



# El dilema entre profundidad, cobertura y costo al momento de elegir un estudio genómico



Como comentamos y analizamos en nuestro newsletter anterior *Ser Gen, Panel o Exoma, ¿esa es la cuestión?*, en la era de la genómica una de las decisiones más importantes al momento de realizar un estudio genómico, buscando un diagnóstico molecular preciso, es determinar ¿qué me conviene secuenciar?. Manteniendo el objetivo de maximizar la eficiencia diagnóstica y buscando minimizar el costo y la complejidad del estudio, si descartamos la secuenciación del genoma completo y/o de uno (o unos pocos) genes, por los motivos previamente mencionados, la solución que emerge para la mayoría de las enfermedades mendelianas como óptima, es la secuenciación exómica (WES, del inglés Whole Exome Sequencing). El WES se ha aplicado exitosamente en la última década al descubrimiento de variantes en contextos de investigación y al diagnóstico molecular en la práctica clínica, siendo actualmente, (junto a los paneles de cáncer), la estrategia más utilizada. Sin embargo, haber decidido realizar un WES no resuelve “del todo” el problema mencionado, ya que en una segunda etapa de toma de decisiones, debemos considerar aspectos relacionados con la profundidad, la cobertura (tanto teórica -kit elegido- como aquella derivada de la estadística previa) y el costo. Analizar estos aspectos es el objetivo del presente artículo.

## Secuenciación exoma completo vs exoma clínico

La Secuenciación del Exoma Completo (WES), es un método para secuenciar de forma selectiva las regiones codificantes del genoma (los genes a partir de los cuales se expresan las proteínas). Este experimento realiza la captura y secuenciación de los exones (y regiones aledañas) de los aproximadamente 20.000 genes que componen el genoma humano. En ausencia de una historia clínica, o un diagnóstico definido, el WES permite escanear todas las regiones codificantes del genoma presentándose como una alternativa más fácil y barata que el genoma completo y aún así eficaz.

Recientemente, ha cobrado relevancia el uso del, comúnmente denominado, “Exoma Clínico” (CES, del inglés Clinical Exome Sequencing). El mismo se basa en la captura de un conjunto acotado de genes (típicamente entre 3000 a 8000 aprox. dependiendo del kit utilizado), con probada asociación con patologías mendelianas. En otras palabras, comprende un panel que busca cubrir un alto porcentaje de las variantes patogénicas o probablemente patogénicas reportadas en ClinVar<sup>1</sup>, lo que lo convierte en un panel adecuado para la investigación clínica y para confirmación de sospecha diagnóstica. El subconjunto de genes incluido en un CES se elige sobre la base de diversos estudios acerca de las asociaciones con distintas enfermedades de

1



<sup>1</sup> ClinVar: <https://www.ncbi.nlm.nih.gov/clinvar/>



## El dilema entre profundidad, cobertura y costo al momento de elegir un estudio genómico

etiología genética, -especialmente obtenidas a partir de Online Mendelian Inheritance in Man (OMIM)<sup>2</sup> y Human Mutation Database (HGMD)<sup>3</sup> - y la prevalencia/frecuencia de las mutaciones en estos genes. Por ejemplo, el panel de secuenciación TruSight One Expanded (Illumina Inc., San Diego, EE. UU.) contiene ~6700 genes y regiones asociadas con fenotipos clínicos conocidos. Sin embargo, la lista de genes asociados a enfermedades se amplía cada semana, en base a nuevas investigaciones y publicaciones y, de esta forma, se sustentan las diferencias entre el número de genes estudiados por distintos laboratorios.

Las ventajas claves de usar el diseño selectivo del CES incluyen la capacidad de restringir el análisis a genes clínicamente relevantes. Interpretar la variación en los genes "desconocidos" o en un número muy grande de genes en algunos casos constituye un desafío y es probable que resulte en una lista extendida de variantes de significado incierto (VUS) de difícil interpretación.

Además, es posible excluir la captura y secuenciación de genes particulares; por ejemplo, genes asociados a enfermedades tardías (de inicio en la edad adulta) en pacientes pediátricos.

La desventaja de usar una lista restringida de genes (o sea usar CES) se debe a su incapacidad de extender el análisis a todo el exoma. Esto a menudo se conoce como "abrir el exoma" y muchas veces propicia el descubrimiento de nuevas causas genéticas asociadas con enfermedades de etiología genética. Los médicos e investigadores aprecian la capacidad de revisar los resultados en todo el exoma e interrogar las secuencias de genes que actualmente no

están asociadas con un fenotipo clínico dado. A medida que los nuevos genes se asocian con fenotipos clínicos, es útil poder volver a analizar los datos resultantes de la secuenciación del ADN de los pacientes, contando con el consentimiento apropiado, para llevar a cabo el reanálisis en función de nuevas hipótesis de trabajo.

En este contexto, uno de los enfoques cada vez más solicitado -y el que solemos utilizar y recomendar en Bitgenia -, consiste en realizar la secuenciación completa del exoma y restringir el análisis a genes específicos, mediante el filtrado bioinformático. De esta manera, es posible crear paneles de genes virtuales: los genes se pueden agregar y eliminar fácilmente, así como ponerlos en una "lista negra", de modo que los datos analizados no estén disponibles para ciertos genes (por ejemplo, BRCA1/BRCA2). Se puede realizar una primera priorización utilizando un conjunto definido de genes (seleccionados en función del fenotipo clínico del paciente). Si ese primer enfoque resulta negativo, se podría ampliar la búsqueda a un conjunto extendido de genes o, alternativamente, a todas las regiones codificantes capturadas.

En resumen, teniendo en cuenta la relación costo-beneficio, la diferencia de precio mínima (en algunos casos, inexistente) y las ventajas del WES con respecto al CES, resulta más provechoso y conveniente secuenciar los ~20 mil genes con muy buena cobertura y no limitarse a la secuenciación de entre 4 a 8 mil genes, incrementando las posibilidades de llegar a un diagnóstico molecular, especialmente en aquellos casos que resultan desafiantes.



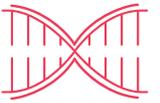
2



<sup>2</sup> Omim: <https://omim.org/>

<sup>3</sup> HGMD: <http://www.hgmd.cf.ac.uk/>





## El dilema entre profundidad, cobertura y costo al momento de elegir un estudio genómico

### Profundidad

El segundo aspecto a considerar al momento de realizar un estudio genómico, ya sea WES, CES, Panel o incluso Genoma Completo (WGS), es la profundidad. Comencemos con algunas definiciones. La **profundidad teórica media** es el único parámetro de un experimento de secuenciación que se puede definir de antemano, se refiere al número promedio de veces que una determinada posición se “lee” y refleja cuál es la relación de lo que voy a “leer” (output generado) en función de la cantidad de bases que quiero secuenciar (tamaño de la muestra). Por ejemplo, si digo (o solicito) un exoma 100X, esto significa que si la cantidad total de bases del exoma es 1Mb, en total en el experimento se secuenciarán 100Mb. Las profundidades usuales para WES son 100, 200 o 300X, para genoma completo 30X, y para paneles en el orden de 400 a 1000X (o más).

En los experimentos de secuenciación, sobre todo aquellos que incluyen captura (WES) o amplificación (Paneles), la profundidad obtenida es heterogénea y depende de cada posición. Para un WES 100X, por ejemplo, puede haber posiciones que fueron leídas 200 veces y, otras, solamente 20. De este resultado se desprende el concepto de **profundidad por posición**. Lo que reporta es cuántas veces fue efectivamente secuenciada cada posición (y se utiliza para el llamado de variantes). A partir de la misma, uno determina finalmente la profundidad promedio obtenida, y al compararla con la

teórica puede tener una estimación del éxito del experimento. Si, por ejemplo, se diseñó un WES 100X y la profundidad media obtenida fue inferior 30X, esto podría indicar que hubo un problema con el experimento (contaminación, muestra degradada etc.) y debe repetirse.

Entonces, habiendo definido con precisión el concepto, pasemos a la pregunta de: ¿cuál es la profundidad teórica media más conveniente? - La respuesta a esta pregunta, como en el caso de la elección de la técnica, no es única y depende del objetivo y del contexto. Sin embargo, podemos dar algunas “reglas del pulgar”. Tomando como punto de partida el genoma completo, y su alto costo, dado que su profundidad suele ser pareja, un genoma 30X suele tener calidad suficiente como para asegurar un eficiente llamado de variantes y determinación de CNVs. Una profundidad superior solo se recomienda en casos particulares en el contexto de investigación, por ejemplo, si uno desea estudiar el genoma tumoral y ver subpoblaciones. Para el caso de WES (o exoma panelizado), para diagnóstico molecular, el estándar suele ser 100X. En caso de que solamente se secuencie al paciente y luego se avance en una verificación y análisis de la segregación familiar de las variantes candidatas por tecnología Sanger en el grupo familiar, no se justifica utilizar una profundidad teórica superior. La experiencia indica que el número de variantes llamadas al usar una profundidad superior, es relativamente pequeño en relación al costo adicional. A la hora de diseñar paneles, es importante considerar una profundidad superior, de 500X o 1000X, para asegurarse una cobertura pareja y completa, que justifique el mayor precio y el objetivo acotado de genes (regiones) a analizar.

3



# El dilema entre profundidad, cobertura y costo al momento de elegir un estudio genómico

Técnica	Profundidad aconsejada
Genoma	30X
Exoma	100X
Panel	500-1000X

Tabla 1. Profundidades Teóricas Recomendadas.

## Cobertura

Finalmente, llegamos al concepto de cobertura, que surge de la combinación de los conceptos y elecciones analizadas en los puntos anteriores, y que en última instancia determinará nuestra eficiencia diagnóstica. La cobertura responde a la pregunta de qué porcentaje de la longitud total que se buscaba secuenciar, fue efectivamente leído. Por ejemplo, si queremos secuenciar un genoma humano entero, unas 3Gb (haploide) y al final del experimento leímos 2.5Gb, nuestra cobertura fue del 83%. Entonces, la elección de que secuenciar, WGS, WES, Panel condiciona cuál es nuestro objetivo (target) de 100% de cobertura y se encuentra definido de antemano. Una vez realizado el experimento de secuenciación se podrá entonces calcular la cobertura obtenida. Ahora bien, para definir la misma, se debe elegir un umbral de profundidad mínima a partir del cual se considera que efectivamente la posición fue secuenciada o cubierta. Volviendo al ejemplo, si hacemos un WGS 30X, y establecemos un umbral de 10X, entonces la cobertura será el porcentaje de bases del total del genoma que fueron leídas con una profundidad superior a 10.

Es importante destacar, en relación a la cobertura, que cuando uno va a realizar un WES/CES o un panel, muchas veces nos

interesa no la cobertura total, sino la de algunos genes o regiones en particular. Aquí aparecen dos cuestiones importantes. En primer lugar existe, cuando se utiliza un método de captura/amplificación una cobertura teórica, que nos dice qué genes o regiones el kit busca secuenciar (o sea cuál es el objetivo). Al momento de elegir entre diferentes opciones, es importante analizar en detalle si el experimento que nos ofrecen, efectivamente, cubre por diseño, las regiones de interés. Es aquí donde se encuentran las mayores diferencias entre WES/CES y paneles, ya que los mismos poseen, por diseño, objetivos de secuenciación diferentes. Si bien la cobertura teórica suele ser un buen punto de partida para decidir si el experimento propuesto cubre (o no) los genes de interés, las coberturas reales suelen variar mucho entre diferentes genes (con valores que pueden ir desde el 30 al 80%), aún cuando la cobertura teórica de los mismos sea superior al 90%. Es por ello que un mejor criterio para la toma de decisión suele ser basarse en la cobertura promedio observada, para los genes de interés, en una muestra representativa de casos realizados previamente.

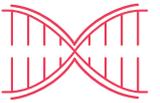
Un punto importante a considerar, sobre todo en casos donde el reporte que se realiza es negativo (cuando no se encuentran variantes diagnósticas), consiste en analizar la cobertura "real" de los genes candidatos, ya que cuando la cobertura es baja, esto indica una alta posibilidad de que la variante diagnóstica se encuentre en alguna de las zonas no leídas, por lo que uno puede optar por buscarlas con otras técnicas (por ejemplo secuenciación por sanger de las regiones no leídas).

Otro punto a mencionar, consiste en reflexionar un momento sobre cómo se



4





# El dilema entre profundidad, cobertura y costo al momento de elegir un estudio genómico

calcula la cobertura (ya sea la teórica o la real), en WES/CES/paneles.

Generalmente, se analiza la misma en función del tamaño de los genes, particularmente si uno desea incluir las regiones 5` y 3` UTR, se puede observar que los valores obtenidos suelen ser bajos (< 50%). Esto es porque las fronteras de los genes usualmente no se encuentran bien definidas y las regiones 5` y 3` UTR son extensas. Además, muchas veces, el exoma "teórico" incluye todos los exones, siendo que un buen porcentaje de los mismos, no son codificantes o se encuentran poco representados en los transcritos predominantes. Es por ello, que cuando lo que se busca es analizar la capacidad de un experimento de lograr un diagnóstico molecular (encontrar la variante responsable), en Bitgenia desarrollamos el concepto de cobertura de variantes con relevancia clínica. En el mismo, la cobertura es determinada como el % de variantes de relevancia clínica que son teórica o efectivamente, cubiertas por un dado experimento de WES/CES/Panel. Esta medida, suele arrojar valores de cobertura superiores (al basado en % de secuencia) y resulta extremadamente útil al momento de elegir el experimento adecuado en función de los genes de interés.

## Bibliotecas

Por último, luego de haber decidido el tipo de secuenciación y la profundidad a la cual realizarla, se debe elegir la biblioteca de captura a utilizar. Para esto es importante, como mencionamos anteriormente, poder realizar un análisis de cobertura teórica de los genes de interés en función del kit de captura.

cubiertas por los distintos kits comerciales.

Tomamos como ejemplo cuatro kits de captura que se suelen ofrecer habitualmente, el SureSelect Human All Exon V7, SureSelect Human All Exon V6, Twist Core Exome Panel y el panel de Illumina, denominado exoma clínico, Trusight One.

Para cada uno ellos se tomó el archivo bed y se intersectaron las regiones teóricamente capturadas por cada uno de ellos. Entre las regiones capturadas por cada kit, se encuentran regiones codificantes, regiones intrónicas, regiones UTR (Untranslated Region), etc.

	TruSight One (6472)	Twist Core Exome Panel (22563)	Sure Select V6 (28668)
Sure Select V7 (26121)	6407	22553	25223
TruSight One (6472)		6324	6441
Twist Core Exome Panel (22563)			22526

5

Tabla 2: Intersecciones por pares

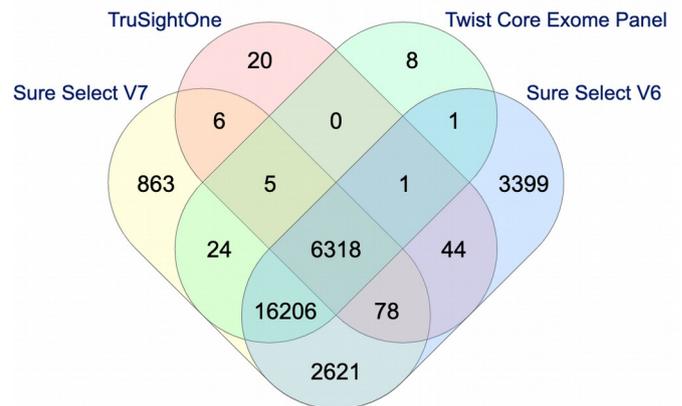


Diagrama de Venn: Intersección de regiones capturadas

Para finalizar este artículo, haremos algunas comparaciones relativas a las regiones



## El dilema entre profundidad, cobertura y costo al momento de elegir un estudio genómico

Aquí podemos ver cómo los diferentes kits enriquecen o cubren distintas regiones, distintos genes en mayor o menor medida. Cada biblioteca, además de los exones, captura determinadas regiones o posiciones de interés, que pueden incluir posiciones puntales, regiones intrónicas, etc. Adicionalmente, cada uno de ellos ofrecen distintas coberturas de las regiones flanqueantes a los exones, algunos se extienden por unas decenas o cientos de bases a ambos lados, otros “cortan” la captura ni bien finaliza el exón.

Con este newsletter, desde Bitgenia, hemos tratado brevemente las cuestiones que se deben tener en cuenta a la hora de solicitar un estudio genómico. De esta manera, priorizamos mantener actualizados a nuestros clientes, difundiendo información de relevancia relacionada con qué biblioteca de captura utilizar y con qué profundidad secuenciar, para tomar decisiones con confianza, priorizando la cobertura de los genes de interés.

*EQUIPO BITGENIA*



6

