Venn, puede haber intersección o no, generalmente no arrojan los mismos resultados.

- Cecilia Moreyra: Con respecto al proceso de filtrado ya del relevamiento decanta claramente pero igualmente para reforzar: Qué columnas son prioritarias? Cuáles opcionales?

Sebastián Menazzi: Las columnas que son opcionales son las que identifican la meta-data de la variante. El resto son todas prioritarias.

- Cecilia Moreyra: Te interesa guardar filtros predeterminados relacionados a un determinado fenotipo?

Sebastian Menazzi: Si, serían conjuntos de genes. Se realiza a través de un panel bio-informático.

Se aplicaría un filtro de fenotipo para que aparezcan todos los genes que son relevantes relacionados con el mismo y el resultado del filtrado se mostraría en genes-fenotipo.

Matias Butti: El nombre del filtro es el nombre del síndrome/patología y eso trae un filtro de genes.

- Cecilia Moreyra: Te interesa seleccionar qué columnas querés ver y el orden?

Sebastián Menazzi: Se necesita visualizar aquellas columnas que son prioritarias, con opción a aquellas identificadoras de la variante que son “metadata” de la misma.

Propuesta de ejemplo de filtros y dinámica de los mismos

- Cecilia Moreyra: Por último se presenta una propuesta de manejo de filtros relacionada a un gran volumen de datos con posibilidad de guardar filtros pre-determinados que servirá como punto de partida para el prototipo de Exoma Browser: <https://www.investing.com/stock-screener/>.

La propuesta fue aceptada por los participantes.

Fecha	21/03/2022
Modalidad	Presencial
Objetivo	Presentación del prototipo finalizado.
Asistentes	Sebastián Menazzi, Cecilia Moreyra.
Duración	1 hora

Se presenta el prototipo finalizado a Sebastián Menazzi quien da la aprobación final del mismo.

Apéndice 4. Cuestionario SUS – Gabriel Ercoli

1. ¿Crees que la usarías frecuentemente?

a. Totalmente de acuerdo

b. De acuerdo

c. Ni de acuerdo, ni en desacuerdo

d. En desacuerdo

e. Totalmente en desacuerdo

2. ¿Crees que es fácil de usar?

a. Totalmente de acuerdo

b. De acuerdo

c. Ni de acuerdo, ni en desacuerdo

d. En desacuerdo

e. Totalmente en desacuerdo

3. En cuanto a los términos utilizados que representan conceptos de genética. ¿Crees que son complejos de comprender?

a. Totalmente de acuerdo

b. De acuerdo

c. Ni de acuerdo, ni en desacuerdo

d. En desacuerdo

e. Totalmente en desacuerdo

4. ¿Necesitarías soporte técnico para utilizar la herramienta?

a. Totalmente de acuerdo

b. De acuerdo

c. Ni de acuerdo, ni en desacuerdo

d. En desacuerdo

e. Totalmente en desacuerdo

5. ¿Agregarías alguna opción a las ya existentes en la herramienta para hacer más completa la solución?

a. Totalmente de acuerdo

b. De acuerdo

c. Ni de acuerdo, ni en desacuerdo

d. En desacuerdo

e. Totalmente en desacuerdo

6. ¿Quitarías alguna opción a las ya existentes en la herramienta para hacer más simple la solución?

a. Totalmente de acuerdo

b. De acuerdo

c. Ni de acuerdo, ni en desacuerdo

d. En desacuerdo

e. Totalmente en desacuerdo

7. ¿Crees que hay alguna inconsistencia?

a. Totalmente de acuerdo

b. De acuerdo

c. Ni de acuerdo, ni en desacuerdo

d. En desacuerdo

e. Totalmente en desacuerdo

8. ¿Crees que sería conveniente integrar las consultas a bases de datos genéticos realizadas por fuera de la herramienta?

a. Totalmente de acuerdo

- b. De acuerdo
- c. Ni de acuerdo, ni en desacuerdo
- d. En desacuerdo

e. Totalmente en desacuerdo

9. ¿Crees que las funciones están bien integradas entre sí?

a. Totalmente de acuerdo

- b. De acuerdo
- c. Ni de acuerdo, ni en desacuerdo
- d. En desacuerdo

e. Totalmente en desacuerdo

10. ¿Crees que necesitarías aprender muchas cosas para ser capaz de usar este sistema?

- a. Totalmente de acuerdo
- b. De acuerdo
- c. Ni de acuerdo, ni en desacuerdo
- d. En desacuerdo

e. Totalmente en desacuerdo

Apéndice 5. Cuestionario SUS – Matias Butti

1. ¿Crees que la usarías frecuentemente?

- a. Totalmente de acuerdo
- b. De acuerdo**
- c. Ni de acuerdo, ni en desacuerdo
- d. En desacuerdo
- e. Totalmente en desacuerdo

2. ¿Crees que es fácil de usar?

- a. Totalmente de acuerdo
- b. De acuerdo**
- c. Ni de acuerdo, ni en desacuerdo
- d. En desacuerdo
- e. Totalmente en desacuerdo

3. En cuanto a los términos utilizados que representan conceptos de genética. ¿Crees que son complejos de comprender?

- a. Totalmente de acuerdo
- b. De acuerdo
- c. Ni de acuerdo, ni en desacuerdo
- d. En desacuerdo**
- e. Totalmente en desacuerdo

4. ¿Necesitarías soporte técnico para utilizar la herramienta?

- a. Totalmente de acuerdo
- b. De acuerdo
- c. Ni de acuerdo, ni en desacuerdo

d. En desacuerdo

e. Totalmente en desacuerdo

5. ¿Agregarías alguna opción a las ya existentes en la herramienta para hacer más completa la solución?

a. Totalmente de acuerdo

b. De acuerdo

c. Ni de acuerdo, ni en desacuerdo

d. En desacuerdo

e. Totalmente en desacuerdo

6. ¿Quitarías alguna opción a las ya existentes en la herramienta para hacer más simple la solución?

a. Totalmente de acuerdo

b. De acuerdo

c. Ni de acuerdo, ni en desacuerdo

d. En desacuerdo

e. Totalmente en desacuerdo

7. ¿Crees que hay alguna inconsistencia?

a. Totalmente de acuerdo

b. De acuerdo

c. Ni de acuerdo, ni en desacuerdo

d. En desacuerdo

e. Totalmente en desacuerdo

8. ¿Crees que sería conveniente integrar las consultas a bases de datos genéticos realizadas por fuera de la herramienta?

a. Totalmente de acuerdo

b. De acuerdo

c. Ni de acuerdo, ni en desacuerdo

d. En desacuerdo

e. Totalmente en desacuerdo

9. ¿Crees que las funciones están bien integradas entre sí?

a. Totalmente de acuerdo

b. De acuerdo

c. Ni de acuerdo, ni en desacuerdo

d. En desacuerdo

e. Totalmente en desacuerdo

10. ¿Crees que necesitarías aprender muchas cosas para ser capaz de usar este sistema?

a. Totalmente de acuerdo

b. De acuerdo

c. Ni de acuerdo, ni en desacuerdo

d. En desacuerdo

e. Totalmente en desacuerdo

Glosario

Diagnóstico Molecular: Proceso que se usa para identificar una enfermedad por medio del estudio de las moléculas, tales como proteínas, ADN y ARN, de un tejido o líquido corporal.

Gen: región del genoma que dará origen a una molécula de ARN (molécula intermedia) y (generalmente) a una proteína resultante.

Exoma: subconjunto del genoma que codifica la información más relevante, aquella parte de los genes que sintetizan proteínas.

Fenotipo: característica clínica, conductual y de comportamiento del paciente asociada con una patología.

Variante en ADN: cambio en alguna “letra” del genoma con respecto al genoma de referencia. Puede ser benigno (existen diferencias entre los humanos que no implican patología, estas son la mayoría de las variantes que identificamos en estudios de exoma) o patogénico (si genera enfermedad).

Enfermedad/patología Monogénica: provocadas por la variante de 1 solo gen y sin factores ambientales.

Enfermedad/patología Multi génica/multifactorial: cuando se agrupan múltiples variantes en genes distintos, y se suma el impacto del medio ambiente.

Archivo VCF: archivo de variantes respecto del genoma de referencia. Cada renglón es una variante.

Proteína: secuencia al igual que el ADN, pero de aminoácidos.

Genoma Humano: secuencia completa de ADN de un organismo, compuesta por millones de “letras” (ATCG) dispuestas de una forma predeterminada.

Técnica Next Generation Sequencing: método de lectura de ADN en forma masiva y paralela, para generar datos como la secuencia del exoma de un individuo en forma rápida.

Cigosis: Conjugación de gametos o unión sexual de dos organismos unicelulares.

Homocigosis: solo una copia del gen tiene la variante.

Heterocigosis: las dos copias del gen, paterna y materna, tienen la variante.

Heterocigosis compuesta: variantes patogénicas diferentes en cada una de las copias del mismo gen.

Incidental Finding: hallazgo independiente del diagnóstico presuntivo.

Término HPO: término que describe un fenotipo.

Predictor bioinformático: predicción de la consecuencia biológica (en la proteína) de la mutación.

Cromosoma: estructura que alberga al ADN en la célula.

Transcripto: distintas formas de combinar los bloques de exones.

Impacto Funcional de la variante: efecto sobre la proteína.

Significancia Clínica de la variante: importancia clínica del efecto de un tratamiento.

Patrón de Herencia: patrón que rige la transmisión de diferentes caracteres causantes de las enfermedades hereditarias monogénicas. Puede ser de origen dominante o recesivo.

Panel bioinformático: panel para generar una priorización de genes de acuerdo con las características clínicas del paciente.

Dato ómico: dato que pueden usarse para desarrollar una prueba basada en la ómica. Ejemplos de tipos de datos ómicos son: genómicos, transcriptómicos, proteínómicos, epigenómicos, metabólicos.

Mendeliano: Monogénico. Cumplen con las leyes de Mendel.

Base de datos OMIN: Repositorio de todas las patologías monogénicas de origen [mendeliano](#).