



LIBRO DE RESÚMENES



3º CURSO DE GENÉTICA HUMANA

24 - 26 de enero de 2008

Instituto de Biomedicina de Valencia,
CSIC y CIBER de Enfermedades Raras

Organiza:



Sociedad Española de Genética (SEG)

La genética estudia y nos informa acerca de la variación y de la herencia de los caracteres de las especies biológicas. Sin embargo, frente al estudio histórico de la genética de las especies y de las poblaciones el abordaje del individuo como tal no se ha planteado hasta fechas muy recientes y éste es un aspecto que importa mucho cuando el objeto de interés somos nosotros, los seres humanos. Podemos afirmar que “no existe ciencia del individuo y la medicina padece una contradicción fundamental: la práctica clínica se hace con los individuos mientras que su teoría es únicamente universal”; hablamos de enfermedades y no de enfermos. Esta observación fue realizada por observadores del pasado, cuando no había opción para tratar al paciente como individuo y no había suficientes herramientas científicas para abordar la cuestión de la individualidad. La genética nos ha abierto las puertas para conocer al individuo, para generar una ciencia del individuo. Hoy sabemos que esta ciencia se define por la unicidad del individuo -genética, embriológica y biográfica-; esta unicidad del individuo es la base de la variación de la especie humana, sea de individuos sanos, sea de personas enfermas. Sabemos que las enfermedades afectan a individuos y que estos padecen modos de enfermar propios de cada uno. La genética humana nos ofrece un abordaje inmediato de la unicidad del individuo al tiempo que también establece las relaciones poblaciones entre ellos.

¿Cómo es esta relación de la genética humana con la medicina, que hace que también hablemos en muchas ocasiones de genética médica? ¿Qué es la genética médica? Empezando por la definición común de genética como la ciencia de la variación biológica, se podría decir que la genética humana es la ciencia de la variación biológica en los seres humanos y que la genética médica es la ciencia de la variación biológica humana en relación con la salud y la enfermedad. La genética clínica es la parte de la genética médica interesada en la salud y modos de enfermar de los individuos humanos y de sus familias. Estos aspectos de la genética como ciencia básica del ser humano, la aplicación de la misma a los modos de enfermar, esto es, a la medicina, y su traslación biunívoca a la medicina clínica, tanto por responder a las necesidades de la práctica clínica como para recibir las preguntas que se generan en el acto clínico, son el leitmotiv de curso sobre genética humana que ofrece la Sociedad Española de Genética en su tercera edición en estas primeras fechas de 2008. Con el diseño de los temas y la selección de los ponentes se ha pretendido ofrecer un panorama amplio de la genética humana, que aúne a la vez cuestiones actuales de gran relevancia científica con aspectos de desarrollo tecnológico y de aplicación en la práctica clínica de las enfermedades genéticas. Esperemos que el curso sea de vuestro agrado y os resulte útil y que, a su vez, os muestre la gran relevancia que esta disciplina científica y clínica tiene en el desarrollo de la medicina moderna, una medicina que en muchos de sus postulados es una medicina genética.

Carmen Espinós
Francesc Palau

Valencia, 24 de enero de 2008

ÍNDICE

F. Palau. Unidad de Genética y Medicina Molecular, Instituto de Biomedicina de Valencia, CSIC, y CIBER de Enfermedades Raras. Valencia.

El marco traslacional de la genética humana

7

1. GENOMA: NUEVOS CONOCIMIENTOS, NUEVAS PREGUNTAS

15

J. Escribano. Área de Genética, Facultad de Medicina, Universidad Castilla-La Mancha. Albacete.

Estructura del genoma humano: perspectivas en biomedicina

17

J. Fernández- Piqueras. Departamento de Biología, Unidad de Genética, Facultad de Ciencias, Universidad Autónoma de Madrid. Madrid.

Los microRNAs como genes de susceptibilidad en cáncer

21

J. Surrallés. Departament Genètica i Microbiologia, Universitat Autònoma de Barcelona y CIBER de Enfermedades Raras. Barcelona.

Síndromes de reparación del DNA y predisposición al cáncer: lecciones desde la anemia de Fanconi

25

2. CARTOGRAFIADO, IDENTIFICACIÓN DE GENES, FISIOPATOLOGÍA MOLECULAR

27

C. Espinós. Unidad de Genética y Medicina Molecular, Instituto de Biomedicina de Valencia, CSIC, y CIBER de Enfermedades Raras. Valencia.

Polimorfismos, mapas y genes en las enfermedades mendelianas

29

J. Pérez-Tur. Unidad de Genética Molecular, Instituto de Biomedicina de Valencia, CSIC. Valencia.

Análisis genético de formas no mendelianas de enfermedades

37

S. Rodríguez de Córdoba. Departamento de Inmunología, Centro de Investigaciones Biológicas (CIB), CSIC, y CIBER de Enfermedades Raras. Madrid.

Fisiopatología molecular y celular en la enfermedad genética

41

J. M. Millán. Unidad de Genética y Diagnóstico Prenatal, Hospital Universitario La Fe y CIBER de Enfermedades Raras. Valencia.

El síndrome de Usher: Diagnóstico molecular en una enfermedad genéticamente heterogénea y la importancia del estudio poblacional

47

F. J. Ramos. Departamento de Pediatría, Facultad de Medicina, Universidad de Zaragoza. Zaragoza.

Diagnóstico y asesoramiento genético: papel de la genética clínica asistencial en los servicios de salud en España

51

3. UTILIDAD DE LOS ORGANISMOS MODELO EN EL ESTUDIO DE LAS ENFERMEDADES GENÉTICAS HUMANAS

55

P. Sanz. Unidad de Señalización por Nutrientes, Instituto de Biomedicina de Valencia, CSIC y CIBER de Enfermedades Raras. Valencia.

La levadura *Saccharomyces cerevisiae* como modelo y herramienta para el estudio de enfermedades genéticas humanas

57

J. C. Rodríguez Aguilera. Centro Andaluz de Biología del Desarrollo (CABD), Universidad Pablo de Olavide-CSIC, y CIBER de Enfermedades Raras. Sevilla.

***Caenorhabditis elegans*, un espejo de 959 células**

61

M. D. Moltó. Departamento de Genética, Facultad de Ciencias Biológicas, Universitat de València. Valencia.

Qué ofrece *Drosophila* como modelo de estudio de las enfermedades neurodegenerativas

65

J. L. Gómez-Skármeta. Centro Andaluz de Biología del Desarrollo (CABD), Universidad Pablo de Olavide, CSIC. Sevilla.

La genómica comparada en el estudio del desarrollo, la evolución y las enfermedades génicas

71

P. Pérez. Laboratorio de Modelos Animales, Instituto de Biomedicina de Valencia, CSIC, y Centro de Investigación Príncipe Felipe (CIPF). Valencia.

Modelos transgénicos para el estudio del síndrome humano Displasia Ectodérmica

75

4. DE GENOMAS Y PROTEOMAS EN LA ENFERMEDAD GENÉTICA

79

I. Marín. Unidad de Bioinformática, Instituto de Biomedicina de Valencia, CSIC. Valencia.

Genómica comparativa de genes implicados en la enfermedad de Parkinson

81

A. Llerena. Centro de Investigación Clínica del Área de Salud de Badajoz (CICAB), SES, Universidad de Extremadura. Badajoz.

Aplicación de la farmacogenética al tratamiento de pacientes psiquiátricos

85

H. Dopazo. Unidad de Farmacogenómica y Genómica Comparativa, Departamento de Bioinformática, Centro de Investigación Príncipe Felipe (CIPF), y CIBER de Enfermedades Raras. Valencia.

Predicción de mutantes deletéreos en el genoma humano

89

V. Rubio. Laboratorio de Enzimopatología Estructural, Instituto de Biomedicina de Valencia, CSIC y CIBER de Enfermedades Raras. Valencia.

Biología estructural y genética humana

93

F. Javier Chaves. Laboratorio de Estudios Genéticos, Fundación para la Investigación del Hospital Clínico Universitario de Valencia.

Nuevas tecnologías actuales en los laboratorios de genética clínica

97

EL MARCO TRASLACIONAL DE LA GENÉTICA HUMANA

Francesc Palau

Unidad de Genética y Medicina Molecular, Instituto de Biomedicina de Valencia, CSIC, y CIBER de Enfermedades Raras. Valencia.

Durante los últimos treinta años la genética humana se ha convertido en una ciencia básica de la moderna investigación biomédica, y el desarrollo de su aplicación a la práctica clínica ha sido fundamental para ampliar nuestro conocimiento de las causas de las enfermedad humana y del desarrollo de nuevas técnicas diagnósticas y aproximaciones terapéuticas, así como del consejo genético, entendido éste como una proceso de información y, también, una herramienta terapéutica. En todo este tiempo reciente llama la atención la celeridad con la que muchos hallazgos de laboratorio de investigación han llegado al marco sanitario, principalmente a los laboratorios clínicos. Este proceso de traslación desde *la poyata a la cabecera del enfermo* supone también que haya aumentado el interés científico por responder a preguntas nacidas de la práctica clínica. En palabras de Jordi Camí “la investigación traslacional es la investigación biomédica que es capaz de trasladar sin solución de continuidad un hallazgo de laboratorio a la práctica clínica y, viceversa, la investigación capaz de trasladar las preguntas que surgen en la observación clínica a respuestas en el laboratorio.” Pues bien, ¿cuál es el marco en el que se ha venido desarrollando la interacción entre la investigación genética y la medicina, especialmente en lo relativo al estudio de las causas y mecanismos que causan enfermedad, el diagnóstico y el pronóstico de las enfermedades genéticas y, finalmente, el tratamiento de las mismas? Para dar una respuesta a esta pregunta son varios los aspectos que conviene tratar: i) conocer el proceso histórico de de la genética humana en el siglo XX, especialmente desde la década de los años 50 hasta nuestros días, y determinar la relación de la genética humana con las especialidades médicas, ii) reconocer las disciplinas de la genética que se han desarrollado en relación con la medicina (también con la biología evolutiva de ser humano), iii) definir el carácter nosológico de la enfermedad genética, iv) remarcar los hitos que han supuesto un cambio fundamental en la investigación en genética humana, y v) revisar los aspectos clínicos aplicados sobre los que ha incidido el conocimiento que ha generado la investigación genética sobre el modo de enfermar.

La genética hunde sus raíces en el siglo XIX en los estudios del Mendel sobre las leyes de la herencia. La primera propuesta sobre el papel de la genética en relación con la enfermedad fue el concepto de “errores congénitos del metabolismo” introducido por Archibald Garrod quien los propuso formalmente en sus *Croonian lectures*, publicadas en 1909, discutiendo sobre la alkaptonuria, la pentosuria, la cistinuria y el albinismo. La aproximación científica de la genética humana se puede datar en 1956 cuando se determinó que el número de cromosomas del complemento diploide humano era de 46. Obviamente también fue fundamental el reconocimiento del DNA como la molécula que constituye el material hereditario por parte del grupo de

Avery y de su estructura de doble hélice por Watson y Crick, y el concepto de un gen, una enzima de Beadle y Tatum.

La genética tiene una especial relación con la medicina. De hecho, es uno de los pilares en los que se fundamenta la moderna medicina y podemos hablar de una verdadera medicina genética que da soporte al reconocimiento del individuo y de su manera específica, individual, de enfermar. La genética humana o médica se ha desarrollado en el contexto de algunas especialidades médicas como la pediatría, la medicina interna o la obstetricia y, más recientemente, de la neurología o la oncología. Sin embargo, al contrario que estas disciplinas clínicas cuyo fundamento son el momento cronológico del desarrollo del ser humano o un sistema, órgano o proceso patológico, que podríamos definir genéricamente como vertical, la genética humana desarrollada como disciplina clínica (genética médica o clínica) tiene un carácter fundamentalmente transversal e involucra al resto de disciplinas médicas.

La genética humana y médica tiene muchos campos de interés, tal como viene indicado por las distintas direcciones en las que se ha desarrollado la genética. Las áreas principales reconocidas de especialización son el estudio de los cromosomas (citogenética), el estudio de la estructura y función de los genes (genética molecular y bioquímica), el estudio del genoma, su organización y función (genómica) el estudio de la variación genética en las poblaciones humanas, los factores que determinan las frecuencias alélicas y el origen de las mutaciones (genética poblacional), el estudio de la distribución poblacional y geográfica de las enfermedades genéticas (epidemiología genética), el estudio del control genético del desarrollo (genética del desarrollo), y la aplicación de la genética al diagnóstico, manejo, tratamiento y cuidado del paciente (genética clínica).

El concepto de enfermedad genética ha variado considerablemente en los últimos tiempos, de tal modo que el papel de los genes en el proceso de enfermar no sólo interesa por su relación causal inmediata y por el carácter hereditario que se da en muchas ocasiones de las enfermedades de base genética, sino también por la participación directa del gen en los mecanismos patogénicos y en la alteración de la homeostasis de las rutas metabólicas y fisiológicas del organismo. En un sentido amplio pero útil, aun a riesgo de ser reduccionista, la nosología de la enfermedad genética puede establecerse en tres grandes grupos: trastornos monogénicos (incluyen las enfermedades mendelianas por mutación en el genoma nuclear y las enfermedades mitocondriales), los síndromes cromosómicos y las enfermedades complejas o multifactoriales. La aproximación a su estudio y análisis y su enfoque clínico es distinta aunque compartan algunos aspectos comunes.

La investigación genética humana tiene puntos específicos que requieren aproximaciones metodológicas diferentes al conjunto de la investigación en genética. Uno de ellos, por otra parte fundamental, es el hecho de que no es posible generar poblaciones humanas con un acervo genético específico; es éticamente y moralmente inadmisibles. No obstante, sí es posible utilizar estrategias que permitan localizar genes en el genoma humano y definir las mutaciones que causan las enfermedades. En este sentido, hemos de indicar los enormes logros que se ha alcanzado en el cartografiado de genes mendelianos, la identificación de genes mutantes y en la patología molecular. La estrategia de la clonación posicional, implementada en los años ochenta del pasado siglo, ha sido básica en este proceso. También es una realidad el cartografiado y la identificación de genes que confieren susceptibilidad a enfermedades complejas. El otro gran avance en la genética humana y la genética en general ha sido el Proyecto Genoma Humano y los distintos proyectos genoma de diferentes especies, algunas clásicas en la historia y el desarrollo de la genética

del siglo XX. La investigación genética biomédica abarca actualmente un abanico muy amplio de campos como son el propio cartografiado e identificación de genes; la aplicación al diagnóstico genético y de disponibilidad de nuevas pruebas y tecnologías genéticas; la genómica funcional, la genómica estructural, la proteómica y la bioinformática; la manipulación genética de células y animales y, finalmente, las nuevas aproximaciones de tratamiento de la enfermedad con el desarrollo de nuevas terapias basadas en el conocimiento genético o en la tecnología genética. La genética es una pieza fundamental en el desarrollo de la biología de sistemas y, por ende, de la patología o la biomedicina de sistemas.

¿Cómo se ha venido trasladando este conocimiento, en muchas ocasiones alcanzado a una alta velocidad, en el campo de la medicina clínica, donde la genética humana tiene uno de sus campos de acción más importantes? En una palabra, ¿cuál ha sido y es el marco traslacional de la genética humana? Tres son las áreas de traslación hacia la clínica: el diagnóstico (y lo que comporta en cuanto a pronóstico), el consejo genético y la terapia. Cada una de estas áreas tiene un desarrollo e implantación dispar en función de las estructuras de servicio de los sistemas de salud y del momento actual de desarrollo del conocimiento y de la tecnología.

Quisiera, pues, para finalizar, exponer algunas cuestiones importantes para dos de estos puntos, el diagnóstico y el consejo genéticos, en donde se refleja la enorme incidencia y capacidad para aplicar en la práctica clínica el conocimiento en genética humana que se ha generado a tiempos recientes.

Diagnóstico genético

El diagnóstico en medicina es un proceso clínico con el que se pretende responder a la pregunta: ¿qué está alterado en el individuo? El análisis genético ha aportado al diagnóstico de las enfermedades una batería de herramientas biológicas de primer orden, no sólo por su capacidad de confirmar el diagnóstico clínico de sospecha mediante el estudio de mutaciones en los genes específicos de cada enfermedad o grupo de enfermedades sino también por ofrecer al médico la capacidad de aplicar el diagnóstico a situaciones clínicas distintas de la que plantea el enfermo, como es el caso del diagnóstico predictivo en individuos sanos o en medicina reproductiva. Este diagnóstico genético o molecular requiere de la disponibilidad de pruebas genéticas y del desarrollo de técnicas moleculares que mejoren la calidad de los análisis genéticos al tiempo que permitan ampliar el espectro de enfermedades a las que ofrecer un diagnóstico. En este sentido, es muy pertinente dejar claro que diagnóstico genético y análisis genético o pruebas genéticas (el *genetic testing* anglosajón) no son la misma cosa y requiere que lo contextualicemos en el marco de la práctica clínica.

El análisis genético se utiliza para identificar cambios en la secuencia del DNA o en los cromosomas que se correlacionen con una enfermedad o riesgo elevado de desarrollar una enfermedad. Este tipo de análisis se puede emplear para el diagnóstico antes de que los síntomas sean reconocibles y para determinar el riesgo personal para ciertas enfermedades multifactoriales o complejas. De ahí que los resultados de un análisis genético puedan tener un efecto directo en la vida de las personas. Existen diversas definiciones con variaciones entre ellas de “análisis genético o pruebas genéticas”, dependiendo de cuál es la amplitud que se quiera abarcar. La definición de trabajo utilizada por la OCDE indica que “las pruebas genéticas son análisis de las variaciones en

las secuencias de DNA de la línea germinal o de los productos génicos o efectos originados por cambios en secuencias heredables, los cuales son predictivos de efectos significativos en la salud”.

En análisis genético que acabamos de definir se puede aplicar en distintas situaciones clínicas y para diferentes propósitos diagnósticos:

- **Diagnóstico biológico clínico:** Posiblemente se trata de la causa más frecuente por la que se solicita una prueba genética para un paciente que presenta síntomas y signos clínicos de sospecha de un trastorno que puede tener causa genética. En este caso, la prueba se practica para confirmar, matizar o excluir un diagnóstico clínico. En muchos casos el análisis genético se emplea ampliamente como una prueba de exclusión con baja probabilidad de que dé un resultado positivo (un ejemplo es la prueba para el síndrome del cromosoma X frágil en niños con problemas de aprendizaje o cognitivos).

- **Diagnóstico predictivo:** El objetivo es estimar el riesgo de desarrollar un trastorno genético en el futuro para una persona asintomática. Este tipo de diagnóstico requiere para su implantación la puesta en marcha de un programa específico en el que participe no sólo los neurólogos (u otros clínicos según la especialidad) y genetistas clínicos, sino también psicólogos, asistentes sociales y personal de apoyo. Se pueden distinguir dos formas de diagnóstico predictivo:

Diagnóstico presintomático, mediante el cual se busca una mutación en un individuo sano la cual, si está presente, casi con total certeza condicionará la aparición de la enfermedad. Este tipo de prueba se aplica a condiciones genéticas de la edad adulta tales como la enfermedad de Huntington. Los adultos pueden no tener síntomas en el momento en que se realiza la prueba pero existe un riesgo alto (50%) de que haya heredado la mutación de un progenitor enfermo y pueda padecer la enfermedad en un futuro. En gran medida el interés de estas personas está en organizar su vida personal y reproductiva en función de la información que se les aporte.

Diagnóstico de predisposición, cuyo objetivo es analizar una mutación que pudiera aumentar la susceptibilidad a padecer un trastorno (ej. las mutaciones en los genes *BRCA1* y *BRCA2* que proveen cierta susceptibilidad al cáncer de mama, pero que un resultado positivo no es indicativo de un 100% de riesgo de desarrollar el tumor mamario). Hay muchas otras enfermedades comunes complejas –en el campo de la neurología el prototipo sería la enfermedad de Alzheimer– que son multifactoriales y cuya etiología está influida tanto por factores genéticos como ambientales. No obstante, en trastornos como la enfermedad de Alzheimer o la enfermedad de Parkinson hay que tener en cuenta la existencia de formas raras monogénicas que tienen una herencia mendeliana y que requieren un tratamiento como el resto de trastornos monogénicos.

- **Diagnóstico de portadores:** En las enfermedades autosómicas recesivas o ligadas al X los individuos portadores son sanos. En esta situación la realización de una prueba genética puede tener valor para la toma de decisiones sobre aspectos reproductivos. En niños en los que la prueba genética no tiene ninguna implicación para su salud, el análisis de portador es controvertido y muchos genetistas están a favor de que la decisión se posponga hasta que el niño alcance la edad adulta y pueda solicitarlo por sí mismo tras un consentimiento informado. Es una situación muy similar a la del diagnóstico predictivo.

- **Diagnóstico prenatal:** Esta clase de diagnóstico tiene como objetivo averiguar si el feto es portador de una mutación y por tanto, de desarrollar la enfermedad en cuestión tras el nacimiento, siempre en función del tipo de herencia. El diagnóstico prenatal para el síndrome de Down y otras alteraciones cromosómicas tiene demanda principalmente a causa del incremento del riesgo asociado a la edad materna. Por comparación, la indicación de diagnóstico prenatal de trastornos monogénicos es menos frecuente. Suele solicitarse a raíz de que una pareja tiene un riesgo elevado por experiencia propia o familiar de alguien que padece la enfermedad.

- **Diagnóstico genético preimplantatorio (DGP):** Es un proceso diagnóstico que se practica en embriones obtenidos mediante técnicas de reproducción asistida en parejas con antecedentes de enfermedad monogénica en ellos mismos o en la familia. Se realiza sobre una o dos células del embrión en fase de blastocisto de modo que se puedan seleccionar para su implantación aquéllos que no son portadores de la mutación en cuestión.

- **Cribado genético:** Los diagnósticos predictivo, prenatal o de portadores se puede ofrecer a la población en general o a una subpoblación particular por un elevado riesgo. La implementación de un cribado genético requiere de una política por parte de las autoridades sanitarias y de desarrollar un programa específico.

Hay otros contextos clínicos en los que el análisis genético tiene indicaciones específicas como puede ser la farmacogenética que permite analizar variantes polimórficas de respuestas a medicamentos o la clasificación en subtipos de las enfermedades, cuestión que tiene relevancia en oncología para caracterizar distintos tipos de cáncer.

En este marco conceptual y práctico, ¿qué hace distinto al análisis o diagnóstico genético de otros análisis clínicos? Las pruebas genéticas difieren de otros tipos de análisis clínicos. Los resultados pueden tener implicaciones para la persona y posiblemente también para sus hijos y otros familiares. Los individuos se hacen el análisis sólo una vez en su vida y un genotipado erróneo puede tener consecuencias importantes incluso años más tarde. En este caso, tanto un resultado negativo como positivo equivocado puede perturbar enormemente a las personas afectadas. Los análisis genéticos tienen además la peculiaridad, poco frecuente en otros tipos de análisis clínicos o bioquímicos, de que necesitan una interpretación de los resultados en el contexto del riesgo de recurrencia y de ofrecer un consejo genético.

Consejo genético

En 1975 un comité de la *American Society of Human Genetics* propuso una definición de consejo o asesoramiento genético que ha venido siendo fundamental en la práctica clínica y en la explicación del significado del consejo genético como herramienta clínica y terapéutica:

“El consejo genético es un proceso de comunicación mediante el cual se abordan problemas humanos asociados con la padecimiento o con el riesgo de recurrencia de un trastorno genético en una familia. Este proceso supone un esfuerzo de una o más personas entrenadas adecuadamente para ayudar al individuo a la familia a: (1) la comprensión de los hechos médicos incluyendo el diagnóstico, el curso probable del trastorno y su manejo disponible, (2) apreciar cómo contribuye la herencia en el riesgo de recurrencia en los familiares, (3) comprensión de las herramientas disponibles para modificar el riesgo de recurren-

cia, (4) elegir un modo de actuar que sea apropiado en función del riesgo, necesidades familiares, creencias religiosas y actuar acorde con tales decisiones, y (5) modificar del mejor modo posible el curso de la enfermedad en los miembros afectados de la familia y/o el riesgo de recurrencia del trastorno en cuestión.”

Esta definición es muy interesante por varias razones. En primer lugar, resalta el hecho de que la naturaleza de la interacción es en ambos sentidos, médico *versus* paciente y viceversa, y no sólo en el de “dar consejo o asesorar”. En segundo lugar, se trata de un proceso que idealmente tiene lugar a lo largo de un período de tiempo de modo que el paciente puede asimilar gradualmente la complejidad de la información relativa al diagnóstico, pronóstico y riesgo, así como ser capaz de formular sus decisiones o estrategias. El tercer punto relevante es el énfasis que manifiesta en la autonomía del paciente en la toma de decisiones relacionadas con las pruebas genéticas, la reproducción o el tratamiento, y el reconocimiento de que tales decisiones serán diferentes dependiendo del contexto personal, familiar o cultural en el que están hechas. El cuarto punto radica en el reconocimiento del impacto familiar que tiene un trastorno genético y que no tiene otras clases de enfermedad, lo que da pie a hablar de un componente psicoterapéutico del consejo genético para ayudar a la gente a manejarse con las implicaciones reproductivas y de cualquier índole que tienen estas enfermedades catalogables en muchas ocasiones como raras. Un aspecto importante viene dado por la frase “personas [profesionales] entrenadas adecuadamente”, lo que implica que el consejo genético requiere unos conocimientos y habilidades especiales distintas de aquellas que se necesitan para otras actuaciones médicas.

El consejo genético es un proceso dinámico y cambiante que ha de adaptarse a las necesidades sociales y los avances en el conocimiento científico y técnico, y así ha venido siendo desde la época de la definición que hemos utilizado. En la práctica actual el consejo genético sigue siendo un proceso de comunicación pero también hay que verlo como una herramienta terapéutica que permite modificar la vida o la manera de ver los problemas a las personas afectadas por trastornos no sólo genéticos sino también adquiridos como es el caso de la exposición a teratógenos durante el embarazo, manteniendo siempre la autonomía del individuo en la toma de decisiones. Hay aspectos fundamentales a tener en cuenta en la filosofía de los servicios de genética y de consejo genético. Entre ellos cabe reseñar la utilización enteramente voluntaria de los servicios, evitándose presiones económicas o eugenésicas de promoción de la prevención de la enfermedad genética. Idealmente, los servicios genéticos, incluyendo asesoramiento, diagnóstico y tratamiento deben alcanzar a toda la población. Parte de las actividades de los servicios de genética deberían estar orientados hacia la educación de los pacientes y de la población general, aunque también a profesionales de disciplinas relacionadas con aspectos médicos, sanitarios, psicológicos y sociales. La mayoría de genetistas clínicos y asesores genéticos (una figura ésta que no existe en nuestro país) son partidarios de ofrecer a los pacientes toda la información relevante. Este es un aspecto importante para la educación y la capacidad de tomar de decisiones por parte de las personas afectadas, aunque hay puntos concretos que puede ser cuestionable si el beneficio es mayor que el posible perjuicio. Posiblemente, la característica que define mejor al consejo genético que se ofrece hoy en día es la aceptación de una aproximación no prescriptiva o no dirigida del asesoramiento, es decir, ofrece el consejo genético sin ánimo de favorecer ninguna acción particular entre las posibles opciones que se plantean. Finalmente, cabe destacar los aspectos de la atención psicosocial y las dimensiones afectivas que hay en un proceso de comunicación como es el asesoramiento genético, así como el criterio básico de confidencialidad y protección de la privacidad de las personas familiares afectadas por enfermedades genéticas.

Referencias

- Antonarakis SE, Beckmann JS. Mendelian disorders deserve more attention. *Nature Rev Genet* 2006;7:277-282
- Camí Morell J. Investigación biomédica. En: Tratado de Medicina Interna Farreras-Rozman. Rozman C ed. 2002
- Childs B, Wiener C, Valle D. A science of the individual: implications for a medical school curriculum. *Annu Rev Genom Hum Genet* 2005;6:313-330
- Espinós C, Palau F. Genética de las enfermedades neurológicas. En: Tratado de Neurología, Pascual J ed., Ars XXI, Barcelona, 2008
- McKusick VA. A 60-year tale of spots, maps, and genes. *Annu Rev Genom Hum Genet* 2006;7:1-27
- Nussbaum RL, McInnes RR, Willard HF. Thompson & Thompson Genetics in Medicine, W.B. Saunders, Filadelfia, 2001
- Rimoin DL, Hirschhorn K. A history of medical genetics in pediatrics. *Pediatr Res* 2004;56:150-159
- Strachan T, Read AP. Human Molecular Genetics 3, Garland Publishing, Londres y Nueva York, 2004



1



1. GENOMA: nuevos conocimientos, nuevas preguntas

ESTRUCTURA DEL GENOMA HUMANO: PERSPECTIVAS EN BIOMEDICINA

Julio Escribano Martínez

Área de Genética, Facultad de Medicina, Universidad Castilla-La Mancha, Albacete.

El proyecto Genoma Humano ha impulsando de una manera espectacular nuestra comprensión de la estructura y función del genoma humano. En paralelo al desarrollo de este proyecto se han secuenciado o se están secuenciando y analizando los genomas de otros organismos. Así, con el nacimiento del siglo XXI, hemos adquirido la capacidad de abordar los problemas biológicos de una manera global, y hemos asistido a la aparición de nuevas disciplinas como la genómica, proteómica, transcriptómica, etc. En esta charla analizaremos cómo se organiza el genoma humano, en buena medida bajo la luz aportada por el proyecto genoma humano, y veremos algunas aplicaciones de este conocimiento en medicina.

Para comenzar, conviene tener presentes algunos datos básicos sobre el genoma:

- El tamaño del genoma nuclear es de unos 3.000 millones de pares de bases y se estima que contiene unos 25.000 genes.
- El tamaño medio de un gen humano es de unos 28.000 pb.
- Las secuencias génicas y relacionadas representan aproximadamente el 37% del ADN genómico.-
- Una pequeña fracción del del genoma nuclear corresponde a ADN codificante (aproximadamente el 1,5%).
- Los genes con función similar pueden estar agrupados, pero es más habitual que estén dispersos por diferentes cromosomas.

El genoma mitocondrial humano está constituido por ADN circular bicatenario, contiene 37 genes: 13 que codifican polipéptidos mitocondriales (complejos respiratorios); 22 que codifican ARNt y 2 que codifican ARNr.

Casi todos los genes están presentes en una única copia por juego haploide de cromosomas, aunque algunos de ellos se encuentran repetidos, estando las distintas copias dispersas por el genoma o bien situadas unas a continuación de otras. Las duplicaciones génicas explican la repetición de estos genes.

Más de la mitad del genoma está integrado por secuencias que no forman parte de los genes y que en su inmensa mayoría está repetido. La función del ADN extragénico repetido es poco conocida. Según la organización de las repeticiones podemos encontrar **ADN repetido en tándem**, y ADN repetido **disperso**.

El ADN repetitivo en tándem está formado por secuencias de longitud variable repetidas unas a continuación de otras. Estas repeticiones pueden estar localizadas en diferentes localizaciones cromosómicas. Dependiendo del tamaño medio de la unidad que se repite, y de la longitud total del bloque repetido puede clasificarse en: satélite, minisatélite y microsatélite.

El ADN satélite esta formado por múltiples repeticiones en tándem de unidades de tamaño medio (5-170 nucleótidos), que pueden alcanzar en total longitudes de cientos de miles de pares de bases. Este tipo de ADN no se transcribe, y constituye el grueso de la heterocromatina. Cuando se centrifuga ADN genómico en un gradiente de densidad de CsCl, aparecen bandas separadas de la banda principal de ADN que corresponde a este tipo de ADN y por ello se denomina satélite.

El ADN minisatélite está compuesto por unas 1000 repeticiones de secuencias cortas (6-24 nucleótidos). Aunque la mayoría de este tipo de ADN se encuentra cerca de los telómeros, existen secuencias minisatélite intercaladas en otras posiciones cromosómicas. El ADN minisatélite se caracteriza su polimorfismo, es decir por la variación del número de repeticiones entre individuos.

El ADN microsatélite esta constituido por pequeños bloques de repeticiones en tándem de una secuencia de 1 a 4 nucleótidos de longitud y es muy polimórfico. Los microsatélites también se denominan STR (secuencias cortas repetidas en tándem, en inglés *short tandem repeats*). El análisis de este tipo de secuencias se ha incorporado de manera rutinaria a la práctica forense.

ADN repetitivo disperso (secuencias derivadas de transposiciones). Existen algunos elementos de ADN capaces de trasladarse de una posición cromosómica a otra, denominados elementos genéticos móviles o transponibles. Estos elementos han originado muchas secuencias repetidas y dispersas que en conjunto representan aproximadamente un **43%** del genoma humano. Existen dos grandes grupos de elementos genéticos móviles los transposones y los retrotransposones. Los primeros utilizan un intermediario de ADN para desplazarse y los segundos de ARN, sintetizado mediante transcripción reversa. En el genoma humano son mucho más numerosos los retrotransposones y elementos relacionados que los transposones. Los elementos de la familia LINE-1 son retrotransposones denominados autónomos por codificar transcriptasa inversa. Con unas 600,000 copias, los LINE representan aproximadamente el 15% del genoma humano. Las secuencias Alu, denominadas así por contener un lugar de corte de la enzima de restricción Alu I, también son muy abundantes en el genoma humano y a diferencia de los LINE no codifican transcriptasa inversa. La función de este tipo de ADN es poco conocida.

Variación del genoma humano. Gran parte de las diferencias genéticas entre individuos corresponden a cambios de un solo nucleótido en una determinada posición (*locus*). Estas diferencias se denominan SNPs (*Single Nucleotide Polymorphisms*) o polimorfismos de un solo nucleótido. Se estima que el genoma de dos individuos difiere en unos 15 millones de nucleótidos. Se han identificado más de 3 millones de estos SNPs que están siendo analizados para determinar su implicación en las enfermedades comunes. Este tipo de variación genética es responsable de una fracción importante de la susceptibilidad genética a la enfermedad. Durante el último año se han identificado variaciones de mas de 50 genes que incrementan el riesgo de padecer enfermedades como la diabetes, la enfermedad de Crohn o la artritis reumatoide. Además, los genomas de los individuos pueden diferenciarse en la duplicación, delección y/o inversión de regiones cromosómicas de miles o millones de pares de bases de longitud. Estos cambios en el ADN pueden afectar al número de copias de los genes o a sus elementos reguladores. La variación en el número de copias de ciertos genes también puede predisponer a padecer algunas enfermedades complejas como el autismo.

La identificación de mutaciones en enfermedades monogénicas ha aumentado la capacidad de realizar diagnóstico prenatal y presintomático. Esta información, además de contribuir a entender los mecanismos moleculares que conducen del genotipo al fenotipo, permitirá el desarrollo de nuevas terapias génicas. Más importante aún, a medida que las técnicas de análisis masivo de la secuencia del ADN continúen su evolución, se producirá un giro en la estrategia actual de asistencia médica, pasando de un modelo sanitario que trata la enfermedad en sus fases avanzadas, a un modelo preventivo, basado en la identificación de los riesgos de cada individuo de padecer distintas patologías. Ello permitirá actuar sobre los factores modificables desencadenantes de la enfermedad.

Referencias

Couzin, J. and Kaiser, J. Genome wide association Closing the Net on Common Disease Genes. **Science** (2007) 316, 820–822.

Estivill, X. and Armengol L. Copy Number Variants and Common Disorders: Filling the Gaps and Exploring Complexity in Genome-Wide Association Studies. **PLoS Genetics** (2007) 3, 1787-1799.

International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. **Nature** (2001) 409, 860-921.

Sebat, J. et al., Strong Association of De Novo Copy Number Mutations with Autism. **Science** (2007) 316, 445.

Stranger, B.E. et al., Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes. **Science** (2007) 315, 848.

T. Strachan y A. P. Read. Human Molecular Genetics. 3rd edition. Garland Publishing 2004. Chapter 9: “Organization of the human genome”

Venter, C. y col. The Sequence of the Human Genome. **Science** (2001) 291, 1304-1351.

Wong, K. K. et al. A Comprehensive Analysis of Common Copy Number Variations in the Human Genome. **Am. J Hum. Genet.** (2007) 80: 91–104.

LOS MICRORNAS COMO GENES DE SUSCEPTIBILIDAD EN CÁNCER

José Fernández Piqueras

Universidad Autónoma de Madrid. Centro de Biología Molecular Severo Ochoa CSIC-UAM.

Introducción

El Premio Nóbel en Fisiología y Medicina del año 2006 recayó en los investigadores Andrew Z. Fire (de la Universidad de Stanford, en California) y Craig C. Mello (del Massachusetts Medical School en Worcester, USA) por el descubrimiento de un nuevo mecanismo de regulación génica en *C. elegans* denominado “Interferencia por RNA” o “RNA interferente” (**RNAi**). Este nuevo mecanismo constituye una nueva vía de silenciamiento mediado por pequeñas moléculas de RNA no codificante de cadena doble (Fire et al., 1998). Se trataba de un tipo de control desconocido hasta la fecha que, sin embargo, ha resultado esencial para el desarrollo normal y la fisiología de los organismos. Más aún, el **RNAi** es también un sistema de protección frente a infecciones virales, especialmente en plantas e invertebrados, y contribuye decisivamente a la estabilidad genómica participando en la organización de la cromatina y el control de los elementos móviles. El descubrimiento de este nuevo mecanismo ha tenido un impacto extraordinario en el campo de la Biomedicina, y se ha transformado en una herramienta experimental de silenciamiento génico (*knock-down*).

Tras su descubrimiento en *C. elegans* la existencia de RNAi fue documentada en una amplia variedad de organismos incluyendo la mosca de la fruta, tripanosomas, plantas, planarias, hidras, pezcebra y mamíferos, con la notable excepción de *Saccharomyces cerevisiae*.

Hay tres tipos de moléculas de RNA no codificante implicadas o, si se prefiere, tres modalidades de RNAi: (1) **siRNA** (short interfering RNA), (2) **miRNA** (micro-RNAs) y (3) **raSiRNA** (relacionado con las secuencias repetidas y la organización de la cromatina) (Huttenhofer & Schattner, 2006). En las modalidades **siRNA** y **raSiRNA** el proceso comienza con la formación de moléculas de RNA de cadena doble que resultan al unirse cadenas sencillas complementarias transcritas por la RNAPolasaII. En el caso de los **miRNA** se trata de genes específicos que se transcriben en moléculas sencillas (monocatenarias) de RNA con una región interior de auto-apareamiento (Southeimer & Carthew, 2005). Las moléculas efectoras del silenciamiento son siempre pequeñas moléculas de RNA de cadena sencilla de entre 21 y 25 nucleótidos que derivan del procesamiento en el núcleo de las moléculas precursoras (de mayor tamaño) por unos complejos enzimáticos denominados DICER. En los casos de **siRNA** y **miRNA** las moléculas de RNA efectoras son utilizadas como sondas para identificar los RNAm blanco inmersas en otro complejo enzimático distinto (RISC). En la modalidad **siRNA** los RNAm blanco son reconocidos por homología y cortados; sin embargo, en la modalidad **miRNAs**, las moléculas efectoras reconocen secuencias específicas de la región 3' UTR impidiendo la traducción (cuando la homología es sólo parcial) o incluso cortando cuando la homo-

logía es completa (Zeng, 2006). En el mecanismo **ra siRNA** el complejo RISC es sustituido por otro diferente (RIST) que los encauza hacia secuencias complementarias del ADN heterocromático.

Los microRNAs como genes de susceptibilidad en cáncer

Desde su descubrimiento en *C. elegans* los microRNAs se manifestaron pronto como una clase abundante de moléculas reguladoras presentes en todos los metazoos. Una buena parte de los genes *microRNA* (no codificantes) se agrupan en “clusters” y pueden tener sus *loci* en las regiones intrónicas de genes codificantes. Las predicciones bioinformáticas vaticinan que cada microRNA podría controlar miles de genes blanco y, puesto que ya se conocen cientos de microRNAs en el genoma de un mamífero, se podría decir que los microRNAs estarían controlando la expresión de una tercera parte del acervo total de genes de cualquier mamífero, incluida la especie humana. Por otro lado, puesto que los microRNAs son capaces de controlar aspectos fundamentales del desarrollo (Plasterk, 2006), la apoptosis (Jovanovic & Hengarter, 2006) y el ciclo de división celular, la hipótesis de su implicación en cáncer no parecía una propuesta descabellada (Caldas & Brenton, 2005; Chen, 2005). Si un microRNA está controlando la expresión de un proto-oncogen manteniendo sus niveles de expresión dentro del equilibrio homeostático, la inactivación o delección de ese microRNA supondría la sobre-expresión del proto-oncogen favoreciéndose el desarrollo de un proceso cancerígeno. Del mismo modo si el microRNA controla un gen supresor de tumor, la sobre expresión del microRNA podría suponer una bajada significativa en los niveles de expresión del gen supresor por debajo de los niveles homeostáticos contribuyéndose también al desarrollo de un tumor. En el primer caso el microRNA se comportaría como un gen supresor y en el segundo como un oncogen.

En los tres últimos años se han acumulado numerosas evidencias de la implicación de decenas de microRNAs en el desarrollo de la práctica totalidad de los cánceres conocidos, controlando la expresión de los más conocidos oncogenes y genes supresores (Esquela-Kerscher & Slack, 2006; Blenkiron & Miska, 2007). Valga como ejemplo el control de la expresión de los oncogenes *Ras*, *Myc* etc por los microRNAs de la familia *let-7* (Kumar et al., 2007). En estos casos las alteraciones de los microRNAs en las células tumorales se producen como consecuencia de reordenaciones cromosómicas, deleciones o inactivaciones de carácter epigenético. No se han descrito mutaciones puntuales. Pero en otros casos, es el propio oncogen o el gen supresor (al ser factores de transcripción) el que induce la expresión de un microRNA. Así, el oncogen *c-Myc* puede inducir la expresión del “cluster” *miR 17-92* produciéndose un curioso “loop” de regulación en el que los microRNAs controlan a su vez la expresión de *E2F-1*, un efector de *c-Myc* (O’Donnell et al 2005). Uno de los hallazgos más recientes ha sido la constatación de que *Tp53*, el gen supresor tumoral por excelencia, es capaz de inducir la expresión de los microRNAs *mir34 a-c* que modulan la proliferación y supervivencia celulares al controlar la expresión de genes clave como *Bcl2*, *CiclinaE*, *Cdk6* etc. (Corney et al., 2007; He et al 2007; Hermeking, 2007). El equipo de Robert Weinberg ha demostrado también que el *microRNA-10b* está implicado en los procesos de invasión y metástasis en cáncer de mama (Ma et al., 2007).

Nuestro equipo, junto al dirigido por el Dr. Malumbres en el CNIO, ha venido trabajando en una región del cromosoma 12 de ratón (de unas 7 Mb) que sufre frecuentes pérdidas alélicas en linfomas linfoblásticos de células T, y contiene el 10% de todos los microRNAs que se han descrito hasta la fecha en los genomas humano y del ratón. Entre ellos destaca el *miR-203* que está dramáticamente silenciado en este tipo de tumores, mediante una combinación de alteraciones

genéticas y epigenéticas. En nuestro trabajo hemos demostrado que este microRNA actúa a través del gen *Ab1* en su versión libre o translocada (*BCR-ABL*). El tratamiento de las células de tumores hematológicos (de ratón y humanos) que sobre-expresan ABL o BCR-ABL con vectores de expresión portadores del gen *miR-203* reduce la expresión de aquellos genes y la proliferación celular. Más aún, este efecto puede ser revertido mediante la transfección con los genes *ABL* o *BCR-ABL1* carentes de la región 3'UTR por la que los reconoce el microRNA. Por consiguiente hemos propuesto que la re-expresión del gen *miR-203* podría tener un efecto terapéutico beneficioso en pacientes con linfomas o leucemias que sobre-expresan la proteína ABL o la proteína de fusión BCR-ALL (Cancer Cell, en prensa).

Teniendo en cuenta su participación en la iniciación y progresión de los cánceres humanos (Dalmay & Edwards, 2006) muchas líneas de investigación en curso se están centrando en la determinación de los “perfiles” genómicos de expresión de los microRNAs en las células cancerígenas, puesto que constituyen “firmas genéticas” de extraordinaria utilidad para la clasificación, el diagnóstico, la estadificación, el pronóstico y la respuesta al tratamiento de los diferentes tipos de cánceres (Lu et al., 2005 ; Calin & Croce, 2006; Blenkiron & Miska, 2007).

Por otro lado, la existencia de los mecanismos de inactivación via RNAi se está aprovechando para el diseño de estrategias “knock-down” que permitan silenciar genes humanos implicados en cualquier tipo de enfermedades genéticas humanas, como son todos los cánceres conocidos (Dykxhoorn et al., 2003; Cullen, 2005; Kim & Rossi, 2007).

Referencias

- Blenkiron C & Miska EA (2007) miRNAs in cancer: approaches, aetiology, diagnostics and therapy. *Human Molecular Genetics* 16, Review Issue 1 R106–R113.
- Caldas C & Brenton JD (2005) Sizing up miRNAs as cancer genes. *Nature Med* 11: 712.
- Calin GA and Croce CM (2006) MicroRNA signatures in human cancers. *Nature Rev Genet* 6: 857-866.
- Chen CZ (2005) MicroRNAs as oncogenes and tumor suppressors. *N Engl J Med* 353: 17.
- Corney DC et al (2007) MicroRNA-34b and MicroRNA-34c are targets of p53 and cooperate in control of cell proliferation and adhesion-independent growth. *Cancer Res* 67: 8433-8.
- Cullen B (2005) RNAi: the natural way. *Nature Genet* 37: 1163.
- Dalmay T & Edwards DR (2006) microRNAs and the hallmarks of cancer. *Oncogene* 25: 6170-6175.
- Dykxhoorn et al (2003) Killing the messenger: short RNAs that silence gene expression. *Nature Rev Mol Cell Biol* 4: 457.
- Fire et al. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391: 806-11.
- He X et al (2007) The guardian's little helper: microRNAs in p53 tumor suppressor network. *Cancer Res* 67: 11069-101.
- He L et al (2007) MicroRNAs join the p53 network: another piece in the tumour-suppression puzzle. *Nature Rev Cancer* 7: 819.

- Hermeking H (2007) p53 enters the microRNA world. *Cancer Cell* 12: 414-8.
- Huttenhofer A and Schattner P (2006) The principles of guiding by RNA: chimeric RNA-protein enzymes. *Nature Rev Genet* 7: 475-482.
- Jovanovic M & Hengartner MO (2006) miRNAs and apoptosis: RNAs to die. *Oncogene* 25: 6176-87.
- Kim DH & Rossi JJ (2007) Strategies for silencing human disease using RNA interference. *Nature Rev Genet* 8: 173-184.
- Lu J et al (2005) MicroRNA expression profiles classify human cancers. *Nature* 435 : 834.
- Ma L et al (2007) Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* 449: 682-9.
- O'Donnell KA et al (2005). C-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 435: 839-843.
- Plasterk RHA (2006) Micro RNAs in animal development. *Cell* 124 : 877-881.
- Esquela-Kerscher A & Slack FJ (2006) Oncomirs: microRNAs with a role in cancer. *Nature Rev Cancer* 6: 259-269.
- Southeimer EJ & Carthew RW (2005) Silence from within: siRNAs and miRNAs. *Cell* 122: 9-12.
- Zeng Y (2006) Principles of micro-RNA production and maturation. *Oncogene* 25: 6156-6162.

SÍNDROMES DE REPARACIÓN DEL DNA Y PREDISPOSICIÓN AL CÁNCER

Jordi Surrallés

Departament Genètica i Microbiologia, Universitat Autònoma de Barcelona y CIBER de Enfermedades Raras. Barcelona.

El genoma contiene la clave para el correcto desarrollo y funcionamiento de los seres vivos. La integridad del material genético y la estabilidad cromosómica están continuamente amenazada por carcinógenos endógenos o exógenos como determinados mutágenos químicos y radiaciones. Altos índices de mutación incrementan del riesgo de cáncer y están relacionados con el envejecimiento celular y las enfermedades hereditarias. Afortunadamente, nuestras células disponen de mecanismos de reparación de las roturas cromosómicas para contrarrestar sus efectos dañinos y evitar la transformación tumoral. Consecuentemente, defectos en estos genes incrementan el riesgo de cáncer y causan síndromes de fragilidad cromosómica con predisposición tumoral como la anemia de Fanconi (FA), la ataxia telangiectasia, el síndrome de Nijmegen y algunos tipos de cáncer de mama hereditario con mutaciones en los genes BRCA1/BRCA2. A modo indicativo, la probabilidad acumulada de contraer cáncer a lo largo de la vida de un paciente FA es del 70% antes de los 50 años y el 85% de los individuos portadores de determinadas mutaciones en BRCA2 desarrollaran cáncer de mama a lo largo de la vida. Por tanto, la investigación de estos síndromes es particularmente importante no sólo para curar a los enfermos sino también porque pueden darnos claves para entender procesos biológicos básicos indispensables para el mantenimiento de la integridad genética y la prevención del cáncer en la población general.

La anemia de Fanconi se considera un modelo biomédico de gran importancia a pesar de su baja frecuencia en la población, aproximadamente 1-5 casos por millón. Con esta enfermedad se hizo el primer trasplante de cordón umbilical, la primera selección preimplantacional de embriones HLA compatibles para obtener células de cordón umbilical óptimas para transplantar a un hermano afecto, y es una de las primeras enfermedades en la que se han hecho ensayos clínicos de terapia génica. Además, los 13 genes relacionados con la enfermedad están implicados en una ruta de reparación de entrecruzamientos en el DNA y, por tanto, son responsables de suprimir el cáncer en la población general y, en muchos casos, modulan la respuesta tumoral a la quimioterapia.

La caracterización diagnóstica y genética de la anemia de Fanconi es compleja debido al elevado número de genes implicados y conlleva estudios de fragilidad cromosómica y de sensibilidad celular a inductores de enlaces cruzados en el DNA, subtipaje por complementación génica mediada por retrovirus, estudios de ciclo celular y análisis mutacional. La clonación y posterior caracterización de los genes y proteínas implicadas ha sido clave para un mejor entendimiento de la biología molecular de esta enfermedad. Los modelos recientes sugieren que la principal función de los genes Fanconi es promover, vía recombinación homóloga, la reparación de horquillas de replicación bloqueadas, aunque es probable que participen en otras funciones clave para el mantenimiento de la estabilidad genómica.

Referencias

Revisiones

Wang W (2007) Emergence of a DNA-response network consisting of Fanconi anemia and BRCA proteins. *Nature Reviews Genetics* 8:735-748

Joenje, H. and K. J. Patel (2001). "The emerging genetic and molecular basis of Fanconi anaemia." *Nat Rev Genet* 2: 446-457

Kutler, D. I., B. Singh, J. Satagopan, S. D. Batish, M. Berwick, P. F. Giampietro, H. Hanenberg and A. D. Auerbach (2003). A 20-year perspective on the International Fanconi Anemia Registry (IFAR). *Blood* 101: 1249-1256

Articulos recientes del grupo

Bogliolo M, A Lyakhovich, E Callén, M Castellà, E Cappelli, MJ Ramírez, A Creus, R Marcos, R Kalb, K Neveling, D Schindler and J Surrallés (2007) Histone H2AX and Fanconi anemia FANCD2 function in the same pathway to maintain chromosome stability. *The EMBO J* 26:1340-1351.

Casado JA, Callén E, Jacomé A, Río P, Castella M, Ferro T, Muñoz A, Sevilla J, Cantalejo A, Cela E, Cervera J, Sánchez-Calero J, Badell I, Estella JM, Dasí A, Olivé T, JJ Ortega, Rodríguez-Villa A, Tapia M, Molinés A, Madero L, Segovia JC, Neveling K, Kalb R, Schindler D, Hanenberg H, Surrallés J, Bueren JA (2007) A Rational Multidisciplinary Strategy for the Genetic Subtyping of Fanconi Anemia Patients: Conclusions from the Spanish Fanconi Anemia Research Network. *Journal of Medical Genetics*. 44:241-249.

Lyakhovich, A. and Surralles. J. New roads to FA/BRCA pathway: H2AX. *Cell Cycle* 1019-1023.

Kalb R, K Neveling, H Hoehn, K Goettsche, SD Battish, C Hunt, M Berwick, E Callén, J Surrallés, JA Casado, JA Bueren, A Dasí, CM Zwaan, R Vervenne, G Pals, J de Winter, H Joenje, M Grompe, AD Auerbach, H Hanenberg and D Schindler (2007) Hypomorphic Mutations in the Central Fanconi Anemia Gene *FANCD2* Sustain a Significant Group of FA-D2 Patients with Severe Phenotype. *Am J Hum Genet* 80: 895-910.

Lyakhovich, A. and Surralles, J. (2007) FANCD2 depletion sensitizes cancer cells repopulation ability in vitro. *Cancer Letters* 256:186-195.



2



2. Cartografiado, Identificación de genes, Fisiopatología molecular

POLIMORFISMOS, MAPAS Y GENES EN LAS ENFERMEDADES MENDELIANAS

Carmen Espinós

Unidad de Genética y Medicina Molecular, Instituto de Biomedicina de Valencia, CSIC y CIBER de Enfermedades Raras. Valencia.

Gregor Mendel denominó carácter hereditario a aquella cualidad susceptible de ser transmitida de generación en generación. La herencia en una enfermedad mendeliana o monogénica está exclusivamente relacionada con un locus/gen que sigue las leyes de Mendel y por tanto, su expresión en el fenotipo depende sobre todo de las relaciones de dominancia/recesividad que existan en cada caso. Se estima que en torno al 1,4% de los niños recién nacidos están afectados por uno de estos trastornos. La base de datos OMIM (*Online Mendelian Inheritance in Man*; <http://www.ncbi.nlm.nih.gov/omim>) recoge más de 10.000 loci implicados en enfermedades que en muchos casos son raras.

Una posible estrategia para identificar el gen responsable de una enfermedad es la clonación de un gen del que se conoce su localización cromosómica si bien, la información disponible sobre su función bioquímica o su papel en la patogénesis puede ser nula o prácticamente nula; se trata de la clonación posicional (Strachan and Read, 1996). La estrategia habitual consiste en construir mapas genéticos y físicos de la región, refinar la localización subcromosómica, y aislar los posibles genes existentes en la región que pudieran ser candidatos para la enfermedad.

La construcción de mapas genéticos que muestre la posición relativa de distintos loci basándose en la frecuencia de recombinación, precisa del análisis de ligamiento. Para ello se procede a la estima del valor LOD o LOD SCORE (Z) que se define como el logaritmo de la probabilidad relativa de que dos loci estén ligados para una determinada fracción de recombinación (q). Existe ligamiento significativo si el LOD SCORE es superior o igual a 3 y se descarta ligamiento si éste es inferior a -2. Para su cálculo se suelen emplear paquetes informáticos como el FASTLINK 2.1 (Cottingham et al. 1993; Schaffer et al. 1994). Un análisis de ligamiento bipuntual refiere al análisis de cada marcador con respecto al locus de la enfermedad, mientras que un análisis multipuntual implica considerar varios marcadores simultáneamente con respecto al locus de la enfermedad. Por este motivo, el análisis multipuntual ha supuesto una herramienta útil para la localización de marcadores y construcción de mapas genómicos. También el cartografiado por homociguidad en familias consanguíneas es una herramienta muy útil cuando se trata de una enfermedad recesiva con una frecuencia muy baja. Si se conoce la localización cromosómica del gen causante de la enfermedad de modo que es posible la construcción de haplotipos, es de esperar que los pacientes con padres emparentados sean homocigotos por ascendencia tanto para la mutación como para los haplotipos que se encuentran en la región de ligamiento.

La construcción de un mapa genético requiere de la existencia de marcadores genéticos distribuidos por todo el genoma. Éstos son polimorfismos de DNA que se heredan siguiendo los patrones mendelianos cuyo análisis permite mediante la detección de puntos de entrecruzamiento o recombinación, delimitar la región dentro de la cual se localizan los genes/*loci* de interés. Los primeros polimorfismos que se emplearon fueron los RFLPs (*Restriction Fragment Length Polymorphisms*) tipados en sus inicios por *Southern blot* y actualmente por PCR. Otro tipo de polimorfismos son los VNTRs (*Variable Number of Tandem Repeats*) o minisatélites y los STRs (*Short Tandem Repeat*) o microsatélites. Más informativos que los anteriores ya que son multialélicos (los RFLPs son bialélicos). Los VNTRs consisten en repeticiones consecutivas de aproximadamente 60 nucleótidos y se analizan por *Southern blot*, y los STRs son repeticiones de hasta 6 nucleótidos y para su análisis se emplea la PCR. También se emplean como marcadores genéticos polimórficos, los SNPs (*Single Nucleotide Polymorphism*) que consisten en la sustitución de un nucleótido por otro. Son las variantes polimórficas más comunes: en la base de datos del *National Center for Biotechnology Information*, <http://www.ncbi.nlm.nih.gov>, se describe un total de 11.811.594 SNPs, de los que se han validado 5.689.586 SNPs, en el genoma humano. Además, su estudio es posible con múltiples metodologías, incluyendo técnicas de última generación de análisis masivo por lo que el empleo de éstos para genotipado a gran escala es indudablemente un área emergente. Por último, durante los últimos años se ha descrito un tipo de polimorfismo al que se llama CNV (*Copy Number Variation*), abundante en el genoma humano y en otros genomas de mamíferos, cuyo tamaño va de algunas kilobases a megabases (Redon et al. 2006). Se conocen cerca de 1.500 regiones CNVs distribuidas por todo el genoma que abarcan unas 360 megabases (12% del genoma).

La enfermedad de Charcot-Marie-Tooth

La enfermedad de Charcot-Marie-Tooth (CMT) es una neuropatía sensitivo-motora hereditaria (NSMH) que está entre las enfermedades hereditarias neurológicas más comunes: 1 de cada 2500 individuos está afectado (Skre 1974; Combarros et al. 1987). Se clasifica en dos tipos atendiendo a estudios electrofisiológicos y de conducción nerviosa: tipo 1 o desmielinizante con una velocidad de conducción nerviosa disminuida, y tipo 2 o axonal con una velocidad normal (Dyck 1968; Harding and Thomas 1980). Se trata además, de una enfermedad rara con una amplia heterogeneidad genética con más de 30 genes/*loci* descritos (Tablas 1 y 2).

Una de las variantes de CMT es el tipo 4A (MIM 214400) que cursa con una herencia autosómica recesiva cuyo gen responsable fue identificado por nuestro grupo en 2002 (Cuesta et al. 2002). La serie objeto de estudio contaba con tres familias, LF20, LF38 y LF249, que presentaban un cuadro clínico muy homogéneo caracterizado por inicio muy precoz en la infancia, neuropatía axonal, disfonía por parálisis de cuerdas vocales y herencia autosómica recesiva (Sevilla et al. 2001; Sevilla et al. 2003). En sus inicios sólo se disponía de DNA de los miembros de dos de estas familias, LF20 y LF249 en las que todos los afectados eran varones. Por ello, primeramente se descartó la posibilidad de que se tratara de un gen ligado al sexo mediante el análisis del gen *GJB1* (responsable de la forma CMTX1, Tabla 1) y mediante un análisis de ligamiento multipuntual abarcando todo el cromosoma X. A continuación se analizaron los *loci* CMT conocidos hasta el momento y se obtuvo ligamiento positivo ($Z_{máx} = 2,01$) para los marcadores D8S286 y D8S167 en el locus CMT4A en 8q21.1 (Sevilla et al. 2001). En esta región se localiza el gen *PMP22* causante del CMT1A (Tabla 1) que fue descartado ya que la búsqueda de mutaciones fue negativa.

Posteriormente, se reclutó a la familia LF38 que cuenta con tres generaciones y varios matrimonios consanguíneos, lo que hacía posible un cartografiado por homocigotidad. Con esta familia, además de confirmar el ligamiento del locus CMT4A ($Z > 3$), se pudo reducir la región candidata a 2 cM. En ésta se localiza entre otros, el gen *GDAP1*, un gen candidato ya que se sobreexpresa en un sistema de diferenciación de las células de la línea Neuro2a, derivada de un neuroblastoma de ratón, transfectadas con la enzima GD3 sintasa (2,8-sialiltransferasa), la cual juega un papel importante en la biosíntesis de los gangliósidos de las series b y c. Además, el gen homólogo de ratón *Gdap1* se expresa fuertemente en cerebro. La búsqueda de mutaciones de los tres propósitos de las familias incluidas en el estudio reveló tres cambios: p.Q163X, p.S194X y p.T288fsX290. Todos ellos conducían a proteínas truncadas. Exactamente el probando de la familia LF38 era homocigoto para p.Q163X, el de LF249 era heterocigoto compuesto para p.Q163X/p.S194X, y el de LF20 era heterocigoto compuesto para p.Q163X/p.T288fsX290. El estudio se hizo extensivo al resto de familiares de cada uno de los probandos y se comprobó la segregación mendeliana de la enfermedad. Adicionalmente, se buscaron estas tres mutaciones en 67 individuos sanos y no se detectaron, lo que apoya la idea de que se trata realmente de cambios patológicos. Estudios posteriores han permitido concluir que mutaciones en el gen *GDAP1* son responsables de un importante número de casos de CMT en nuestra población y que entre éstas, la más frecuente es p.Q163X (Claramunt et al. 2005).

También en nuestro laboratorio identificamos que el gen *SH3TC2* es responsable de una variante de CMT asociada a población gitana (Claramunt et al. 2007). En este caso la serie objeto de estudio consistía en 20 familias de etnia gitana españolas que presentaban CMT desmielinizante con un patrón de herencia autosómico recesivo. La enfermedad en estos pacientes cursaba con una edad de inicio temprana, caídas frecuentes, deformidades en los pies, escoliosis, atrofia muscular distal y ataxia sensorial prominente (Kalaydjieva et al. 2005; Vilchez et al. 2006). En este estudio primero se descartaron los loci CMT asociados a etnia gitana: CMT4D o NSMH-Lom (MIM 601455) y CMT4G o NSMH-Russe (MIM 605285) (Tabla 1). Este estudio permitió identificar la mutación fundadora p.R148X en el gen *NDRG1* responsable de NSMH-Lom en cuatro familias y detectar posible ligamiento al locus NSMH-Russe en otras tres familias. Llegados a este punto el estudio se recondujo prestando atención a los aspectos clínicos.

Trece de estas familias eran supervisadas por un único equipo de neurólogos y el fenotipo de éstas semejaba al descrito para CMT4C (MIM 601596) (Vilchez et al. 2006). Por este motivo se planteó la posibilidad de que el gen causante de la enfermedad en esta serie fuera el mismo que causa CMT4C, el gen *SH3TC2* localizado en 5q32. Se incluyeron en el estudio 10 de éstas 13 porque los análisis genéticos previos habían mostrado que tres familias eran NSMH-Lom (CMT-394) o NSMH-Russe (CMT-502 y CMT-444). Finalmente, la serie de estudio en esta nueva aproximación comprendía diez familias. El análisis de ligamiento bipuntual mostró que existía ligamiento positivo con marcadores de esta región ($Z_{\text{máx}} = 8.45$, *D5S413*). El análisis de haplotipos mostró que todas las familias excepto una (CMT-421), portaban un haplotipo core común 3-C para los marcadores *D5S413*-IVS14+69C/T. De estas nueve familias con un haplotipo core común, pacientes de ocho familias diferentes lo llevaban en homocigosis y en una familia CMT-235, se observaron dos haplotipos diferentes, lo que sugería que dos mutaciones distintas eran las causantes de la enfermedad en esta serie. La búsqueda de mutaciones en el gen *SH3TC2* mostró que los pacientes portadores del core 3-C eran homocigotos para la mutación p.R1109X, mientras que los pacientes de la familia CMT-235 eran homocigotos compuestos para p.R1109X/p.C737_P738delinsX. Consecuentemente de las diez familias incluidas en el análisis

del gen *SH3TC2* se concluye que éste es el responsable de la enfermedad en nueve de ellas. En la familia CMT-421 fueron descartados los tres loci CMT asociados a la etnia gitana, lo que subraya la heterogeneidad genética asociada a la enfermedad.

En nuestra serie de veinte familias de etnia gitana los análisis genéticos realizados de los genes *NDRG1* (NSMH-Lom) y *SH3TC2* (CMT4C) y del locus NSHM-Russe permitieron llegar a alguna conclusión en 17 de ellas:

4 familias eran NSMH-Lom.

3 familias eran probablemente NSHM-Russe.

9 familias eran CMT4C.

1 familia no era ni NSMH-Lom ni NSHM-Russe ni CMT4C.

Así, se llevó a cabo la búsqueda de mutaciones en el gen *SH3TC2* en las tres familias restantes en las que previamente se había descartado tanto NSMH-Lom como NSHM-Russe. Los resultados mostraron que el propositus de una de ellas presentaba el core 3-C y que era homocigoto para la mutación *SH3TC2* p.R1109X. Consecuentemente la mutación *SH3TC2* p.R1109X estaba presente en 20 cromosomas que llevaban el core 3-C, indicando un origen común. Con el fin de investigar la historia genética de esta mutación, procedimos al cálculo de la edad alélica mediante dos aproximaciones matemáticas diferentes y éste mostró que esta mutación probablemente apareció a finales del siglo XVIII, atendiendo a los datos incluidos en este estudio. Se postula que un pequeño grupo de emigrantes portando la mutación *SH3TC2* p.R1109X se separó de su grupo tribal o casta. Este pequeño grupo podría ser el origen de la actual distribución en España de modo que habría actuado como un cuello de botella. La edad aquí estimada correspondería al análisis de los datos desde que este cuello de botella tuvo lugar.

Referencias

Claramunt R, Pedrola L, Sevilla T, Lopez de Munain A, Berciano J, Cuesta A, Sanchez-Navarro B, Millan JM, Saifi GM, Lupski JR, Vilchez JJ, Espinos C, Palau F (2005) Genetics of Charcot-Marie-Tooth disease type 4A: mutations, inheritance, phenotypic variability, and founder effect. *J Med Genet* 42:358-65

Claramunt R, Sevilla T, Lupo V, Cuesta A, Millan JM, Vilchez J, Palau F, Espinos C (2007) The p.R1109X mutation in *SH3TC2* gene is predominant in Spanish Gypsies with Charcot-Marie-Tooth disease type 4. *Clinical Genetics* 71:343-349

Combarros O, Calleja J, Polo JM, Berciano J (1987) Prevalence of hereditary motor and sensory neuropathy in Cantabria. *Acta Neurol Scand* 75:9-12

Cottingham RW, Jr., Idury RM, Schaffer AA (1993) Faster sequential genetic linkage computations. *Am J Hum Genet* 53:252-63

Cuesta A, Pedrola L, Sevilla T, Garcia-Planells J, Chumillas MJ, Mayordomo F, LeGuern E, Marin I, Vilchez JJ, Palau F (2002) The gene encoding ganglioside-induced differentiation-associated protein 1 is mutated in axonal Charcot-Marie-Tooth type 4A disease. *Nat Genet* 30:22-5

Dyck P, Lambert , EH (1968) Lower motor and primary sensory neuron diseases with peroneal

muscular atrophy. I. Neurologic, genetic, and electrophysiologic findings in hereditary polyneuropathies. *Arch Neurol* 18:603-618

Harding AE, Thomas PK (1980) Genetic aspects of hereditary motor and sensory neuropathy (types I and II). *J Med Genet* 17:329-36

Kalaydjieva L, Lochmuller H, Tournev I, Baas F, Beres J, Colomer J, Guergueltcheva V, Herrmann R, Karcagi V, King R, Miyata T, Mullner-Eidenbock A, Okuda T, Milic Rasic V, Santos M, Talim B, Vilchez J, Walter M, Urtizberea A, Merlini L (2005) 125th ENMC International Workshop: Neuromuscular disorders in the Roma (Gypsy) population, 23-25 April 2004, Naarden, The Netherlands. *Neuromuscul Disord* 15:65-71

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, et al. (2006) Global variation in copy number in the human genome. *Nature* 444:444-54

Schaffer AA, Gupta SK, Shriram K, Cottingham RW, Jr. (1994) Avoiding recomputation in linkage analysis. *Hum Hered* 44:225-37

Sevilla T, Cuesta A, Chumillas MJ, Mayordomo F, Pedrola L, Palau F, Vilchez JJ (2003) Clinical, electrophysiological and morphological findings of Charcot-Marie-Tooth neuropathy with vocal cord palsy and mutations in the GDAP1 gene. *Brain* 126:2023-33

Sevilla T, Cuesta, A., Chumillas, M.J., Mayordomo, F., García-Planells, J., Palau, F., Vilchez, J.J. (2001) Clinical and genetic studies in three Spanish families with severe autosomal recessive Charcot-Marie-Tooth axonal neuropathy. *Acta Myol* 20:49-52

Skre H (1974) Genetic and clinical aspects of Charcot-Marie-Tooth's disease. *Clin Genet* 6:98-118

Strachan T, Read AP (1996) *Human Molecular Genetics*. BIOS Scientific Publishers Ltd, Oxford

Vilchez J, Sevilla T, Garcés M, Mayordomo F, Chumillas MJ, Claramunt R, Palau F (2006) A cluster of hereditary motor and sensory neuropathy CMT4C in Spanish gypsies with a founder mutation in KIAA1985 gene. *Neurology* 66:A86

Tabla 1. Genes/Loci implicados en CMT desmielinizante

Autosómico Dominante (CMT1)

CMT1A	17p11.2	<i>Duplicación 17p11.2-12 PMP22</i>
CMT1B	1q22-q23	<i>PO ó MPZ</i>
CMT1C	16p13.1-p12.3	<i>LITAF</i>
CMT1D	10q21-q22	<i>EGR2</i>
CMT1F	8p21	<i>NEFL</i>

Intermedio Dominante (CMT1-DI)

DI-CMTA	10q24.1-q25.1	—
DI-CMTB	19p12-13.2	<i>DNM2</i>
DI-CMTC	1p34-p35	<i>YARS</i>

Autosómico Recesivo (CMT4)

CMT4A	8q13-q21.1	<i>GDAP1</i>
CMT4B1	11q22	<i>MTMR2</i>
CMT4B2	11p15	<i>SBF2</i>
CMT4C	5q21-q33	<i>SH3TC2</i>
CMT4D	8q24	<i>NDRG1</i>
(HMSN-Lom)		
CMT4E	10q21	<i>EGR2</i>
	1q22	<i>PO</i>
	17p11.2	<i>PMP22</i>
CMT4F	19q13.3	<i>PRX</i>
CMT4G	10q22-q23	—
(HMSN-Russe)		
CMT4H	12p11.21-q13.11	—

Recesivo ligado al cromosoma X

CMTX1	Xq13.1	<i>GJB1</i>
-------	--------	-------------

Dominante ligado al cromosoma X

CMTX3	Xq26	—
CMTX4	Xq24-q26	

Tabla 2. Genes/Loci implicados en CMT axonal

Autosómico Dominante (CMT2)

CMT2A1	1p36.2	<i>KIF1Bb</i>
CMT2A2		<i>MFN2</i>
CMT2B	3q13-q22	<i>RAB7</i>
CMT2C	12q23-q24	—
CMT2D	7p14	<i>GARS</i>
CMT2E	8p21	<i>NEFL</i>
CMT2F	7q11-q21	<i>HSP27</i>
CMT2G	12q12-q13.3	—
CMT2H	8q	—
CMT2I	1q22-q23	<i>MPZ</i> <i>PO</i>
CMT2J		
CMT2K	8q13-q21.1	<i>GDAP1</i>
CMT2L	12q24	<i>HSPB8</i> <i>HSP22</i>

Autosómico Recesivo (AR-CMT2)

ARCMT2A (CMT2B1)	1q21	<i>LMNA</i>
ARCMT2B (CMT2B2)	19q13.3	<i>MED25</i>
ARCMT2C	8q21.3	<i>GDAP1</i>

Recesivo ligado al cromosoma X

CMTX2	Xp22.2
CMT5X	Xq21-q24

ANÁLISIS GENÉTICO DE FORMAS NO MENDELIANAS DE ENFERMEDADES

Jordi Pérez-Tur

Unitat de Genètica Molecular, Instituto de Biomedicina de Valencia, CSIC. Valencia.

El gran desarrollo de la genética molecular humana en relación con la identificación de genes implicados en la aparición de enfermedades tiene que ver con la aplicación de las técnicas de análisis genético a formas mendelianas de estas enfermedades. Sin embargo, es frecuente que una fracción mayoritaria de los pacientes de enfermedades muy prevalentes no presenten historia familiar o ésta no pueda ser calificada como mendeliana. Tal es el caso de la enfermedad de Alzheimer, en la que aproximadamente un 5-10% de los casos presenta una historia familiar de tipo mendeliano mientras que en el resto se ha identificado una importante contribución genética al riesgo frente a la enfermedad. Así, las técnicas de ligamiento genético aplicadas a las formas monogénicas han permitido identificar 3 genes causantes de la enfermedad [1]. Por otra parte, tanto los análisis de segregación como los análisis de gemelos y los análisis poblacionales, demuestran que existen una importante contribución genética al riesgo de padecer enfermedad de Alzheimer no mendeliana (por ejemplo, ver [2]).

El objetivo del análisis genético no tiene porqué estar dirigido hacia la identificación de aquellos factores implicados en el riesgo de desarrollar una enfermedad, sino que también permite el análisis de factores que influyen alguna característica específica de la misma, el grado de progresión de la misma o su gravedad e, incluso, la potencial respuesta frente a tratamientos terapéuticos. Es decir, la aplicación de técnicas de análisis genético no mendeliano al estudio de enfermedades puede permitir disponer de una imagen precisa de la contribución de los genes a los diversos aspectos en los que una enfermedad concreta se manifiesta.

Históricamente, los primeros métodos de análisis genéticos desarrollados se encontraban condicionados por las tecnologías existentes. Así, se desarrollaron métodos basados en el análisis de microsatélites como los estudios conocidos como de *parejas de hermanos* (del inglés *sib-pairs*). Estos estudios implicaban el análisis de un elevado número de marcadores genéticos (entre 400 y 800) en poblaciones de tamaño relativamente grande para la época (cerca de 700-800 individuos) y presentaban un elevado coste económico así como en tiempo. Estos estudios presentaban la ventaja de que, por hacer un barrido del genoma completo, permitían identificar varios *loci* distintos potencialmente relacionados con la enfermedad en estudio. En concreto, este método se basa en comparar la distribución esperada de alelos compartidos en cada fratría con la observada en realidad. En una situación de independencia marcador/enfermedad encontramos la siguiente distribución: en el 25% de las fratrías los componentes no comparten ningún alelo entre ellos, en un 50% de fratrías se comparte un alelo y en el 25% restante se comparten los dos alelos. Si un marcador se desvía de esta distribución será indicativo de que el *locus* en el que dicho mar-

gador se encuentra está asociado con la enfermedad. Por ejemplo, véase el trabajo de Blacker y colaboradores en enfermedad de Alzheimer [3]. En este trabajo se analizaron 381 microsatélites en 437 familias (1.439 individuos entre afectados y no afectados) lo que permitió la identificación de un *locus* con un valor de ligamiento altamente significativo y que se correspondía con un gen previamente conocido y otros 12 *loci* en los que podría existir algún gen que contribuyera a la aparición de la enfermedad en la población. Alguno o varios de estos *loci* sería, posiblemente, un falso positivo, sin embargo los resultados de este y otros estudios similares ponen en evidencia el grado de complejidad de la contribución genética a esta enfermedad. El riesgo frente a la enfermedad no viene dado por un único gen, sino que son combinaciones determinadas de polimorfismos las que contribuyen al riesgo. Esto hace que en individuos distintos, la combinación de alelos que se puede presentar y que aumenta el riesgo frente a la enfermedad sea distinta lo que complica su identificación.

Un problema asociado a este tipo de estudios lo constituye el hecho de que la información obtenida cuando se da ligamiento implica regiones excesivamente grandes lo que dificulta sobremedida la identificación del factor genético subyacente al ligamiento encontrado. Esto es así debido a la existencia de un elevado número de individuos que pueden compartir el marcador ligado por mero azar.

Los desarrollos tecnológicos de los últimos años han permitido que el análisis genético de formas no mendelianas de las enfermedades humanas se desplace a técnicas basadas en el desequilibrio de ligamiento y que se desarrollen de manera importante las técnicas de asociación genética. Con el conocimiento derivado del proyecto del Genoma Humano y de sus continuaciones HapMap (I y II) se está llegando a disponer de un mapa de variabilidad en el genoma basada en un tipo diferente de marcadores, los polimorfismos de base simple (del inglés *SNP: single nucleotide polymorphisms*). En el genoma se estima que existen varios millones de estos polimorfismos y existe un proyecto destinado a identificar cuales son aquellos que permiten capturar el máximo de información genética con el mínimo esfuerzo de laboratorio. Por otra parte, el desarrollo de plataformas tecnológicas de alto rendimiento permite el genotipado rápido de un número muy importante de estos marcadores con lo que el cuello de botella se desplaza ahora a la capacidad de análisis de la información generada. Este cuello de botella no impide que se vayan publicando trabajos en los que se llevan a cabo estudios que utilizan estas nuevas aproximaciones. Por ejemplo, recientemente se ha publicado un estudio de asociación sobre 14.000 pacientes y 3.000 controles que incluía 7 enfermedades comunes (trastorno bipolar, enfermedad arterial coronaria, Chron, hipertensión, artritis reumatoide, diabetes de tipo I y diabetes de tipo II) [4]. En este estudio se replican resultados obtenidos en estudios anteriores y se apunta a nuevos *loci* para varias de estas enfermedades. Con independencia de la magnitud del estudio, resulta especialmente relevante el que el análisis producido no introduzca grandes avances metodológicos, se sigue con el análisis tradicional que busca regiones sobre-representadas en la población con la enfermedad comparando marcador (sea éste el genotipo o un haplotipo) y enfermedad cuando, como ya hemos comentado anteriormente, estas enfermedades tendrán un origen en el que varios factores genéticos han de confluír para que la enfermedad pueda llegar a darse si se dan las condiciones ambientales adecuadas.

Finalmente, hay que señalar también la existencia de nuevos desarrollos en el campo del análisis genético de enfermedades complejas. Por un lado, los recientemente identificados polimorfismos de variaciones en el número de copias puede suponer un nuevo factor a considerar a la hora de llevar a cabo estudios de este tipo [5]. Por otro, la existencia de cambios epigenéticos heredables

o no en el genoma y, finalmente el estudio de poblaciones aisladas o de caracteres cuantitativos, además de los cualitativos tradicionales, permitirán avanzar en el futuro próximo en el conocimiento de los factores genéticos implicados en las enfermedades humanas.

Referencias

- [1] OMIM (<http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=104300>).
- [2] Martinez, M.; Campion, D.; Brice, A.; Hannequin, D.; Dubois, B.; Didierjean, O.; Michon, A.; Thomas-Anterion, C.; Puel, M.; Frebourg, T.; Agid, Y.; Clerget-Darpoux, F. : Apolipoprotein E epsilon-4 allele and familial aggregation of Alzheimer disease. Arch. Neurol. 55: 810-816, 1998.
- [3] Blacker, D.; Bertram, L.; Saunders, A. J.; Moscarillo, T. J.; Albert, M. S.; Wiener, H.; Perry, R. T.; Collins, J. S.; Harrell, L. E.; Go, R. C. P.; Mahoney, A.; Beaty, T.; and 9 others : Results of a high-resolution genome screen of 437 Alzheimer's disease families. Hum. Molec. Genet. 12: 23-32, 2003.
- [4] The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661 - 678 (07 Jun 2007).
- [5] Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Månér S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. Large-scale copy number polymorphism in the human genome. Science. 2004 Jul 23;305(5683):525-8.

FISIOPATOLOGÍA MOLECULAR Y CELULAR EN LA ENFERMEDAD GENÉTICA

Santiago Rodríguez de Córdoba

Departamento de Inmunología, Centro de Investigaciones Biológicas (CIB), CSIC y CIBER de Enfermedades Raras. Madrid.

La secuenciación del Genoma Humano, la anotación del mismo y la disponibilidad de tecnologías para el genotipado y secuenciación a gran escala han permitido en los últimos años la identificación de numerosos genes responsables de enfermedades hereditarias. En muchas ocasiones los genes identificados codificaban para proteínas de función conocida, lo que ha facilitado trasladar el conocimiento genético a la interpretación de las bases moleculares de la enfermedad a la que esos genes estaban asociados. En otras muchas ocasiones, sin embargo, los genes identificados como responsables de la enfermedad han sido genes “nuevos” que codificaban proteínas cuya función en la célula o en el organismo eran desconocidas. En estos casos han sido necesarios numerosos estudios adicionales para entender la función de estas proteínas a nivel molecular y en la fisiología celular antes de poder profundizar en el conocimiento de las bases moleculares de la enfermedad en cuestión. En mi exposición presentaré ejemplos para ilustrar estas dos situaciones y describiré cómo hemos planteado en mi laboratorio la caracterización funcional de los genes responsables de dos enfermedades genéticas: el Síndrome Hemolítico Urémico y la enfermedad de Lafora.

El Síndrome Hemolítico Urémico (SHU) es la principal causa de insuficiencia renal aguda en niños. Es un trastorno de la microvasculatura, clínicamente definido por anemia hemolítica microangiopática y trombocitopenia, que afecta preferentemente a los riñones y se manifiesta con hematuria, oligoanuria y fracaso renal. Tradicionalmente se distinguen dos formas de SHU. La forma más frecuente se denomina SHU-típico o SHU asociado a diarrea, ya que ésta suele preceder a la aparición del síndrome. Esta forma se asocia con frecuencia a infecciones por la cepa de *E. coli* O157:H7, que secreta una potente toxina capaz de unirse a receptores en la superficie de las células endoteliales y provocar la destrucción de las mismas de forma directa o a través de la activación de mecanismos inflamatorios y procoagulantes. La gran mayoría de los pacientes con SHU-típico evolucionan satisfactoriamente al cabo de 2 o 3 semanas y recuperan completamente la función renal. En contraste a los casos de SHU-típico, existe un grupo de pacientes (5-10% del total) con SHU no asociado a diarrea, comienzo insidioso, evolución muy mala y elevada mortalidad (54%). Entre los supervivientes de este grupo la mitad tiene recurrencias y la mayoría evoluciona a insuficiencia renal terminal. Así mismo, el porcentaje de recidivas de la patología en los enfermos SHU-atípico a los que se les realiza trasplante renal es muy elevado. Esta forma “atípica” de SHU afecta tanto a niños como adultos y tiene un componente genético muy importante.

Se sospechaba desde hace tiempo que el sistema del Complemento podría estar implicado en la patogénesis del SHU-atípico, pues se habían identificado algunos pacientes con déficit de factor

H, pero solo recientemente se ha demostrado que existe una estrecha asociación entre el SHU-atípico y mutaciones en genes del Complemento. Diversos estudios en varios laboratorios han establecido que aproximadamente un 50% de los pacientes con SHU-atípico son portadores de mutaciones en heterocigosis en genes del Complemento. Estos genes codifican proteínas reguladoras como factor H (15-25% del total de pacientes con SHU-atípico), MCP (10-15%) y factor I (5%) o proteínas que participan en la formación de la C3-convertasas de la vía alternativa como factor B (2-4%) y C3 (5%).

De todos estos estudios genéticos, el hallazgo tal vez más importante en relación a los mecanismos patogénicos del SHU-atípico fue ver que las mutaciones en factor H asociadas a SHU-atípico no se distribuían de manera uniforme a lo largo de la molécula de factor H sino que se agrupaban en la región C-terminal, sugiriendo que es la alteración funcional de esta parte de la molécula, implicada en la protección de superficies, lo que es crítico para el desencadenamiento de la patología. El análisis funcional posterior de varias de las mutaciones en factor H, MCP y factor I asociadas a SHU-atípico mostró que, efectivamente, las mutaciones asociadas a SHU-atípico alteraban fundamentalmente la capacidad de estas proteínas reguladoras de proteger a las células y tejidos del organismo frente a la acción del Complemento, sin afectar, en cambio, su capacidad para controlar la homeostasis del Complemento en plasma. Los resultados de estos estudios funcionales explican por qué los enfermos de SHU-atípico en remisión tienen, en general, niveles normales de C3 y C4 en plasma y confirman que el SHU-atípico no es el resultado de una deficiencia clásica del Complemento –que se traduce normalmente en una situación de hipocomplementemia (ausencia de C3 y pérdida de actividad de Complemento)-, sino más bien una situación de «autolesión» por el sistema del Complemento debido a una pérdida de regulación de la activación del Complemento sobre las superficies celulares.

Es importante señalar que la desregulación de la vía alternativa del Complemento que caracteriza al SHU-atípico puede producirse tanto por una disminución en la actividad de las proteínas reguladoras como por una actividad anormalmente elevada de las C3-convertasas. Así, mientras que las mutaciones en factor H, MCP y factor I incapacitan a estas proteínas para realizar su función reguladora, las mutaciones en factor B o en C3 son mutaciones que resultan en una C3-convertasa más activa.

En resumen, la pérdida de actividad de factor H o factor I en plasma, de MCP en las superficies celulares, la generación de autoanticuerpos anti-factor H o la presencia de mutaciones “activadoras” en Factor B o en C3, predisponen a SHU-atípico porque impiden una regulación adecuada del Complemento sobre las superficies celulares. En este contexto, una situación que dispare la activación del Complemento en la microvasculatura no se podrá controlar de modo apropiado, lesionando las células del endotelio vascular y ocasionando la destrucción del tejido.

El SHU-atípico y la glomerulonefritis membranoproliferativa de tipo 2 (MPGN2) son dos patologías diferentes que se asocian con mutaciones y polimorfismos en el gen CFH, lo que indica que la desregulación de la vía alternativa del Complemento juega un papel importante en la patogénesis de ambas condiciones. Sin embargo, las variantes genéticas de factor H que se asocian con una u otra enfermedad son diferentes, poniendo igualmente de manifiesto la existencia de una relación entre distintas alteraciones funcionales en factor H y una u otra patología. El depósito de C3 y la acumulación de material electrón-denso a lo largo de la membrana basal glomerular (GBM) son características de MPGN2. En contraste, SHU-atípico es una microangiopatía trombótica que se dispara por el daño al endotelio renal. El déficit de factor H en humanos y en animales se aso-

cia con MPGN2, mientras que las mutaciones en la región C-terminal de la molécula de factor H se asocian con SHU-atípico. La estrecha relación entre las mutaciones en el extremo C-terminal de factor H y SHU-atípico se ha demostrado recientemente en un modelo murino. Los ratones con déficit de factor H (Cfh^{-/-}) desarrollan MPGN2 como consecuencia de la activación masiva de C3 que se produce en plasma por la falta de factor H. Estos ratones deficientes de factor H tienen niveles muy bajos de C3 y carecen de un Complemento activo en plasma (son hipocomplementemicos). Introduciendo en estos ratones Cfh^{-/-} un transgén que produce una proteína factor H modificada, que simula las moléculas de factor H que presentan los pacientes con SHU-atípico, los ratones Cfh^{-/-} recuperan los niveles de C3 (y la actividad del Complemento), desarrollando espontáneamente SHU-atípico en lugar de MPGN2. Estos resultados son extraordinariamente importantes. Por un lado proporcionan un modelo murino de SHU-atípico que será muy útil en el desarrollo de estrategias para el tratamiento de los pacientes. Por otro, el cambio de fenotipo de MPGN2 a SHU-atípico como consecuencia de la recuperación de un Complemento activo en los ratones Cfh^{-/-} valida las hipótesis formuladas sobre la patogénesis del SHU-atípico y establece definitivamente que la combinación de un Complemento activo en plasma con una protección deficiente de las superficies celulares es crítica en la patogénesis de SHU-atípico.

Como he comentado, en los últimos años se ha producido un avance considerable en nuestro conocimiento de las bases moleculares del SHU-atípico. Hoy sabemos que la patología se produce por la lesión del endotelio glomerular debido a la desregulación del sistema del Complemento y que mutaciones en los genes codificando proteínas del Complemento, así como la presencia de auto anticuerpos anti-factor H, son factores de riesgo importantes que predisponen a SHU-atípico. En la actualidad los estudios del Complemento y los análisis genéticos (Tabla 2) explican el 50% de los casos de aHUS, facilitando su diagnóstico y permitiendo un pronóstico de la evolución de la enfermedad y de un posible trasplante renal. Sin embargo, todavía existe un 50% de pacientes con SHU-atípico en los que se desconoce el gen(es) responsable(s) de una posible susceptibilidad genética y en los que no se detecta la presencia de anticuerpos anti-factor H. Estos pacientes no son diferentes de los que llevan mutaciones en genes de Complemento, por lo que es posible que otros genes de Complemento no analizados todavía puedan ampliar la lista de genes asociados a SHU-atípico en años sucesivos. No cabe duda que el conocimiento actual de las bases moleculares del SHU-atípico es ya útil para el manejo y tratamiento de un número importante de pacientes. Los avances futuros serán determinantes para definir con precisión la relación que existe entre el tipo de mutación(es) que llevan los pacientes, la presentación clínica de la enfermedad y la respuesta a diversos tratamientos. Es también previsible que en un futuro próximo se vayan definiendo mejor los factores ambientales que disparan el SHU-atípico en individuos susceptibles y que se desarrollen terapias eficaces que corrijan los defectos del Complemento que caracterizan a estos pacientes. En resumen, el SHU-atípico es una patología de base genética compleja modulada por factores ambientales, para la que se empieza vislumbrar una solución a partir del conocimiento generado por la identificación de factores de predisposición genética.

La epilepsia mioclónica progresiva del tipo Lafora (LD) es una enfermedad grave autosómica recesiva caracterizada por un deterioro progresivo neurológico, mioclonía y epilepsia. La enfermedad se inicia durante la adolescencia (normalmente entre los 10 y 17 años de edad) en forma de crisis generalizadas tónico-clónicas o crisis visuales que suelen describirse como visión de luces o estrellas. Después aparecen las mioclonías que conforme avanza la enfermedad aumentan en frecuencia y se convierten en constantes. Posteriormente aparece una demencia que progresa rápi-

damente y va acompañada de apraxia, afasia y ceguera. La mayoría de los enfermos mueren antes de los 10 años del comienzo de la enfermedad tras evolucionar a un estado vegetativo terminal. La enfermedad de Lafora (LD) fue descrita por Gonzalo R. Lafora en 1911 al observar la presencia de depósitos intracelulares de poliglucosanos, denominados cuerpos de Lafora, en el cerebro y en la médula espinal de un paciente adolescente. Estos cuerpos de inclusión son característicos de la LD y se encuentran en cualquier tipo celular, no sólo en el sistema nervioso central.

LD es una enfermedad genética que se hereda con un patrón autonómico recesivo y penetrancia completa. Aparece en cualquier parte del mundo con una incidencia muy baja, aunque es relativamente más frecuente en los países mediterráneos. Se han identificado dos genes asociados a esta enfermedad: EPM2A situado en el cromosoma 6q24, y EPM2B en el cromosoma 6p22.3. EPM2A codifica para laforina y se encuentra mutado en un 50-60% de los casos de LD. La laforina es una proteína de 331 aminoácidos con actividad fosfatasa de especificidad dual, y presenta un dominio funcional de unión a carbohidratos en el extremo amino-terminal. La laforina interacciona con R5 (también denominada PTG), que es una de las subunidades reguladoras de la proteína fosfatasa 1 (PP1). R5 actúa como una proteína adaptadora que facilita la acción catalítica de PP1 sobre glucógeno sintasa, glucógeno fosforilasa y la fosforilasa quinasa, lo que da lugar a la acumulación intracelular de glucógeno. La eliminación del gen EPM2A en ratones produce un fenotipo similar al que se observa en los pacientes de LD, incluyendo la presencia de cuerpos de Lafora, neurodegeneración y trastornos neurológicos graves. Es interesante destacar que en este modelo animal, la neurodegeneración y la aparición de los cuerpos de Lafora, anteceden en varios meses a los trastornos de conducta, a la ataxia y a las mioclonías epilépticas. El gen EPM2B que codifica para la malina, se encuentra mutado en el 30-40% de los enfermos de Lafora. La malina es una E3-ubiquitin ligasa de 395 aminoácidos y que posee un dominio RING finger y seis dominios NHL. Se ha descrito una forma de LD en perros que está causada por mutaciones en EPM2B. Es importante señalar que los pacientes con mutaciones en laforina o en malina son indistinguibles, lo que sugiere que estas dos proteínas actúan a través de la misma vía fisiológica.

En los años 60, basándose en la similitud de la composición de los cuerpos de Lafora con la amilopectina (glucógeno sin ramificaciones), se había postulado que la acumulación de los cuerpos de Lafora podría ser consecuencia de defectos en el metabolismo del glucógeno generados por alteraciones en la regulación de la síntesis o degradación de esta molécula. Incluso antes, la acumulación intracelular de poliglucosanos (cuerpos de Lafora) llevó a G. R. Lafora a considerar la enfermedad como algún tipo de glucogenosis. A partir de 1998/99 y de 2003, cuando los genes responsables de la enfermedad de Lafora fueron identificados, numerosos trabajos han analizado la función enzimática de las proteínas laforina y malina e investigado qué proteínas interaccionaban con ellas, tratando de establecer un papel para estas proteínas en la fisiología celular. Estos resultados permitieron concluir que laforina es una fosfatasa dual, malina una E3-ubiquitin ligase y que laforina interacciona con malina y con PTG entre otras proteínas, como ya he señalado anteriormente. Estas observaciones, además de la identificación de un sitio de unión de carbohidratos en laforina que permitía la interacción de esta con poliglucosanos y la demostración de que laforina se traslocaba a los sitios intracelulares donde se sintetizaba glucógeno, se interpretaron como una confirmación de que efectivamente laforina y malina estaban implicadas en el metabolismo del glucógeno. Sin embargo, como era esta implicación y porqué en ausencia de laforina y malina se formaban los cuerpos de lafora era una incógnita difícil de resolver. Además, era intrigante que se formaran cuerpos de Lafora en las neuronas, cuando se sabía que este tipo celular era incapaz de sintetizar glucógeno. Trabajando con líneas celulares de neuroblastoma y cultivos

primarios de neuronas de rata hemos demostrado recientemente que las neuronas tienen la maquinaria para sintetizar glucógeno, pero que esta actividad está silenciada por la acción conjunta de laforina y malina. Este silenciamiento ocurre a través de un nuevo mecanismo de regulación de la síntesis del glucógeno que requiere la formación del complejo laforina-malina y que implica la ubiquitinalización y degradación vía proteosoma de dos proteínas clave en la síntesis del glucógeno, la glucógeno sintasa (GS) y PTG. Sabemos que un mecanismo similar opera en otros tejidos en los que también aparecen cuerpos de lafora. Mientras que estos resultados ofrecen una explicación a la formación de cuerpos de lafora en ausencia de laforina o malina, plantean nuevas preguntas relevantes a la fisiopatología de LD como son: entender las consecuencias de la acumulación de glucógeno en las neuronas y cómo se relaciona este proceso con la epilepsia y neurodegeneración que caracteriza a esta enfermedad, en este sentido es importante determinar si el complejo laforina-malina está implicado en la regulación de otras proteínas o procesos potencialmente implicados en la patogénesis de la LD y finalmente, ya que parece lógico pensar que la formación del complejo laforina-malina es un proceso regulado, determinar cuáles son las condiciones que determinan la formación de este complejo.

En resumen, la epilepsia mioclónica progresiva del tipo Lafora es una patología de base genética compleja cuyas bases moleculares empiezan a entenderse a partir del conocimiento generado por la caracterización funcional de las proteínas codificadas por los genes responsables.

Referencias

- Solaz-Fuster MD, Gimeno-Alcañiz JV, Ros S, Fernandez-Sanchez ME, Garcia-Fojeda B, Garcia OC, Vilchez D, Dominguez J, Garcia-Rocha M, Sanchez-Piris M, Aguado C, Knecht E, Serratosa J, Guinovart JJ, Sanz P, Rodriguez de Córdoba S. Regulation of glycogen synthesis by the laforin-malin complex is modulated by the AMP-activated protein kinase pathway. *Hum Mol Genet.* 2007 Nov 20; [Epub ahead of print]
- Vilchez D, Ros S, Cifuentes D, Pujadas L, Vallès J, García-Fojeda B, Criado-García O, Fernández-Sánchez E, Medraño-Fernández I, Domínguez J, García-Rocha M, Soriano E, Rodríguez de Córdoba S, Guinovart J. Novel regulatory mechanism for glycogen synthesis involving the ubiquitin-proteasome pathway is disrupted in progressive myoclonus epilepsy. *Nature Neurosciences* 10:1407-1413 (2007).
- Fernández-Sánchez M.E., Criado-García O., Heath K.E., García-Fojeda B., Medraño-Fernández I., Gomez-Garre P., Sanz P., Serratosa J.M. and Rodríguez de Córdoba S. Laforin, the dual-phosphatase responsible for Lafora disease, interacts with R5 (PTG), a regulatory subunit of protein phosphatase-1 that enhances glycogen accumulation. *Hum Mol Genet* 12:3161-3171 (2003).
- Serratosa, J. M., Gómez-Garre, P., Gallardo, M. E., Berta Anta, B, Daniel Beltrán-Valero de Bernabé, D., Lindhout, D., Tassinari, C. A., Michelucci, R., Malafosse, A., Topcu, M., Grid, D., Dravet, C., Berkovic, S. F. and Rodríguez de Córdoba, S. A novel protein tyrosine phosphatase gene is mutated in progressive myoclonus epilepsy of the Lafora type (EPM2). *Hum. Mol. Genet.* 8:345-352 (1999).
- Pickering MC, Goicoechea de Jorge E, Martinez-Barricarte R, Recalde S, Garcia-Layana A, Rose KL, Moss J, Walport MJ, Cook HT, Rodriguez de Córdoba S, Botto M*. Spontaneous haemolytic uraemic syndrome triggered by complement factor H lacking surface recognition domains. *J Exp Med.* 204:1249-56 (2007).

E. Goicoechea de Jorge, C.L. Harris, J. Esparza-Gordillo, L. Carreras, E. Aller Arranz, C. Abarrategui Garrido, M. López-Trascasa, P. Sánchez-Corral, B.P. Morgan and S. Rodríguez de Córdoba. Gain-of-function mutations in complement factor B are associated with atypical hemolytic uremic syndrome. *Proc Natl Acad Sci U S A*.104:240-245 (2007).

Sánchez-Corral P, Pérez-Caballero D., Huarte, O., Simckes, A.M., Goicoechea de Jorge, E., Lopez-Trascasa M. and Rodríguez de Córdoba, S. Structural and functional characterization of factor H mutations associated with atypical Hemolytic Uremic Syndrome. *Am. J. Human Genet.* 71:1285-1295 (2002).

Pérez-Caballero D., Gonzalez Rubio C., Gallardo M E, Vera M, Lopez-Trascasa M, Rodríguez de Córdoba, S and Sánchez-Corral P. Clustering of missense mutations in the C-terminal region of factor H in atypical Hemolytic Uremic Syndrome. *Am. J. Human Genet.* 68:478-484 (2001).

EL SÍNDROME DE USHER: DIAGNÓSTICO MOLECULAR DE UNA ENFERMEDAD GENÉTICAMENTE HETEROGÉNEA Y LA IMPORTANCIA DEL ESTUDIO POBLACIONAL

José M. Millán Salvador

Unidad de Genética. Hospital La Fe y CIBER de Enfermedades Raras. Valencia

El síndrome de Usher es un buen ejemplo de enfermedad clínica y genéticamente heterogénea. Atendiendo a su prevalencia, constituye la primera causa de sordoceguera hereditaria y, desde el punto de vista clínico, se distinguen tres tipos. El Usher tipo I presenta hipoacusia neurosensorial prelingual profunda o severa, retinosis pigmentaria (RP) de aparición pre o peripuberal y alteración del equilibrio. El síndrome de Usher tipo II se caracteriza por una hipoacusia neurosensorial leve o moderada, la RP aparece más tarde que en el tipo I y los pacientes preservan la función vestibular. Por último, el tipo III asocia una pérdida auditiva progresiva, RP de inicio variable y, en ocasiones, función vestibular alterada.

Este síndrome presenta siempre un patrón de herencia autosómico recesivo y, hasta la fecha, se han descrito 6 loci responsables del Usher tipo I y se han identificado los genes responsables para 5 de ellos (*MYO7A*, *USH1C*, *CDH23*, *PCDH15* y *SANS*). Existen tres genes implicados en el Usher tipo II (*USH2A*, *VLGR1* y *WHRN*) y uno responsable del Usher tipo III (*USH3A*). Se estima que el 25% de los casos de síndrome de Usher, en cualquiera de sus formas clínicas, se debe a genes todavía no identificados.

Las proteínas codificadas por estos genes pertenecen a familias proteicas funcionalmente distintas pero, recientemente, se ha puesto de manifiesto que estas proteínas interactúan entre sí formando un entramado que se expresa en la retina y el órgano de Corti, en la cóclea, denominado "interactoma" Usher, fundamental para el desarrollo y mantenimiento de estos dos órganos.

El diagnóstico genético molecular de los pacientes con este síndrome se hace difícil debido a la mencionada heterogeneidad, a la complejidad de los genes implicados (la mayoría de ellos presentan varias isoformas) el gran tamaño de estos genes y su expresión específica en los órganos alterados.

Como suele ser habitual en muchas enfermedades de origen genético, las mutaciones responsables de la enfermedad son exclusivas o casi exclusivas de cada familia. Esto implica que, para el estudio genético de un nuevo caso o una nueva familia diagnosticada clínicamente de síndrome de Usher, es necesario el rastreo de mutaciones en, como mínimo, una región codificante de más de 20 kilobases.

Los estudios de poblaciones han permitido detectar mutaciones recurrentes en algunas poblaciones endogámicas, como pueden ser las mutaciones Y176X y N48K en el gen *USH3A* en poblaciones finlandesa y ashkenazi respectivamente. En estas dos poblaciones el síndrome de Usher tipo III es especialmente prevalente, sin embargo, está prácticamente ausente en el resto.

La única mutación con una prevalencia relevante entre los pacientes Usher cualquiera que sea su origen es la mutación 2299delG en el gen *USH2A* responsable de Usher tipo II y cuya frecuencia alélica oscila entre el 45% en poblaciones del norte de Europa, al 14% en España. Estas cifras reflejan el problema añadido que aparece cuando se estudian poblaciones del sur de Europa y de la cuenca mediterránea, que son genéticamente mucho más heterogéneas que las poblaciones norte-europeas.

La complejidad de los estudios individuales para este síndrome se puede reducir mediante el estudio de la enfermedad en la población de origen del individuo. Estos estudios nos van a permitir:

Estimar el porcentaje de casos de los distintos subtipos clínicos en nuestra población.

Establecer, si fuese posible, una correlación genotipo-fenotipo que permita, ante un determinado fenotipo, rastrear el gen que con más probabilidad sea el responsable.

Determinar el porcentaje de casos en el que está implicado cada uno de los genes responsables. Esto permitirá establecer un algoritmo de estudio cuando llegue un paciente nuevo.

Así, el estudio poblacional del síndrome de Usher en España ha permitido establecer que el Usher tipo II es el doble de frecuente que el tipo I, suponiendo cerca del 66% de los casos. El Usher tipo III, en España, al igual que en la mayoría de las poblaciones estudiadas a excepción de las mencionadas anteriormente, presenta una prevalencia muy baja.

Respecto al Usher tipo 1, hemos determinado que el gen *MYO7A* es responsable de alrededor del 45% de los casos de origen español, *CDH23* entre un 20-25% y *PCDH15* entre el 15-20% serían los siguientes en frecuencia, siendo los genes *USH1C*, y *SANS* junto a otro u otros genes no identificados, responsables del resto de los casos, con porcentajes muy residuales para cada uno de ellos.

En el caso del Usher tipo 2 todavía no se ha llevado a cabo un screening mutacional tan exhaustivo como para el Usher 1. A pesar de que el Usher 2 está causado por tres genes frente a cinco para el Usher 1, la complejidad de los genes *USH2* ha evitado, de momento, un análisis epidemiológico con detalle del tipo II. En cualquier caso, estimamos que más del 75% de los casos Usher tipo II están causados por mutaciones en el gen *USH2A*, alrededor del 5% podría estar causado por el gen *VLGR1* y el resto de casos vendrían causados por mutaciones en el gen *WHRN* y otros aún no identificados.

El Usher tipo III como se ha mencionado anteriormente es muy poco prevalente en nuestra población. El gen *USH3A* no es responsable de más del 2% de los casos.

Bajo estas condiciones se puede establecer un algoritmo para el estudio de estas enfermedades complejas y plantearse la utilidad de determinadas herramientas basadas en la tecnología de microchips de ADN (microarrays para la detección de mutaciones o microchips de SNPs para casos familiares o consanguíneos) o la creación de una herramienta nueva específica para cada población.

Así, hemos detectado algunas mutaciones que aparecen en nuestra población con mayor frecuencia que en otras poblaciones, como las mutaciones 6049+1G>A y 6050-9G>A en *CDH23* e incluso alguna mutación como la delección 2431_2432delAA en el gen *USH2A* que es exclusiva de población española y que aparece en varias familias afectadas de una determinada región geográfica. Estas familias presentan un mismo haplotipo asociado a la mutación lo que implica un origen común para todas ellas.

Otro problema que el estudio poblacional puede, en ocasiones, resolver es la interpretación del carácter patológico de los cambios encontrados. Las mutaciones que producen un cambio de aminoácido pueden ser verdaderamente patológicas o simples polimorfismos que no alteran la función de la proteína. En estos casos, es fundamental llevar a cabo un rastreo del cambio en una población sana del mismo origen que la población afectada para estimar si se trata de un polimorfismo presente en dicha población sana, o un polimorfismo raro o una mutación patológica, lo que requeriría posteriores estudios funcionales para determinar su patogenicidad.

En conclusión, los estudios poblacionales son fundamentales a la hora de plantear el diagnóstico molecular de una enfermedad genéticamente heterogénea ya que permite crear un algoritmo de estudio que optimice los recursos técnicos y económicos de los que se dispone y ayuda a determinar qué cambios son verdaderamente patológicos. Esto es especialmente importante en poblaciones como la nuestra, cuyo *background* genético es también heterogéneo.

Referencias

Ahmed ZM, Riazuddin S, Riazuddin S, Wilcox ER. The molecular genetics of Usher syndrome. Clin Genet. 2003 63(6):431-44. Review.

El-Amraoui A, Petit C. Usher I syndrome: unravelling the mechanisms that underlie the cohesion of the growing hair bundle in inner ear sensory cells. J Cell Sci. 2005 15;118(Pt 20):4593-603. Review.

Espinos C, Millan JM, Beneyto M, Najera C. Epidemiology of Usher syndrome in Valencia and Spain. Community Genet. 1998 1(4):223-8.

Kremer H, van Wijk E, Marker T, Wolfrum U, Roepman R. Usher syndrome: molecular links of pathogenesis, proteins and pathways. Hum Mol Genet. 2006 15;15 Spec No 2:R262-70. Review.

Cremers FP, Kimberling WJ, Kulm M, de Brouwer A, van Wijk E, Te Brinke H, Cremers CW, Hoefsloot LH, Banfi S, Simonelli F, Fleischhauer JC, Berger W, Kelley PM, Haralambous E, Bitner-Glindzicz M, Webster AR, Saihan Z, De Baere E, Leroy BP, Silvestri G, McKay G, Koenekoop RK, Millan JM, Rosenberg T, Joensuu T, Sankila EM, Weil D, Weston MD, Wissinger B, Kremer H. Development of a genotyping Microarray for usher syndrome. J Med Genet. 2007 44:153-160.

DIAGNÓSTICO Y ASESORAMIENTO GENÉTICO: PAPEL DE LA GENÉTICA CLÍNICA ASISTENCIAL EN LOS SERVICIOS DE SALUD EN ESPAÑA

Feliciano. J. Ramos

Departamento de Pediatría, Facultad de Medicina, Universidad de Zaragoza. Zaragoza.

La Genética es una de las disciplinas científicas que más ha contribuido, y probablemente contribuirá, a la comprensión de las causas de las enfermedades y con ello a encontrar medios para su prevención y tratamiento. Su potencial en el suministro de servicios en la asistencia sanitaria y las repercusiones que tendrá en la mejora de la salud de los ciudadanos serán un tema capital en la Medicina del siglo XXI.

Las enfermedades de base genética constituyen un problema de salud de primer orden. Según los datos disponibles actualmente, aproximadamente el 3% de los recién nacidos presentan alguna anomalía o enfermedad genética y el 8% de la población desarrollará alguna enfermedad de origen genético antes de los 25 años, constatando que los factores genéticos desempeñan un papel predominante en aproximadamente 1/3 de los trastornos crónicos en la edad adulta. Un estudio reciente (2004) ha puesto de manifiesto el impacto real de las enfermedades genéticas en el sistema de salud de EE.UU. analizando el número de admisiones hospitalarias en la edad pediátrica. Los resultados fueron clarificadores: se encontró una causa genética subyacente (única o multifactorial) en el 71% de los ingresos en hospitales infantiles, incrementándose al 96% en el casos de las enfermedades crónicas. En relación al gasto sanitario generado, la atención a estos pacientes representó el 81% del presupuesto asistencial total. Incluso considerando exclusivamente las enfermedades monogénicas, éstas supusieron el 34% de las admisiones y el 50% del gasto.

En la Unión Europea (U.E.) la enfermedades genéticas son la tercera causa de mortalidad infantil después de los accidentes y el cáncer y la causa del 50% de las muertes antes de los 15 años. Es decir, las enfermedades genéticas son crónicas, a menudo mortales, y la supervivencia suele ir asociada a discapacidades graves, dando lugar a una gran carga familiar, sociosanitaria y económica. También tienen un gran impacto sobre la salud reproductiva.

Actualmente, en la sanidad pública, las pruebas genéticas son las que más se están expandiendo en el conjunto de las pruebas diagnósticas *in vitro*. Se diferencian de otras pruebas en cuatro aspectos fundamentales: 1) los resultados de dichas pruebas son “para toda la vida”, 2) los resultados obtenidos pueden tener implicaciones importantes para otros miembros de la familia, 3) pueden ser predictivas de la aparición de enfermedad clínica en un individuo aparentemente asintomático, y 4) pueden ser claves para el manejo y pronóstico de un embarazo, influenciando de forma decisiva la toma de decisiones en relación al mismo. Por todo ello, aparte de necesidad de la calidad de los servicios, éstas deben ir siempre acompañadas de información previa y asesoramiento genético posterior, haciendo del consejo genético parte integral de las pruebas genéticas.

La Genética también ha sido fundamental para comprender y abordar el cáncer. En primer lugar para reconocer los tumores que tienen su origen en un gen defectuoso y siguen un patrón hereditario determinado, ya que cuando existen antecedentes familiares positivos, las pruebas genéticas pueden contribuir a conocer las posibilidades de que otros miembros de la familia aparentemente sanos lo desarrollen. En segundo lugar, el estudio de las mutaciones adquiridas en células somáticas, que son el origen de la mayoría de los cánceres, han abierto el camino para la mejor comprensión de su patogénesis y de su abordaje terapéutico.

La mayoría de las enfermedades llamadas “comunes”, como las cardiovasculares y la diabetes, presentan un componente genético importante, cuya investigación en relación con los factores medioambientales implicados, permitirá avanzar en su diagnóstico precoz, en el diseño de tratamientos más efectivos, y en la implementación de medidas preventivas eficaces. Asimismo, hoy conocemos que la respuesta a numerosos fármacos de uso común depende en parte de factores genéticos, por lo que la farmacogenética es un área con una gran proyección de futuro en relación al desarrollo de pruebas que contribuyan a la mejora de la salud de la población.

España es, junto con Grecia, el único país miembro de la U.E. en el que la Genética no está oficialmente reconocida como especialidad sanitaria. Tampoco disponemos de programas formativos en Genética (ver informe de la Sociedad Europea de Genética Humana: www.eshg.org). Ambas carencias son, probablemente, el motivo principal de que no haya habido un desarrollo coherente de los servicios de genética, muchos de los cuales carecen de la organización y planificación adecuadas y están integrados por profesionales con titulaciones diversas, bien sea médicos o de otras profesiones sanitarias, con formación adecuada en Genética, pero sin haber recibido una formación reglada en este campo. Ello condiciona una escasez insostenible de personal cualificado para atender a una demanda asistencial exponencialmente creciente. Por otro lado, los servicios de genética, a pesar de que la mayoría disponen de la tecnología necesaria, muchos carecen de una planificación global de las necesidades asistenciales, dando lugar a una falta de optimización en la oferta de análisis genéticos. Al tratarse de estudios complejos, de coste elevado y con gran repercusión para el futuro del paciente y su familia, la necesidad de planificación y coordinación es ya ineludible.

En nuestro país, el actual sistema sanitario ofrece servicios asistenciales de genética que en muchos casos no se ajustan al modelo de los países de nuestro entorno en los que la Genética está reconocida como especialidad desde hace muchos años. Esta situación da lugar a disfunciones importantes en el sistema que se traducen en una falta de optimización de los recursos, con una repercusión negativa en la atención al ciudadano y el consiguiente incremento (muchas veces innecesario) del gasto sanitario. Según los datos disponibles de una encuesta realizada por la AEGH, en 2005 se realizaron en toda España cerca de 25.000 cariotipos en sangre periférica, con una tasa de patológicos inferior al 6%; 43.000 cariotipos prenatales (por amniocentesis), con una tasa de patológicos inferior al 3%; también se completaron más de 53.000 estudios moleculares, de los cuales sólo el 18% fueron patológicos. El problema es que muchos de esos estudios genéticos no se acompañaron del correspondiente asesoramiento genético.

En resumen, los 3 principales problemas que justifican la necesidad de iniciar acciones concretas que regulen la actividad asistencial en Genética Humana en nuestro país, son:

1. **Servicios asistenciales insuficientes**, compuestos mayoritariamente por personal con formación adecuada pero sin acreditación oficial.

2. **Ausencia de programas de formación y acreditación**, lo que agrava progresivamente la situación con el paso del tiempo si no se toman las medidas adecuadas.

3. **Ausencia de planificación estratégica a nivel nacional y autonómico**, que impide la optimización de la oferta asistencial y del gasto sanitario generado por las enfermedades de base genética.

Referencias

McCandless SE, Brunger JW, Cassidy SB. The burden of genetic disease on inpatient care in a children's hospital. *Am J Hum Genet* 2004; 74:121.

Royal College of Physicians. Commissioning clinical genetic services. London, RCP, 1998.

Provision of genetic services in Europe: current practices and issues. Recommendations of the European Society of Human Genetics. *Eur J Hum Genet* 2003; 11(sup. 2): S2-S4.

Encuesta sobre Servicios de Genética Asistenciales en España durante el año 2005. AEGH, 2006.

(*)Resumen extraído del documento elaborado por la AEGH para la propuesta de creación de la Especialidad de Genética Clínica en España (Enero 2007)

3

3. Utilidad de los organismos modelo en el estudio de las enfermedades genéticas humana

LA LEVADURA *SACCHAROMYCES CEREVISIAE* COMO MODELO Y HERRAMIENTA PARA EL ESTUDIO DE ENFERMEDADES GENÉTICAS HUMANAS

Pascual Sanz

Unidad de Señalización por Nutrientes, Instituto de Biomedicina de Valencia, CSIC y CIBER de Enfermedades Raras. Valencia.

La levadura *Saccharomyces cerevisiae* es un organismo eucariota unicelular que comparte con eucariotas multicelulares más complejos, muchos de los sistemas de detección de señales externas, transmisión de las mismas y su transformación en los ajustes metabólicos y de expresión génica necesarios para adaptarse a los cambios en el ambiente externo. Por ser un organismo unicelular, la levadura *S. cerevisiae* tiene la ventaja de su rápido crecimiento, la facilidad de utilización de técnicas genéticas clásicas y su facilidad para la introducción de DNA exógeno en su interior. El genoma de esta levadura fue el primero en secuenciarse completamente, convirtiéndose así en un organismo clave en investigación genómica, incluyendo extensos estudios en transcritómica, análisis global de la función de cada uno de los genes (se dispone de mutantes de disrupción de cada uno de los alrededor de 6.000 genes que contiene el genoma de *S. cerevisiae*), de la localización subcelular de las proteínas que codifican, de las interacciones proteína-proteína (tanto por doble híbrido como con cromatografía de afinidad en tandem), así como del estado de fosforilación de las mismas. Todos estos datos, unidos a la valiosa información sobre su bioquímica, biología molecular y celular, hacen de *S. cerevisiae* el organismo de elección para el estudio de procesos de regulación complejos (1).

Dado que la mayoría de los procesos fisiológicos de eucariotas superiores también están presentes en *S. cerevisiae*, esta levadura se ha convertido en el organismo de elección para estudiar la función de las proteínas relacionadas con tales procesos. En un primer paso hay que identificar el posible gen ortólogo en la levadura relacionado con el gen humano a estudiar. Una vez identificado, procederemos a recabar toda la información posible sobre ese determinado gen. En este sentido, existe una base de datos de libre acceso muy útil que recopila toda la información relativa a un determinado gen y su correspondiente proteína (www.yeastgenome.org). A continuación podemos plantearnos el trabajar con mutantes de disrupción de ese gen ortólogo y complementar su función mediante la introducción del gen humano correspondiente. De esta manera podemos estudiar si el gen humano cumple una función similar a su ortólogo de levadura. Si este es el caso, la funcionalidad de formas mutantes del gen humano relacionadas con una determinada patología, pueden ser estudiadas fácilmente.

Un ejemplo de lo que se acaba de exponer es el caso del gen de la glucocinasa pancreática humana (GCK). Esta enzima participa en la primera etapa de la glicólisis, fosforilando la glucosa, que entra en la célula pancreática a través del transportador de glucosa de baja afinidad, GLUT2, para formar glucosa-6P. Se ha descrito que alteraciones en el gen de la glucocinasa se correlacionan

con diferentes patologías. Así, mutaciones que conducen a una pérdida de la actividad de la enzima se encuentran en un tipo particular de diabetes monogénica denominada MODY2 (maturity onset diabetes of the young type 2). En cambio, mutaciones que conducen a formas hiperactivas de la enzima se encuentran en la hiperinsulinemia hipoglucémica persistente de la infancia (PHHI, persistent hyperinsulinemic hypoglycemia of the infancy). El gen ortólogo de GCK en levaduras se denomina HXK2 y codifica para la enzima hexokinasa 2, de función similar en el inicio de la ruta glicolítica. Nuestro grupo pudo comprobar que en mutantes que carecían de HXK2, la introducción del gen GCK humano complementaba la función fosforiladora de glucosa, permitiendo el crecimiento de las levaduras recombinantes en medios conteniendo glucosa. Además, dado que se ha descrito que Hxk2 cumple una función reguladora modulando el estado de activación de la protein kinasa Snf1, implicada en la asimilación de fuentes de carbono distintas a la glucosa, pudimos comprobar que la glucokinasa humana era capaz también de ejercer esta función reguladora, lo que plantea la posibilidad de que esta enzima pueda mantener esta función reguladora adicional en células de mamífero (2).

La estructura molecular de la Hxk2 se resolvió en el año 2000 (3). Dado que la secuencia proteica de Hxk2 y la glucokinasa humana presentan un elevado grado de homología (alrededor del 33% de similitud), esto permitió identificar en la estructura de la Hxk2, residuos críticos de la glucokinasa asociados bien a MODY2 o a PHHI y comprender el por qué de las alteraciones en los parámetros cinéticos. Recientemente, la estructura de la glucokinasa humana se ha resuelto, lo que ha permitido definir una extraordinaria similitud estructural entre ambas enzimas (4).

Otro ejemplo interesante es el caso de la protein kinasa activada por AMP (AMPK). La AMPK es un sensor energético celular que cuando se activa es capaz de inhibir rutas anabólicas y activar rutas catabólicas con la finalidad de mantener el estado energético celular. AMPK es un complejo trimérico formado por una subunidad catalítica (a) y dos subunidades reguladoras (b y g). La levadura *S. cerevisiae* posee ortólogos para cada una de las subunidades de AMPK, que forman el complejo SNF1, constituido por la subunidad catalítica Snf1 (ortólogo de AMPKa) y las subunidades reguladoras Snf4 (ortólogo de AMPKg) y Gal83/Sip1/Sip2 (ortólogos de AMPKb). Estudios en levadura han permitido comprender como interaccionan entre si las subunidades dentro del complejo. Éste se activa por fosforilación de la subunidad catalítica por kinasas específicas. Estudios en la levadura también han permitido la identificación de estas kinasas, y definir así los correspondientes ortólogos en células de mamífero, lo que ha supuesto un gran avance en el estudio de la regulación de la actividad de AMPK en mamíferos (1).

AMPK juega un papel importante en la fisiología celular. De hecho, se ha observado recientemente que fármacos de uso habitual en el tratamiento de la diabetes de tipo 2, como la metformina o las tiazolidinedionas (rosiglitazona), ejercen su acción por activación de AMPK. Además, se ha descrito que mutaciones en el gen de la subunidad reguladora g, están relacionadas con la aparición de determinadas cardiomiopatías (síndrome de Wolf-Parkinson-White). La introducción de estas mutaciones en los residuos correspondientes de la subunidad reguladora Snf4 de levadura, ha permitido conocer que estas mutaciones provocan un estado de activación constitutivo de AMPK.

Un ejemplo adicional de utilización de la levadura *S. cerevisiae* en la elucidación de procesos fisiológicos complejos es su utilización en el estudio de la regulación de la actividad de la protein fosfatasa de tipo 1 (PP1), por subunidades reguladoras específicas. La protein fosfatasa PP1 se localiza mayoritariamente en el núcleo celular, donde ejerce su función reguladora a nivel del

ciclo celular. La levadura *S. cerevisiae* tiene un ortólogo de PP1 denominado Glc7, que cumple funciones similares. Estudios en levadura han demostrado que Glc7 se une a dos subunidades reguladoras, denominadas Ypi1 y Sds22, que posibilitan su internalización nuclear y median en la acción que Glc7 tiene a nivel del ciclo celular (transición G2-M) (5). Ypi1 y Sds22 tienen ortólogos en mamíferos denominados Inhibidor 3 (Inh3) y hSds22, respectivamente. Gracias a los estudios en levadura, se ha descrito recientemente que Inh3 y hSds22 tienen una función similar a la de sus ortólogos de levadura y participan en la internalización nuclear de PP1 (6). Recientemente también se ha encontrado que la expresión conjunta de Inh3 y hSds22 es capaz de complementar defectos en Ypi1, por lo que la levadura puede ser utilizada para comprender la función de estas subunidades en cuanto a ciclo celular se refiere. Todavía no se conocen patologías humanas asociadas a defectos en Inh3 o hSds22, posiblemente porque mutaciones en estos reguladores sean muy deletéreas. No obstante, la información que se obtenga de la función de sus ortólogos en levadura será muy valiosa para comprender el funcionamiento de estas proteínas de mamíferos.

Además del papel de *S. cerevisiae* como organismo modelo en el que estudiar procesos fisiológicos complejos, esta levadura también puede servir de herramienta para comprender el funcionamiento de genes de mamíferos no conservados. En 1989, Fields y Song (7) describieron el sistema de doble híbrido en levadura, que ha sido ampliamente utilizado para estudiar las interacciones proteína-proteína de mamífero. Un ejemplo de esta utilización de la levadura como herramienta la encontramos en la regulación de la interacción de dos proteínas relacionadas con la epilepsia mioclónica progresiva de tipo Lafora, denominadas laforina y malina. Estas proteínas no tienen un ortólogo en levadura, pero gracias a la utilización de este organismo como herramienta se ha demostrado que la laforina y la malina interactúan entre sí, y que mutaciones, tanto en el gen que codifica para la laforina como en el que codifica para la malina, alteran negativamente la interacción entre ellas. Por tanto la enfermedad puede ser explicada tanto por defectos en las actividades enzimáticas de las dos proteínas, como en alteraciones en la interacción entre ellas.

En conclusión, la levadura *S. cerevisiae* es un organismo muy útil para estudiar el funcionamiento de procesos fisiológicos complejos, así como para servir de herramienta para definir en el funcionamiento de procesos fisiológicos no conservados.

Referencias

1. Sanz, P. (2007) Yeast as a model system to study glucose-mediated signalling and response. *Front Biosci*, **12**, 2358-2371.
2. Mayordomo, I. and Sanz, P. (2001) Human pancreatic glucokinase (GlcB) complements the glucose signalling defect of *Saccharomyces cerevisiae* *hvk2* mutants. *Yeast*, **18**, 1309-1316.
3. Kuser, P.R., Krauchenco, S., Antunes, O.A. and Polikarpov, I. (2000) The high resolution crystal structure of yeast hexokinase PII with the correct primary sequence provides new insights into its mechanism of action. *J Biol Chem*, **275**, 20814-20821.
4. Kamata, K., Mitsuya, M., Nishimura, T., Eiki, J. and Nagata, Y. (2004) Structural basis for allosteric regulation of the monomeric allosteric enzyme human glucokinase. *Structure (Camb)*, **12**, 429-438.

5. Pedelini, L., Marquina, M., Arino, J., Casamayor, A., Sanz, L., Bollen, M., Sanz, P. and Garcia-Gimeno, M.A. (2007) YPI1 and SDS22 proteins regulate the nuclear localization and function of yeast type 1 phosphatase Glc7. *J Biol Chem*, **282**, 3282-3292.
6. Lesage, B., Beullens, M., Pedelini, L., Garcia-Gimeno, M.A., Waelkens, E., Sanz, P. and Bollen, M. (2007) A complex of catalytically inactive protein phosphatase-1 sandwiched between Sds22 and inhibitor-3. *Biochemistry*, **46**, 8909-8919.
7. Fields, S. and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature (London)*, **340**, 245-246.

CAENORHABDITIS ELEGANS, UN ESPEJO DE 959 CÉLULAS

J. C. Rodríguez Aguilera

Centro Andaluz de Biología del Desarrollo (CABD), Universidad Pablo de Olavide, CSIC. Sevilla.

Historia de *C. elegans* como organismo modelo

La investigación sobre *C. elegans* se inicia en 1965 de la mano de Sydney Brenner. Su simpleza anatómica (unas 1000 células) y facilidad de observación al ser transparente, junto con su corto tiempo de generación (3 días), y alimentación sencilla a base de *E. coli*, hicieron que pronto se reconociera su idoneidad para trabajo en laboratorios económicamente modestos.

Durante la primera década de uso de *C. elegans* se establecieron sus bases anatómicas y ultra-estructurales, además de descubrir algunos patrones básicos de comportamiento del animal. Aunque el primer mapa genético de Brenner vio la luz en 1974 con unos 100 loci basados en marcadores de comportamiento, eran sin duda tiempos de ciencias morfológicas y del microscopio electrónico como tecnología punta.

Los años 80 marcaron la explosión de la biología molecular, la secuenciación del ADN y la descripción de mutantes basados sobretodo en aspectos bioquímicos. Así se desarrollaron mucho más los mapas genéticos, aportando en ellos información sobre rutas metabólicas. En este periodo se describieron detalles del funcionamiento del sistema nervioso y su influencia en el comportamiento, además del linaje celular completo durante el desarrollo embrionario.

Los años 90 trajeron las secuenciaciones masivas y los bancos de datos genéticos, las alineaciones de secuencias y la anotación de secuencias candidatas de ser genes. Pero coincidiendo con la culminación de la secuenciación del genoma de *C. elegans*, las observaciones de Craig Mello y Andrew Fire sobre la posibilidad de silenciar específicamente la expresión génica con ARN bicatenario sentaron los pilares de la genética inversa.

En el nuevo siglo irrumpió con fuerza la potente tecnología de silenciamiento génico, o ARN de interferencia, mediante un sencillo procedimiento alimentario basado en ciertas estirpes de *E. coli*. Se volvieron a describir nuevos genes, muchos de ellos genes de ARN, comenzando los ensayos sistemáticos y las librerías de ARN de interferencia. En paralelo aparecieron los ensayos a gran escala sobre expresión génica, los arrays trajeron datos masivos de difícil interpretación, la época de la biología *in silico* y de la bioinformática estaba servida.

Todas estas profundas revoluciones en las tecnologías aplicables a la biomedicina nos han dejado 3 premios Nobel en 40 años de *C. elegans*. Sydney Brenner (2002) por sus contribuciones al entendimiento de la muerte celular programada y la organogénesis. Craig Mello y Andrew Fire (2006) por el descubrimiento del silenciamiento génico mediado por ARN bicatenario.

¿Por qué *C. elegans*?

C. elegans se ha convertido en un modelo muy popular debido a la sencillez de su cultivo ya que no requiere de inversión significativa en infraestructura más allá de una lupa estereoscópica con transiluminación.

C. elegans se alimenta de un césped de bacterias (en general la estirpe *E. coli* OP50) crecidos sobre una base de agar tipo microbiológico algo modificada. El mantenimiento de las poblaciones de *C. elegans* es extremadamente simple y sólo requiere transferirlos mediante el recorte de un pequeño trozo de agar de la placa de origen a la de destino. La dedicación es pues mínima y flexible en el tiempo, como para adaptarlo a las agendas más exigentes.

Al igual que *S. cerevisiae*, *C. elegans* se puede mantener como poblaciones propagadas de modo asexual o sexual, ya que los dos sexos existentes (machos y hermafroditas) permiten la generación de estirpes genéticamente homogéneas, y de cruzarlas a voluntad. El genoma de *C. elegans* está completamente secuenciado y el número de genes establecidos en unos 20.000, divididos en 6 cromosomas distintos.

Biología de *C. elegans*

No se puede concebir la utilidad de un organismo modelo basándose sólo en la existencia de datos genéticos, sino que se requiere el conocimiento profundo de los procesos celulares y fisiológicos del organismo en condiciones saludables. Sólo así se pueden establecer los efectos y consecuencias de las alteraciones genéticas inducidas o existentes.

En el caso de *C. elegans*, la biología del organismo se conoce el desarrollo embrionario a nivel celular desde la fecundación hasta la eclosión del huevo, incluido el papel que juegan todos y cada uno de los blastómeros en la formación de las poblaciones celulares que originarán determinados órganos y tejidos. El desarrollo embrionario culmina generando una larva tipo L1 de 550 células y habiendo muerto 113 por apoptosis. El desarrollo del organismo no concluye aún. Serán necesarias 4 fases larvianas con sus correspondientes mudas hasta llegar a un organismo adulto de 959 células, 302 de las cuales son neuronas) y 18 células más habrán muerto por apoptosis.

El organismo adulto está muy bien estudiado en términos fisiológicos. Sirva como curiosidad que se conoce que cada huevo necesita de unos 90 a 120 segundos para ser puesto y que el intervalo entre puestas consecutivas es unos 20 minutos. La defecación tiene lugar cada 46-49 segundos. La puesta de huevos es de unos 350 huevos en un plazo de 4 días de vida, de los 21 días de promedio que componen el ciclo vital.

Herramientas y recursos

El estudio de *C. elegans* ha sido posible gracias al desarrollo de diversos tipos de herramientas y recursos. Entre ellos hay que destacar lo siguiente:

- Wormbase.org contiene la secuencia completa del genoma de *C. elegans*, siendo posible el alineamiento y comparación de secuencias (BLAST) con otros organismos, la búsqueda y predicción de genes, intrones, exones, promotores y secuencias regulatorias. También se presentan las estirpes mutantes conocidas y sus correspondientes fenotipos.

- Caenorhabditis Genetics Center, es un banco de estirpes cuidadosamente clasificadas y mantenidas. Este recurso es gratuito por el momento y suministra casi todos los mutantes existentes utilizados en la literatura mundial de este organismo.
- Wormimage.org es un banco de datos de imágenes para buscar patrones de expresión de genes por tejidos, sexos, desarrollo, patrones temporales, etc.
- WORMATLAS.org es un atlas anatómico de preparaciones de microscopía electrónica y óptica que permite diseccionar y reconstruir tridimensionalmente los tejidos y órganos de *C. elegans*.
- Knockout consortium es un consorcio para la generación de nuevos mutantes en genes de interés. Este recurso permite ahorrar tiempo y subcontratar este aspecto tan problemático y que tanto tiempo consume. Generan estirpes homocigotos mutantes o heterocigotos balanceadas en el caso de genes letales.
- Textpresso es un robot de búsqueda a texto completo en la literatura publicada sobre *C. elegans*. La literatura es requerida y anotada cada vez que un miembro de la comunidad investigadora genera una nueva referencia.

C. *elegans* como modelo de enfermedades genéticas humanas

Existen múltiples ejemplos del uso de *C. elegans* como organismo modelo. Cabe destacar la enorme contribución que este organismo ha aportado al conocimiento sobre la ruta de señalización de la insulina y sus consecuencias sobre el metabolismo de grasas y la oxidación mitocondrial de metabolitos.

Para ilustrar este apartado se mostrarán conocimientos obtenidos utilizando *C. elegans*, en modelos de enfermedades mitocondriales y metabólicas, que tienen sus equivalentes humanas en clínica.

Referencias

- Vázquez-Manrique et al., 2007, The frataxin-encoding operon of *Caenorhabditis elegans* shows complex structure and regulation. *Genomics* 89, 392-401
- Vázquez-Manrique et al., 2006, Reduction of *Caenorhabditis elegans* frataxin increases sensitivity to oxidative stress, reduces lifespan, and causes lethality in a mitochondrial complex II mutant. *FASEB Journal* 20, 172-174
- Nani et al., 2006, Genetics of primary CoQ₁₀ deficiency. *Curr. Genomics* 7, 343-349
- Asencio et al., 2006, Differential expression pattern of *coq-8* gene during development in *Caenorhabditis elegans*, *Gene Expression Patterns* 6, 433-439
- Houthoofd et al. 2005, Metabolism, physiology and stress defense in three aging Ins/IGF-1 mutants of the nematode *Caenorhabditis elegans*, *Aging Cell* 4, 87-95

QUÉ OFRECE DROSOPHILA COMO MODELO DE ESTUDIO DE LAS ENFERMEDADES NEURODEGENERATIVAS

M. Dolores Moltó

Departamento de Genética, Facultad de Ciencias Biológicas, Universitat de València, Valencia.

El género *Drosophila*, y en particular la especie *Drosophila melanogaster*, tiene una larga y fructífera historia como organismo modelo en la investigación genética. Desde que Thomas Hunt Morgan observara, a principios del siglo XX, un mutante de ojos blancos en su laboratorio de la Universidad de Columbia, *Drosophila* ha sido el organismo de elección de muchos genetistas interesados en problemas tan diversos como el desarrollo, el comportamiento, la proliferación celular y últimamente la salud humana y la investigación biomédica. Esta elección obedece a características biológicas y de naturaleza práctica, al desarrollo de un gran número de herramientas genéticas muy sofisticadas por parte de una comunidad científica muy activa a lo largo del tiempo, y más recientemente al conocimiento de la secuencia de su genoma.

D. melanogaster es un pequeño insecto cuyo genoma es unas 18 veces más pequeño que el genoma humano. En él se han descrito un total de 13.600 genes, lo que representa aproximadamente la mitad del número de genes de la especie humana, en parte debido a la mayor redundancia del genoma humano en comparación con el genoma de *Drosophila*. La razón más importante que justifica la explotación de la mosca como organismo modelo en las enfermedades humanas, se basa en la presunción de que los aspectos fundamentales de la biología celular de este organismo, se han conservado en organismos superiores como el humano. De hecho, alrededor de la mitad de las proteínas de *Drosophila* muestran una importante similitud con proteínas de mamífero. De acuerdo con la base de datos Homophila (<http://superfly.ucsd.edu/homophila>), cerca del 77% de los genes humanos que son responsables de enfermedades, tienen un ortólogo en *Drosophila*. Pero la conservación va más allá de las entidades proteicas, de manera que algunos procesos como el establecimiento de los ejes corporales, la organización de estructuras complejas como el corazón, los ojos y el sistema inmune innato, el control de la proliferación celular, el crecimiento de los axones, así como el control de ciertos comportamientos básicos, como el sueño y el abuso de sustancias, están muy conservados entre la mosca y el hombre.

Drosophila constituye un modelo genético sencillo, que ocupa un nicho intermedio entre los modelos unicelulares, como los creados en levadura, y los modelos generados en organismos vertebrados, como el ratón. Para determinadas enfermedades humanas, *Drosophila* puede reproducir *in vivo* algunos de los procesos patológicos en un número elevado de individuos de organización pluricelular y en una escala temporal mucho menor. No obstante, el protagonismo actual de *Drosophila*, dentro de los modelos biomédicos, se asienta principalmente en su gran utilidad para identificar los genes involucrados en un mismo proceso celular y, no menos importante, en la identificación de nuevos fármacos mediante cribado de compuestos a gran escala.

Estrategias de estudio de la función génica en *Drosophila*

Drosophila dispone de una panoplia de técnicas que permite el estudio de la función de un gen cuyo ortólogo en humanos causa una enfermedad hereditaria. Estas técnicas permiten la creación de fenotipos alterados en la mosca, tanto por pérdida como por ganancia de función génica. Dentro de las metodologías más clásicas destaca la mutagénesis química, con el empleo de compuestos como etilnitrosourea, metilmetanosulfonato o etilmetanosulfonato entre otros, y el empleo de radiaciones ionizantes como los rayos X, los rayos γ y la radiación ultravioleta¹. Estas técnicas se caracterizan por la aleatoriedad de las lesiones generadas en el DNA y forman parte de lo que ha venido designándose como estrategias de genética directa, la cual se caracteriza por la clonación de un gen a partir del fenotipo mutante.

A estas técnicas le han seguido otras desarrolladas para facilitar el estudio molecular de los fenotipos generados por las mutaciones, destacando en este sentido la utilización de elementos transponibles modificados. Entre ellos el más utilizado ha sido el elemento P, un transposón no retroviral del genoma de *D. melanogaster*². Actualmente, se dispone de una amplia colección de cepas cada una de ellas con una sola inserción de un elemento P localizada en una región concreta, fruto del Proyecto de Disrupción de Genes mediante Mutagénesis Insercional del Elemento P. Ello permite tener al alcance un posible mutante del gen de interés con solo consultar la Flybase (<http://flybase.bio.indiana.edu/>). Una mejora de la mutagénesis insercional ha sido la creación del sistema *P-homing*³, al aumentar la especificidad de inserción de este elemento en posiciones muy cerca o dentro del gen cuya actividad se desea modificar.

Una de las metodologías que ha revolucionado en cierta medida la investigación con *Drosophila* ha sido la aplicación, en este organismo, del sistema GAL4-UAS propio de levadura⁴. Éste es un sistema muy versátil ya que permite tanto la sobreexpresión de genes concretos como la reducción de su expresión, en éste último caso al combinarse con la técnica basada en la interferencia del RNA⁵. Una gran ventaja del sistema es la capacidad que ofrece de poder controlar espacial y temporalmente cualquier actuación. De esta forma, es posible dirigir la expresión de un gen en un momento del desarrollo que no le corresponde, o en un tejido donde no suele expresarse. Alternativamente, se puede silenciar la expresión de un gen en un estadio y/o tejido donde su producto es necesario. Su aplicabilidad en la generación de moscas transgénicas portadoras de genes humanos con mutaciones patológicas, se ha puesto también de manifiesto en un importante número de enfermedades neurodegenerativas. Con el tiempo se han ido introduciendo algunas modificaciones que han permitido flexibilizar aún más este sistema e inducir la expresión de la proteína GAL4 bajo el control hormonal o por la presencia de antibióticos⁶.

A diferencia de otros organismos, en *Drosophila* no se ha descrito el mecanismo de recombinación homóloga entre un ADN exógeno y su homólogo genómico, tan utilizado en la levadura o en el ratón para generar *knock-outs*. Dado que la mutagénesis insercional no ha tenido éxito en algunos genes, y que la interferencia de RNA no suele generar alelos nulos, se han desarrollado nuevas técnicas para cubrir este hueco. Son técnicas basadas en el sistema FLP-FRT, importado también de levadura, que han permitido la inserción de copias génicas modificadas o el reemplazamiento de la copia genómica⁷.

Modelos de enfermedades neurodegenerativas humanas desarrollados en *Drosophila*

El anuncio del proyecto genoma de *Drosophila* estuvo, además, acompañado por la predicción de cual sería su utilidad como modelo en la comprensión de las enfermedades genéticas humanas. De todas las áreas de aplicación, la neurogenética constituye una de las más estudiadas, por el hecho de que el sistema nervioso de *Drosophila* ha sido analizado exhaustivamente, tanto desde el punto de vista anatómico como fisiológico. Para este tejido, el tamaño no es relevante, de hecho se estima que el cerebro de la mosca de la fruta tiene más de 300.000 neuronas y está organizado, similarmente al de mamíferos, en áreas con funciones especializadas como el aprendizaje, la memoria, el olfato y la visión. Recientemente, se han desarrollado varios modelos de enfermedades neurodegenerativas en la mosca⁸ (Tabla 1), tras la identificación de los genes responsables de las formas mendelianas en humanos. Algunos de estos modelos han generado resultados muy prometedores, mientras otros no han podido reproducir las características esenciales de la enfermedad. Las enfermedades neurodegenerativas son, generalmente, de inicio tardío y muestran complejas alteraciones como la pérdida de memoria, defectos cognitivos, desórdenes en el movimiento, que conjuntamente llevan a la muerte del individuo. Estas alteraciones son debidas en todos los casos a la pérdida o la disfunción de neuronas especializadas y en un gran número de ellos, a mutaciones que afectan al plegamiento de ciertas proteínas. Esto lleva a una acumulación de las mismas formando grandes agregados que son tóxicos para las neuronas, conocidos como inclusiones en la enfermedad de Huntington, placas de amiloide en el caso de la enfermedad de Alzheimer y cuerpos de Lewy en la enfermedad de Parkinson. Los modelos de *Drosophila* para estas enfermedades han sido los pioneros, al tratarse de enfermedades más prevalentes, especialmente las dos últimas. Actualmente, se están desarrollando modelos para enfermedades genéticas raras, como la ataxia de Friedreich y las ataxias dominantes espinocerebelosas (SCA). Estos trastornos están causados por expansiones anómalas de trinucleótidos, que en el caso de las ataxias SCA originan tramos de poliglutaminas en la proteína correspondiente, como ocurre en la enfermedad de Huntington. Para la ataxia de Friedreich, la expansión provoca la pérdida de función del gen que lleva a una significativa disminución de la proteína frataxina⁹.

Una vez se dispone de un modelo, en *Drosophila*, de una enfermedad neurodegenerativa es posible su explotación a través de dos vías principalmente. Una de ellas es la búsqueda de genes cuyas mutaciones puedan modificar, bien aumentando o disminuyendo, el fenotipo mutante de nuestro modelo en la mosca (Tabla 1). Esto se realiza mediante rastreos genéticos utilizando colecciones de cepas de *Drosophila* portadoras de deficiencias, inserciones derivadas de elementos transponibles, y más recientemente, colecciones en las que se han silenciado genes concretos mediante interferencia del RNA. Esta metodología puede aplicarse a gran escala y se parte de un fenotipo mutante que sea visible, fácil de identificar y que afecte a estructuras que no sean esenciales para la viabilidad de las moscas en las condiciones de laboratorio. En el caso de las enfermedades neurodegenerativas, muchos de los rastreos genéticos se han llevado a cabo expresando el gen mutado en el ojo de *Drosophila*, debido a que esta estructura tolera bien las alteraciones que afectan los procesos biológicos básicos. Otras estructuras elegidas han sido las alas y las quetas de las moscas.

Los genes modificadores que han sido identificados en los rastreos genéticos han proporcionado información muy importante sobre los mecanismos patogénicos del fenotipo neurodegenerativo en la mosca, y potencialmente de la enfermedad humana. Tales genes pueden constituir, además, genes causantes de dicha enfermedad (casos de heterogeneidad genética) o de enfermedades clínicamen-

te similares, o factores genéticos de riesgo (en las enfermedades multifactoriales). Asimismo, se convierten en nuevas dianas que explorar en la determinación de los tratamientos terapéuticos.

En segundo lugar, *Drosophila* permite la búsqueda de fármacos eficaces en el tratamiento de la enfermedad. Una vez conocida parcial o totalmente la ruta de patogénesis, se pueden ensayar, en este organismo, compuestos con una actividad biológica conocida para comprobar su efectividad, así como evaluar el efecto de nuevas drogas. Un número importante de compuestos del mercado han sido ya testados en modelos de mosca. En la mayoría de los casos este organismo ha respondido de la forma esperada según el modo de acción del compuesto en humanos, lo que permite especular que los compuestos que son identificados en los rastreos de nuevos fármacos funcionen de forma similar en *Drosophila* y en las células de mamífero. De los compuestos identificados, se exige que mejoren aspectos concretos de una enfermedad humana que han sido reproducidos en la mosca. Además, se puede llegar a determinar un régimen combinado de fármacos con la menor dosis posible y el mejor efecto terapéutico. A diferencia de los modelos en ratón, que son caros de mantener, es posible mantener grandes poblaciones de moscas con un coste razonable. Ello permite que la realización de los cribados farmacológicos sea a gran escala, a partir de colecciones con un número elevado de compuestos químicos diferentes¹⁰. Hoy en día, los modelos basados en animales invertebrados como *Drosophila* constituyen, por sus características, el nexo entre los ensayos *in vitro* y los ensayos preclínicos que permiten evaluar las moléculas con propiedades terapéuticas en las enfermedades humanas.

Referencias

- Ashburner, M., Golic, K.G., and Scout, R. *Drosophila a Laboratory Handbook*. 2nd Ed. Cold Spring Harbour Press (2004).
- Rubin, G.M. *Mobile Genetic Elements*. Academic Press Orlando FL (1983).
- Taillebourg, E and Dura, J.M. (1999). A novel mechanism for P element homing in *Drosophila*. *Proc. Natl. Acad. Sci. USA*, 96: 6856-6861.
- Brand, A.H. and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development*, 118: 401-415.
- Kennerdell, J.R. and Carthew, R.W. (2000). Heritable gene silencing in *Drosophila* using double-stranded RNA. *Nat. Biotechnol*, 18: 896-898.
- McGuire, S.E., Roman, G and Davis, R.L. (2004). Gene expression system in *Drosophila*: a synthesis of time and space. *Trends Genet.*, 20: 384-391.
- Rong, Y.S. and Golic, K.G. (2000). Gene targeting by homologous recombination in *Drosophila*. *Science*, 288: 2013-2018.
- Cauchi, R.J. and van den Heuvel M. (2006). The fly as a model for neurodegenerative diseases: Is it worth the jump? *Neurodegenerative Dis.*, 3: 338-356.
- Gatchel, J.R., Zoghbi, H.Y. (2005). Diseases of unstable repeat expansion: mechanisms and common principles. *Nat. Rev Genet.*, 6: 743-755.
- Desai UA, 10. Desai UA, Pallos J, Ma AA, Stockwell BR, Thompson LM, Marsh JL, Diamond MI. (2006). Biologically active molecules that reduce polyglutamine aggregation and toxicity. *Hum Mol Genet.*, 15: 2114-2124.

Tabla 1. Modelos desarrollados en *Drosophila* de algunas enfermedades neurodegenerativas humanas

Enfermedades poliQ

<i>Genes homólogos en humanos</i>	HD, SCA1, SCA2, SCA6
<i>Degeneración neuronal</i>	Neuronas fotorreceptoras, quetas sensoriales
<i>Déficit motor</i>	Pérdida de la coordinación motora en larvas, menor capacidad de escalada en adultos.
<i>Agregación proteica</i>	Nuclear, peri-nuclear y citoplasmática
<i>Letalidad prematura</i>	Si
<i>Genes modificadores</i>	Implicados en el plegamiento de las proteínas

Enfermedad de Parkinson

<i>Genes homólogos en humanos</i>	UCHL1, PARK2, PARK6, PARK7
<i>Degeneración neuronal</i>	Pérdida de neuronas dopaminérgicas y degeneración retiniana
<i>Déficit motor</i>	Alteraciones en la capacidad de vuelo de los adultos
<i>Agregación proteica</i>	Inclusiones intraneuronales de α -synucleína
<i>Letalidad prematura</i>	Si en los mutantes nulos de parkina <i>Genes modificadores</i> Chaperonas moleculares modifican la toxicidad de α -synucleína Parkina suprime el fenotipo inducido por α -synucleína

Enfermedad de Alzheimer

<i>Genes homólogos en humanos</i>	PS1, PS2, TAU, APP
<i>Degeneración neuronal</i>	Dependiente de la edad
<i>Déficit de aprendizaje</i>	Inducidos por la misexpresión del péptido A β
<i>Agregación proteica</i>	Placas amiloideas (A β) y ovillos neurofibrilares de la proteína tau
<i>Letalidad prematura</i>	Si en los casos de transgénesis del péptido A β y de la proteína tau
<i>Genes modificadores</i>	Mutaciones en neprilisina-2 suprimen la toxicidad inducida por el péptido A β . Varias quinasas y fosfatasas modifican la patología generada por la proteína tau

Ataxia de Friedreich

<i>Genes homólogos en humanos</i>	FXN
<i>Degeneración neuronal</i>	Sistema nervioso periférico, preferentemente neuronas sensoriales.
<i>Déficit motor</i>	Disminución de la capacidad de escalada de los adultos
<i>Letalidad prematura</i>	Si cuando la disminución de la proteína FH es muy drástica
<i>Genes modificadores</i>	ND

APP, amyloid precursor protein; A β , beta-amyloid, HD, Huntington disease; FH, *Drosophila* frataxin protein; FXN, Frataxin gene; ND, no determinado; PARK: parkinson; polyQ, poli-glutaminas ; PS, presenilin; SCA, spinocerebellar ataxia; UCHL1, ubiquitin carboxyl-terminal esterase L1.

LA GENÓMICA COMPARADA EN EL ESTUDIO DEL DESARROLLO, LA EVOLUCIÓN Y LAS ENFERMEDADES GÉNICAS

J. L. Gómez-Skármeta

Centro Andaluz de Biología del Desarrollo (CABD), Universidad Pablo de Olavide, CSIC. Sevilla.

Todas las células de un organismo contienen el mismo genoma, pero no todas transcriben los mismos genes. Eso las diferencia en células hepáticas, renales, musculares, cardíacas, etc. El proceso de transcripción está altamente regulado a través de proteínas, denominadas factores de transcripción, que se unen a regiones de ADN no codificantes (regiones reguladoras) y favorecen la transcripción desde el ADN codificante al correspondiente ARNm. Esas secuencias reguladoras de ADN no codificantes son pequeñas regiones distribuidas en distintos lugares, a diferentes distancias de la región codificante. En función de la complejidad de transcripción de un ARNm, éste tendrá más o menos regiones reguladoras. El gran problema es que, a diferencia de las regiones codificantes (5-10% del genoma) cuyo código se conoce y por lo tanto podemos “leer” a partir de la secuencia de un genoma, el lenguaje y función de las regiones reguladoras (90-95% del genoma) se desconoce.

Desarrollo y evolución

¿Cómo se explica la gran diversidad de forma que existen en el reino animal? Actualmente, pensamos que uno de los motores esenciales que han operado durante la evolución para generar distintas formas radica en la variación de las regiones reguladoras no codificantes (Carroll, 2005). El símil podría ser: las proteínas son como los materiales (ladrillos, vigas, cables, etc) de una construcción; las regiones reguladoras serían los planos cuyas instrucciones hay que seguir. En función del conjunto de instrucciones los mismos materiales pueden construir un adosado o el museo Guggenheim de Bilbao. Por lo tanto, la evolución ha ido variando el número y tipo de órdenes que operan sobre las mismas regiones codificantes y ha generado la inmensa variedad de formas del reino animal. El desafío científico de este siglo es el de conocer y comprender el lenguaje de estas órdenes.

Comenzamos a saber algo. Así, todos los vertebrados tenemos estructuras comunes y, de hecho, durante el desarrollo embrionario existe un estadio, denominado filotípico, en el que todos los embriones de los distintos vertebrados son muy parecidos. Este estadio, cuando se manifiesta el plan corporal de los vertebrados, es el resultado de una regulación similar de los mismos genes. Posteriormente en el desarrollo, la regulación diferencial de los mismos genes, generará las diferencias morfológicas. La comparación de los genomas completos de varios vertebrados ha permi-

tido descubrir que existe unas 3000 regiones no codificantes conservadas (RNCs) en todos los vertebrados (Bejerano et al., 2004; Sandelin et al., 2004; Woolfe et al., 2005). Puesto que la conservación evolutiva implica función, es lógico pensar que parte de estas RNCs contengan elementos reguladores. Ensayos realizados por diversos laboratorios, entre ellos el nuestro, lo han demostrado (de la Calle-Mustienes et al., 2005; Gottgens et al., 2000; Nobrega et al., 2003; Spitz et al., 2003; Woolfe et al., 2005). Curiosamente, los distintos linajes animales contienen un conjunto diferente de RNCs. Así, las moscas y los gusanos tienen un conjunto independiente de RNCs diferente entre ellos y al de los vertebrados. Una posibilidad es que cada grupo de RNCs contenga el conjunto de instrucciones para construir un plan corporal distinto.

Enfermedades

Las enfermedades génicas son debidas a un mal funcionamiento de un determinado gen durante el desarrollo embrionario o adulto de un individuo. Este mal funcionamiento puede ser debido a mutaciones en la región codificante del gen o a defectos en las regiones reguladoras. Múltiples estudios a escala genómica recientes, en los cuales se ha determinado la asociación de marcadores cromosómicos a enfermedades génicas, indican que muchas de estas enfermedades no se deben a mutaciones en regiones codificantes sino a mutaciones en regiones reguladoras. Sin embargo, en la gran mayoría de los casos, no se ha determinado aún con precisión la posición precisa de las lesiones en el ADN ni las causas por las que dichas lesiones causan enfermedades en humanos.

Nuestro trabajo

El proyecto de nuestro laboratorio, en colaboración con varios grupos europeos, pretende identificar el conjunto de regiones reguladoras necesarias para generar un vertebrado y determinar la implicación de estas regiones en distintas enfermedades humanas. Para ello, mediante transgénesis en pez cebra, analizaremos la actividad funcional de las 3000 RNACs presentes en todos los vertebrados. Estos estudios serán esenciales para descifrar el lenguaje que opera en la regulación de la expresión génica y por tanto entender los procesos que operan en el desarrollo y en la evolución. Además, nuestros estudios nos permitirán identificar regiones reguladoras candidatas a estar afectadas en enfermedades génicas humanas. La secuenciación de estas regiones en pacientes con dichas enfermedades y el análisis funcional de las regiones mutadas identificadas en animales modelo, nos permitirán determinar las causas de dichas enfermedades y establecer métodos sencillos de diagnóstico.

Referencias

- Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S. and Haussler, D. (2004). Ultraconserved elements in the human genome. *Science* 304, 1321-5.
- Carroll, S. B. (2005). Evolution at two levels: on genes and form. *PLoS Biol* 3, e245.
- de la Calle-Mustienes, E., Feijoo, C. G., Manzanares, M., Tena, J. J., Rodríguez-Seguel, E., Letizia, A., Allende, M. L. and Gómez-Skarmeta, J. L. (2005). A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate *Iroquois* cluster gene deserts. *Genome Res.* 15, 1061-72.

Gottgens, B., Barton, L. M., Gilbert, J. G., Bench, A. J., Sanchez, M. J., Bahn, S., Mistry, S., Grafham, D., McMurray, A., Vaudin, M. et al. (2000). Analysis of vertebrate SCL loci identifies conserved enhancers. *Nat Biotechnol* 18, 181-6.

Nobrega, M. A., Ovcharenko, I., Afzal, V. and Rubin, E. M. (2003). Scanning human gene deserts for long-range enhancers. *Science* 302, 413.

Sandelin, A., Bailey, P., Bruce, S., Engstrom, P. G., Klos, J. M., Wasserman, W. W., Ericson, J. and Lenhard, B. (2004). Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics* 5, 99.

Spitz, F., Gonzalez, F. and Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* 113, 405-17.

Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K., Vavouri, T., Smith, S. F., North, P., Callaway, H., Kelly, K. et al. (2005). Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol* 3, e7.

MODELOS TRANSGÉNICOS PARA EL ESTUDIO DEL SÍNDROME HUMANO DISPLASIA ECTODÉRMICA

Paloma Pérez

Laboratorio de modelos animales, Centro de Investigación Príncipe Felipe (CIPF) e Instituto de Biomedicina de Valencia, CSIC. Valencia.

El ratón de laboratorio se ha convertido durante el siglo pasado en el modelo animal por excelencia para la investigación biomédica (1). Son numerosas las razones que justifican este uso masivo. En primer lugar, la similitud de los genomas de ratón y humano ha permitido abordar estudios genéticos y fisiológicos en el ratón, que pueden ser extrapolados al hombre. A diferencia de otros modelos animales (levaduras, gusanos y moscas), idóneos para abordar otro tipo de estudios, los ratones, como muchos otros mamíferos, sufren diversas enfermedades de modo natural, como el cáncer, la arteriosclerosis, diabetes y glaucoma. Además, el ciclo vital del ratón, su fácil mantenimiento y estabulación a un coste comparativamente bajo, explican el uso tan extendido de este modelo. Sin duda alguna, un hecho que ha facilitado los estudios con ratones de laboratorio es el enorme conocimiento acumulado durante muchos años de las cepas puras e híbridas de laboratorio, en la actualidad disponibles comercialmente. Además, los análisis de ratones con mutaciones espontáneas han posibilitado abordar estudios fenotípicos que, en general, se debían a una mutación en un solo gen. Sin embargo, los genes responsables no estaban inicialmente identificados y la identificación fenotípica quedaba claramente restringida a los fenotipos visibles. En ese sentido, es indudable que la modificación genética de ratones ha supuesto uno de los mayores logros en la investigación biomédica, reconocido recientemente con la concesión del Premio Nobel de Fisiología y Medicina 2007 a los Dres M. Capecchi, O. Smithies y M. Evans por sus aportaciones pioneras en este campo (2). Los métodos desarrollados por estos investigadores han hecho posible abordar estudios funcionales exquisitamente detallados de cualquier gen en un animal vivo, diseñando modelos animales para una plétora de enfermedades humanas.

Bajo el epígrafe de transgénesis, se incluyen diversas aproximaciones (3). Nos referiremos brevemente a la generación de ratones transgénicos, *knock out* y *knock in*, así como a los sistemas inducibles, ejemplificados con el uso de la técnica *cre/lox*. La generación de ratones transgénicos se consigue mediante la inyección pronuclear del DNA exógeno de interés, denominado transgén, que se integra al azar en el genoma, en multicopia. Puesto que la integración del transgén no es controlada, existe la posibilidad de que ocurran mutaciones de inserción; ello hace indispensable el análisis de, al menos, dos líneas transgénicas generadas de forma independiente. Una de las aplicaciones más habituales de los transgénicos en el estudio de las consecuencias de la ganancia de función de un producto génico (sobrexpresión). Sin embargo, también se han realizado con éxito transgénicos que son dominantes negativos de función. La expresión del transgén puede estar dirigida por un promotor general o específico de tejido que, a su vez, puede expresarse de forma constitutiva o inducible. También es frecuente el uso de genes “reporter” (lacZ, GFP), que

permiten abordar estudios de expresión y de regulación de promotores. El uso de la recombinación homóloga ha permitido desarrollar herramientas donde los genes “normales” son reemplazados por versiones “alteradas” que inactivan la función génica (mutaciones *null/Knock out*) o bien se generan alteraciones génicas específicas sin alterar el contexto del gen (*Knock in*). El proceso de generación de estos animales se basa en producir mutaciones dirigidas en el gen de interés en células madre embrionarias (ES cells), que al transferirse a blastocistos permiten la producción de ratones que portan la mutación (quimeras) que, una vez transmitida a la línea germinal, resulta en un ratón *Knock out*. Además, el uso de los sistemas inducibles, que ilustraremos aquí con la técnica *cre/lox*, ha permitido un grado de refinamiento aún mayor, al posibilitar el control espacio-temporal de la expresión de un gen determinado. Esta técnica se basa en el uso de la recombinasa del bacteriófago cre, que reconoce específicamente las secuencias *loxP*, de manera que si estos sitios se sitúan flanqueando un gen de interés, la actividad cre es capaz de escindir el fragmento contenido entre los sitios *loxP*, originando las alteraciones genéticas deseadas.

En conjunto, el uso de estas técnicas ha permitido abordar numerosos estudios de función génica y, en concreto, generar numerosos modelos murinos de enfermedades humanas. Aunque existen limitaciones técnicas para la aplicación de estos modelos, la progresión en los últimos años hace previsible anticipar la creación de modelos de enfermedades poligénicas, como la diabetes o la hipertensión. Aún desconocemos si será o no factible extender estos hallazgos a la clínica, restaurando por ejemplo un gen “anormal” en un tejido determinado. Para entender el papel crucial que ha jugado el desarrollo de modelos murinos en la investigación biomédica, resulta imprescindible señalar la existencia de repositorios de ratones públicos, entre los que destaca por ser pionero y por su trayectoria, constituyendo una referencia obligada *The Jackson Laboratory*, en Bar Harbor, Maine. La optimización de distintas técnicas de reproducción asistida, incluyendo la criopreservación de embriones, fertilización *in vitro* o trasplante de ovarios, hace posible propagar líneas transgénicas con dificultad reproductiva y, sobre todo, permite la difusión y distribución de los modelos animales entre distintos investigadores de la comunidad científica sin límite geográfico. La disponibilidad de numerosas páginas *web* con información relativa a diversas patologías humanas está facilitando enormemente el avance en la investigación biomédica (4). Existen modelos murinos de enfermedades tan relevantes como el síndrome de Down, fibrosis quística, diabetes (tipo I y II), distrofia muscular de Duchenne...

A continuación nos referiremos a alguno de los modelos del síndrome humano Displasia Ectodérmica (DE) generados en la última década, y de cuyo análisis ha derivado una mejor comprensión de la enfermedad. Las DEs constituyen un grupo grande y heterogéneo de enfermedades humanas que engloban aproximadamente 200 condiciones clínicas. Hasta la fecha, tan solo 30 de estas patologías han sido estudiadas a nivel molecular con la consiguiente identificación del gen causal. Las DEs se originan como consecuencia de alteraciones en el desarrollo de epitelios que se son derivados del ectodermo embrionario, de ahí su denominación. En clínica, una enfermedad se diagnostica como DE si existen defectos primarios en, al menos, uno de los siguientes tejidos: pelo (hipotricosis, alopecia parcial o total), uñas (distróficas, hipertróficas), anomalías dentarias (dientes ausentes o anormales) y glándulas sudoríparas (hipoplásicas o aplásicas). Las DEs se consideran enfermedades raras con una incidencia estimada de 7 en 10.000 nacimientos y siguen todos los modos de herencia mendeliana (autosómica dominante o recesiva, ligada a cromosoma X dominante o recesiva) aunque también se han descrito casos esporádicos. Recientemente, se ha propuesto una clasificación considerando la función de los genes implicados en los distintos tipos de DE, que establece cuatro grandes subgrupos que incluyen genes

implicados en la comunicación célula-célula, adhesión, regulación transcripcional y desarrollo (5). En algunos casos, como la displasia ectodérmica anhidrótica, la ausencia de glándulas sudoríparas causa hipertermia en el recién nacido y se baraja como posible causa de muertes súbitas sin causa identificada en el área de la neonatología. Incluso en aquellos casos en que la vida del paciente no está comprometida, los defectos en desarrollo dental y de paladar, entre otros, hacen necesarias frecuentes intervenciones quirúrgicas, acarreado perjuicios en la calidad de vida del paciente y familiares, puesto que éstas son malformaciones congénitas que requieren un seguimiento exhaustivo desde el nacimiento y durante muchos años de la vida del paciente, con el consiguiente impacto socioeconómico a nivel hospitalario.

El desarrollo de los apéndices ectodérmicos, como el folículo piloso o el diente, ambos tejidos afectados en el síndrome DE, se regula por una inducción secuencial de interacciones entre dos tejidos adyacentes, el epitelio y el mesénquima, que constituyen el mecanismo más importante de organogénesis específica de tejido y regula numerosas funciones celulares, a pesar de que las moléculas señaladoras sean comunes en diferentes órganos (6). En la clasificación funcional de las DEs de Lamartine, se señalan dos factores de transcripción, NF- κ B y p63, como genes causales de varias de las enfermedades humanas DE. En el caso de NF- κ B, se ha demostrado que la vía de señalización mediada por el complejo I κ B quinasa (IKK), responsable de la activación de NF- κ B, es esencial en la homeostasis epidérmica y el desarrollo de apéndices cutáneos y se han identificado varias patologías con mutaciones de falta de función total o parcial de la subunidad reguladora del complejo IKK, IKK γ , que incluyen la Incontinencia pigmenti (IP) y diversas DEs asociadas a inmunodeficiencia (ED-ID) (7). Por otro lado, el gen p63, homólogo del supresor tumoral p53, es esencial para el desarrollo epitelial y mutaciones en este gen causan varios síndromes humanos DE, que incluyen: “ectodermal dysplasia-ectrodactyly-cleft lip/palate” (EEC), “ankyloblepharon-ED-clefting síndrome” (AEC), “acro-dermatol-ungual-lacrimal-tooth” (ADULT), “limb-mammary síndrome” (LMS) y “non-syndromic split hand/foot malformation” (SHFM).

Una de las líneas de investigación de nuestro grupo aborda el estudio de las bases moleculares del síndrome DE a través del análisis funcional de diversos modelos transgénicos en los que un desbalance de la acción mediada por el receptor de glucocorticoides (GR) induce numerosos defectos morfogenéticos similares al cuadro clínico de pacientes afectados por DE. GR es una proteína perteneciente a la superfamilia de receptores nucleares que puede actuar mediante funciones dependientes e independientes de su unión al DNA (revisado en 8). Nuestro trabajo reciente, a través de la caracterización de ratones transgénicos de sobreexpresión del receptor de glucocorticoides (K5-GR) ha demostrado que la función de GR en epidermis y otros epitelios estratificados derivados del ectodermo embrionario es crucial para el desarrollo normal del organismo, y que un desbalance en dicha función se acompaña de numerosas malformaciones congénitas, acorde con el papel de los GCs como teratógenos (9-11). El análisis fenotípico de los ratones K5-GR se ha realizado en distintas líneas transgénicas heterocigotas y homocigotas. Estos animales exhiben múltiples defectos epiteliales en la epidermis, folículo piloso, desarrollo dental y de paladar, además de ausencia de glándulas de Meibomio y prepucciales, constituyendo por tanto, un valioso modelo para el estudio de DE. Se han descrito defectos oculares en muchas de las enfermedades DE y, es importante señalar que la sobreexpresión de GR induce una tremenda disfunción ocular en embriones tardíos y recién nacidos transgénicos, causando alteraciones en la proliferación de los epitelios de la córnea y los párpados (10). Los defectos fenotípicos en los transgénicos K5-GR se ejercen, al menos en parte, mediante la interferencia con IKK/NF- κ B, con una disminución de la actividad quinasa IKK, que correlaciona con una reducción de los niveles de la proteína IKK γ

tanto en epidermis como en ameloblastos dentarios. Además, hemos detectado una disminución dramática en los niveles de expresión del marcador de células epiteliales p63. Recientemente, hemos generado un nuevo modelo transgénico que sobreexpresa un mutante de GR defectivo en transactivación, pero que conserva las funciones de transrepresión intactas (K5-GR-TR, 12). Nuestros datos previos indican que los transgénicos K5-GR-TR exhiben un fenotipo epitelial que solapa parcialmente con K5-GR, resultando de máximo interés discriminar los genes diana y las vías reguladas diferencialmente en estos dos modelos. El estudio conjunto de K5-GR-TR en el contexto del modelo K5-GR aportará una información muy valiosa relativa a la función de GR en desarrollo epitelial. Además, datos muy recientes de nuestro grupo indican que los ratones deficientes en GR, GR^{-/-}, que mueren perinatalmente debido a múltiples defectos en distintos órganos vitales, muestran afectación epitelial al menos, en el epitelio del párpado, subrayando la importancia de la acción de GR en la morfogénesis epitelial (13). Nuestro trabajo está encaminado a identificar y caracterizar en detalle las vías de señalización implicadas en los defectos morfogénicos inducidos por GR, con objeto de contribuir a descifrar las bases moleculares subyacentes a esta patología humana, avanzar en la posible diagnosis precoz de la enfermedad y, eventualmente, paliar las lesiones epiteliales características de la enfermedad.

Referencias

1. Pinkert CA 2003 Transgenic Animal Technology: Alternatives in Genotyping and Phenotyping. *Comparative Medicine* 53(2), 126-139
2. Manis JP 2007 Knock Out, Knock In, Knock Down — Genetically Manipulated Mice and the Nobel Prize. *N Engl J Med* 357, 2426-2429
3. Hogan B, Beddington R, Costantini F, Lacy E 1997 *Manipulating the Mouse Embryo*. Cold Spring Harbor Press: New York
4. www.jax.org/jaxmice
5. Lamartine J. Towards a new classification of ectodermal dysplasias 2003 *Clin Exp Dermatol*. 28:351-355
6. Thesleff I, Vaahtokari A, Partanen AM 1995 Regulation of organogenesis. Common molecular mechanisms regulating the development of teeth and other organs. *Int J Dev Biol* 39: 35-50
7. [Smahi A, Courtois G, Rabia SH, Doffinger R, Bodemer C, Munnich A, Casanova JL, Israel A.](#) 2002 The NF-kappaB signalling pathway in human diseases: from incontinentia pigmenti to ectodermal dysplasias and immune-deficiency syndromes. *Hum Mol Genet* 11(20):2371-2375
8. De Bosscher K, Vanden Berghe W, Haegeman G 2003 The Interplay between the Glucocorticoid Receptor and Nuclear Factor-kB or Activator Protein-1: Molecular Mechanisms for Gene Repression *Endocrine Reviews* 24(4):488-522
9. Pérez P, Page A, Bravo A, Del Rio M, Gimenez-Conti I, Budunova IV, Slaga TJ, Jorcano JL 2001 Altered skin development and impaired proliferative and inflammatory responses in transgenic mice overexpressing the glucocorticoid receptor. *FASEB J* 15:2030-20



4



4. De Genomas y Proteomas en la enfermedad genética

GENÓMICA COMPARATIVA DE GENES IMPLICADOS EN LA ENFERMEDAD DE PARKINSON

Ignacio Marín

Unidad de Bioinformática, Instituto de Biomedicina de Valencia, CSIC. Valencia.

¿Cuál es la utilidad de los datos evolutivos en biomedicina?

Si consideramos que nuestro interés principal es aportar nuevos conocimientos relevantes para paliar o curar una enfermedad genética humana, puede parecer que el estudio evolutivo de los genes o productos génicos implicados en dicha enfermedad será en general irrelevante. ¿Es posible que conocer si dichos genes están o no en una anémona nos pueda servir de algo?

La respuesta es que a menudo sí, dado que dichos estudios nos pueden proporcionar información significativa sobre tres puntos fundamentales:

- **Organismos modelo:** si conocemos que un gen, o incluso el sistema celular en el que el producto de dicho gen funciona, está conservado evolutivamente, podemos explorar las funciones de dicho gen o producto génico en organismos donde la experimentación sea más sencilla que en humanos. Para ello, debemos estar seguros de que los genes que consideramos son ortólogos y no parálogos, lo que comúnmente requerirá análisis filogenéticos.

- **Predicción de funciones:** es frecuente que la información acerca de la función de un gen o proteína humanos sea escasa o inexistente mientras que disponemos de información de genes o proteínas relacionados en otras especies. La pregunta es: ¿podemos esperar que dicha información sea útil? Pues dependerá de lo cercana que sea dicha relación, algo que nuevamente sólo podemos comprender mediante procedimientos comparativos. La información comparativa nos puede sugerir en qué módulos funcionales está implicado nuestro gen, con qué moléculas puede interaccionar la proteína que estudiamos, etc.

- **Predicción del efecto de mutaciones:** la comparación de secuencias entre especies puede permitir saber si una determinada mutación afecta a aminoácidos críticos para la función de la proteína o a nucleótidos críticos para la expresión correcta de un gen, que suelen estar conservados evolutivamente. Por otro lado, el modelado tridimensional a partir de proteínas de otras especies permite pronosticar cómo se pliega una proteína humana y así puede contribuir a discriminar si cierta mutación puede alterar dicho plegamiento.

¿Cuáles son las estrategias para obtener y comprender los datos comparativos?

La conjunción de experimentos dirigidos y de datos obtenidos mediante técnicas de alto rendimiento (secuenciación masiva, datos de *microarrays*, datos masivos de interacción proteína-proteína, etc.) genera una cantidad ingente de información. El problema es cómo utilizarla para nuestros fines. Para ello se requieren técnicas que reúnan dos condiciones: 1) que puedan manejar grandes cantidades de información; 2) que puedan extraer regularidades de dicha información. Aunque a menudo nos encontraremos con que deberemos generar nuestras propias herramientas, lo cierto es que hay ya un conjunto de métodos estandarizados, todos ellos computacionales, para realizar parte de los análisis. Algunos de los más importantes son:

- Métodos para manipular y comparar secuencias: incluyen poderosos editores que pueden ser utilizados para visualizar y manipular múltiples secuencias, programas que comparan secuencias entre sí, detectando si son similares o no, programas que permiten alinear múltiples secuencias, programas que detectan motivos comunes en múltiples secuencias, etc. Estos procedimientos se puede decir que están en gran parte estandarizados hoy en día, con sólo ocasionales mejoras. Sin embargo, es problemático usar un número de secuencias extremadamente grande. También es un problema no resuelto el cómo alinear eficientemente secuencias muy poco similares.
- Métodos filogenéticos: permiten establecer las relaciones evolutivas de las secuencias nucleotídicas o proteicas, utilizando los datos que proporcionan dichas secuencias para generar dendrogramas, árboles que muestran el parecido entre las secuencias. Estos árboles se pueden usar para establecer ortologías. Los métodos filogenéticos están en gran parte estandarizados, pero tienen limitaciones serias si el número de secuencias es muy grande. Tampoco es obvio en muchos casos cómo demostrar que las topologías encontradas son las correctas.
- Métodos para utilizar la información de interacciones proteína-proteína: podemos querer usar el conjunto de interacciones proteína-proteína conocidos en una especie (su *interactoma*) para deducir sus funciones o para pronosticar interacciones en humanos. Estos métodos están todavía en buena parte en desarrollo.
- Métodos para utilizar los datos de expresión generados con *microarrays*. Además de los datos humanos existentes, podemos asimismo utilizar datos de especies modelo. Los procedimientos están también en desarrollo, aunque hay ya algunos programas útiles para casos muy sencillos (por ejemplo, en algunas especies, establecer qué puede significar funcionalmente que un grupo de genes haya cambiado de expresión).
- Comparación de estructuras proteicas: comprender qué significan los dominios proteicos presentes en nuestra proteína por comparación con proteínas de otros organismos o modelar su estructura tridimensional a partir de datos de proteínas no humanas puede darnos pistas significativas acerca de su función, los efectos estructurales de mutaciones, etc. Estos métodos están bastante bien desarrollados hoy en día.

Un punto importante es que todos estos tipos de información deben luego integrarse a fin de proporcionar una visión conjunta de los resultados de los que se dispone. Dicha integración es uno de los aspectos más dificultosos, dado que en general, no existen procedimientos estandarizados para realizarla.

La enfermedad de Parkinson como ejemplo

Nuestro grupo lleva varios años concentrado en la caracterización de la evolución y patrones de diversificación de familias génicas que incluyen genes implicados en enfermedades humanas. Buena parte de nuestro trabajo se ha centrado en la enfermedad de Parkinson. Los motivos son dos. En primer lugar, es la segunda más frecuente enfermedad neurodegenerativa humana, tras la enfermedad de Alzheimer. En segundo lugar, está bien establecido que tiene un importante componente genético. Si bien es cierto que la mayor parte de los casos de Parkinson no parecen tener una causa obvia (clasificándose así como esporádicos, idiopáticos), se ha demostrado que existen casos de Parkinson familiar, asociados a mutaciones en diversos genes. Además, incluso los casos esporádicos tienen un componente genético: estudios en amplias poblaciones humanas han puesto de manifiesto que los individuos emparentados con alguien que sufre Parkinson esporádico tienen una probabilidad mayor de la esperada de sufrir también la enfermedad. Además, se ha comprobado recientemente que alteraciones en los genes implicados en Parkinson familiar pueden contribuir en ocasiones a la aparición de Parkinson esporádico.

¿Cómo pueden los datos comparativos contribuir a comprender una enfermedad tan estudiada? Pues, como ya se ha mencionado, proporcionan diversas informaciones que pueden ser de interés:

1. Hay genes implicados en Parkinson que son evolutivamente muy antiguos pero también hay otros de origen muy reciente. El conjunto de la información disponible sugiere que la mayoría de procesos celulares patológicos asociados a Parkinson no pueden estudiarse en modelos eucariotas unicelulares e incluso es imposible estudiar algunos de ellos en organismos modelo invertebrados, como *Drosophila melanogaster* o *Caenorhabditis elegans*, pues los genes implicados no existen en esos organismos. Sin embargo, algunos procesos sí podrían analizarse, y de hecho ya se están analizando con éxito, en modelos simples, tales como levaduras, *Dictyostelium* o invertebrados.
2. A menudo se puede comprender el impacto de mutaciones que afectan a las proteínas codificadas por los genes implicados en Parkinson familiar usando información de proteínas relacionadas de otros organismos.
3. La predicción de funciones es dificultosa, dada la escasa conservación evolutiva de parte del sistema, pero pueden obtenerse algunos datos relevantes al considerar combinadamente información en humanos, ratón, *Drosophila* y la levadura *Saccharomyces cerevisiae*.

Algunas referencias útiles

Introducirse en el campo de la genómica comparativa y la integración de datos aplicada a la biomedicina es dificultoso, pues la información está muy dispersa. Sin embargo hay algunos libros básicos de bioinformática que pueden ser un buen punto de partida, al explicar buena parte de los métodos básicos que deben conocerse. Entre ellos, destacaría estos:

J. Pevsner (2003) *Bioinformatics and functional genomics*. Wiley-Liss.

D. Baxevanis y B. F. F. Ouellette, eds. (2005) *Bioinformatics. A practical guide to the analysis of genes and proteins*. Tercera edición. Wiley-Interscience.

M. Campbell y L. J. Heyer (2007) *Discovering genomics, proteomics and bioinformatics*. Segunda edición. CSHL Press.

M. R. Barnes, ed. (2007) *Bioinformatics for geneticists*. Segunda edición. John Wiley and sons.

En cuanto a los métodos propiamente evolutivos, que tienen sus propias complicaciones, quizás las mejores introducciones son:

M. Nei y S. Kumar (2000) *Molecular evolution and phylogenetics*. Oxford University Press.

J. Felsenstein (2004) *Inferring phylogenies*. Sinauer Associates, Inc.

P. G. Higgs y T. K. Attwood (2005) *Bioinformatics and molecular evolution*. Blackwell publishing.

Hay innumerables revisiones que discuten la enfermedad de Parkinson y su componente genético. Una visión rápida y razonablemente completa se puede obtener aquí:

OMIM, Parkinson disease:

<http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=168600>

Finalmente, nuestros artículos sobre ésta y otras enfermedades pueden encontrarse en nuestra página web: <http://www.uv.es/~genomica/>

APLICACIÓN DE LA FARMACOGENÉTICA AL TRATAMIENTO DE PACIENTES PSIQUIÁTRICOS

Adrián Llerena

Centro de Investigación Clínica del Área de Salud de Badajoz (CICAB), SES, Universidad de Extremadura. Badajoz.

Cytochrome P450 (CYP) enzymes are involved in the metabolism of several clinically important drugs, as well as in the metabolism of endogenous substrates which so far is not well understood. Thus CYPs genetic polymorphism may play a role in the interindividual variability of drug response as well as in the vulnerability to certain psychiatric disorders.

1. Pharmacogenetics and vulnerability to psychiatric disorders.

Previously we reported the high frequency of **CYP2D6** Poor Metabolizers (PMs) among psychiatric patients.¹ **CYP2C9** is involved in the metabolism of several important psychoactive substances (tetrahydrocannabinol, amitriptyline, phenytoin, etc.). Moreover, it is known that CYP enzymes are expressed in the brain. We have shown that *CYP2C9*3* allele is related to the risk of major depressive disorder,^{2,3} and also that CYP2D6 PM phenotype is related to personality and the participation in clinical trials.^{4,5} We described the relationship of CYP2D6 activity and personality traits, suggesting that CYP2D6 enzyme may have an endogenous neuroactive substrate or product, such as a biogenic neurotransmitter amine.⁶ Later, it was reported that a serotonin metabolite seems to be a CYP2D6 substrate.⁷ It has also been reported that CYP2C9 activity is modulated by endogenous substrates such as adrenaline, serotonin and melatonin. In addition, a linkage between *CYP2C9* and *CYP2C8* genetic polymorphisms has been suggested.⁸ These enzymes are implicated in the metabolism of arachidonic acid, which may explain the relationship described with depression.⁹ In addition, we have recently found the implication of both *CYP2D6* and also *CYP2C9*, in the metabolism of fluoxetine to his active metabolite norfluoxetine.¹⁰

The frequency of *CYP2C9* and *CYP2D6* alleles among psychiatric patients suffering from **major depressive disorder** (n=70), or **schizophrenia** (n=89) and among another group of **healthy volunteers** (n=138) was also studied. We are show in a study a lower frequency of PMs in schizophrenic patients than in healthy volunteers supporting the hypothesis of a potential role of CYP2D6 in the vulnerability to schizophrenia.¹¹ In depressive patients, the frequency of the *CYP2C9*3* allele was higher (p<0.01) among the depressive patients than in the population of schizophrenic patients (odds ratio=3.3) and healthy volunteers (odds ratio=2.8).^{2,3}

2. Pharmacogenetics and variability to drug treatment.

CYP enzymes CYP2D6 and CYP2C9 play an important role in the metabolism of widely used psychotropic drugs¹². The genetic polymorphism of CYP enzymes might explain the interindividual variability on drug **plasma concentration** and side effects, such as the risk of **cardiotoxicity** due to *Torsade de Pointes* Type arrhythmia during treatment with antipsychotic drugs. In several studies we have analysed the relevance of CYP2D6 and CYP2C9 on plasma concentrations and **QTc interval lengthening** during treatment in clinical regular conditions. Psychiatric patients during antipsychotic treatment with haloperidol, thioridazine or risperidone monotherapy were studied.¹³ Depressive patients treated with fluoxetine monotherapy were also studied. Drug and metabolites plasma concentration, CYP2D6 phenotype, CYP2D6 and CYP2C9 genotypes and QTc interval lengthening were determined for each patient.

-Haloperidol: the dose-corrected plasma concentration was related to ($r=0.39$, $p<0.05$) CYP2D6 metabolic capacity (debrisoquine MR).^{14,15} 11% of the patients were identified as having a QTc interval longer than 456 msec, which may be considered as the threshold value for the risk of cardiac arrhythmia.¹⁶

-Thioridazine: the plasma concentration was correlated to ($r=0.51$, $p<0.001$) debrisoquine MR, and to the *CYP2D6* active genes ($r=-0.37$, $p<0.01$).^{17,18} QTc interval lengthening was correlated with the plasma concentration ($r=0.29$, $p<0.05$), and debrisoquine MR ($r=0.31$, $p<0.01$).¹⁹

-Risperidone: *CYP2D6* active genes were correlated to the dose-corrected plasma concentration ($p<0.05$), which is also related to debrisoquine MR ($r=0.79$, $p<0.001$).²⁰ *CYP2D6* active genes were related to QTc interval ($p<0.05$), and also to the dose-corrected plasma concentration ($p<0.05$), which is also correlated to debrisoquine MR ($r=0.79$, $p<0.001$).^{21,22}

-Fluoxetine: the dose corrected plasma concentrations ($p<0.01$, $r=-0.36$) and fluoxetine/norfluoxetine ratio were correlated ($p<0.01$, $r=-0.39$) to CYP2D6 genotypes.¹⁰ Among patients with two CYP2D6 active genes, dose corrected plasma concentrations of fluoxetine and active moiety were ($p<0.05$) higher in the *CYP2C9*1/*2* and *CYP2C9*1/*3* genotype groups compared to *CYP2C9*1/*1*.¹⁰

-Drug Interactions: We have also shown the effect of concomitant treatment with drugs metabolised by CYP2D6 on the plasma concentration of the studied antipsychotic drug (risperidone, haloperidol and thioridazine).¹²

-Effect of smoking: The relevance of smoking in the plasma concentration of thioridazine and the active metabolite mesoridazine has also been demonstrated.^{18,23}

3. Conclusions:

The present results support the potential implication of CYP2D6 and CYP2C9 in the vulnerability to psychiatric disorders,

The influence of CYP2D6 in the plasma concentration of haloperidol, thioridazine, risperidone and fluoxetine has been demonstrated. Thus, higher (among PMs) or lower (among Ultrarapids) plasma concentrations than expected may occur when patients are treated at regular recommended dose.

It has been demonstrated the relevance of CYP2D6 genetic polymorphism for the risk of cardiotoxicity (QTc interval lengthening) during treatment with antipsychotic drugs.

CYPs genotype and phenotype must be taken into consideration during treatment with several psychotropic drugs. Moreover, in patients with risk factors (i.e. cardiological disease, inherited long QT syndrome, elderly Poor or Ultrarapid metabolizers) cautious dosage and ECG monitoring are suggested due to the increased risk of developing cardiac arrhythmias.

References

1. Llerena A, Herraiz AG, Cobaleda J, Johansson I, Dahl ML. Debrisoquin and mephenytoin hydroxylation phenotypes and CYP2D6 genotype in patients treated with neuroleptic and antidepressant agents. *Clin Pharmacol Ther* 54:606-11. 1993.
2. Llerena A, Berecz R, Dorado P, Gonzalez AP, Penas-Lledo EM, De La Rubia A. CYP2C9 gene and susceptibility to major depressive disorder. *J Pharmacogen* 3:30-2. 2003.
3. Dorado P, Peñas-Lledó EM, González AP, Cáceres MC, Cobaleda J, Llerena A. Increased risk for major depression associated with the short allele of the serotonin transporter promoter region (5-HTTLPR-S) and the CYP2C9*3 allele. *Fundam Clin Pharmacol*. 21(4):451-3. 2007.
4. Llerena A, Edman G, Cobaleda J, Benitez J, Schalling D, Bertilsson L. Relationship between personality and debrisoquine hydroxylation capacity. Suggestion of an endogenous neuroactive substrate or product of the cytochrome P4502D6. *Act Psychiatr Scand* 87:23-8. 1993.
5. Llerena A, Cobaleda J, Benitez J. Debrisoquine hydroxylation phenotypes in healthy volunteers. *Lancet* (8651):1398. 1989.
6. Yu AM, Idle JR, Herraiz T, Kupfer A, Gonzalez FJ. Screening for endogenous substrates reveals that CYP2D6 is a 5-methoxyindolethylamine O-demethylase. *Pharmacogenetics* 13: 307-19. 2003.
7. Fradette C, Yamaguchi N, Du Souich P. 5-Hydroxytryptamine is biotransformed by CYP2C9, 2C19 and 2B6 to hydroxylamine, which is converted into nitric oxide. *Br J Pharmacol*. 141:407-14. 2004.
8. Yasar U, Lundgren S, Eliasson E, Bennet A, Wiman B, de Faire U, Rane A Linkage between the CYP2C8 and CYP2C9 genetic polymorphisms. *Biochem Biophys Res Commun*. 22;299:25-8. 2002.
9. Adams PB, Lawson S, Sanigorski A, Sinclair AJ. Arachidonic acid to eicosapentaenoic acid ratio in blood correlates positively with clinical symptoms of depression. *Lipids*. 31:S157-61. 1996.
10. Llerena A, Dorado P, Berecz R, Gonzalez AP, Penas-Lledo EM. Effect of CYP2D6 and CYP2C9 genotypes on fluoxetine and norfluoxetine plasma concentrations during steady-state conditions. *Eur J Clin Pharmacol* 59:869-73. 2004.
11. Llerena A, Dorado P, Peñas-Lledó EM, Cáceres MC, De la Rubia A. Low frequency of CYP2D6 poor metabolizers among schizophrenia patients. *Pharmacogenomics J*. 7(6):408-10. 2007.

12. Dorado P, Berecz R, Peñas-Lledó EM, Cáceres MC, Llerena A. Clinical implications of CYP2D6 genetic polymorphism during treatment with antipsychotic drugs. *Curr Drug Targets*. 7(12):1671-80. 2006.
13. Dorado P, Berecz R, Peñas-Lledó EM, Llerena A. Antipsychotic drugs and QTc prolongation: the potential role of CYP2D6 genetic polymorphism. *Expert Opin Drug Metab Toxicol*. 3(1):9-19. 2007.
14. Bertilsson L, Dahl ML, Ekqvist B, Llerena A. Disposition of the neuroleptics perphenazine, zuclopenthixol, and haloperidol cosegregates with polymorphic debrisoquine hydroxylation. *Psychopharmacol Ser* 10:230-7. 1993.
15. Llerena A, de la Rubia A, Berecz R, Dorado P. Relationship between haloperidol plasma concentration, debrisoquine metabolic ratio, CYP2D6 and CYP2C9 genotypes in psychiatric patients. *Pharmacopsychiatry* 37:69-73. 2004.
16. Llerena A, Berecz R, de la Rubia A, Dorado P. QTc interval lengthening and debrisoquine metabolic ratio in psychiatric patients treated with oral haloperidol monotherapy. *Eur J Clin Pharmacol* 58: 223-4. 2002.
17. Llerena A, Berecz R, de la Rubia A, Fernandez-Salguero P, Dorado P. Effect of thioridazine dosage on the debrisoquine hydroxylation phenotype in psychiatric patients with different CYP2D6 genotypes. *Ther Drug Monit* 23:616-20. 2001.
18. Berecz R, De La Rubia A, Dorado P, Fernandez-Salguero P, Dahl ML, Llerena A. Thioridazine steady-state plasma concentration are influenced by tobacco smoking and CYP2D6, but not to CYP2C9 genotype. *Eur J clin Pharmacol* 59:45-50. 2003.
19. Llerena A, Berecz R, De La Rubia A, Dorado P. QTc interval lengthening is related to CYP2D6 hydroxylation capacity and plasma concentration of thioridazine in patients. *J Psychopharmacol* 16: 361-4. 2002.
20. Berecz R, Llerena A, De La Rubia A, Gomez J, Kellermann M, Dorado P, Degrell I. Relationship between risperidone and 9-hydroxy-risperidone plasma concentrations and CYP2D6 enzyme activity in psychiatric patients. *Pharmacopsychiatry* 35:231-4. 2002.
21. Llerena A, Berecz R, Dorado P, De La Rubia A. QTc interval, CYP2D6 and CYP2C9 genotypes, and risperidone plasma concentrations in patients. *J Psychopharmacol* 18:189-93. 2004.
22. Llerena A, Berecz R, Dorado P, de la Rubia A. QTc interval, CYP2D6 and CYP2C9 genotypes and risperidone plasma concentrations. *J Psychopharmacol*. 18(2):189-93. 2004.
23. Dorado P, Berecz R, Peñas-Lledó EM, de la Rubia A, Llerena A. No effect of the CYP1A2*1F genotype on thioridazine, mesoridazine, sulforidazine plasma concentrations in psychiatric patients. *Eur J Clin Pharmacol*. 63(5):527-8. 2007.

FUNCTIONAL PREDICTION OF NSSNPS IN THE HUMAN GENOME

Hernán J. Dopazo

Unidad de Farmacogenómica y Genómica Comparativa, Departamento de Bioinformática, Centro de Investigación Príncipe Felipe (CIPF). Valencia.

Since the earlier works of J. B. S. Haldane on sickle-cell anemia, biologists recognize the power of natural selection on genetic variation and its association to human diseases. Further developments demonstrated that most of the genetic changes occurring in a population do not affect the phenotype, or more accurately, the reproductive capacity (fitness) of the genotypes carrying genetic variants (Kimura, 1983). Recently 3.1 million single nucleotide polymorphisms (SNPs) were found in the human genome (IHM, 2007) and a major goal on biomedical research is to understand the role of the common genetic variants in susceptibility to common diseases in human populations.

A world wide survey on the genetic variation in genes associated to common human diseases concluded that SNPs occur at a frequency of 1 out of 346 bp and are roughly equally divided between synonymous and nonsynonymous changes (Cargill, et al, 1999). As approximately, two-thirds of random mutations in coding sequences alter an amino acid, the fact that nsSNPs compromise one half the total SNPs, implies strong selection against amino acid altering changes. The force of selection is also evident when comparing nsSNPs causing non-conservative amino acid substitutions with those causing a conservative change. Non-conservative nsSNPs represent only 36% of all nsSNPs, whereas randomly distributed mutations would be expected to produce a higher proportion (52%) of non-conservative changes (Cargill, et al, 1999). Currently, the NCBI SNP database (dbSNP, build 127) collects 5,689,286 validated human SNPs out of which 78,845 are nonsynonymous coding SNPs (nsSNPs). That means that about only 1% of the human validated SNPs could probably affect gene function. One of the most important questions in human genetics is to deduce which of these genetic variants are functionally relevant for human health. In other words, which of these 1% of genetic variants are targets of selective or neutral evolutionary processes in the human genome. Far from the theoretical interest of this enquire, this prediction would help in the genotyping process of SNPs probably associated to disease in classical genotype-phenotype association studies in human populations.

The earliest studies on the functional prediction of human nsSNPs were pioneered by Sunyaev et al. (2000, 2001), Chasman & Adams, (2001), Wang & Moutl (2001), Miller & Kumar (2001), Saunders & Baker (2002) and Santibañez Koref, et al (2003).

Sunyaev et al. (2000) estimated that around 70% of disease-causing mutations occur at structurally and functionally important sites with well defined properties such as less than 5% of solvent accessibility, sometimes located in beta-strands, active sites, disulphide bonds, or

evolutionary conserved sites. Moreover, they found that most of the allelic variants map to the same structurally and functionally important regions of the proteins suggesting that many of them probably have negative effects on the phenotype. In a subsequent study Sunayev et al. (2001) estimated that around 20% of the human nsSNPs affect protein function and that an average human genotype carries about 2,000 of such nsSNPs. They observed that of the majority of disease-causing mutations were at low frequencies in human populations (1% - 20%), which was considered as a validation of their method.

Wang & Moulton (2001) found that by far the largest proportion (83%) of disease-nsSNPs affects protein stability, 5% maps on binding sites, and approximately 10% correspond to cases where their 3D structural model gives a false negative result. They predicted that 70% of nsSNPs studied in hypertension, cardiovascular, endocrinology, and neuropsychiatric diseases correspond to cases of neutral polymorphisms while the remaining 30% affect the stability of the protein. Alternatively, Chasman & Adams (2001) used a combination of statistical methods to define structural and evolutionary parameters with significant association to disease. From the knowledge of the effects of about 6,000 mutations from the Lac repressor and the T4 lysozyme protein they estimated that around 26-32% of nsSNPs have deleterious effects on human protein function.

While most of these studies mainly focused on structural parameters of proteins, Miller & Kumar (2001) explicitly studied the role of the evolutionary conservation in the functional prediction of nsSNPs emphasizing the risk of using concepts like conservation profile and similarity cutoff percentage values used in the previous models. They pointed out that evolutionary data cannot be treated as independent observations for use in statistics because they share a non-random structure of dependence defined in the historical relationships of the species. That is, model approaches based on similarity could overestimate the variability of a given site if an identical residue appears in multiple species due to phylogenetic constraints. Therefore, using an explicit method of phylogenetic reconstruction on 7 human disease proteins, Miller & Kumar (2001) demonstrated that human nsSNPs mutations are over-abundant at amino acid positions most conserved throughout the long-term history of metazoans. Human polymorphic replacement mutations and silent mutations were found randomly distributed across sites with respect to the level of conservation of amino acid sites within genes. They concluded that disease-causing amino acid changes are those that are not observed among species probably because they are not accepted by natural selection in long term evolutionary time.

In the same vein, an explicit statistical phylogenetic model was developed by Santibañez Koref, et al (2003). The method indicates the probability of a given mutation being pathological considering the evolutionary conservation and the variability associated to each protein. Although the method they developed was outstanding in the fields of comparative genomic and human health, the necessary calibration of the model for each human protein is a major drawback for its use in large scale analysis.

Saunders & Baker (2002) evaluated the behaviour of alternative variables in the functional prediction of nsSNPs. When using a combination of evolutionary and structural variables, they concluded that the prediction is better than when - a single kind of variable is used on its own. When fewer than five to ten homologs are available, they emphasized that the prediction of deleterious mutation should include structural information, suggesting that the evolutionary data is more informative than the structural data when a high number of sequences are used for prediction.

Finally, Arbiza et al, (2006) introduced, and Capriotti et al, (2007) extend, an explicit evolutionary measure of selective pressures at a codon level as a direct functional predictor of nsSNPs. Using maximum likelihood (ML) models to estimate the well established ratio (?) between nonsynonymous to synonymous rates of evolution in mammals, they concluded that codons with $\omega < 0.1$ maps residues where mutations producing disease are frequent in human.

In this talk, I will overview many of the main methods developed to predict the functional properties of nsSNPs in the human genome. I will insist that most of these methods are based on features derived from protein structures and/or the evolutionary conservation which represent proxies in order to infer its cost on fitness. Next, I will focus on the measure and description of the selective constraints associated to nsSNPs in the human genome as a direct way to search for selective pressures acting at the codon level. We will see that this evolutionary approach based on the maximum likelihood estimation of evolutionary rates evaluates the fitness of each nsSNPs at the codon level at the same time that improves previous attempts to differentiate deleterious alleles from neutral polymorphisms in the human genome. Finally I will present the PupaSNPs suite tool (Conde L., et al. 2006, <http://www.pupasnp.org>) where all the predictions for nsSNPs on coding sequence are collected for the human genome.

References

- Arbiza, L., et al. 2006. Selective pressures at a codon-level predict deleterious mutations in human disease genes. *J. Mol. Biol.* 358, 1390-1404.
- Cargill, M. et al. 1999. Characterization of single nucleotide polymorphisms in coding regions of human genes. *Nature Genet.* 22, 231-238.
- Capriotti, E., et al. 2007. Use of estimated evolutionary strength at the codon level improves the prediction of disease-related protein mutations in human. *Human Mutation* (in press).
- Chasman, D. and R. M. Adams. 2001. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. *J Mol Biol* 307(2): 683-706.
- Conde L., et al. 2006. PupaSuite: Finding functional SNPs for large-scale genotyping purposes. *Nucleic Acids Res.* 34: W621-5.
- IHMC, 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449 (7164):851-61.
- Kimura, M. 1983. *The neutral theory of molecular evolution*. Cambridge Univ. Press, NY.
- Miller, M. P. and S. Kumar. 2001. Understanding human disease mutations through the use of interspecific genetic variation. *Hum. Mol. Genet.* 10, 2319-2328.
- Santibanez Koref, M. F., et al 2003. A phylogenetic approach to assessing the significance of missense mutations in disease genes. *Hum. Mutat.* 22, 51-58.
- Saunders, C. T. and Baker D. 2002. Evaluation of structural and evolutionary contributions to deleterious mutation prediction. *J Mol Biol.*322(4):891-901.
- Sunyaev S. R. et al. 2000. SNP frequencies in human genes an excess of rare alleles and differing modes of selection. *Trends Genet.* 16(8):335-7.

Sunyaev, S. et al. 2001. Prediction of deleterious human alleles. *Hum. Mol. Gen.*, 10: 591-597.

Wang, Z. and J. Moulton. 2001. SNPs, Protein Structure, and Disease. *Hum Mut* 17:236-270

Further reading

Ng & Henikoff (2006). Predicting the effects of amino Acid substitutions on protein function. *Annu Rev Genomics Hum Genet.* 2006; 7:61-80

Dopazo, H. (2008). Selective Constraints on Human Disease Mutations and Polymorphisms. In *Handbook of Human Molecular Evolution*. D. N. Cooper Ed. Wiley & Sons Ltd. UK. In press.

BIOLOGÍA ESTRUCTURAL Y GENÉTICA HUMANA

Vicente Rubio

Laboratorio de Enzimopatología Estructural. Instituto de Biomedicina de Valencia, CSIC y CIBER de Enfermedades Raras (grupo 7-039)

La biología estructural y la genética humana están relacionadas desde hace ya mucho tiempo [1]. La hemoglobina es la primera proteína para la que se definió una enfermedad molecular, la anemia falciforme, y también una de las dos primeras para las que se determinó su estructura, por lo que desde el principio del conocimiento molecular de esta enfermedad, la misma se asoció al cambio estructural. Sin embargo, durante mucho tiempo la obtención de nuevas estructuras era un proceso extremadamente costoso, lo que hizo de la anemia falciforme una mera prueba de principio de la relación entre genética (en realidad entre enfermedad genética) y estructura. Sin embargo, hoy las condiciones han cambiado radicalmente, ya que el ritmo de avance de la biología estructural ha aumentado extraordinariamente, y en ~10% de los trabajos publicados en revistas prestigiosas de genética se hace referencia a datos estructurales [1].

La técnica básica para la determinación de estructuras macromoleculares a alta resolución sigue siendo la cristalografía de proteínas. A ella se ha unido la resonancia magnética nuclear, principalmente para macromoléculas de masa reducida, aunque cada vez dentro de un rango de tamaño más elevado. La microscopía electrónica de macromoléculas provee datos estructurales de baja resolución, aunque progresivamente se va acercando a los límites de la cristalografía [2].

En cristalografía se han producido avances extraordinarios desde los años 90 del siglo pasado (revisados en [2]). La introducción de la cristalografía a baja temperatura (“criocristalografía”) ha sido un salto de gigante. También lo ha sido la producción de proteínas recombinantes, que ha facilitado extraordinariamente la generación de grandes cantidades de la proteína de interés, así como su modificación, por ejemplo introduciendo mutaciones que mejoran su cristalizabilidad, incorporando etiquetas que facilitan la purificación, o incorporando selenometionina (en vez de metionina) para resolver el problema de la determinación de las fases, clave en cristalografía, usando la señal anómala del selenio. También ha sido muy importante [2] la introducción de matrices comerciales para la exploración de las condiciones de cristalización, el aumento de la potencia de los ordenadores, el desarrollo de programas para el procesamiento de datos, para el cálculo de las fases y para la obtención del mapa de densidad electrónica, así como para el trazado del modelo molecular (la estructura es el modelo molecular construido dentro de la envoltura de electrones que constituye el mapa de densidad electrónica; los electrones difractan los rayos X). Finalmente, la generalización del uso de sincrotrones como fuentes de rayos X ha mejorado la productividad y las resoluciones alcanzadas.

No hay que olvidar los avances que han venido de la mano del conocimiento genómico. El conocimiento de genomas completos de muchos organismos ha puesto al alcance de los biólogos estructurales la clonación a voluntad de los genes deseados y la expresión heteróloga de muchos de dichos genes, permitiendo plantear desde una perspectiva quizá visionaria lo que se ha dado en llamar “genómica estructural” [3], idealmente la determinación de las estructuras de todos los productos génicos de un organismo, o, más realísticamente, la clasificación de todos los productos génicos en superfamilias, familias y subfamilias estructurales, determinando la estructura de al menos un elemento de cada subfamilia. Un beneficio derivado ya de esta aproximación ha sido el desarrollo de instrumentos que permiten la determinación estructural masiva con alta eficacia (“high throughput”): automatización de la clonación, expresión génica y purificación del producto, exploración masiva de condiciones de cristalización (“granjas de cristales”) mediante robotización de las pruebas de cristalización y del examen de los posibles cristales, automatización de la manipulación de muestras para la difracción, del procesamiento de datos de difracción, del cálculo de las fases y del trazado del modelo molecular. Aunque el proceso entero es en general más un desideratum que una realidad para la mayoría de macromoléculas sometidas al mismo, hoy lo usual es que los laboratorios cristalográficos hayan puesto en marcha el proceso de robotización para la preparación e identificación de los cristales. Otra consecuencia de las iniciativas de genómica estructural es el gran aumento de estructuras conocidas (ver el Protein DataBank, <http://www.rcsb.org/pdb>), que hace frecuente el poder calcular las fases de nuestro cristal por similitud con una estructura conocida (“reemplazo molecular”), y que ha aumentado las posibilidades de modelización, cada vez con más precisión, de estructuras (Swiss-Model en <http://www.expasy.ch>) de productos génicos no cristalizados pero con un grado razonable de identidad de secuencia con proteínas de estructura conocida. Así, aunque no somos capaces de predecir la estructura “ab initio” desde la secuencia, si podemos hacerlo por homología, existiendo instrumentos bioinformáticos que realizan este trabajo de forma automática para genomas enteros (ModBase; <http://modbase.compbio.ucsf.edu>).

Por tanto, cada vez es más común que dispongamos de conocimiento estructural o que podamos inferir datos estructurales con relativamente poco esfuerzo para los productos génicos en que podamos estar interesados en nuestro quehacer genético. ¿Qué valor tiene este conocimiento?

En primer lugar, este conocimiento tiene *valor racionalizador* [1]. Entendemos muy bien por qué la mutación Glu6Val en la cadena beta de la hemoglobina desoxigenada causa la formación de polímeros ordenados que son la base de la patología en la anemia falciforme. En este caso se trata en realidad de una enfermedad que podríamos llamar de plegamiento: un cambio de residuo causa un cambio en la tendencia de la molécula de hemoglobina a agregar. El conocimiento del mecanismo estructural del defecto sugiere formas de tratarlo. Así, la carbamilación de la hemoglobina S, que afecta a lisinas de la superficie, previene la polimerización. Del mismo modo, la inducción de la expresión de hemoglobina fetal disminuye la probabilidad de polimerización extensa de la hemoglobina S a pesar de que dicha hemoglobina siga siendo mayoritaria en los homocigotos, ya que, así como las impurezas disminuyen el tamaño de los cristales de proteína o incluso previenen completamente la cristalización, la heterogeneidad molecular debe prevenir la formación de polímeros de gran tamaño de hemoglobina S. Igualmente, la estructura de enzimas y de sus centros activos hacen posible entender los cambios inducidos por mutaciones clínicas sobre, por ejemplo, la actividad del enzima o la cinética del mismo. Cuando estas mutaciones afectan a sitios para cofactores o para activadores que pueden administrarse como medicamentos, es posible utilizar la información estructural para predecir si la terapia sustitutiva o megavitamínica va a ser eficaz.

El conocimiento estructural puede tener también *valor diagnóstico*, en el sentido de hacer posible la discriminación sobre la base de la estructura entre mutaciones patogénicas y polimorfismos triviales [1,4,5]. Recientemente hemos insistido en el valor de esta aproximación en el déficit de ornitina transcarbamilasa [5], el déficit más común del ciclo de la urea, en el que se puede predecir a partir de cálculos simples que no más de un 25% de las mutaciones de cambio de sentido son patogénicas. Es cierto que la expresión “in vitro” del enzima mutado y el ensayo de su estabilidad y de su actividad pueden aclarar experimentalmente las consecuencias del cambio de aminoácido. Sin embargo, cuando como en el caso del déficit de ornitina transcarbamilasa, las mutaciones no muestran recurrencia significativa, siendo muy numerosas las descritas, la aproximación basada en los estudios de expresión, aunque factible, es tediosa, exigente y cara, siendo mucho más simple una aproximación “in silico” basada en las predicciones derivadas del conocimiento estructural.

También tienen *valor* en genética humana, en el contexto de los dos epígrafes anteriores, *las estructuras de proteínas bacterianas, de arqueas o de plantas*, homólogas a proteínas humanas. Con frecuencia la producción y expresión recombinante de proteínas animales o humanas plantea gravísimas dificultades que hacen imposible tanto la introducción en las mismas de mutaciones clínicas para determinar sus efectos, como la determinación estructural. Estos problemas pueden resolverse muchas veces usando la proteína equivalente bacteriana o de arqueas (preferentemente termófilas, ya que las proteínas termo-resistentes exhiben menor flexibilidad y suelen cristalizar mejor). Estas proteínas pueden representar excelentes modelos de proteínas humanas causantes de errores congénitos, como hemos podido demostrar para carbamil fosfato sintetasa I y, más recientemente, para glutamato 5-quinasa (uno de los dos componentes del enzima humano pirrolin 5-carboxilato sintetasa) [6, 7]. Incluso la determinación de la estructura de un enzima homólogo bacteriano que cataliza una reacción distinta ha hecho posible elaborar el primer modelo estructural de acetilglutamato sintasa, otro enzima causal de hiperamonemia congénita en el ser humano [8].

El conocimiento de la estructura puede *arrojar luz sobre la función de productos génicos identificados en el estudio genético* [9] Desde la introducción de la clonación posicional es común identificar genes causantes de patología de cuya función no sabemos nada. En estos casos la determinación experimental (o, en su defecto, la modelización) de la estructura del producto génico puede suministrar claves importantes para explorar experimentalmente la función de dicho producto, como en el caso de la proteína DJ-1 [10], cuyas mutaciones se asocian a enfermedad de Parkinson. Hay que decir, además, que las similitudes estructurales pueden ser un valioso elemento para la anotación de nuevos genes en proyectos de secuenciación completa de genomas, ya que es frecuente la conservación de estructura en un grado mucho mayor que la conservación de secuencia.

Junto a estos usos ya consolidados de la biología estructural en el ámbito de la valoración de la función y de las disfunciones de los productos génicos, hay que considerar también la *necesidad de más conocimiento estructural* del aparato genético mismo, *para entender los efectos epigenéticos* [1] *así como los de mutaciones por expansión trinucleotídica* en regiones no codificantes (y también quizá en regiones codificantes) [11]. En la distrofia miotónica 1 y en el síndrome del X frágil se ha observado la participación de RNA no codificante, de factores remodeladores de la cromatina y de modificaciones histónicas, además de la repetición trinucleotídica y la metilación de CpGs, en la determinación de la estructura de la cromatina en los loci correspondientes. Dado el impacto de la estructura de la cromatina sobre la transcripción, y de ésta sobre los síntomas de enferme-

dad en las expansiones trinucleotídicas no codificantes, es evidente la necesidad de un conocimiento estructural detallado de la cromatina en la región del gen causante de patología, para lograr un entendimiento completo de la expresión génica anormal en estas enfermedades [1, 11].

En resumen, cada vez va a ser más difícil prescindir de la biología estructural en el quehacer diario en genética, sea humana, médica o de otros organismos.

Referencias

1. Laskowski RA, Thornton JM. (2007) Understanding the molecular machinery of genetics through 3D structures. *Nat Rev Genet.* 14:1-11.
2. Rubio V (2003) La cristalización de proteínas. En *Estructura de Proteínas* (C Gómez-Moreno y J Sancho Sanz, eds.). Editorial Ariel SA, Barcelona, pp. 233-252.
3. Editorial (2007) Looking ahead with structural genomics. *Nat Struct Mol Biol.* 14:1.
4. Ng PC, Henikoff S. (2006) Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet.* 2006;7:61-80.
5. Arranz JA, Riudor E, Marco-Marín C, Rubio V. (2007) Estimation of the total number of disease-causing mutations in ornithine transcarbamylase (OTC) deficiency. Value of the OTC structure in predicting a mutation pathogenic potential. *J Inherit Metab Dis.* 30:217-226.
6. Yefimenko I, Fresquet V, Marco-Marín C, Rubio V, Cervera J. (2005) Understanding carbamoyl phosphate synthetase deficiency: impact of clinical mutations on enzyme functionality. *J Mol Biol.* 349:127-141.
7. Marco-Marín C, Gil-Ortiz F, Pérez-Arellano I, Cervera J, Fita I, Rubio V. (2007) A novel two-domain architecture within the amino acid kinase enzyme family revealed by the crystal structure of *Escherichia coli* glutamate 5-kinase. *J Mol Biol.* 367:1431-1446.
8. Ramón-Maiques S, Fernández-Murga ML, Gil-Ortiz F, Vagin A, Fita I, Rubio V. (2006) Structural bases of feed-back control of arginine biosynthesis, revealed by the structures of two hexameric N-acetylglutamate kinases, from *Thermotoga maritima* and *Pseudomonas aeruginosa*. *J Mol Biol* 356:695-713.
9. Lee D, Redfern O, Orengo C. (2007) Predicting protein function from sequence and structure. *Nat Rev Mol Cell Biol.* 8:995-1005.
10. Lee SJ, Kim SJ, Kim IK, Ko J, Jeong CS, Kim GH, Park C, Kang SO, Suh PG, Lee HS, Cha SS. (2003) Crystal structures of human DJ-1 and *Escherichia coli* Hsp31, which share an evolutionarily conserved domain. *J Biol Chem.* 278:44552-44559.
11. Wang YH. (2007) Chromatin structure of repeating CTG/CAG and CGG/CCG sequences in human disease. *Front Biosci.* 12:4731-4741.

TECNOLOGÍAS ACTUALES EN LOS LABORATORIOS DE GENÉTICA CLÍNICA

F. Javier Chaves

Laboratorio de Estudios Genéticos, Fundación para la Investigación del Hospital Clínico Universitario de Valencia.

Los laboratorios para el estudio genético de pacientes han sufrido una revolución en los últimos años basadas en numerosas mejoras técnicas y desarrollo del conocimiento del genoma humano (Venter et al, 2001). Estos desarrollos han permitido pasar de analizar unos pocos estudios genéticos al día o la semana a realizar miles o millones en un día. Estos cambios vienen del desarrollo de numerosas metodologías que aprovechan diferentes características del ADN y numerosas técnicas para su estudio. Algunos ejemplos pueden ser el aumento de la capacidad de los sistemas de PCR, procesado de muestras, de secuenciación, desarrollo de los sistemas de PCR cuantitativo a tiempo real y de los programas para el análisis de los resultados obtenidos.

En los últimos años también se han incorporado diferentes sistemas de análisis de microchips que permiten analizar miles o millones de sondas simultáneamente y, por lo tanto, el análisis de gran número de mutaciones y polimorfismos, alteraciones cromosómicas o el estudio de los niveles de ARNm de miles de genes, pudiéndose analizar simultáneamente los niveles de ARNm de todos los genes conocidos (Heller et al, 2002). Finalmente, uno de los mayores avances ha sido el desarrollo de sistemas comerciales de secuenciación masiva de ADN. Estos sistemas tienen pueden analizar en un solo experimento millones de bases de la secuencia de ADN. Esta habilidad permite obtener gran cantidad de información de un solo individuo, de mezclas de ácidos nucleicos, identificación de mutaciones en baja proporción de la muestra, secuenciación de genes completos y de numerosos fragmentos en gran cantidad de individuos. Finalmente, este sistema nos permite la identificación de posibles patógenos en una muestra, dado que no se necesita conocer la secuencia de lo que se quiere analizar (ver información en www.roche-applied-science.com o www.appliedbiosystems.com).

En conclusión, hay numerosas herramientas para el estudio de los ácidos nucleicos que permiten abordar estudios de gran envergadura con numerosas muestras o analizar gran cantidad de datos en cada muestra.

Referencias

Venter JC, et al. The sequence of the human genome. Science. 2001;291(5507):1304-51.

Heller MJ. DNA microarray technology: devices, systems, and applications. Annu Rev Biomed Eng. 2002;4:129-53.

