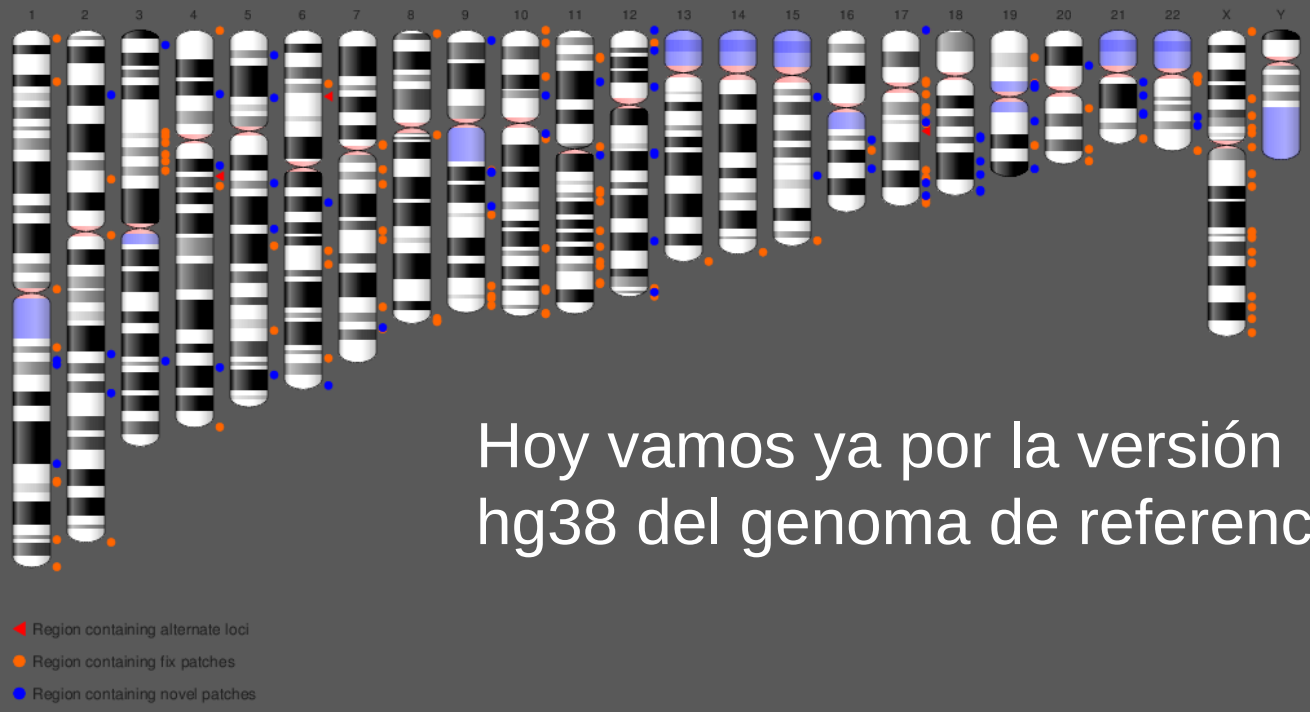
Grandes proyectos genómicos

- El proyecto Genoma Humano
- El proyecto 1000 Genomas
- The 100,000 Genomes Project
- Roadmap Epigenomics Project
- Encode
- The 4D Nucleome Project
- …

El Proyecto Genoma Humano

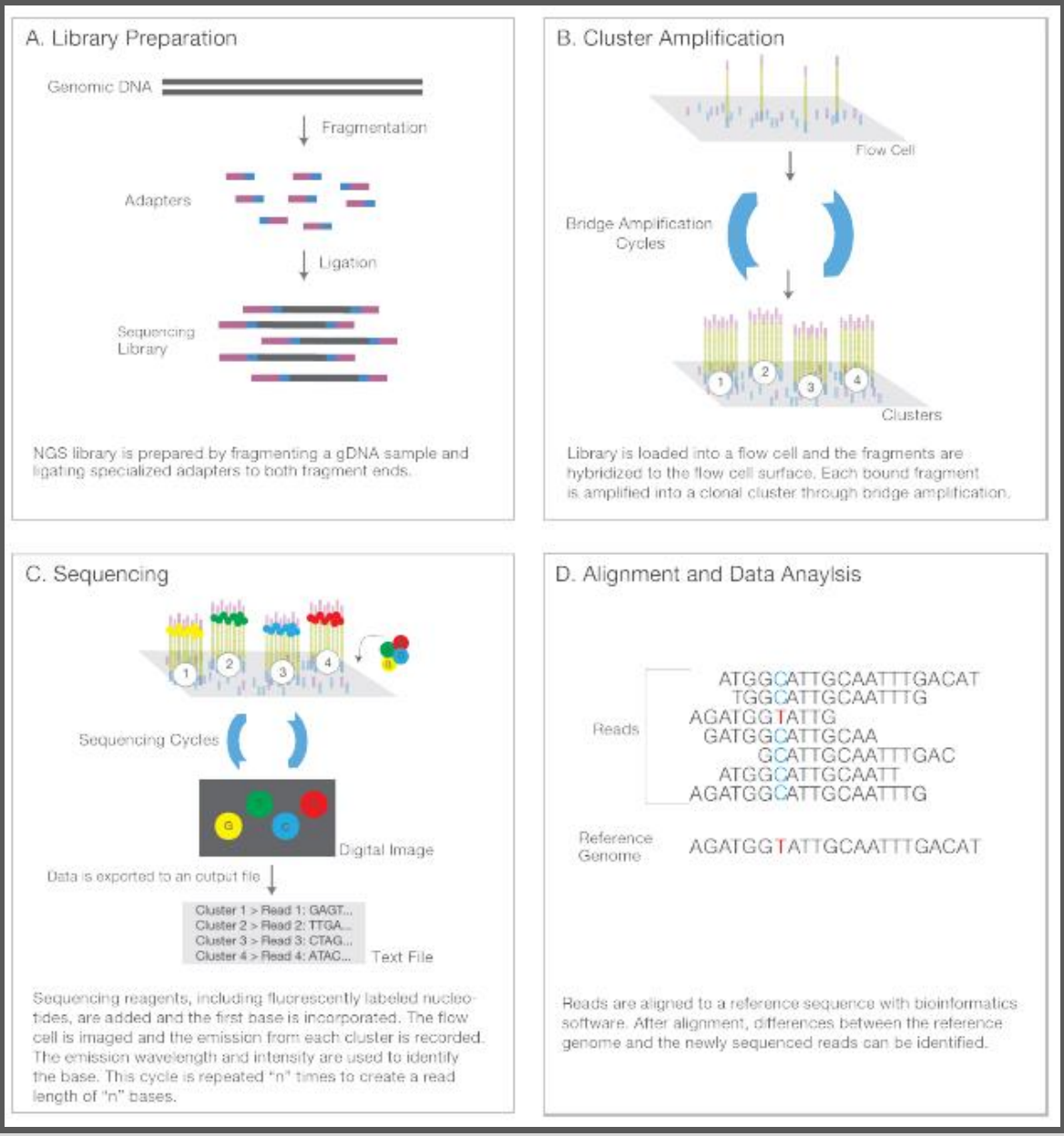
- Fue una iniciativa internacional lanzada en la década de los 90 del pasado siglo para mapear y secuenciar el conjunto de genes del ser humano (genoma)

- Completado en 2003 con la publicación de la primera **secuencia de referencia** del genoma humano.



Hoy vamos ya por la versión hg38 del genoma de referencia

Secuenciación masiva o NGS: Next Generation Sequencing



454
Pyrosequencing (PS)

454 SEQUENCING

ion torrent
Semiconductor Sequencing

illumina

Illumina
Reversible Termination (RT)

life technologies | applied biosystems

SOLID
Sequencing by Ligation (SBL)

PacBio RS II with touch screen
RS Remote for run design
SMRT Cells
DNA Sequencing Kit

Secuenciación de tercera generación: Single Molecule Real Time (tecnología SMRT, PacBio).

Lecturas (reads) muy largas: Esta tecnología consigue lecturas de longitudes de hasta 30 kbp con un tamaño medio de entre 4,200 y 8,500 pb.

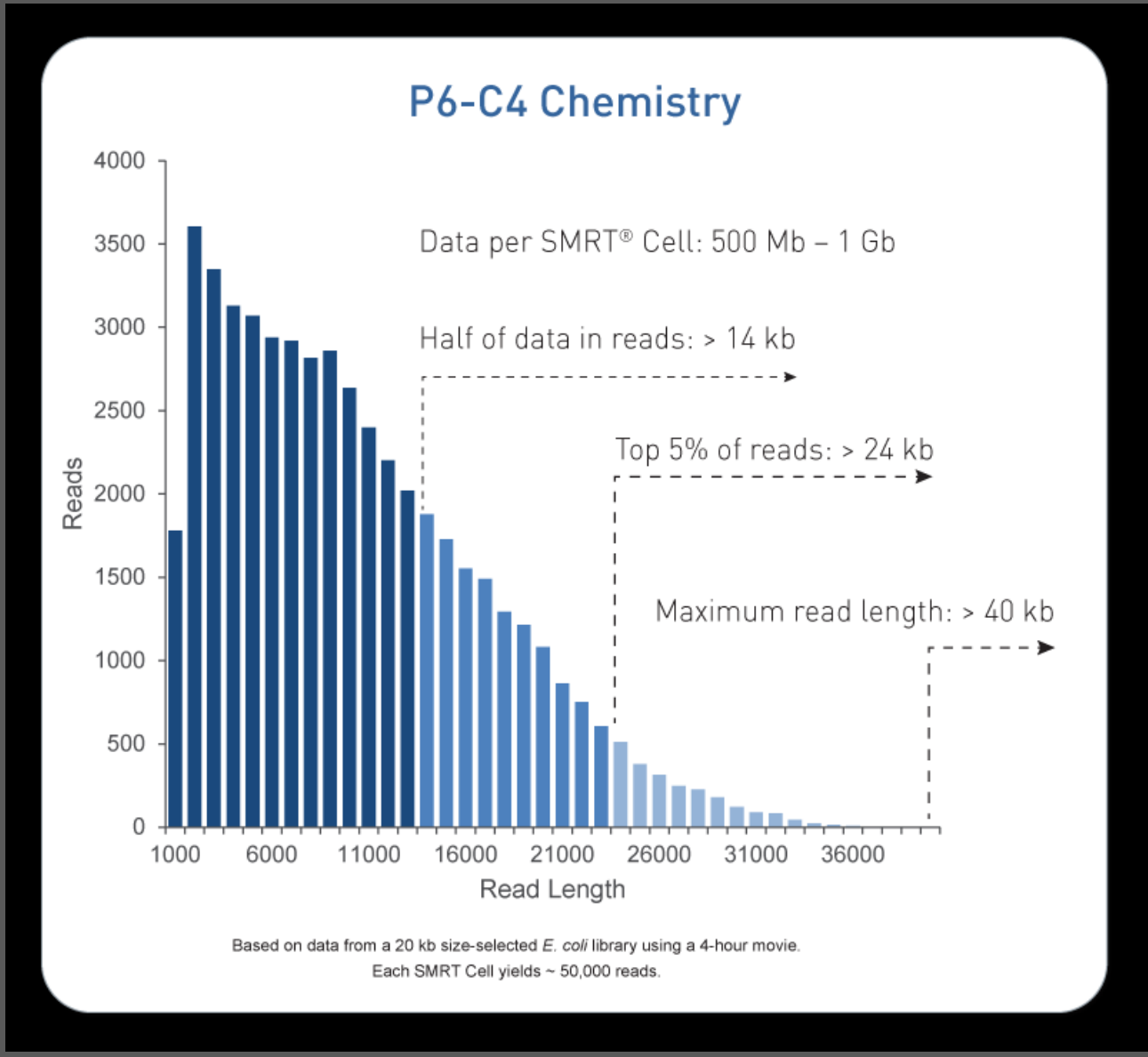
PacBio

Sequencing with the PacBio RS II system based on single molecule, real-time (SMRT) technology:

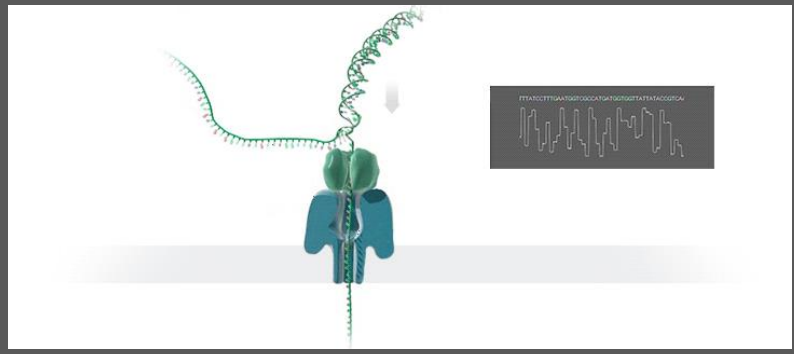
Long reads: Depending upon starting library, half of the data are in reads >14,000 base pairs long with the longest reads over 40,000 base pairs.

High accuracy: Perform de novo assembly of genomes and detect variants with greater than 99.999% accuracy. Sequence individual molecules with 99% accuracy at greater than Sanger lengths.

High sensitivity: Detect minor variants that are present at a frequency less than 0.1%.



MinION nanopore: *a miniaturised single-molecule analysis system, designed for single use and to work through the USB port of a laptop or desktop computer*

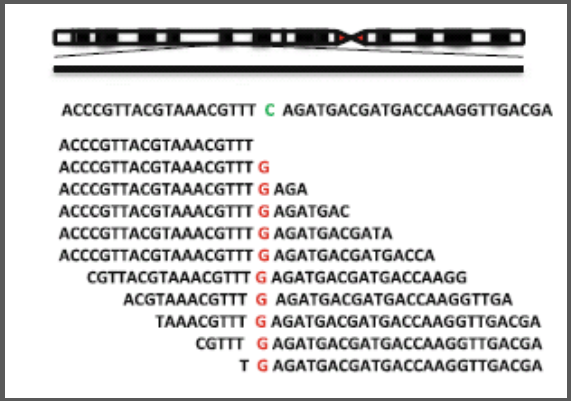


- Algo impreciso aún
- Pero muy adecuado para estudios de campo: biodiversidad, epidemias (EBOLA), etc.

Secuenciación masiva

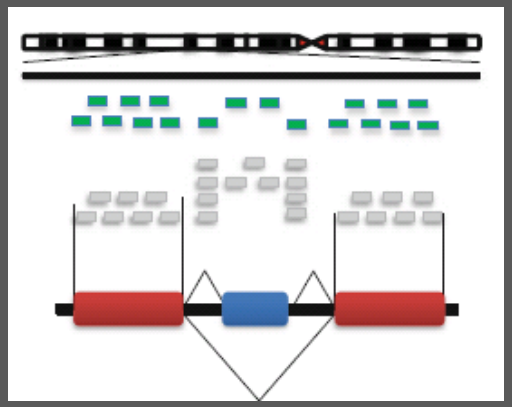
APLICACIONES

Variabilidad



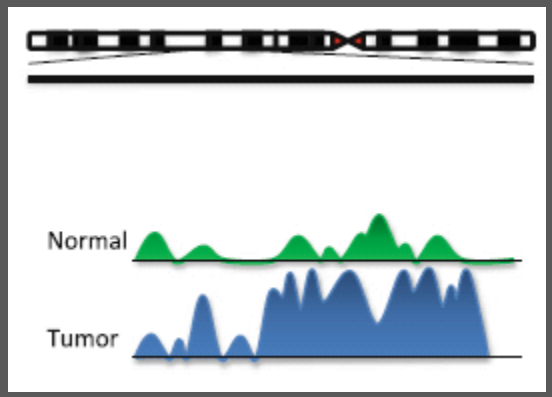
Detección de variantes: SNVs y CNVs, inserciones y deleciones

Regulación

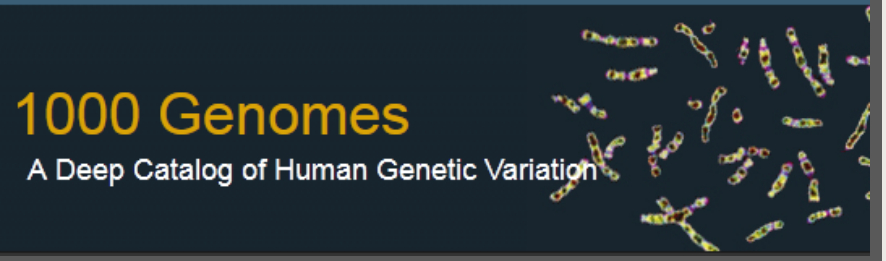


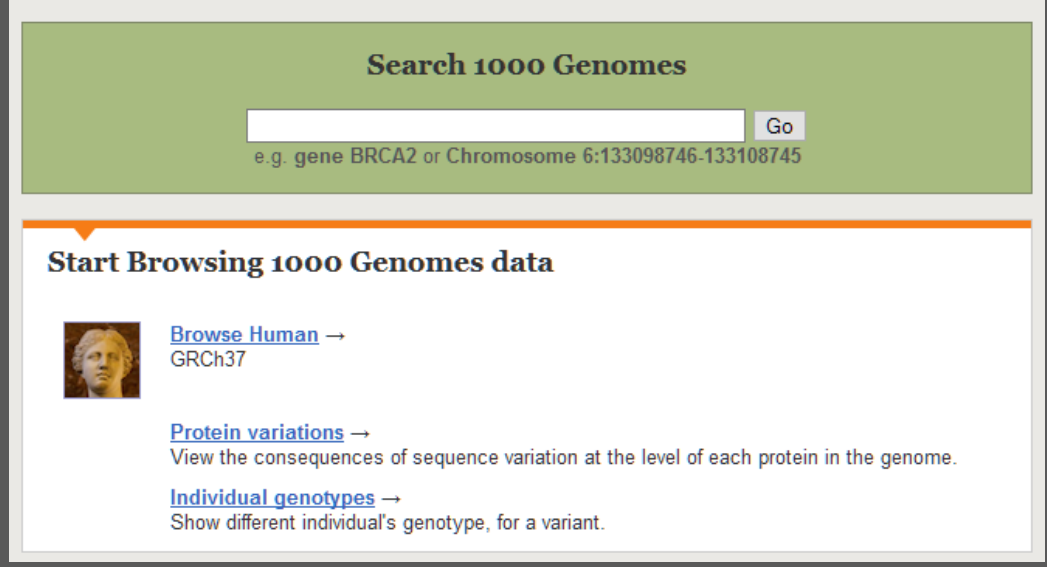
RNA-seq: expresión génica

Epigenómica



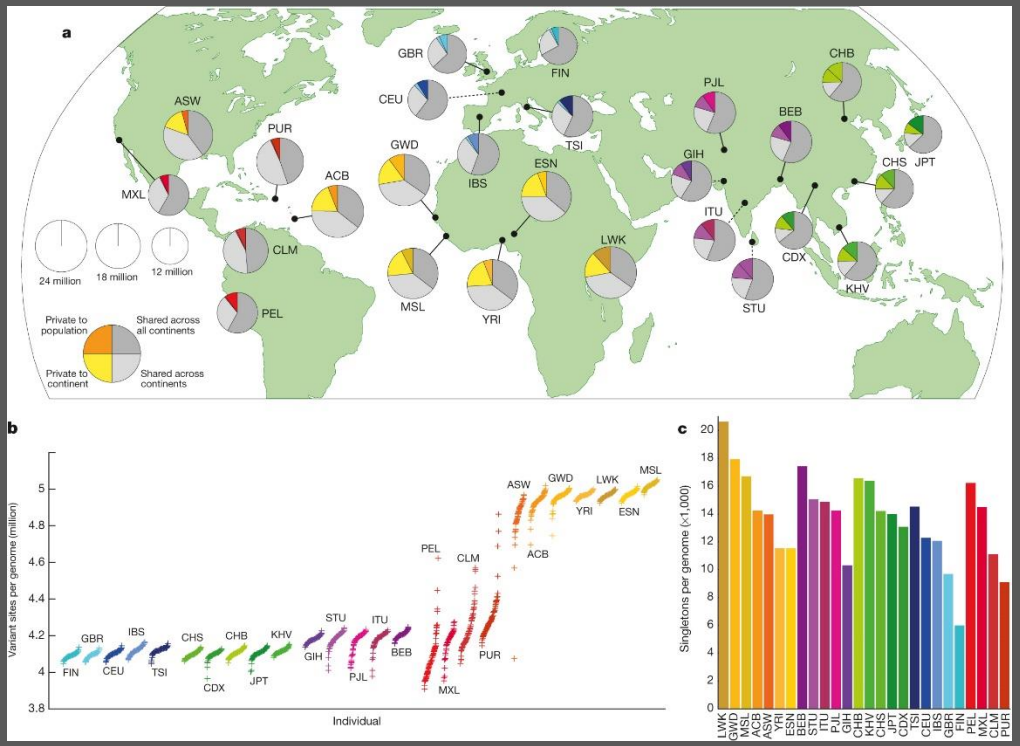
- Metilación diferencial del ADN
- Metilación de histonas
- TFBSs





- Caracterización de la variación genética en el genoma humano
- La mayor parte de las variantes genéticas se localizan en las regiones no-codificadoras del genoma → elementos reguladores
- Cada uno de nosotros lleva por término medio entre 200 y 300 variantes de pérdida de función (LOF)

Population sampling

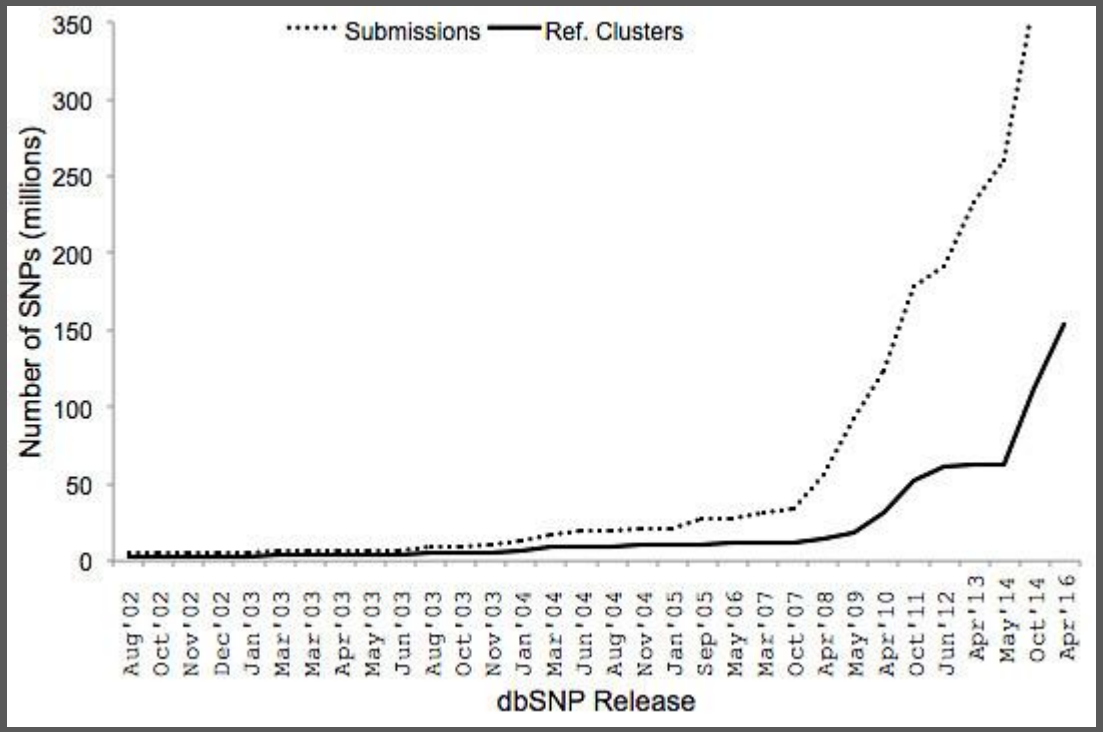


a, Polymorphic variants within sampled populations. The area of each pie is proportional to the number of polymorphisms within a population. Pies are divided into four slices, representing variants private to a population (darker colour unique to population), private to a continental area (lighter colour shared across continental group), shared across continental areas (light grey), and shared across all continents (dark grey). Dashed lines indicate populations sampled outside of their ancestral continental region. **b,** The number of variant sites per genome. **c,** The average number of singletons per genome.

A Auton *et al*. *Nature* **526**, 68-74 (2015) doi:10.1038/nature15393

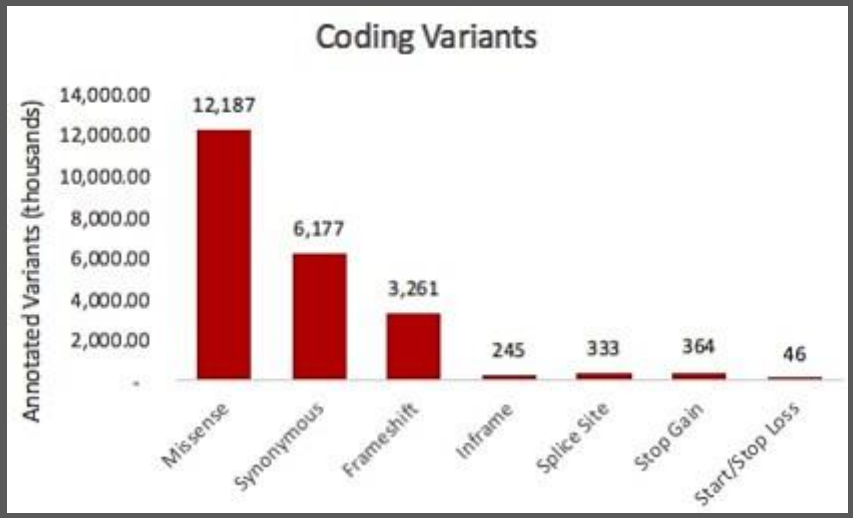
Population labels: https://catalog.coriell.org/1/NHGRI/About/Guidelines-for-Referring-to-Populations

dbSNP contains 152.7 million human variants



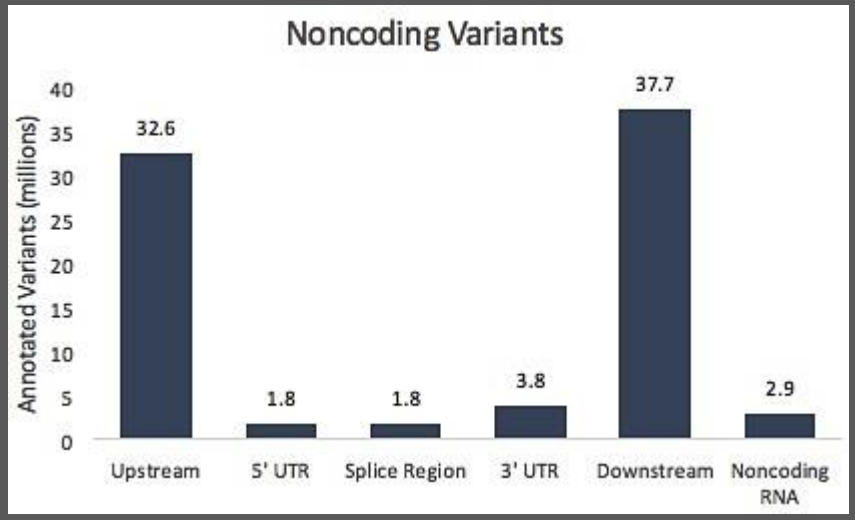
One variant for every 20.5 base pairs in the human genome

6+ million coding variants



One variant every 5 or 6 base pairs in coding regions

The vast majority of known variants in our genome lie outside of protein-coding exons



Genomics england

About us | 100,000 Genomes Project | GeCIP | GENE Consortium

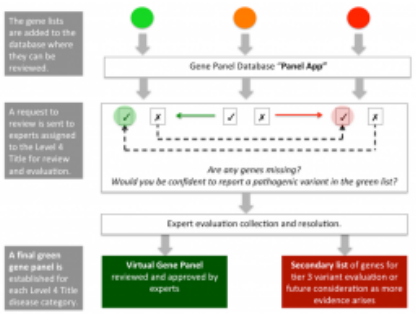
Home › The 100,000 Genomes Project

The 100,000 Genomes Project

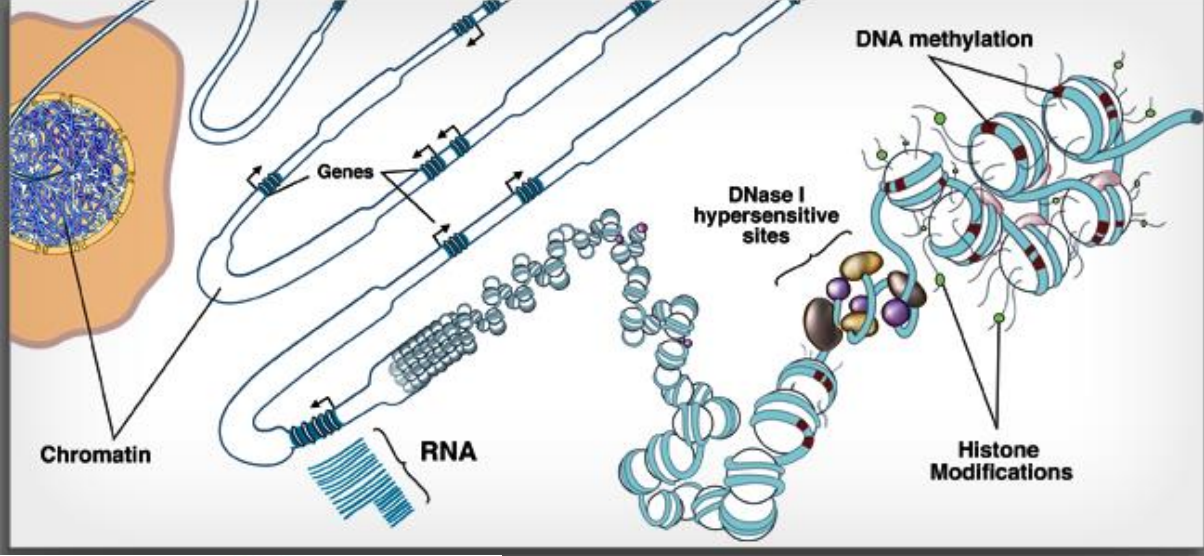
The project will sequence 100,000 genomes from around 70,000 people. Participants are NHS patients with a rare disease, plus their families, and patients with cancer.

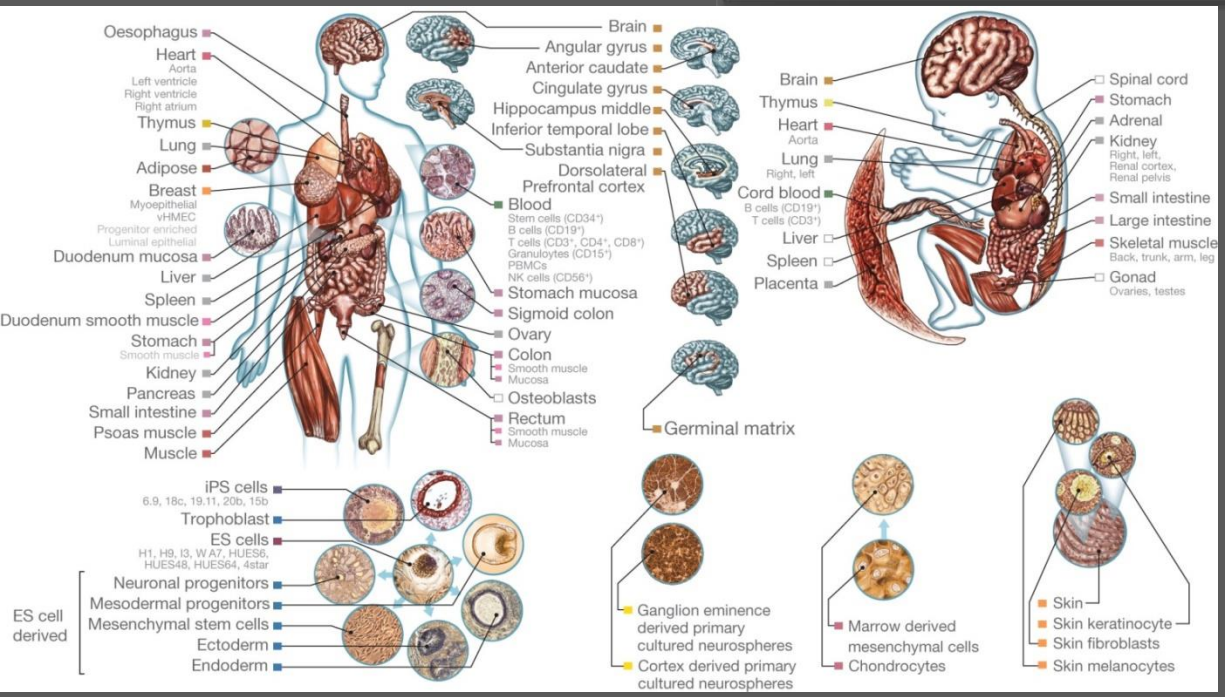
New rare disease gene tool launched – PanelApp

Genomics England has developed a unique resource, the 'PanelApp'. It is a new crowdsourcing tool for the scientific community, allowing knowledge of rare disease genetics to be shared and evaluated. This will create comprehensive evidence-based gene panels for rare diseases. The resource is publically available for anyone who would like to view and download the gene panels. Experts can register as a reviewer to make evaluations of the gene panels.



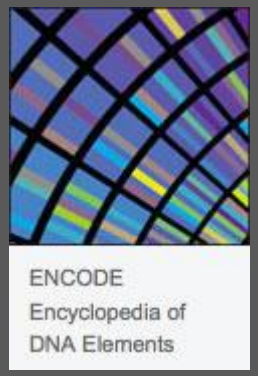
ROADMAP epigenomics PROJECT





Roadmap Epigenomics Consortium *et al. Nature* **518**, 317-330 (2015) doi:10.1038/nature14248

nature



El proyecto ENCODE

- La ‘Encyclopedia of DNA Elements’ (ENCODE) surge de una colaboración internacional iniciada en 2003.
- El objetivo de ENCODE es elaborar un catálogo exhaustivo de todos los elementos funcionales en el genoma humano, incluyendo tanto ARNs como proteínas, asi como aquellos elementos reguladores que controlan el tipo celular y el momento del desarrollo en que un gen es activo.
- La cuestión es: la suma de los exones de los aprox. 21.000 genes humanos no llegan al 2% del genoma ¿para que sirve el 98% restante? ¿es ADN basura?

Principales hallazgos de ENCODE

La mayor parte del genoma (80.4%) se puede asociar con al menos una función en alguno de los 147 tipos celulares analizados. Puesto que puede haber hasta 2.000 tipos celulares, este porcentaje podría llegar a ser mucho más alto!

Los elementos específicos de primates están sometidos a selección natural → deben ser funcionales.

Se han descubierto 399.124 enhancers y 70.292 promotores.

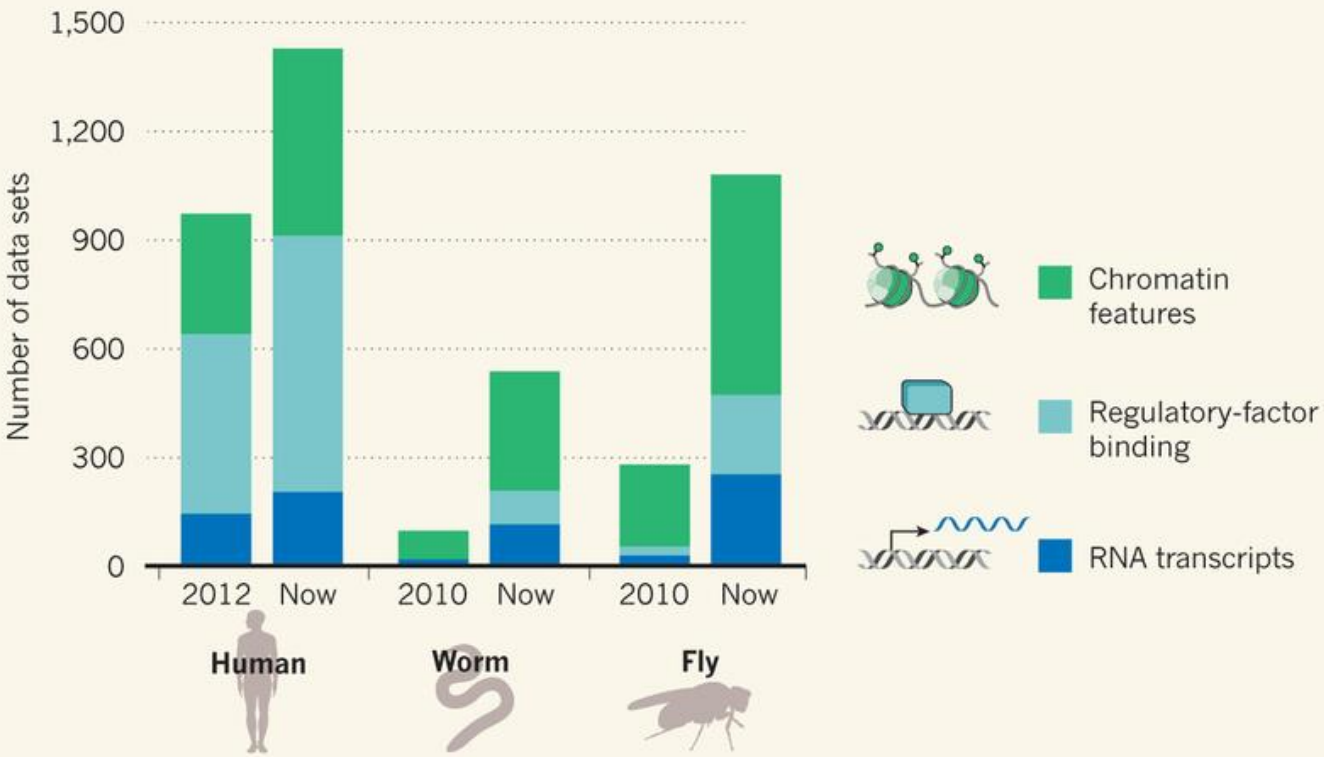
Muchas de los elementos funcionales encontrados se localizan en las regiones no-codificadoras (fuera de los genes).

Los SNPs asociados con enfermedades mediante GWAS abundan en las regiones no-codificadoras.

Muchas enfermedades se asocian con un determinado factor de transcripción que varía entre tipos celulares.

Felix Muerdter & Alexander Stark, Nature 512, 374–375 (28 August **2014**)

Más de 1600 nuevos conjuntos de datos, lo que hace un total de 3300 entre ENCODE y modENCODE



Cautelas sobre el proyecto ENCODE (extraidas de las publicaciones de 2014):

"…although they are extremely data-rich, the papers expose how data sets that are created to catalogue all functional elements under standardized conditions are not sufficient for understanding the regulation of transcription, chromatin biology and enhancer function, nor the evolution of these mechanisms."

Según **Dan Graur** esto quiere decir que:

- Not every piece of chewing gum attached to the soles of your shoes is functional.
- Moreover, the function of the sole of your shoe to which the chewing gum stuck is NOT to bind chewing gum.

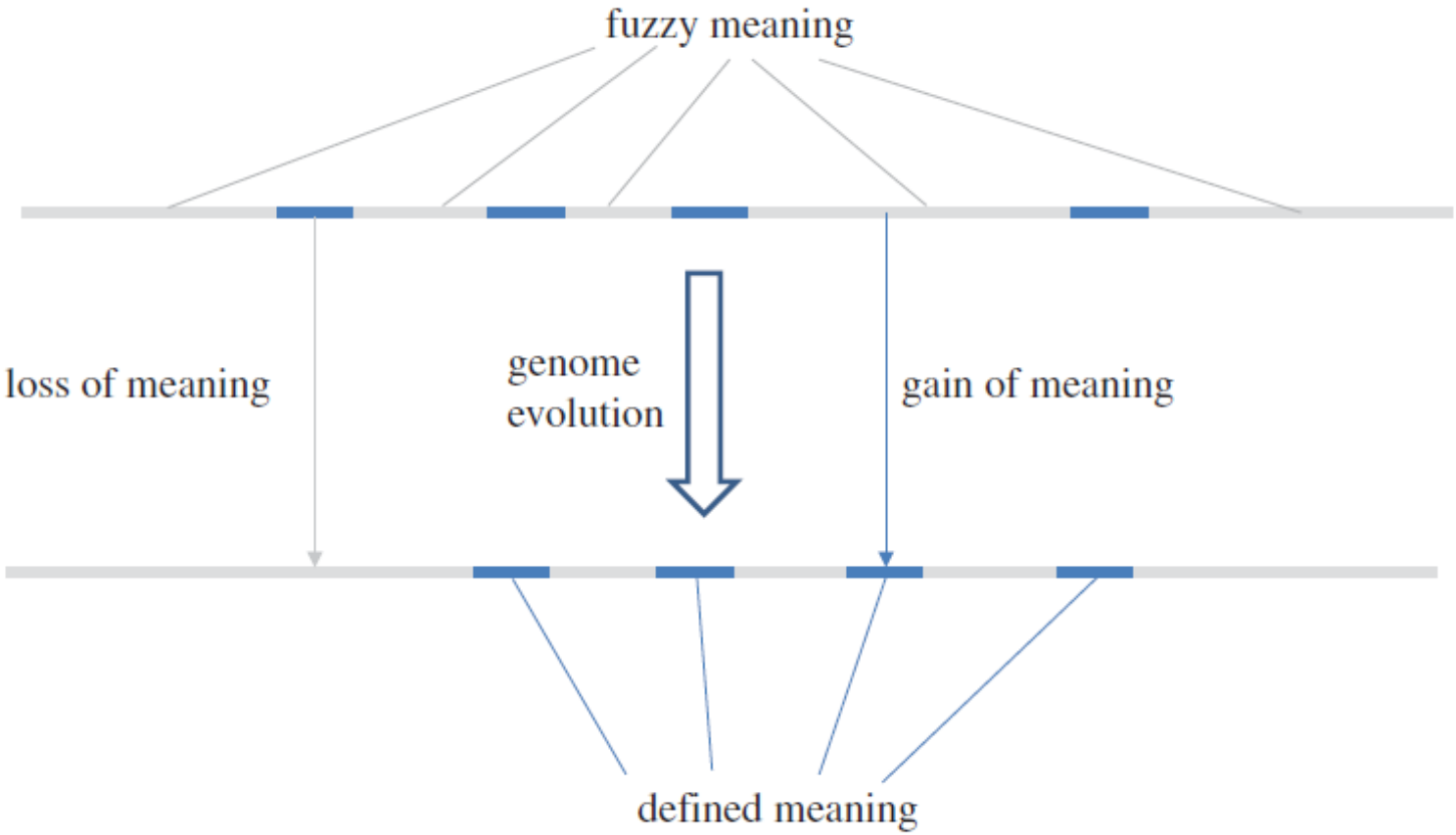
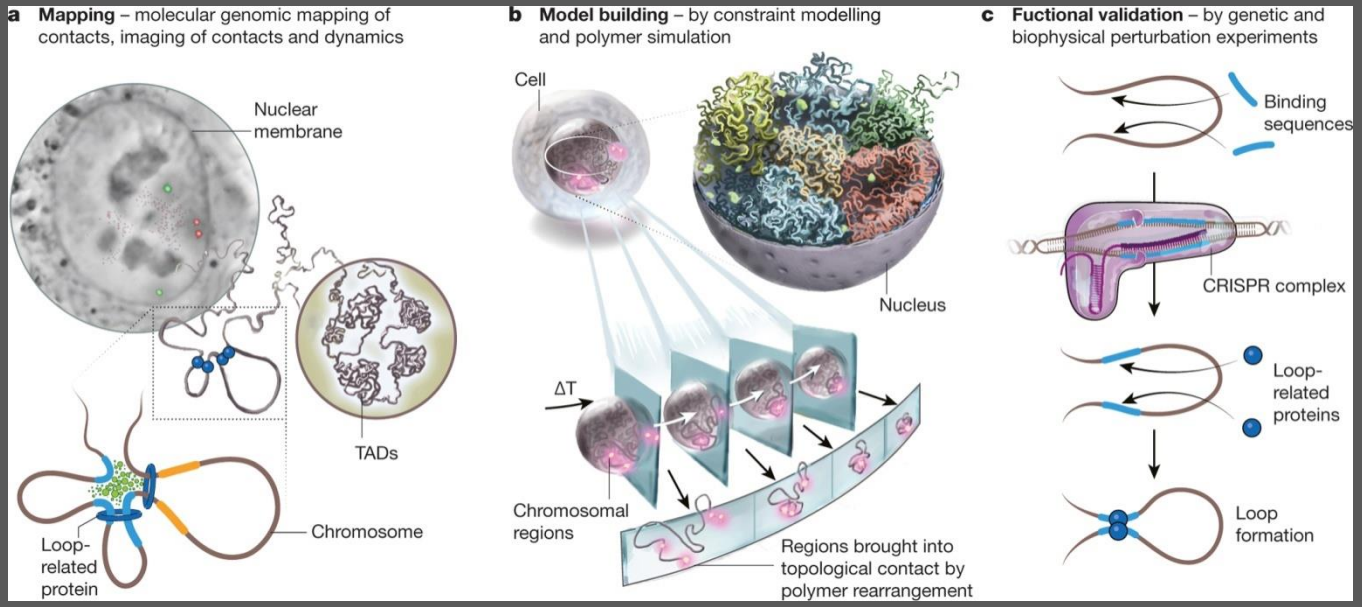


Figure 2. The fuzzy meaning concept and gain and loss of meaning. The cartoon schematically shows a fragment of a genome of a complex multicellular organism (animal or plant) that consists mostly of sequences with fuzzy meaning, interspersed with 'islands' of defined meaning such as genes (exons) encoding structural RNAs and proteins as well as evolutionarily conserved regulatory elements. (Online version in colour.)

Eugene V. Koonin (2016) The meaning of biological information. Phil. Trans. R. Soc. A 2016 374 20150065; DOI: 10.1098/rsta.2015.0065.

The 4D Nucleome project



Data obtained with imaging and chromosome conformation capture-based assays can be used for building spatial and dynamic models of chromosomes using two main approaches. In the data-driven approach, experimental data are used directly to generate ensembles of conformations that reproduce the experimental observations. In the *de novo* approach, ensembles of conformations are built according to known or hypothesized physical or biological processes. Models are then selected based on their agreement with experimental data.

J Dekker *et al. Nature* **549,** 219–226 (14 September 2017) doi:10.1038/nature23884

nature

El futuro de la genómica…

- Muchos **hospitales** tendrán pronto departamentos de **medicina genómica**
- Los **secuenciadores** serán del tamaño de los USBs (ya los hay!)
- Miles de **test genéticos** personales estarán disponibles… en los supermercados
- Para 2025, se habrán secuenciado ya **millones de genomas**

Dawn Field, 2015. Perfect genetic knowledge, AEON magazine