



Enfermedades Infecciosas y Microbiología Clínica

www.elsevier.es/eimc



Secuenciación masiva para el diagnóstico y la epidemiología de tuberculosis

Iñaki Comas^{a,b,c,*} y Ana Gil^d

^aUnidad de Genómica de la Tuberculosis, Instituto de Biomedicina de Valencia, Consejo Superior de Investigaciones Científicas, Valencia, España

^bUnidad de Genómica de Tuberculosis, Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana (FISABIO), Valencia, España

^cCIBER (Centro de Investigación Biomédica en Red) en Epidemiología y Salud Pública, Instituto de Salud Carlos III, Madrid, España

^dServicio de Microbiología, Hospital Universitario La Fe, Valencia, España

RESUMEN

Palabras clave:

Secuenciación genómica
Mycobacterium tuberculosis
Epidemiología genómica
Genotipado
Resistencia a antibióticos

La tuberculosis (TB) se ha convertido en la enfermedad infecciosa que más personas mata mundialmente por delante del VIH (virus de la inmunodeficiencia humana) y la malaria. La incidencia de TB apenas se reduce en un 2% cada año, por lo que no se podrá eliminar hasta dentro de 200 años. Parte del problema es que las actuales herramientas de control de la TB son antiguas y ya no son suficientes para acelerar la erradicación de la TB. Nuevos métodos son, por tanto, necesarios en los campos de diagnóstico, tratamiento y prevención. La secuenciación genómica masiva puede convertirse en una de esas herramientas. La caracterización genómica de aislados de TB está demostrando su utilidad tanto para entender la epidemiología de la enfermedad como para su diagnóstico, en especial para la detección de resistencias a antibióticos. Sin embargo, las técnicas genómicas y el análisis informático asociado están todavía lejos de usarse de rutina en la mayor parte del mundo. En esta revisión describiremos las actuales tecnologías de secuenciación en el contexto del complejo *Mycobacterium tuberculosis*, las aportaciones de dichas tecnologías al diagnóstico y la epidemiología de la enfermedad, así como los esfuerzos que se están haciendo para dar el salto a su aplicación en el contexto clínico.

© 2016 Elsevier España, S.L.U. Todos los derechos reservados.

Next generation sequencing for the diagnostics and epidemiology of tuberculosis

ABSTRACT

Keywords:

Whole genome sequencing
Mycobacterium tuberculosis
Genomic epidemiology
Genotyping
Antibiotic resistance

Tuberculosis (TB) has overtaken HIV (human immunodeficiency virus) and malaria as the leading cause of death by an infectious disease worldwide. The reduction in the TB incidence is a modest 2% of cases per year, thus we will need 200 years to eradicate the disease. Part of the problem is that TB control tools are decades old and cannot anymore contribute to accelerate eradication of TB. New diagnostics, treatments and vaccines are urgently needed. Next generation sequencing has the potential to become one of these new tools. Genomic characterization of TB isolates is already showing its potential for epidemiology and diagnostics, particularly to identify drug resistance mutations. However, the experimental and bioinformatics skills needed are still far from being standardized and are not easy to incorporate as a routine in clinical laboratories. In this review we will describe current next generation sequencing approaches applied to the *Mycobacterium tuberculosis* complex, their contribution to the diagnostics and epidemiology of the disease and the efforts that are being undertaken to make the technology accessible to public health and clinical microbiology laboratories.

© 2016 Elsevier España, S.L.U. All rights reserved.

Introducción

La identificación de micobacterias por técnicas de biología molecular se inició a finales de los años ochenta del siglo pasado¹. A principios de los noventa, varios grupos de investigación presentaron los resultados obtenidos tras la incorporación de estos métodos para la detección de micobacterias englobadas en el complejo *Mycobacte-*

*Autor para correspondencia.

Correo electrónico: inaki.comas@uv.es (I. Comas).

rium tuberculosis (MTBC)². La conclusión fue que las técnicas moleculares permitían adelantar el diagnóstico de enfermedad tuberculosa frente a los métodos tradicionales de cultivo, aunque nunca sustituirlos como técnica de referencia, debido a que no permitían determinar la infectividad o viabilidad del bacilo, ni las resistencias a los fármacos antituberculosos³. A pesar de que este último aspecto se ha suplido en gran parte con las nuevas tecnologías que permiten detectar genes relacionados con mutaciones asociados a resistencias en un alto porcentaje de cepas, no se ha solucionado por completo¹.

El método de referencia para el diagnóstico de la tuberculosis (TB) es el cultivo, aunque en algunos países la observación directa de bacilos ácido-alcohol resistentes es la única prueba diagnóstica disponible. La microscopía, aunque barata, no es sensible, especialmente en niños y personas que conviven con pacientes positivos para el virus de la inmunodeficiencia humana (VIH); no diferencia entre micobacterias no tuberculosas y especies englobadas en el MTBC ni proporciona información acerca de la sensibilidad. En cultivo en medios habituales de Löwenstein-Jensen y Middlebrook, las micobacterias del MTBC pueden tardar semanas en crecer, retrasándose así los resultados y la determinación de presencia o no de resistencia a los fármacos antituberculosos. Por todo ello, las técnicas moleculares se han ido incorporando al algoritmo diagnóstico de la TB en los diferentes hospitales⁴.

Estas técnicas, además, están teniendo y tendrán un papel clave en la vigilancia y el control epidemiológico de la enfermedad a nivel global. Al ritmo al que se reduce la incidencia global de la enfermedad tardaremos más de 200 años en erradicar la TB. Solo la aparición de nuevas herramientas para el diagnóstico, la prevención y el tratamiento nos hará cumplir los objetivos de erradicación para 2050⁵. En su nueva estrategia, End TB Strategy (2016-2035), la Organización Mundial de la Salud (OMS) incluye como una de sus metas el diagnóstico temprano de la TB, así como un estudio de sensibilidad anti-tuberculosa universal. Los laboratorios tienen, por tanto, un papel crítico en la era posterior a 2015 y en la adaptación de las aproximaciones moleculares. En este último informe mencionan que el nuevo cartucho en desarrollo Xpert Ultra[®] podría incluso reemplazar al cultivo tradicional como primera herramienta en el diagnóstico de la TB. La aparición de tecnologías de secuenciación masiva puede superar las limitaciones de los diagnósticos moleculares actuales y en un futuro ser un método de referencia, como ya lo está empezando a ser para la epidemiología de la enfermedad⁶. En esta revisión se repasará la historia, pero sobre todo el papel presente y futuro de la secuenciación genómica en el diagnóstico y la epidemiología molecular.

Tecnologías de secuenciación masiva aplicadas al bacilo de la tuberculosis

En 1905 una cepa de TB conocida como H37 se aisló asociada a un brote infeccioso en Estados Unidos. Dicha cepa (conocida como H37Rv) y su hermana avirulenta (conocida como H37Ra) se convirtieron en las cepas de referencia más usadas en laboratorios de investigación de todo el mundo. En 1998, H37Rv fue uno de los primeros genomas bacterianos completamente secuenciados y abrió las puertas de la era genómica en el estudio de la TB⁷.

El genoma de H37Rv y su manipulación han permitido un gran número de descubrimientos relevantes para el desarrollo de nuevos antibióticos, vacunas y diagnósticos⁸. La determinación de la secuencia genómica del bacilo de la TB ha permitido estudiar los determinantes genéticos de la resistencia a antibióticos y construir bases de datos asociadas⁹, así como el desarrollo de todo un nuevo conjunto de pruebas moleculares para el diagnóstico de MTBC y para la determinación genotípica de resistencia a los fármacos antituberculosos¹⁴. Por último, el estudio de las diferentes regiones codificantes y no codificantes del genoma ha permitido entender su papel en la biología y la virulencia de la bacteria, siendo el origen de muchas de las vacunas candidatas que están en estudio¹⁰.

Desde el punto de vista de la epidemiología molecular y la microbiología clínica, su impacto ha sido muy importante. Por ejemplo, aunque muchos de los marcadores moleculares que se usan ahora ya existían antes de conocerse la secuencia genómica¹¹, solo a posteriori se han podido caracterizar y poner en el contexto adecuado. Entre los más conocidos se encuentra la región del espoligotipado conocida como CRISPR (*clustered regularly interspaced short palindromic repeats*). Los espoligotipos se usan a veces para definir posibles casos de transmisión, pero sobre todo para clasificar los aislados en diferentes familias génicas¹². Actualmente, el marcador más empleado en epidemiología molecular se conoce como MIRU-VNTR (*mycobacterial interspersed repetitive units-variable number tandem repeats*)¹³. Con este método se analizan diferentes regiones del genoma que se caracterizan por tener un número variable de repeticiones en tándem conocidas como minisatélites. Aunque existe un gran número de minisatélites, por motivos de reproducibilidad se usan 2 grupos estándar conocidos por el número de regiones que interrogan, como MIRU 15 y MIRU 24. Debido a que la frecuencia de cambio genético de dichas zonas (duplicaciones y deleciones) es muy alta, la combinación del número de copias de esos *loci* permite identificar cepas de estrecha relación genética y diferenciarlas de otras que, aunque puedan estar próximas genéticamente, no pertenezcan al mismo grupo de transmisión.

En el año 2005 aparecieron las primeras tecnologías de secuenciación masiva¹⁴. Al contrario que las técnicas usadas hasta ese momento, conocidas como secuenciación por perdigonada (*shotgun sequencing*), estas tecnologías son mucho menos laboriosas al no requerir un paso previo de clonación en *Escherichia coli*. La otra característica fundamental es la cantidad de nucleótidos capaces de determinar en una sola carrera¹⁵. Por poner un ejemplo, actualmente las plataformas de Illumina son las más empleadas en secuenciación bacteriana. El tamaño medio de un genoma de *M. tuberculosis* es de algo más de 4,4 millones de pares de bases (pb) nucleotídicas, o lo que es lo mismo 4,4 megabases. Si solo quisiéramos secuenciar un genoma en una carrera de Illumina MiSeq leeríamos cada base del genoma como media 1.700 veces, por lo que su cobertura sería 1.700×. Con la plataforma de Illumina NextSeq *rapid run* sería 13.600× y en uno de Illumina HiSeq completo 67.700×. Por usar otro ejemplo, en la plataforma de Ion Torrent se podría leer hasta 455 veces cada base del genoma analizado (tabla 1). Por lo tanto, todas ellas generan mucha más cantidad de secuencia que la realmente necesaria para analizar un solo genoma. Lo que realmente se hace es secuenciar decenas o cientos de genomas a la vez, dependiendo de la cantidad de datos generados por cada plataforma. Esta estrategia se denomina *multiplexado* y es la que ha permitido disminuir, en gran medida, los precios de secuenciación genómica. En la tabla 1 se puede consultar el número aproximado de genomas de diferentes cepas del MTBC que se pueden analizar a la vez. Si tuviéramos una colección de 800 cepas la podríamos analizar en una sola carrera con un HiSeq 1500/2500 y el genoma de cada una de ellas se leería como media 80 veces. La cuestión está en que cada cadena de ADN genómico es fragmentada en trozos de entre 500 y 1.000 pb, añadiéndose a cada uno posteriormente una etiqueta nucleotídica que hace que bioinformáticamente se pueda asignar posteriormente cada lectura a la cepa correspondiente. En el laboratorio de FISABIO (Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana) usamos rutinariamente Illumina MiSeq con 20 cepas por carrera. Si solo contamos el coste de las librerías y su secuenciación, sin tener en cuenta la mano de obra, cuesta secuenciar cada cepa entre 80 y 100 euros, aunque va bajando conforme se automatizan.

Otra de las características básicas a la hora de elegir una plataforma de secuenciación es el tamaño de las lecturas que es capaz de generar. Así, se pueden clasificar las plataformas en las que se generan lecturas cortas (entre 100-300 pb) y en las que se generan lecturas largas de hasta varias kilobases (kb) (tabla 1). Las más usadas en el campo de la microbiología clínica y epidemiología de la TB son las

Tabla 1Descripción de las plataformas de secuenciación genómica usadas comúnmente para el análisis del complejo de *Mycobacterium tuberculosis*

Plataforma de secuenciación	Número de lecturas	Tamaño de la lectura	Cobertura media para un genoma	Aislados por carrera (80×)
HiSeq 1500/2500 V4	2,00 E + 09	150	68.164	852
HiSeq 1500/2500 <i>rapid run</i>	3,00 E + 08	150	10.227	128
NextSeq 500 <i>high output</i>	4,00 E + 08	150	13.636	170
NextSeq 500 <i>mid output</i>	1,30 E + 08	150	4.432	55
MiSeq V3	2,50 E + 07	300	1.705	21
Ion Torrent 314	5,00 E + 05	400	45	1
Ion Torrent 316	2,50 E + 06	400	227	3
Ion Torrent 318	5,00 E + 06	400	455	6

de lecturas cortas, por razones varias. Así, por una parte, pueden generar una mayor cantidad de secuencias, lo cual significa que en una misma carrera podemos analizar muchos aislados a la vez, por lo que se reducen los costes. Por otra parte, en el caso del MTBC, el análisis bioinformático posterior no intenta reconstruir el genoma de la cepa (lo que se conoce como ensamblado de novo), sino que se mapean las lecturas a un genoma de referencia de secuencia ya conocida: estrategia conocida como *resecuenciación*. Las estrategias de mapeo a una referencia son muy útiles para analizar bacterias que tienen poca diversidad tanto a nivel de cambios puntuales cromosómicos como a nivel de contenido génico, como es el caso de las de la MTBC.

En bacterias más diversas, como puede ser *E. coli*, la estrategia de mapeo a una referencia no es posible debido a que las cepas pueden ser muy diferentes entre sí. Se sabe desde hace tiempo que 2 cepas de *E. coli* pueden llegar a tener hasta un 60% de genes diferentes¹⁶. En esos casos es difícil encontrar un genoma de referencia ya secuenciado y se requiere determinar el genoma de novo de las cepas estudiadas. Las tecnologías que generan secuencias muy largas son mucho más útiles, pues permiten ensamblar el genoma como si fuera un puzle, buscando fragmentos (conocidos como *contigs*) que se solapan parcialmente. Actualmente, la plataforma de lecturas largas más usada es PacBio (Pacific Biosciences), que genera lecturas de rango entre 10-15 kb y es capaz de finalizar un genoma completo de *M. tuberculosis* en una única carrera. Esta tecnología tiene la ventaja de que no necesita amplificar el material genómico de referencia (PCR [*polymerase chain reaction*]-free), sino que trabaja directamente sobre la molécula de ADN original (por eso es una tecnología *single-molecule*), de tal manera que puede evitar los sesgos introducidos por la PCR, que sí está presente en las plataformas de Illumina y de Ion Torrent. Su desventaja es que no permite analizar más de una cepa a la vez, por lo que los precios y los tiempos de respuesta son demasiado altos para usarla de rutina en epidemiología o diagnóstico clínico.

Aportaciones de la genómica a la microbiología clínica y a la epidemiología de la tuberculosis

Diversidad genómica global del complejo *Mycobacterium tuberculosis*

La secuenciación masiva ha permitido un gran número de descubrimientos en la biología y evolución de la micobacteria⁸. La genómica de poblaciones ha permitido identificar a nivel global 8 grandes grupos genéticos, conocidos como linajes¹⁷. De esta manera, el MTBC está compuesto por 7 linajes asociados a TB humana. Dichos linajes incluyen los 2 conocidos como *M. africanum* y un nuevo linaje detectado recientemente en Etiopía¹⁸. Aparte de estos linajes, existen también las especies asociadas a animales, entre las que se encuentra *M. bovis*, que es la causa de entre el 1 y el 3% de los casos de TB en humanos y que tiene un gran impacto económico. La secuenciación genómica de cepas representativas de los linajes animales y huma-

nos demuestra que la distancia entre todos ellos no es mayor de 2.500 cambios nucleotídicos, con una identidad nucleotídica mayor del 99%¹⁷. Esto indica que, según la taxonomía actual, todas las especies del complejo serían en realidad una sola¹⁹. Los diferentes linajes de humanos demuestran diferencias fenotípicas tanto a nivel de la bacteria como a nivel clínico y epidemiológico²⁰. Muchas de estas diferencias se deben, probablemente, a que los diferentes linajes han evolucionado en paralelo a las poblaciones humanas, lo que ha llevado a que la bacteria no sea ni genética ni fenotípicamente homogénea¹⁷.

Desde el punto de vista del genotipado, la información genómica ha permitido una definición objetiva de grupos jerárquicos (desde linajes a sublinajes), así como el desarrollo de paneles de polimorfismos de un solo nucleótido (SNP) para detectar dichos grupos^{12,21,22}. Por ejemplo, en el Hospital La Fe de Valencia el genotipado de muestras de los últimos 4 años revela que un 90% pertenece al linaje 4, como era de esperar. Una robusta caracterización de estos linajes y sublinajes es el primer paso para futuros estudios sobre la influencia del genotipo en la transmisión o en la manifestación clínica de la enfermedad⁶. En San Francisco se ha demostrado que dichos linajes se transmiten preferentemente en poblaciones humanas de su mismo origen geográfico²³. Este es el caso del linaje 2, que es el más común en Asia y se transmite mejor entre individuos de origen asiático. Además, estas diferencias no pueden explicarse solamente por la red social de contactos. Cuando se analizan esas preferencias entre individuos positivos para el VIH, las diferencias entre linajes desaparecen de una manera dependiente de la dosis, de forma que cuanto menor es el recuento de linfocitos CD4, menor es la asociación²⁴. Esto sugiere que hay un papel importante de interacción entre el sistema inmune y el genotipo de la micobacteria. Seguramente, las nuevas aproximaciones permitirán comprender en el futuro qué determinantes genéticos influyen en la interacción entre el hospedador y la micobacteria^{25,26}.

El análisis del genoma de cepas de todo el mundo también ha proporcionado información importante acerca de la biología del bacilo de la TB. Se ha demostrado que los epítomos de *M. tuberculosis* están hiperconservados²⁷. Dicho de otra manera, al contrario que otros muchos patógenos, el bacilo de la TB no acumula mutaciones para escapar del sistema inmune. La hipótesis actual es que la micobacteria necesita ser reconocida por el sistema inmune porque precisa de la reacción inmunitaria para su posterior transmisión. El análisis genómico también ha permitido determinar la tasa de mutación genómica (la frecuencia con la que se acumulan cambios con el tiempo), así como que esta es ligeramente diferente según el linaje^{28,29}. Así, las cepas conocidas como Beijing (que forman parte del linaje 2) tienen una tasa de mutación mayor, lo que seguramente explica por qué se han asociado en todo el mundo a la adquisición de resistencias a antibióticos³⁰.

En resumen, un gran número de estudios ha demostrado el impacto de las diferencias genéticas entre cepas de TB en diferentes

aspectos fenotípicos, desde la fisiología de la bacteria hasta la epidemiología o la manifestación de la enfermedad²⁰. La secuenciación genómica está ayudando a entender cuáles son los determinantes genéticos de dichos cambios fenotípicos, lo que a su vez permitirá determinar si deben o no tenerse en cuenta en el desarrollo de nuevos antibióticos, diagnósticos y vacunas para controlar la enfermedad²⁵.

El genoma como marcador epidemiológico

El primer estudio que demostró el poder de la secuencia genómica como marcador epidemiológico se publicó en 2009³¹. En dicho estudio, 3 cepas prácticamente idénticas en relación con otros marcadores —RFLP (*restriction fragment length polymorphism*), MIRU— eran totalmente diferentes, con lo que se demostró el mayor poder resolutivo del genoma. Poco después Gardy et al³² usaron las técnicas genómicas para resolver un brote de transmisión que había empezado en los años noventa en la Columbia Británica. Al combinar los resultados genómicos con datos epidemiológicos, los autores demostraron que dentro de los grupos de transmisión se puede identificar muchas veces la cadena de eventos individuales. El trabajo ya pone sobre la mesa la posibilidad de que muchos de los grupos que se conocen como cadenas de transmisión podrían estar causados por un solo individuo que habría infectado a varias personas (en inglés, *super spreader*)³².

Además, la ventaja del análisis de genomas completos es que permite descubrir y desarrollar nuevos marcadores para llevar a cabo la vigilancia epidemiológica o el diagnóstico clínico. Un claro ejemplo es el uso de polimorfismos únicos de secuencia para detectar cepas de especial interés. En los últimos 2 años se han publicado varios estudios con esta orientación. Stucki et al³³ secuenciaron las cepas de un brote de TB que ocurrió en el cantón suizo de Berna. Dicho brote se dio entre 1991 y 1992, con la mayor parte de los casos encontrados asociados a factores de riesgo conocidos. Los autores compararon las secuencias genómicas de los primeros aislados del brote con las de cepas control circulantes actualmente, lo que les permitió encontrar diferencias genéticas exclusivas de las cepas asociadas al brote. A partir de ahí, desarrollaron una técnica de tipado de PCR a tiempo real y analizaron todos los aislados de casos ocurridos en Berna desde 1992. De esta manera, fueron capaces de identificar 68 casos adicionales asociados al brote³³. Al determinar la secuencia genómica de dichos casos corroboraron que se debían al brote y que podían obtener mucha más resolución en la relación entre los casos que la que se consigue con métodos tradicionales. En un estudio similar, Pérez-Lago et al³⁴ identificaron marcadores específicos de cepas circulantes en Almería asociadas previamente a brotes. Desarrollaron un ensayo conocido como TRAP (*targeted regional allele-specific oligonucleotide PCR*) para detectar prospectivamente nuevos casos asociados. La diferencia en este último caso es que se trata de una técnica de bajos requerimientos técnicos, con empleo de geles, lo que permitió descentralizar el ensayo y poder aplicarlo en los mismos laboratorios de Almería³⁴.

Epidemiología genómica

En 2012 se publica el primer estudio a escala poblacional de la utilidad de los genomas completos para epidemiología³⁵, realizado por la agencia de salud pública británica Public Health England. Para ello, se secuenciaron 247 cepas de *M. tuberculosis* aisladas en Oxfordshire durante los años 2007-2012. Puesto que en el Reino Unido el genotipado de todos los aislados se centraliza y se hace de forma rutinaria, se disponía del perfil de MIRU 24 para la comparación de los resultados genómicos. En primer lugar, el análisis demuestra que los genomas excluyen casos de transmisión que se habían considerado parte de una misma agrupación siguiendo el patrón de MIRU. En segundo lugar, se establece que una distancia de entre 10-12 cambios

nucleotídicos (SNP) es el límite para definir casos de transmisión reciente y, por último, demuestra que los pacientes con relaciones epidemiológicas claras (p. ej., por estudio de contactos) suelen tener < 5 SNP de diferencias, posiblemente reflejando una transmisión mucho más reciente. Estudios llevados a cabo con colecciones retrospectivas en los Países Bajos³⁶ y prospectivas en Reino Unido³⁷ han demostrado que estos umbrales se mantienen mientras que otros demuestran que la diversidad intrapaciente podría representar un problema para la identificación de grupos de transmisión³⁸.

En comparación con marcadores estándar de epidemiología molecular, como pueden ser los MIRU o los RFLP, queda claro que el uso de genomas va a permitir descartar casos que antes se creían asociados o, por el contrario, incluir otros nuevos previamente descartados³⁹. Walker et al³⁷ encontraron que, para aquellos casos con un vínculo epidemiológico conocido, el número de SNP de diferencia no solía ser > 5. A una mayor distancia genética, pero todavía compatible con transmisión reciente, también se encontraban casos sin relación conocida y separados por varios kilómetros de distancia. Seguramente, la mayor distancia genética (5-10 SNP) refleja el hecho de que en los casos detectados con vínculo epidemiológico es más probable que se trate de transmisiones muy recientes. El mismo estudio reveló que solo 11 de los 26 eventos de transmisión detectados por genomas completos se habían detectado también por estudio de contactos.

Quedan por explicar los casos que se observan en muchos estudios y en los que el número de mutaciones detectadas es > 15, incluso cuando existe un vínculo epidemiológico conocido³⁸. Es difícil saber por qué es tan alto el número de SNP en estos casos, pero existen varios factores que podrían explicarlo⁴⁰. Uno de los más importantes es que los tiempos de latencia son un factor de confusión, pues todavía no está claro si la tasa de mutación de la bacteria durante la latencia es o no la misma que cuando se replica activamente^{28,41}.

Casi todos los estudios de que se dispone se han llevado a cabo en países de baja incidencia; queda por ver cómo de generalizables son algunas de estas conclusiones cuando se trasladan a países de alta incidencia. El único estudio en este sentido se ha llevado a cabo en Malawi, donde se ha demostrado que los umbrales para definir transmisión son parecidos (0-10 SNP), aunque existen factores de confusión como las coinfecciones⁴². Tampoco se dispone de muchos ejemplos de países con alta incidencia en TB resistente a múltiples fármacos (MDR-TB). En un estudio prospectivo realizado en Rusia, en el que se secuenciaron más de 1.000 cepas, se demostró que más del 50% de las cepas circulantes eran MDR⁴³. Es más, muchas de ellas llevaban como mutación la resistencia a rifampicina (RIF), la S450L en el gen *rpoB* (o S531L si usamos la notación de *E. coli*), que es también la más típica en todo el mundo. Asociadas a dichas mutaciones se encontraron mutaciones secundarias en las otras subunidades de la polimerasa (*rpoA* y *rpoC*). Estas mutaciones se identificaron por primera vez en nuestro laboratorio, donde demostramos que, si bien no confieren resistencias a antibióticos, sí compensan por la pérdida de eficacia asociada a las mutaciones de resistencia a antibióticos en el gen *rpoB*⁴⁴. La combinación de la mutación de resistencia a RIF con las mutaciones compensatorias forma parte de la explicación del éxito de las cepas MDR en Rusia⁴³.

El verdadero impacto del uso del genoma como marcador epidemiológico no es tanto la delimitación de los grupos de transmisión como el poder obtener una mayor resolución de la topología de la transmisión. En el citado estudio de Malawi, un 30% de los casos de transmisión no tenían ningún SNP de diferencia, mientras que el 70% restante tenía más de uno⁴² y, por lo tanto, permiten definir la direccionalidad de la transmisión⁴⁵. La capacidad de establecer vínculos individuales de transmisión mejorará la comprensión de los factores del hospedador, la bacteria o el ambiente asociados a un mayor número de casos secundarios, poniéndose el énfasis en el número de transmisiones individuales asociadas a un caso y no en el grupo, como se derivaba hasta ahora de marcadores de baja resolución. Esto

a su vez permitirá mejorar la vigilancia, así como los modelos epidemiológicos predictivos.

Heterogeneidad intrapaciente

Como ocurre con cualquier patógeno en constante evolución, en el caso de *M. tuberculosis* la población bacteriana que infecta a un paciente no es genéticamente homogénea; la cuestión es cuál es la magnitud de dicha heterogeneidad. Esta pregunta se ha estudiado sobre todo a nivel de muestras seriadas de pacientes con MDR-TB. Dichas muestras han permitido definir con gran detalle la aparición durante el tratamiento de diferentes resistencias asociadas a antibióticos. El grado de precisión de estas técnicas es tal que se puede detectar una subpoblación o variante minoritaria al nivel del 10% o incluso del 5% siempre y cuando esté presente en el cultivo estudiado. Este tipo de aproximación ha demostrado que hay pacientes en quienes inicialmente coexisten diferentes mutaciones que confieren resistencia a un antibiótico^{46,47}. Por lo general, al inicio del tratamiento dichas subpoblaciones y sus mutaciones asociadas están en baja frecuencia con respecto a la población sensible para la TB, aunque conforme avanza el tratamiento alguna de ellas puede ser seleccionada eventualmente^{46,48,49}.

Sin embargo, es también interesante intentar entender cuál es la heterogeneidad de la bacteria en ausencia de una presión de selección tan fuerte como la de los antibióticos. En un estudio reciente se ha demostrado que en cultivos positivos de esputos consecutivos de un mismo paciente se puede detectar heterogeneidad genética en la población que infecta³⁸. Dichas diferencias también se pueden observar cuando se analizan los aislados provenientes de muestras de diferentes localizaciones dentro del cuerpo. La media de diferencias genéticas en estos casos suele estar por debajo de 5 SNP. Sin embargo, son también muchos los casos donde dicha variación no se detecta⁵⁰. Las causas de estas discrepancias no están claras. Puede ser que la infección sea reciente y no haya diversificado o que, en algunos casos, el inóculo infectivo sea diverso de partida y, por tanto, la infección que se establece también lo sea. Cabe también recordar que la presencia de diferencias genéticas no está ligada necesariamente al fenotipo encontrado. Muchas de las variaciones observadas son neutrales o casi neutrales y solo muy pocos cambios nucleotídicos tienen una verdadera relevancia funcional. Por tanto, el reto en el futuro será definir qué pequeña fracción de la diversidad intrapaciente que encontramos tiene de verdad un impacto en cualquiera de las características de la bacteria o de su interacción con el hospedador.

Queda por demostrar cómo de generalizable es la diversidad observada en algunos casos en un mismo paciente tanto en el contexto de tratamientos antibióticos como fuera de él. Una consecuencia importante es que todavía no es posible evaluar fehacientemente el impacto que dicha variación intrapaciente puede tener en las inferencias epidemiológicas que realizamos, que normalmente solo tienen en cuenta una muestra por paciente^{38,39}.

¿Coinfección, recidiva o reinfección?

Otra de las aplicaciones importantes del estudio comparado de genomas es poder determinar la frecuencia con la que los episodios recurrentes de TB en un mismo paciente se deben a casos de coinfección o si se trata de la misma cepa que dio lugar al episodio original⁵¹. A nivel clínico, este tipo de información es muy importante para evaluar la capacidad esterilizadora de nuevos tratamientos. Los métodos actuales basados en MIRU tienen una resolución limitada y son los datos genómicos los que están permitiendo una evaluación mucho más exacta del número de casos con coinfección iniciales, recidivas o reinfección⁵¹.

Como se ha visto en el apartado anterior, la coinfección se puede detectar en los datos de secuenciación masiva si una posición variable muestra evidencia de la coexistencia de poblaciones con diferen-

tes alelos. Es lo que se conoce como posiciones heterocigotas. Si el número de estas posiciones detectadas en un genoma es mayor de 15 SNP se considera que el caso muestra evidencia de coinfección: 2 cepas diferentes coexistiendo⁵². Si es menor de 10 SNP consideramos que lo que tenemos es una población clonal que ha diversificado. Una aproximación similar se puede realizar para analizar cultivos de pacientes con recidiva. En ese caso, al comparar el aislado original con el del segundo episodio, si el número de SNP encontrados es compatible con evolución desde la cepa original (< 15 SNP) entonces se puede clasificar como una recidiva; por el contrario, si el segundo caso está a más de 15 SNP de distancia, probablemente se esté hablando de un caso de reinfección^{52,53}.

Como ya se ha comentado, en el caso de la epidemiología genómica estos umbrales de SNP son todavía orientativos y se debe seguir investigando con un mayor número de muestras para establecer criterios definidos de coinfección, recidiva y reinfección. Entre otras cosas, no está claro si todos los linajes de TB tienen la misma tasa de mutación²⁹ o si hay factores durante la infección que influyen en la tasa de cambio genético de la bacteria²⁸ y, por lo tanto, si estos umbrales serían aplicables en cualquier escenario.

Diagnóstico clínico

La secuenciación genómica todavía tiene un uso limitado a nivel de diagnóstico clínico. Uno de los problemas es que hay que esperar a tener un cultivo positivo, y generalmente bien crecido, para obtener suficiente ADN⁵⁴. Sin embargo, los actuales métodos de preparación de librerías requieren cada vez menos cantidad de ADN (alrededor de 0,2 ng/μl o incluso menos). Esto ha hecho que se haya podido extraer suficiente ADN de cepas de TB obtenidas a partir de medio líquido de Middlebrook (MGIT) recién positivizado^{55,56}. Incluso se ha podido secuenciar directamente del glicerinado⁵⁷.

La secuenciación genómica tiene el potencial de proporcionar un perfil completo de cambios nucleotídicos asociados a resistencias. Esto contrasta con el limitado número de pruebas moleculares que pueden realizar diagnósticos moleculares comúnmente empleadas. Recientemente, cruzando la información genómica de más de 2.000 cepas con sus perfiles de resistencia, se ha podido determinar un panel de más de 100 mutaciones asociadas a resistencias contenidas en 24 genes⁵⁸. Como validación usaron un segundo grupo de casos confirmados microbiológicamente y predijeron la resistencia de dichos casos. Al compararlo con el método fenotípico llegaron a la conclusión de que la sensibilidad era de un 92,3% y la especificidad de un 98,4%. De hecho, la información genómica predecía mejor el perfil de resistencias que otras pruebas moleculares comunes⁵⁸. Debido a las limitaciones que se comentan más adelante, el análisis genómico aún no llega a los mismos estándares que el cultivo. Sin embargo, en muchos casos la información molecular obtenida es suficiente para ajustar el tratamiento del paciente; algo que ya ocurre en la práctica en muchos casos para la RIF cuando se detecta mediante pruebas moleculares⁵⁸.

Los nuevos fármacos en uso o en proceso de aprobación por la FDA (Food and Drug Administration) también representan un desafío para las pruebas diagnósticas actuales, pues ninguna de ellas recoge mutaciones de resistencia asociadas a dichos fármacos⁵⁹. Recientemente se ha publicado un claro ejemplo⁴⁷; en este caso se trataba de un paciente inmigrante en Suiza que inicialmente era MDR y presentaba mutaciones asociadas a resistencias a RIF, isoniazida, pirazinamida (PZA), etionamida, linezolid, moxifloxacino y estreptomina. Asimismo, presentaba una mutación de compensación para la RIF, lo que sugiere que la cepa estaba especialmente adaptada⁴⁴. El paciente tuvo un episodio recurrente después de considerarse curado y se le administró bedaquilina y, más tarde, delamanida. El paciente desarrolló resistencias fenotípicas a todas ellas, así como a clofazimina y capreomicina. La secuenciación genómica profunda de aislados consecutivos de dicho paciente ha permitido documentar en detalle la

adquisición de mutaciones a todos estos fármacos, incluyendo la bedaquilina y después la delamanida. El estudio tiene importantes implicaciones para el tratamiento de pacientes MDR y extremadamente resistentes, pero además refleja la importancia de la comunicación cruzada entre el ensayo fenotípico y la genómica.

Existen todavía varias limitaciones para establecer solamente el uso de la secuencia genómica. Primero, no todas las mutaciones que confieren resistencia a antibióticos son conocidas. Por ejemplo, para la RIF se sabe que el 95% de las mutaciones ocurren en lo que se conoce como *rifampicin resistance determining region*. Este porcentaje se reduce mucho cuando se observan los determinantes genéticos de la isoniazida (70% de mutaciones conocidas) o de otros fármacos. A esto hay que unir además que las pruebas de sensibilidad de fármacos como la PZA⁶⁰ o el etambutol (EMB)⁶¹ no están 100% estandarizadas, lo que hace la asociación estadística entre resistencia fenotípica y mutación más complicada. Por otro lado, a no ser que se consigan coberturas de secuenciación muy altas, es muy difícil determinar poblaciones resistentes minoritarias por debajo de un 10%, lo que se conoce como heteroresistencia. Este valor está claramente por debajo del valor del cultivo, que puede llegar a determinar hasta un 1%⁶². Sin embargo, es importante mencionar que la secuenciación genómica como rutina se está evaluando en el Reino Unido, lo que ha llevado a recomendar que las pruebas de sensibilidad solo se hagan en caso de que la secuencia genómica no haya identificado ninguna de las mutaciones de resistencia claramente identificadas⁵⁸.

La verdadera revolución genómica en el campo de la microbiología clínica vendrá con la identificación de la secuencia genómica, no de un cultivo, sino directamente de la muestra clínica⁴⁵. Esto se puede hacer usando técnicas de secuenciación y análisis parecidos a los que ya se usan para estudiar la microbiota para otras enfermedades⁶³. Se basa en extraer todo el ADN de la muestra y secuenciarlo directamente. Bioinformáticamente se clasifican posteriormente las lecturas como proveniente del MTBC o no. Los primeros intentos de hacer metagenómica de esputo solo han conseguido identificar partes del genoma de la bacteria⁶⁴. Esto se debe a que la mayor parte del ADN extraído es humano o de la microbiota asociada al paciente. Estos primeros resultados indican que será necesario el empleo de estrategias de enriquecimiento en la micobacteria para ser realmente capaces de usarlo como método diagnóstico⁶⁵. Alternativamente, en lugar de intentar determinar el genoma completo, se puede intentar determinar solo regiones de interés como son los genes conocidos asociados a resistencia a antibiótico. Esta aproximación se ha utilizado recientemente y ha permitido determinar incluso la presencia de heteroresistencias a un nivel del 1%, muy cerca por tanto del nivel de detección del cultivo⁶⁶. Sin embargo, todas estas técnicas están lejos de tener la estandarización necesaria para que se las incluya como parte de la rutina diagnóstica.

Limitaciones y futuro

En el caso de la epidemiología genómica, muchos de los estudios comentados se han desarrollado en países con baja incidencia. El único aplicado en un país de alta incidencia fue en Malawi, de manera retrospectiva⁴²; si bien el número de SNP entre casos de transmisión era comparable al observado en países de baja incidencia, hay que tener en cuenta que la comunidad donde dicho estudio se ha llevado a cabo es rural, donde no está presente un factor de riesgo clave como la masificación. Los ejemplos en países con alta incidencia en MDR son también limitados⁴³.

En el caso del diagnóstico molecular de resistencias, uno de los grandes problemas es que todavía existe un alto porcentaje de mutaciones asociadas a resistencias que no se conocen. De hecho, el problema es más amplio puesto que la asociación de mutaciones a resistencias depende de una buena caracterización fenotípica. Para fármacos como la PZA —donde el estudio fenotípico arroja muchas veces resultados contradictorios— es difícil asociar nuevas mutacio-

nes a la resistencia. Esta es la razón por la que la gran mayoría de dichas resistencias se han encontrado en el gen *pncA*, que es el único que claramente se ha asociado a resistencias a PZA⁶⁷. La razón por la que es difícil clasificar los casos de TB en resistentes o sensibles a PZA queda fuera de esta revisión, pero está relacionada con el tipo de medio acidificado necesario para la actuación del fármaco. El caso del EMB también es paradigmático; el valor definido de concentración crítica por el EUCAST (European Committee on Antimicrobial Susceptibility Testing) es menor que el valor epidemiológico (ECOFF [*epidemiological cut-off value for resistance*], la concentración mínima inhibitoria más alta observada entre cepas sensibles) observado en cepas sensibles a EMB, por lo que una fracción de aislados sensibles es clasificada como resistentes o viceversa⁶¹. Problemas como estos, pero en otros fármacos (p. ej., la capreomicina o incluso la rifabutinina), hacen que no siempre sea fácil asociar robustamente cambios nucleotídicos a perfiles de resistencias⁶⁸.

Para paliar estos problemas han surgido varias iniciativas que permiten el análisis en la nube de datos de secuenciación masiva, de tal manera que el usuario podrá subir a la red las lecturas resultantes de la secuenciación y el servidor lo analizará siguiendo unos estrictos criterios de control de calidad^{69,70}. Una de esas iniciativas se llama ReSeqTB (Relational Sequencing TB Data Platform)⁵⁹ y es el resultado de una cooperación única fundada por la Bill and Melinda Gates Foundation para que la desarrolle la Foundation for Innovative New Diagnostics y el Critical Path Institute involucrando a los Centers for Disease Control and Prevention (CDC) y a la Organización Mundial de la Salud (OMS). Todo el desarrollo está siendo guiado por 2 grupos expertos mundiales en diferentes aspectos del proceso, desde bioinformáticos o expertos genómicos hasta expertos en análisis clínico de resistencias o epidemiología. El resultado para el usuario medio será una lista de variantes nucleotídicas encontradas con respecto a la referencia, incluyendo un informe sobre la presencia de mutaciones conocidas de resistencia a antibióticos. Uno de los aspectos más interesantes es que, junto con la información genómica, se podrá también incluir la información fenotípica de resistencias. Esto permitirá agregar poco a poco evidencia para mutaciones que globalmente son de baja frecuencia pero que están asociadas a resistencias, de tal manera que la sensibilidad del genotipado aumentará conforme aumenta el número de cepas y de metadatos incorporados a la base de datos. Además, uno de los objetivos principales es obtener información necesaria para el desarrollo de nuevos métodos diagnósticos moleculares que puedan rápidamente detectar resistencias a antibióticos.

Si bien iniciativas como ReSeqTB son un paso importante para acercar la genómica de *M. tuberculosis* a la clínica, todavía quedan muchos pasos por andar⁷¹. Por una parte, la generación de librerías y la extracción de ADN no están estandarizadas, aunque se están dando pasos para ello guiados, sobre todo, por el interés de las agencias de salud pública británicas y norteamericanas. Por otra parte, no existe un consenso en la metodología de análisis de los datos, lo que lleva por ejemplo a que la misma cepa analizada por diferentes grupos pueda no dar siempre el mismo número de mutaciones, factor que puede tener un impacto en las inferencias epidemiológicas o de resistencias. Además se requiere cierta infraestructura informática y el trabajo de un bioinformático experto en el análisis de este tipo de datos. Si bien algunos hospitales de referencia están dando pasos para incorporar este tipo de perfiles en las plantillas, está claro que todavía se está muy lejos de la incorporación de la información genómica a la rutina diagnóstica. A esto se añade que la interpretación de los resultados no siempre es sencilla y que todavía no hay mucho consenso, como se ha ilustrado para los casos que se consideran o no de transmisión, o para determinar qué mutaciones están asociadas a resistencias.

En cualquier caso, el hecho de que agencias como los CDC o el NHS (National Health System) británico estén evaluando la secuenciación genómica para incorporarla en su rutina indica que, en un

tiempo no muy lejano, el uso de la genómica en la epidemiología y microbiología clínica será una realidad y no solo para la TB⁷². Así lo indica el hecho de que el NHS haya iniciado un proceso de certificación de la secuenciación genómica para el diagnóstico de resistencias. Sin embargo, de momento la secuenciación genómica debe ser un complemento de los actuales protocolos diagnósticos. En el futuro se verá si puede o no sustituir al cultivo en ciertas situaciones y si es o no coste-efectiva. Más que una amenaza, la microbiología clínica debe ver en estas técnicas una oportunidad para expandir las disciplinas de su especialidad y hacerlas suyas, asegurándose así seguir siendo una parte fundamental del diagnóstico y la epidemiología de la TB en el futuro.

Financiación

El trabajo en el laboratorio de Iñaki Comas está financiado por un proyecto del MINECO SAF2013-43521-R.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Bibliografía

- Alcaide F, Coll P. Advances in rapid diagnosis of tuberculosis disease and anti-tuberculous drug resistance. *Enferm Infecc Microbiol Clin.* 2011;29:34-40.
- Abe C, Hirano K, Wada M, Kazumi Y, Takahashi M, Fukasawa Y, et al. Detection of *Mycobacterium tuberculosis* in clinical specimens by polymerase chain reaction and Gen-Probe Amplified *Mycobacterium tuberculosis* Direct Test. *J Clin Microbiol.* 1993;31:3270-4.
- Nucleic Acid Amplification Tests for Tuberculosis. MMWR 1996.
- Schluger NW. Changing approaches to the diagnosis of tuberculosis. *Am J Respir Crit Care Med.* 2001;164:2020-4.
- Dye C, Glaziou P, Floyd K, Raviglione M. Prospects for tuberculosis elimination. *Annu Rev Public Health.* 2013;34:271-86.
- Comas I, Gagneux S. The past and future of tuberculosis research. *PLoS Pathog.* 2009;5:e1000600.
- Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature.* 1998;393:537-44.
- Galagan JE. Genomic insights into tuberculosis. *Nat Rev Genet.* 2014;15:307-20.
- Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB. Tuberculosis drug resistance mutation database. *PLoS Med.* 2009;6:e1000002.
- Marinova D, Gonzalo-Asensio J, Aguilo N, Martín C. Recent developments in tuberculosis vaccines. *Expert Rev Vaccines.* 2013;12:1431-48.
- Barnes PF, Cave MD. Molecular epidemiology of tuberculosis. *N Engl J Med.* 2003;349:1149-56.
- Comas I, Homolka S, Niemann S, Gagneux S. Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS One.* 2009;4:e7815.
- Kato-Maeda M, Metcalfe JZ, Flores L. Genotyping of *Mycobacterium tuberculosis*: application in epidemiologic studies. *Future Microbiol.* 2011;6:203-16.
- Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotech.* 2008;26:1135-45.
- Loman NJ, Constantinidou C, Chan JZ, Halachev M, Sergeant M, Penn CW, et al. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol.* 2012;10:599-606.
- Rasko DA, Rosovitz MJ, Myers GSA, Mongodin EF, Fricke WF, Gajer P, et al. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J Bacteriol.* 2008;190:6881-93.
- Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet.* 2013;45:1176-82.
- Firdessa R, Berg S, Hailu E, Schelling E, Gumi B, Erenso G, et al. Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. *Emerg Infect Dis.* 2013;19:460-3.
- Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi M, Sentausa E, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. *Int J Syst Evol Microbiol.* 2014;64:384-91.
- Coscolla M, Gagneux S. Consequences of genomic diversity in *Mycobacterium tuberculosis*. *Semin Immunol.* 2014;26:431-44.
- Coll F, McNeerney R, Guerra-Assunção JA, Glynn JR, Perdigo J, Viveiros M, et al. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Commun.* 2014;5:4812.
- Stucki D, Malla B, Hostettler S, Huna T, Feldmann J, Yeboah-Manu D, et al. Two new rapid SNP-typing methods for classifying *Mycobacterium tuberculosis* complex into the main phylogenetic lineages. *PLoS One.* 2012;7:e41253.
- Gagneux S, DeRiemer K, Van T, Kato-Maeda M, De Jong BC, Narayanan S, et al. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A.* 2006;103:2869-73.
- Fenner L, Egger M, Bodmer T, Furrer H, Ballif M, Battegay M, et al; Swiss HIV Cohort and Molecular Epidemiology of Tuberculosis Study Groups. HIV infection disrupts the sympatric host-pathogen relationship in human tuberculosis. *PLoS Genet.* 2013;9:e1003318.
- Comas I, Gagneux S. A role for systems epidemiology in tuberculosis research. *Trends Microbiol.* 2011;19:492-500.
- Gonzalo-Asensio J, Malaga W, Pawlik A, Astarie-Dequeker C, Passemar C, Moreau F, et al. Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. *Proc Natl Acad Sci U S A.* 2014;111:11491-6.
- Comas I, Chakravarti J, Small PM, Galagan J, Niemann S, Kremer K, et al. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat Genet.* 2010;42:498-503.
- Ford CB, Lin PL, Chase MR, Shah RR, Iartchouk O, Galagan J, et al. Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. *Nat Genet.* 2011;43:482-6.
- Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, et al. *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat Genet.* 2013;45:784-90.
- Parwati I, Van Crevel R, Van Soolingen D. Possible underlying mechanisms for successful emergence of the *Mycobacterium tuberculosis* Beijing genotype strains. *Lancet Infect Dis.* 2010;10:103-11.
- Niemann S, Köser CU, Gagneux S, Plinke C, Homolka S, Bignell H, et al. Genomic diversity among drug sensitive and multidrug resistant isolates of *Mycobacterium tuberculosis* with identical DNA fingerprints. *PLoS One.* 2009;4:e7407.
- Gardy JL, Johnston JC, Sui SJ, Cook VJ, Shah L, Brodtkin E, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med.* 2011;364:730-9.
- Stucki D, Ballif M, Bodmer T, Coscolla M, Maurer AM, Droz S, et al. Tracking a tuberculosis outbreak over 21 years: strain-specific single-nucleotide polymorphism typing combined with targeted whole-genome sequencing. *J Infect Dis.* 2015;211:1306-16.
- Pérez-Lago L, Martínez Lirola M, Herranz M, Comas I, Bouza E, García-de-Viedma D. Fast and low-cost decentralized surveillance of transmission of tuberculosis based on strain-specific PCRs tailored from whole genome sequencing data: a pilot study. *Clin Microbiol Infect.* 2015;21:249.
- Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis.* 2013;13:137-46.
- Bryant JM, Schürch AC, Van Deutekom H, Harris SR, De Beer JL, De Jager V, et al. Inferring patient to patient transmission of *Mycobacterium tuberculosis* from whole genome sequencing data. *BMC Infect Dis.* 2013;13:110.
- Walker TM, Lalor MK, Broda A, Ortega LS, Morgan M, Parker L, et al. Assessment of *Mycobacterium tuberculosis* transmission in Oxfordshire, UK, 2007-12, with whole pathogen genome sequences: an observational study. *Lancet Respir Med.* 2014;2:285-92.
- Pérez-Lago L, Comas I, Navarro Y, González-Candelas F, Herranz M, Bouza E, et al. Whole genome sequencing analysis of inpatient microevolution in *Mycobacterium tuberculosis*: potential impact on the inference of tuberculosis transmission. *J Infect Dis.* 2013;1-11.
- Walker TM, Monk P, Smith EG, Peto TE. Contact investigations for outbreaks of *Mycobacterium tuberculosis*: advances through whole genome sequencing. *Clin Microbiol Infect.* 2013;19:796-802.
- Schurch AC, Kremer K, Daviena O, Kiers A, Boeree MJ, Siezen RJ, et al. High resolution typing by integration of genome sequencing data in a large tuberculosis cluster. *J Clin Microbiol.* 2010;48:3403-6.
- Colangeli R, Arcus VL, Cursons RT, Ruthe A, Karalus N, Coley K, et al. Whole Genome sequencing of *Mycobacterium tuberculosis* reveals slow growth and low mutation rates during latent infections in humans. *PLoS One.* 2014;9:e91024.
- Guerra-Assunção J, Crampin C, Houben RM, Mzembe T, Mallard K, Coll F, et al. Large-scale whole genome sequencing of *Mycobacterium tuberculosis* provides insights into transmission in a high prevalence area. *Elife.* 2015;4:1-17.
- Casali N, Nikolayevskiy V, Balabanova Y, Harris SR, Ignatyeva O, Kontsevaya I, et al. Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat Genet.* 2014;46:279-86.
- Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, et al. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet.* 2012;44:106-10.
- Takiff HE, Feo O. Clinical value of whole-genome sequencing of *Mycobacterium tuberculosis*. *Lancet Infect Dis.* 2015;15:1077-90.
- Sun G, Luo T, Yang C, Dong X, Li J, Zhu Y, et al. Dynamic population changes in *Mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients. *J Infect Dis.* 2012;206:1724-33.
- Bloembergen GV, Keller PM, Stuckia D, Trauner A, Borrell S, Latshang T, et al. Acquired resistance to bedaquiline and delamanid in therapy for tuberculosis. *N Engl J Med.* 2015;373:1986-8.
- Eldholm V, Norheim G, Von der Lippe B, Kinander W, Dahle UR, Caugant DA, et al. Evolution of extensively drug-resistant *Mycobacterium tuberculosis* from a susceptible ancestor in a single patient. *Genome Biol.* 2014;15:490.
- Black PA, De Vos M, Louw GE, Van der Merwe RG, Dippenaar A, Streicher EM, et al. Whole genome sequencing reveals genomic heterogeneity and antibiotic purification in *Mycobacterium tuberculosis* isolates. *BMC Genomics.* 2015;16:857.
- Pérez-Lago L, Navarro Y, Montilla P, Comas I, Herranz M, Rodríguez-Gallego C, et al. Persistent infection by a *Mycobacterium tuberculosis* strain that was theorized to have advantageous properties, as it was responsible for a massive outbreak. *J Clin Microbiol.* 2015;53:3423-9.

51. Cohen T, Wilson D, Wallengren K, Samuel EY, Murray M. Mixed-strain *Mycobacterium tuberculosis* infections among patients dying in a hospital in KwaZulu-Natal, South Africa. *J Clin Microbiol.* 2011;49:385-8.
52. Bryant JM, Harris SR, Parkhill J, Dawson R, Diacon AH, van Helden P, et al. Whole-genome sequencing to establish relapse or re-infection with *Mycobacterium tuberculosis*: a retrospective observational study. *Lancet Respir Med.* 2013;1:786-92.
53. Guerra-Assunção JA, Houben RM, Crampin AC, Mzembe T, Mallard K, Coll F, et al. Recurrence due to relapse or reinfection with *Mycobacterium tuberculosis*: a whole-genome sequencing approach in a large, population-based cohort with a high HIV infection prevalence and active follow-up. *J Infect Dis.* 2015;211:1154-63.
54. Köser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog.* 2012;8:e1002824.
55. Köser CU, Bryant JM, Becq J, Török ME, Ellington MJ, Marti-Renom MA, et al. Whole-genome sequencing for rapid susceptibility testing of *Mycobacterium tuberculosis*. *N Engl J Med.* 2013;369:290-2.
56. Votintseva AA, Pankhurst LJ, Anson LW, Morgan MR, Gascoyne-Binzi D, Walker TM, et al. Mycobacterial DNA extraction for whole-genome sequencing from early positive liquid (MGIT) cultures. *J Clin Microbiol.* 2015;53:1137-43.
57. Bjorn-Mortensen K, Zallet J, Lillebaek T, Andersen AB, Niemann S, Rasmussen EM, et al. Direct DNA Extraction from *Mycobacterium tuberculosis* frozen stocks as a reculture-independent approach to whole-genome sequencing. *J Clin Microbiol.* 2015;53:2716-9.
58. Walker TM, Kohl TA, Omar S V, Hedge J, Del Ojo Elias C, Bradley P, et al. Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect Dis.* 2015;15:1193-202.
59. Starks AM, Avilés E, Cirillo DM, Denkingier CM, Dolinger DL, Emerson C, et al. Collaborative effort for a centralized worldwide tuberculosis relational sequencing data platform. *Clin Infect Dis.* 2015;61 Suppl 3:S141-6.
60. Horne DJ, Pinto LM, Arentz M, Lin SY, Desmond E, Flores LL, et al. Diagnostic accuracy and reproducibility of WHO-endorsed phenotypic drug susceptibility testing methods for first-line and second-line antituberculosis drugs. *J Clin Microbiol.* 2013;51:393-401.
61. Ångeby K, Juréen P, Kahlmeter G, Hoffner SE, Schön T. Challenging a dogma: antimicrobial susceptibility testing breakpoints for *Mycobacterium tuberculosis*. *Bull World Heal Organ.* 2012;90:693-8.
62. Folkvardsen DB, Thomsen V, Rigouts L, Rasmussen EM, Bang D, Bernaerts G, et al. Rifampin heteroresistance in *Mycobacterium tuberculosis* cultures as detected by phenotypic and genotypic drug susceptibility test methods. *J Clin Microbiol.* 2013;51:4220-2.
63. Loman NJ, Constantinidou C, Christner M, Rohde H, Chan JZ, Quick J, et al. A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic *Escherichia coli* O104:H4. *JAMA.* 2013;309:1502-10.
64. Doughty EL, Sergeant MJ, Adetifa I, Antonio M, Pallen MJ. Culture-independent detection and characterisation of *Mycobacterium tuberculosis* and *Mycobacterium africanum* in sputum samples using shotgun metagenomics on a benchtop sequencer. *PeerJ.* 2014;2:e585.
65. Brown AC, Bryant JM, Einer-Jensen K, Holdstock J, Houniet DT, Chan JZ, et al. Rapid whole-genome sequencing of *Mycobacterium tuberculosis* isolates directly from clinical samples. *J Clin Microbiol.* 2015;53:2230-7.
66. Colman RE, Schupp JM, Hicks ND, Smith DE, Buchhagen JL, Valafar F, et al. Detection of low-level mixed-population drug resistance in *Mycobacterium tuberculosis* using high fidelity amplicon sequencing. *PLoS One.* 2015;10:e0126626.
67. Miotto P, Cabibbe AM, Feuerriegel S, Casali N, Drobniewski F, Rodionova Y, et al. *Mycobacterium tuberculosis* pyrazinamide resistance determinants: a multicenter study. *MBio.* 2014;5:e01819-14.
68. Bottger EC. The ins and outs of *Mycobacterium tuberculosis* drug susceptibility testing. *Clin Microbiol Infect.* 2011;17:1128-34.
69. Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, Cirillo DM, et al. PhyResSE: a web tool delineating *Mycobacterium tuberculosis* antibiotic resistance and lineage from whole-genome sequencing Data. *J Clin Microbiol.* 2015;53:1908-14.
70. Coll F, McNerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Med.* 2015;7:51.
71. Fricke WF, Rasko DA. Bacterial genome sequencing in the clinic: bioinformatic challenges and solutions. *Nat Rev Genet.* 2013;15:49-55.
72. Köser CU, Ellington MJ, Cartwright EJP, Gillespie SH, Brown NM, Farrington M, et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog.* 2012;8:e1002824.