

## Anexo 2

# Nomenclatura

## Genes

La nomenclatura de los genes inicialmente parecía simple para organismos sencillos, como bacterias, levaduras. Se designaban, en principio, siguiendo dos criterios: no poner el nombre de su descubridor y nombrarlos con una abreviación (de tres letras) descriptiva de la función del gen. Normalmente con letras en MAYÚSCULA y *cursiva*. El problema vino en organismos superiores. Tres letras no parecían suficientes, pero se mantenía la tendencia de que el nombre fuera descriptivo. Respecto al número del gen, también hay varios criterios. A veces los genes tienen varias copias en el genoma, y a medida que se van identificando se numeran de forma correlativa. Otras veces tienen una función similar y se suelen bautizar correlativamente. (En ocasiones el número hace referencia al peso molecular de la proteína que codifica). Y si el número no quiere decir nada, se pone 1.

Un recurso para consultar información sobre genes es <https://www.genecards.org/>. Se encuentran recogida la información de genes (20692) que codifican proteínas.

También se recoge información sobre genes en <https://www.ncbi.nlm.nih.gov/gene>

<https://www.genenames.org/about/guidelines/> recoge las directrices actuales para nombrar los genes.

## Nucleótidos

En la numeración de nucleótidos de la secuencia de ADN de referencia, la A del codón de iniciación (ATG) se considera el nucleótido +1. No hay ningún nucleótido que se enumere con el cero. Los nucleótidos que se encuentran en posición corriente arriba del +1 deben ir precedidos por el signo menos; por ejemplo, -1, -2, -3, etc. Los nucleótidos situados corriente abajo del +1 no van precedidos de ningún signo.

## Mutaciones

Hay cierta confusión en la terminología. El término mutación se utiliza en ocasiones para indicar un cambio y en otras para indicar un cambio que causa cierta enfermedad. De forma similar, el término polimorfismo se utiliza tanto para describir un cambio que no causa enfermedad como para señalar un cambio que se encuentra con una frecuencia  $\geq 1\%$  de la población. Para evitar esa confusión, se pueden emplear términos tales como variaciones de secuencia, alteración o variante alélica.

Según el tipo de nucleótido sustituido, las mutaciones puntuales se denominan: *a)* transición, que consiste en la sustitución de una pirimidina por otra o de una purina por otra, de tal forma que un par G-C es reemplazado por un par A-T o viceversa, y *b)* transversión, cuando una purina es reemplazada por una pirimidina o viceversa, en cuyo caso un par A-T es reemplazado por un par T-A o C-G.

Para nombrar las mutaciones se recomienda utilizar una serie de reglas elaboradas por el Human Genome Organization (HUGO) Gene Nomenclature Committee (HGNC), que se detallan en su página web ([www.hgvs.org/mutnomen/recs](http://www.hgvs.org/mutnomen/recs)) y resumimos a continuación:

- . La regla más importante es que las variantes deben ser descritas en el nivel más básico posible, es decir a nivel de ADN. La descripción siempre debe referirse a una secuencia de referencia, ya sea la secuencia genómica o la codificante de ADN.
- . Para genes/proteínas, sólo se utilizarán los símbolos recomendados por el HGNC (disponibles en: [www.genenames.org/guidelines.html](http://www.genenames.org/guidelines.html)). Cuando se hace referencia a nivel de ARN o proteína en un determinado texto, la primera vez que se menciona debe utilizarse el siguiente formato: en primer lugar, ADN y entre paréntesis proteína o ARN. Ejemplo: c.48G>C (p.Trp16Cys).
- . Cuando se describen varios cambios, es conveniente proporcionar una tabla con la lista de los hallazgos, utilizando columnas separadas para ADN, ARN y proteínas, e indicar si los cambios fueron deducidos de forma experimental o teórica.

. Para hacer referencia a un nucleótido, deberá precederle la letra g en minúscula si se trata del ADN genómico, la c minúscula si se refiere a un ADN codificante (ADN-c), la m minúscula si se refiere a un ADN mitocondrial, la r minúscula si se refiere a una secuencia de ARN ribosomal y la p minúscula si se refiere a una proteína. Para evitar posibles confusiones entre el número del nucleótido y la g, la c, etc., se intercalará un punto. Por ejemplo g.576, significa que se trata del nucleótido que está situado en la posición 576 del ADN genómico.

Para la descripción de cambios específicos:

- . El símbolo > indica una sustitución a nivel de ADN; por ejemplo, c.75A>T.
- . El símbolo \_ indica un cambio que afecta a varios residuos, que separa el primero y el último de los residuos afectados; por ejemplo, c.77\_79delACT.
- . Con del se indica una delección; por ejemplo, c.77delA.
- . Con dup se indica una duplicación; por ejemplo, c.dupA.
- . Con ins se indica una inserción; por ejemplo, c.75\_76insG.

Los cambios de nucleótido deben empezar por el número del nucleótido seguido del cambio que tiene lugar en esa posición; por ejemplo, 576G>C significa que el nucleótido de la posición 576 que en la secuencia de referencia es una G ha sido reemplazado por una C.

Al describir dos variantes de secuencia en un mismo individuo:

- . Dos cambios en la secuencia en diferentes alelos se representan entre corchetes, separados por el signo +; por ejemplo, c.[76A>C]+[87delG].
- . Dos variantes de secuencia en el mismo alelo se ponen entre corchetes separados por punto y coma; por ejemplo, c.[76A>C; 83G>C].
- . Dos variantes de secuencia en alelos desconocidos se ponen entre corchetes separadas por (+); por ejemplo, c.[76A>C(+)]83G>C].

Para designar la localización de una variación en un intrón, se puede utilizar la numeración de los nucleótidos correspondientes al ADN. Por ejemplo, c.1232+1G>C significa la sustitución de G por C del primer nucleótido del intrón que sigue a la posición 1232 del ADN.

Debido a la gran cantidad de posibles variantes que pueden existir de un único SNP, su nomenclatura es bastante confusa. En la mayoría de los casos cuando se hace referencia a un SNP es sobre la base de su número rs (RefSnps) de la base de datos dbSNP <https://www.ncbi.nlm.nih.gov/snp/>. Así, el número rs es un número de acceso utilizado por investigadores y clínicos para referirse a SNP específicos. Este número de variación es inequívoco.

## Alelos

Las diferentes versiones de un gen (alelo) se designan escribiendo, a continuación del nombre del gen, asterisco (\*) y un número. Normalmente la variación salvaje original se le asigna el número 1, y las demás variantes, números correlativos. Existen genes muy polimórficos, de los cuales se han descritos muchas variantes, como el gen CYP2D6.

Una base de datos donde consultar variantes del CYP450 es <https://www.pharmvar.org/gene/> donde se encuentran recogidas todas sus variantes.

La base de datos <https://www.omim.org/> recoge la información de 16.000 genes relacionados con el fenotipo.