



# genoma humano

El siglo xx estuvo marcado por el desarrollo tecnológico en varias áreas del conocimiento, especialmente en la física, lo cual se vio reflejado en la conquista del espacio y en otras áreas como las comunicaciones (con la creación de materiales superconductores). Precisamente en la transición del siglo xx al XXI, un evento capturó la atención del mundo: el anuncio de la completa secuenciación del genoma humano en junio de 2000, cuya importancia científica es equivalente a la llegada del hombre a la Luna. Así como las famosas palabras pronunciadas por Neil Armstrong “un pequeño paso para el hombre pero un gran paso para la humanidad” simbolizan el enorme avance tecnológico y filosófico que este viaje representó para el ser humano, de la misma manera la descripción de la serie de letras que representan al genoma humano está abriendo paso a una nueva era para la ciencia. Pero, ¿realmente se ha terminado de entender la genética humana?

Se denomina genoma a toda la información genética que un organismo posee, la cual está contenida en el ADN o ácido desoxirribonucleico, cuyos pilares químicos fundamentales son las bases conocidas como adenina, guanina, timina y citosina (A, G, T y C, respectivamente). Éstas se encuentran aparejadas A con T y G con C, de esta manera series sucesivas de ellas conforman el material genético de un organismo; por tanto, las unidades utilizadas para describir el tamaño del genoma son pares de bases. Este material genético está organizado en cromosomas, cada molécula de ADN que los conforma debe poseer tres elementos importantes para su posterior propagación: un centrómero, que es el ancla del cromosoma al uso que se forma cuando la célula se está dividiendo, esto es la mitosis; dos telómeros, que son los componentes de las puntas distales del cromosoma, y varios orígenes de re-

plicación, a partir de los cuales se inicia la replicación de los cromosomas. Al igual que en química la unidad fundamental de la materia es el átomo, de la misma forma la unidad mínima del genoma es el gen.

Los cimientos del Proyecto Genoma Humano se fundan en 1977, año en que Sanger (premio Nobel de Química en 1958) diseñó una técnica que hizo posible la secuenciación completa del genoma de un virus. Este hecho marcó el inicio de una serie de descubrimientos, la implementación de nuevas técnicas y el desarrollo de pro-



El tamaño total del genoma nuclear humano estimado es de 3.2 gigas de bases (algo así como 3 200 000 000 pares de bases) del cual 2.5 gigas de bases es eucromatina y el resto está compuesto por heterocromatina. Es veinticinco veces más grande que cualquier genoma secuenciado hasta ahora y ocho veces mayor que la suma de todos ellos. Otro dato curioso es que el *Homo sapiens* tiene un genoma doscientas veces más grande que el de la levadura, pero doscientas más pequeño que el de la ameba *Amoeba dubia*.

Sólo de 1.1 a 1.4% es secuencia que realmente codifica en proteínas (el ADN que finalmente es traducido a proteínas), 24% equivale a regiones no codificantes que se encuentran entre las codificantes y 75% corresponde a regiones intergénicas. Se ha calculado que en total las secuencias codificantes comprenderán 5% del genoma,



mientras que las secuencias repetidas serán al menos 50%. La especie humana posee 22 cromosomas sexuales (x,y), el más grande es el 2 y el más pequeño es el y. Los cromosomas 17, 19 y 22 son los que poseen una mayor densidad de genes, y los de menor son el x, 4, 18, 13 y el y. En cuanto a la estructura de los pares de bases el porcentaje de guanina y citosina es de 41% en promedio, pero varía en distintas regiones; el más alto registrado es de 49% y el más bajo de

36%. Regiones ricas en guanina y citosina tienden a ser las de mayor densidad génica, en cambio las ricas en adenina y timina poseen pocos genes. Por ejemplo, el porcentaje de guanina y citosina en el cromosoma 19 es de 49%, lo que coincide en que este es el cromosoma que posee la mayor densidad génica.

Las tasas de recombinación se incrementan al disminuir el tamaño del brazo del cromosoma, y tienden a bajar o suprimirse cerca de los centrómeros. Al realizar comparaciones entre secuencias asociadas al cromosoma x y al y entre humanos y otros grupos de primates, se ha visto que la tasa de mutación o cambio es 2.1 veces mayor en el y que en el x. Se ha propuesto que esto se debe a que en la línea germinal masculina se llevan a cabo un mayor número de divisiones y diferentes mecanismos de reparación del ADN.

### Algunas características del genoma humano

gramas de análisis computarizado, que culminaron en la secuenciación del genoma del primer organismo vivo en 1995: la bacteria *Haemophilus influenza*, responsable de la enfermedad de la influenza, con un tamaño de genoma de 1 800 000 pares de bases. La conjunción de todos estos avances tanto en las técnicas de biología molecular como en la automatización de los análisis, generaron la aparición de una nueva ciencia: la genómica. Esta nueva área es producto de la unión de dos disciplinas diferentes, la computación y la biología. Dicha unión ha tenido un impacto impresionante en el desarrollo de nuevas disciplinas que han traído consigo grandes avances, y que sobre todo favorecen el futuro desarrollo de la medicina. La secuenciación completa del genoma podrá tener beneficios incalculables para la población humana.

Para llevar a cabo este proyecto, debido a la complejidad del genoma humano (ya que éste es muy grande y posee secuencias que se repiten mucho, lo cual compli-

ca su análisis), fue necesario analizar otros organismos "modelo", con genomas más sencillos, que permitieran afinar las técnicas que se usan en la secuenciación completa. De esta forma, en 1995 el proyecto progresaba en dos etapas; la primera fue la construcción del mapa físico (estudio de genética "clásica" que permite ubicar la posición de los genes dentro de los cromosomas) y genético (el análisis propiamente dicho de las secuencias y el orden fino de los genes) del genoma humano y del ratón. La segunda etapa fue la secuenciación de los genomas de la levadura *Saccharomyces cerevisiae*, responsable de la producción de la cerveza, y el nemátodo *Caenorhabditis elegans*, una especie de gusano pequeño, así como la secuenciación de regiones específicas de genomas de mamíferos. Lo que estos proyectos demostraron fue que era posible secuenciar a gran escala con efectividad, rapidez y a costos reducidos. Dos grupos independientes secuenciaron el genoma humano, uno estaba constituido por di-

versos grupos internacionales y se denominó Consorcio internacional, el otro estaba conformado por la compañía Celera, con el controversial y agresivo investigador Craig Venter a la cabeza.

La obtención del ADN, necesario para la secuenciación del genoma humano, fue hecha, en el caso del Consorcio internacional, por donadores anónimos (cinco a diez personas). Celera escogió cinco personas de categorías étnicas definidas: un individuo afroamericano, uno chino, uno mexicano y dos caucásicos. En total se usaron muestras de tres mujeres y dos hombres, todos donadores anónimos. La obtención de la muestra se realizó a partir de sangre (130 mililitros), y en el caso de los hombres incluía cinco muestras de semen tomadas en un periodo de seis semanas.

La descripción de la secuencia no es sino el principio de una serie de análisis conducidos para determinar el número de genes que están contenidos (genoma), su posición dentro de los cromosomas (mapas físicos y genéticos), la posible función que realizan, es decir, qué proteína codifican, y para reconocer los elementos reguladores de la expresión de los genes. También permite analizar otras características propias del genoma humano, como serían los sitios variables o polimórficos, es decir, las letras que cambian en cada individuo y que finalmente son la base de la diferencia observable entre los seres humanos. En la actualidad, estos análisis secundarios son los que se encuentran menos desarrollados, y realmente resultan los más importantes. Basta con señalar que el anuncio inicial de la secuenciación completa del genoma humano se refería, en realidad, a que sólo se tenía secuenciado (sin espacios o *gaps* y sin errores) 95% de la secuencia total; sólo de la parte de la eucromatina humana (1 000 000 000 pares de bases = 1 giga de bases) que corresponde a la porción del genoma que se expresa de manera activa; la contraparte es la heterocromatina que es la parte del genoma que no se expresa de esta manera. Para julio de 2001 lo que se tiene perfectamente terminado y sin problemas técnicos es sólo 47.1% de la secuencia del genoma completo (eucromatina y heterocromatina), la secuencia preliminar, que no ha sido revisada ni confirmada, comprende 51.4%. La suma del total de ambos será 98.5% del genoma humano. Estos datos traducidos a pares de bases de secuencia son alrededor de 1 660 078 000 pares de bases ya terminados y 3 547 899 000 pares de bases en versión preliminar. En realidad, se piensa finalizar exitosamente la secuencia completa para el

año 2003. Aun así, el genoma humano está lejos de ser comparable con cualquiera de los organismos secuenciados con anterioridad, como la levadura o la bacteria *Escherichia coli*, no sólo en la parte inicial que implica la descripción de la secuencia completa, sino en su análisis posterior. Este estudio final será el sustrato real para el posible desarrollo de terapias génicas y para la identificación de las bases genéticas de enfermedades como la diabetes y las neurodegenerativas.

### Los genes

El primer paso en la caracterización del genoma es la identificación de los genes, los cuales están formados por regiones codificantes (que son traducidas en proteína) denominadas exones y no codificantes llamadas intrones, que se encuentran intercalados entre los exones. Para que la información de un gen pueda ser procesada y traducida en proteínas, primero se tienen que separar los exones, que poseen información de los intrones que





no son traducidos en proteína; a este proceso se le denomina escisión alternativa de intrones. En el caso del ser humano este proceso es de suma importancia, ya que es el responsable directo de la diversidad de proteínas observada. Además, los genes poseen otros componentes estructurales que no son traducidos en proteínas, estas zonas son las que indican en dónde comienza y termina el gen, y a partir de qué momento puede ser expresado o transcrito. Dentro del ADN encontramos de manera intrínseca ciertos elementos denominados reguladores, cuyo objetivo principal es regular la expresión de los genes, es decir, cuándo y cómo se van a expresar. Existen dos clases de elementos reguladores, los denominados *cis* y los *trans*. Los *cis* son secuencias señalizadoras contenidas dentro del mismo ADN. En cambio, los elementos *trans* se caracterizan por ser proteínas que son transcritas a partir del ADN y que funcionan como reguladores de la transcripción de otros genes.

Para la identificación y predicción de genes se utilizan actualmente tres metodologías: 1) la evidencia directa experimental; 2) la evidencia indirecta relacionada con el parecido a otros genes previamente identificados, y 3) los métodos nuevos (*ab initio*) de reconocimiento de exones, basados en modelos que combinan información estadística acerca de los sitios de escisión de los exones,



del uso de codones (tripletes de bases que al ser leídas corresponden a un aminoácido), así como información del tamaño de intrones y exones.

El proyecto internacional estima que existen de 32 000 a 38 000 genes en total en el genoma humano. Este valor corresponde al doble de los detectados en la mosca de la fruta o en el nemátodo y a más o menos el mismo número que en la planta *Arabidopsis*. Por el contrario, Celera ha identificado 26 383 genes.

Se han descrito unas 1 300 familias de genes en nuestro genoma, de las cuales 98 son específicas de los vertebrados. Los genes humanos tienden a poseer exones pequeños, con un contenido promedio de 150 pares de bases, separados por intrones muy grandes, algunos de más de 10 000 pares de bases. Esto hace muy difícil su identificación, la cual depende básicamente de la disponibilidad experimental del ADN o de su parecido con genes de otros organismos. Este último paso es la parte inicial para la anotación de todos los genomas, o sea, la descripción de los genes. En general, se trata de una de las etapas más difíciles, pero que tiene las aplicaciones directas más importantes, ya que la base de cualquier estudio posterior se encuentra en la identificación de todos los genes que componen al genoma.

En promedio, como ya mencionamos, el tamaño de los exones es de alrededor de 150 pares de bases, su número por gen es de 8.8, el tamaño de los intrones es de 3 365 y el tamaño de la secuencia codificante promedio es de 1 340. Podemos encontrar que algunos de los genes son mayores de 100 000 pares de bases como el de la distrofina (relacionado con la actividad muscular, que mide 240 000). El gen de la titina (inmunoglobulina con dominios o regiones de fibronectina) posee la secuencia codificante más grande, con 80 780 pares de bases, el mayor número de exones (178), y el exón más grande, de 17 106. Al parecer, el tamaño típico de una secuencia codificante en el humano (1 340) es similar al del gusano (1 311) y al de la mosca de la fruta (1 497); la mayoría de los exones comprenden de 50 a 200 pares de bases, lo que sugiere cierta conservación de la maquinaria de escisión entre estos organismos. Sin embargo, el tamaño de los intrones es muy variable en los humanos, lo que provoca una gran variación en el de los genes.

Los genes humanos difieren de los otros organismos en aspectos importantes: se encuentran sobre regiones mucho más grandes de ADN y son usados para la construcción de muchos transcritos alternativos, lo que se tradu-



ce en la creación de aproximadamente cinco veces más productos diferentes (proteínas) en el humano que en la mosca de la fruta, el gusano o la planta *Arabidopsis*.

Las proteínas se encuentran compuestas por dominios equivalentes a módulos que pueden ser utilizados para la construcción de distintas proteínas. En el humano no se ha observado una gran cantidad de dominios nuevos, es decir, que no existe innovación en el surgimiento de otros módulos. La innovación evolutiva consiste en el arreglo lineal de estos dominios en una proteína, en las distintas posiciones que pueden tomar los dominios para constituirlos. Estos arreglos pueden ser formados por medio de la adición, deleción o uso alternativo de los distintos dominios presentes, lo que da como resultado la formación de nuevas proteínas a partir de dominios antiguos. Muchos de estos módulos pueden ser rastreados hasta el antepasado común de todos los animales. El dominio que se encuentra representado en mayor número en las proteínas humanas es el de los “dedos de zinc”, que caracteriza a una familia de factores de transcripción necesarios para el inicio de la copia del ADN, y que se encuentra descrito en 564 proteínas. Estos factores son un ejemplo de elementos reguladores *trans*.

Mientras que el proteoma, las proteínas identificadas en el genoma humano, contiene alrededor de 1.8 veces más arquitecturas protéicas distintas (formas diferentes de proteínas) que la mosca de la fruta o el gusano, y 5.8 más que la levadura. En especial, las proteínas extracelulares son las que muestran una mayor innovación —el humano posee 23 veces más tipos distintos que la mosca de la fruta y 2 veces más que el gusano.

Comparado con los invertebrados, los humanos parecen tener un mayor número de proteínas involucradas en funciones celulares, como la formación del citoesqueleto, la defensa y la inmunidad, la transcripción y la traducción, el desarrollo neural, la hemostasis y la apoptosis.

Aproximadamente 60% de las familias de proteínas identificadas son más numerosas en el humano que en los otros organismos. Es decir, que dentro de la misma familia de proteínas, el humano posee un mayor número de copias que las que presentan otros organismos como la mosca de la fruta y el nemátodo. Se piensa que este incremento es producto de la duplicación de los genes dentro de la familia, lo cual muestra que la duplicación génica ha sido una fuerza evolutiva importante en la historia de los vertebrados. Entre algunas de estas familias encontramos a las inmunoglobulinas, en especial los dominios protéicos que las componen, los cuales no se encuentran representados en la levadura ni en la planta *Arabidopsis*, pero que están presentes en las inmunoglobulinas, en anticuerpos y en el complejo mayor de histocompatibilidad, elementos fundamentales del sistema inmune que nos protege de las infecciones. La expansión que se observa en los grupos de proteínas pertenecientes a esta familia muestran la versatilidad de la respuesta a la infección que presentan los vertebrados. Otro grupo de proteínas que se encuentra en un número considerablemente mayor al descrito en otros organismos son los factores de crecimiento. Dentro del genoma humano se han identificado treinta factores de crecimiento fibroblástico, en cambio, en la mosca de la fruta y el nemátodo sólo existen dos. En el factor de crecimiento transformante



beta se observaron cuarenta y dos tipos, comparados con nueve y seis en la mosca y el nemátodo, respectivamente. Estos factores están relacionados de manera directa con procesos de diferenciación y crecimiento a lo largo de distintas etapas del desarrollo. Finalmente, la familia de los receptores olfativos comprende alrededor de 1 000 genes y pseudogenes (genes no funcionales), aunque en los homínidos la mayor parte de estos genes están interrumpidos o son pseudogenes, lo que sugiere una pérdida masiva en los últimos 10 millones de años, lo cual se refleja en que hemos perdido mucha capacidad olfativa.

La complejidad del genoma humano no radica esencialmente en su tamaño, sino en la multiplicidad de proteínas que se pueden construir a partir del procesamiento de la información que contiene. De acuerdo con esto la complejidad no involucra sólo el tamaño del genoma, sino sobre todo las innovaciones en los dominios y arquitectura de las proteínas a gran escala. En conclusión podemos decir que los humanos poseemos, comparados con todos los otros genomas analizados, un mayor número de genes, de dominios, de familias de proteínas, y de genes duplicados, así como de proteínas con múltiples funciones y formas.

#### Los sitios polimórficos

Estos sitios son variables en las poblaciones humanas, y, como ya mencionamos, la base de las diferencias entre nosotros. Los seres humanos se distinguen entre sí aproximadamente en una base de cada 1 250 (es decir, el cambio de una letra en cada 1 250 o 1 300). Se han identificado de 1.42 a 2.1 millones de sitios polimorfos en la secuencia preliminar, los cuales se encuentran mayormente en zonas no codificantes (sólo 1% en las codificantes). Estos datos sugieren que los niveles de variación son moderados en comparación con otras especies de animales, plantas o bacterias. Los sitios polimorfos son de gran interés médico y comercial, ya que son los que eventualmente permitirán los análisis para identificar las bases genéticas de la enfermedad y los que sugerirán su posible tratamiento.

#### Elementos repetidos

Una característica relevante del genoma humano es que posee secuencias que se repiten mucho. Esto se refiere a que existen repeticiones de ciertas secuencias de ba-



ses y que las encontramos continuamente en el ADN humano. Éstas pueden ir desde dos bases, por ejemplo, de ATATATATAT hasta secuencias un poco más complejas como ATAAAGGATAAAGG o de muy baja complejidad como AAAAAA. Este tipo de secuencias son denominadas secuencias simples y son producto de una especie de tartamudeo de la proteína que replica al ADN (ADN polimerasa). Pero además existen elementos repetidos más complejos, tal es el caso de copias grandes de segmentos pertenecientes al mismo cromosoma y secuencias repetidas en bloques.

Los elementos repetidos, algunas veces denominados parasíticos, pueden ser clasificados en cinco tipos, y corresponden a más de la mitad del ADN humano: a) derivados de eventos de transposición, donde se mueve una región de ADN y se inserta en otro lugar, este ADN puede pertenecer al mismo organismo o ser extranjero; b) copias inactivas de pseudogenes; c) repeticiones de secuencias simples (aproximadamente 3% del genoma humano); d) duplicaciones de segmentos de entre 10 000 y 300 000 pares de bases (copias de una región del genoma a otra); e) secuencias repetidas en grupo (en bloques como en los centrómeros, telómeros, brazos cortos de los cromosomas y genes ribosomales). La mayor parte de las secuencias repetitivas (45%) son derivadas de eventos de transposición. Algunos de los elementos que sufren este fenómeno que se pueden reconocer en el genoma humano son los LINES (elementos transponibles grandes), SINES (elementos pequeños que se transponen), LTR (retrotransposones, que se copian a partir de ARN y no de ADN, como es lo usual) y transposones de ADN. Al parecer no existe evidencia alguna dentro del genoma humano de la actividad de transposones de ADN (no se han movido) en los pasados cincuenta millones de años. Los elementos que se transponen y que más abundan en el genoma humano son las secuencias conocidas como ALU (pertenecen a los SINES) y algunos elementos LINE que representan 60% del total.

La mayoría de los elementos parasíticos se originaron por transcripción reversa, llevada a cabo por retrovirus a partir de ARN. Los transposones han traído consigo nuevos elementos reguladores y han dado origen a nuevos genes. En el genoma humano han sido identificados al menos veinte genes que se consideran como producto de eventos de transposición. Un ejemplo clásico sería el de las telomerasas, enzimas que replican las puntas de los cromosomas durante la replicación.



Dentro de los elementos repetitivos destaca el que corresponde a las duplicaciones de segmentos. Éstas involucran la transferencia de bloques de 1 000 a 200 000 pares de bases a uno o más sitios de los cromosomas. Estas duplicaciones se pueden dividir en dos categorías: inter-cromosomales (entre cromosomas distintos) e intracromosomales (dentro del mismo cromosoma). La secuencia final reportada tiene al menos 3.5% segmentos duplicados, de los cuales 1.5% corresponde a duplicaciones inter-cromosomales y 2% a duplicaciones intracromosomales. Este tipo de datos muestran que es difícil sostener la propuesta de que los genomas de los vertebrados han evolucionado por duplicaciones genómicas totales, que es la explicación clásica que se da al incremento del tamaño de genoma en los vertebrados. Los datos genómicos demuestran que una gran porción del genoma está caracterizado por segmentos muy grandes duplicados, pero no de cromosomas totales.

Es interesante observar que las cuatro regiones que poseen un menor número de elementos repetitivos dentro del humano son las zonas conocidas como cajas homeóticas HOX A, B, C, y D. Éstas pertenecen al grupo de los reguladores *cis*, es decir, son señales que se encuentran en el mismo ADN. Su importancia radica en que son reguladores fundamentales de muchos procesos de desarrollo. También se han reconocido muchas familias de

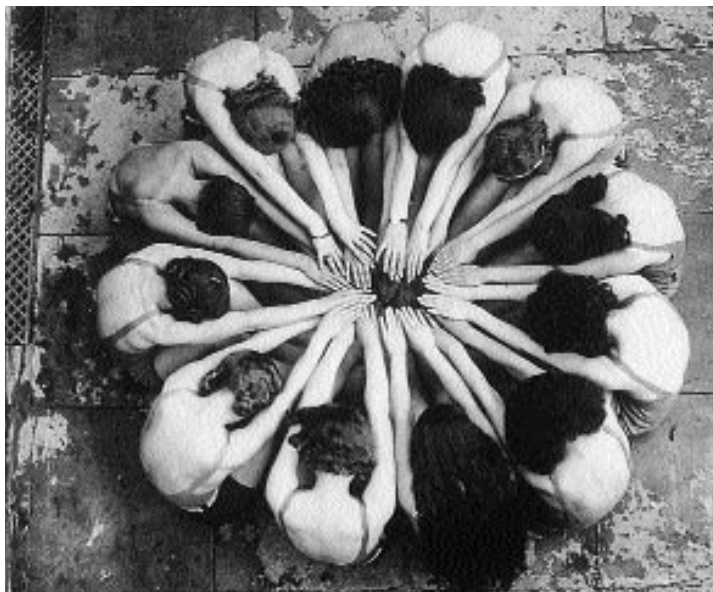


proteínas (reguladores *trans*) que presentan un dominio llamado homeodominio que se une directamente al ADN, todas ellas son reguladoras de la expresión en procesos fundamentales de diferenciación y desarrollo. Se puede decir que son una especie de interruptor que desencadenan una cascada de expresión de distintos genes que culmina en el desarrollo o diferenciación de alguna estructura, como puede ser el caso de un órgano o un tejido en específico.

### ¿El fin o el principio?

El objetivo final del proyecto genoma humano es la compilación de una lista de todos los genes humanos, sus proteínas y su ubicación exacta, con el fin de ser utilizadas

como una especie de “tabla periódica” para la investigación biomédica. En la actualidad se tienen secuenciados adicionalmente los genomas de cuarenta y un bacterias, nueve arqueobacterias, una planta y tres animales. Pero el reto de la biología, más allá de la simple secuenciación de los genomas, consiste en el entendimiento evolutivo y fisiológico de cómo los genes están organizados para la construcción y mantenimiento de las especies, y aún más allá, como Craig Venter menciona, intentar explicarnos cómo se organizan nuestras mentes, lo que nos lleva a investigar acerca de nuestra propia existencia. De esta forma podemos asegurar, casi sin temor a equivocarnos, que así como el siglo xx se denominó el siglo de la física, el siglo xxi será reconocido como el siglo de la biología. Indudablemente nos tocará vivir tiempos interesantes. 🌀



Amanda Cobián y Luis E. Eguarte  
Instituto de Ecología,  
Universidad Autónoma de México.

#### REFERENCIAS BIBLIOGRÁFICAS

International Human Genome Sequencing Consortium. 2001. "Initial Sequencing and Analysis of the Human Genome", en *Nature*, 409, pp. 860-921.  
Ridley, M. 1999. *Genome*. Harper Collins, Gran Bretaña.  
Venter, JC. et al. 2001. "The Sequence of the Human Genome", en *Science*, 291, pp. 1304-1351.  
<http://www.ncbi.nlm.nih.gov/>. Sitio sobre el proyecto del genoma humano.

#### IMÁGENES

P. 56: Eadweard J. Muybridge, secuencia de movimiento de una mano, ca. 1884-1887. P. 57: Gibson

Ralph, de la serie *El sonámbulo*, 1968. P. 58: Atribuido a Charles Hugo y Auguste Vacquerie, *La mano de Madame Hugo*, 1850-1854. P. 59: P. Horst, *Manos, manos*, 1941. P. 60: Herbert Bayer, *Soledad del ciudadano*, 1932. P. 61: Autor desconocido, *La mano de un hombre y la de un niño*, 1875; James Hall Nasmyth, *El dorso de la mano*, 1874; Dieter Appelt, *La liberación de los dedos*, 1979. P. 62: Francis Bruguière, *Mujer con brazo contra la pared*, ca. 1925; Alfred Stieglitz, *Georgia O'Keeffe*, 1933. P. 63: Kishin Shinoyama, *La morena Sily*, 1968. P. 64: Autor desconocido *Hulton Getty Picture Collection*, 1935.

#### AGRADECIMIENTOS

A la Dra. Valeria Souza Saldivar quien cuidadosamente revisó el manuscrito de este artículo. Los autores agradecen el apoyo del proyecto Conacyt 27983-N y el proyecto Genómico Conacyt 0028.