

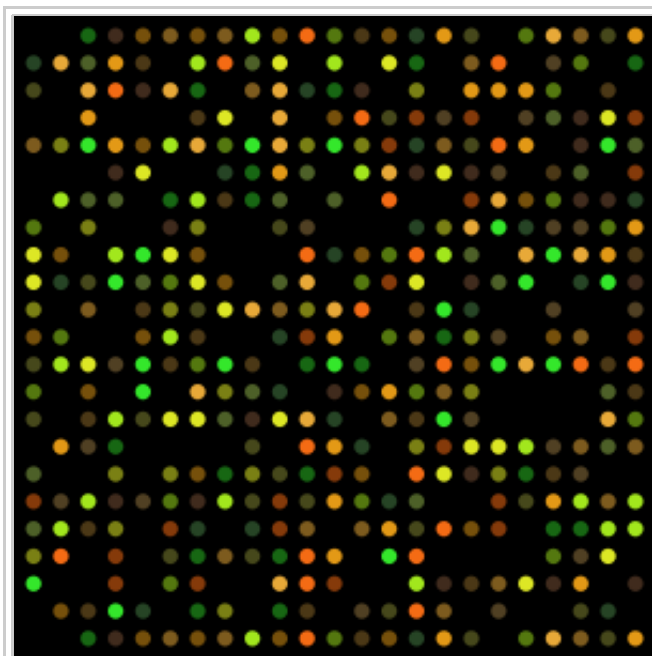
# Functional genomics

From Wikipedia, the free encyclopedia

**Functional genomics** is a field of molecular biology that attempts to make use of the vast wealth of data produced by genomic projects (such as genome sequencing projects) to describe gene (and protein) functions and interactions. Unlike genomics, functional genomics focuses on the dynamic aspects such as gene transcription, translation, and protein–protein interactions, as opposed to the static aspects of the genomic information such as DNA sequence or structures. Functional genomics attempts to answer questions about the function of DNA at the levels of genes, RNA transcripts, and protein products. A key characteristic of functional genomics studies is their genome-wide approach to these questions, generally involving high-throughput methods rather than a more traditional “gene-by-gene” approach.

## Contents

- 1 Goals of functional genomics
- 2 Techniques and applications
  - 2.1 At the DNA level
    - 2.1.1 Genetic interaction mapping
    - 2.1.2 The ENCODE project
  - 2.2 At the RNA level: transcriptome profiling
    - 2.2.1 Microarrays
    - 2.2.2 SAGE
  - 2.3 At the protein level: protein–protein interactions
    - 2.3.1 Yeast two-hybrid system
    - 2.3.2 AP/MS
  - 2.4 Loss-of-function techniques
    - 2.4.1 Mutagenesis
    - 2.4.2 RNAi
  - 2.5 Functional annotations for genes
    - 2.5.1 Genome annotation
    - 2.5.2 Rosetta stone approach
  - 2.6 Functional genomics and bioinformatics
- 3 See also
- 4 References
- 5 External links



A DNA microarray

## Goals of functional genomics

The goal of functional genomics is to understand the relationship between an organism's genome and its phenotype. The term functional genomics is often used broadly to refer to the many possible approaches to understanding the properties and function of the entirety of an organism's genes and gene products. This definition is somewhat variable; Gibson and Muse define it as "approaches under development to ascertain the biochemical, cellular, and/or physiological properties of each and every gene product",<sup>[1]</sup> while Pevsner includes the study of nongenic elements in his definition: "the genome-wide study of the function of DNA (including genes and nongenic elements), as well as the nucleic acid and protein products encoded by DNA".<sup>[2]</sup> Functional genomics involves studies of natural variation in genes, RNA, and proteins over time (such as an organism's development) or space (such as its body regions), as well as studies of natural or experimental functional disruptions affecting genes, chromosomes, RNA, or proteins.

The promise of functional genomics is to expand and synthesize genomic and proteomic knowledge into an understanding of the dynamic properties of an organism at cellular and/or organismal levels. This would provide a more complete picture of how biological function arises from the information encoded in an organism's genome. The possibility of understanding how a particular mutation leads to a given phenotype has important implications for human genetic diseases, as answering these questions could point scientists in the direction of a treatment or cure.

## Techniques and applications

Functional genomics includes function-related aspects of the genome itself such as mutation and polymorphism (such as single nucleotide polymorphism (SNP) analysis), as well as measurement of molecular activities. The latter comprise a number of "-omics" such as transcriptomics (gene expression), proteomics (protein expression), and metabolomics. Functional genomics uses mostly multiplex techniques to measure the abundance of many or all gene products such as mRNAs or proteins within a biological sample. Together these measurement modalities endeavor to quantitate the various biological processes and improve our understanding of gene and protein functions and interactions.

### At the DNA level

#### Genetic interaction mapping

Systematic pairwise deletion of genes or inhibition of gene expression can be used to identify genes with related function, even if they do not interact physically. Epistasis refers to the fact that effects for two different gene knockouts may not be additive; that is, the phenotype that results when two genes are inhibited may be different from the sum of the effects of single knockouts.

#### The ENCODE project

The ENCODE (Encyclopedia of DNA elements) project is an in-depth analysis of the human genome whose goal is to identify all the functional elements of genomic DNA, in both coding and noncoding regions. To this point, only the pilot phase of the study has been completed, involving hundreds of assays performed on 44 regions of known or unknown function comprising 1% of the human genome. Important results include evidence from genomic tiling arrays that most nucleotides are transcribed as coding transcripts, noncoding RNAs, or random transcripts, the discovery of additional transcriptional regulatory sites, further elucidation of chromatin-modifying mechanisms.

## At the RNA level: transcriptome profiling

### Microarrays

Microarrays measure the amount of mRNA in a sample that corresponds to a given gene or probe DNA sequence. Probe sequences are immobilized on a solid surface and allowed to hybridize with fluorescently labeled “target” mRNA. The intensity of fluorescence of a spot is proportional to the amount of target sequence that has hybridized to that spot, and therefore to the abundance of that mRNA sequence in the sample. Microarrays allow for identification of candidate genes involved in a given process based on variation between transcript levels for different conditions and shared expression patterns with genes of known function.

### SAGE

SAGE (serial analysis of gene expression) is an alternate method of gene expression analysis based on RNA sequencing rather than hybridization. SAGE relies on the sequencing of 10–17 base pair tags which are unique to each gene. These tags are produced from poly-A mRNA and ligated end-to-end before sequencing. SAGE gives an unbiased measurement of the number of transcripts per cell, since it does not depend on prior knowledge of what transcripts to study (as microarrays do).

## At the protein level: protein–protein interactions

### Yeast two-hybrid system

A yeast two-hybrid (Y2H) screen tests a "bait" protein against many potential interacting proteins ("prey") to identify physical protein–protein interactions. This system is based on a transcription factor, originally GAL4,<sup>[3]</sup> whose separate DNA-binding and transcription activation domains are both required in order for the protein to cause transcription of a reporter gene. In a Y2H screen, the "bait" protein is fused to the binding domain of GAL4, and a library of potential "prey" (interacting) proteins is recombinantly expressed in a vector with the activation domain. In vivo interaction of bait and prey proteins in a yeast cell brings the activation and binding domains of GAL4 close enough together to result in expression of a reporter gene. It is also possible to systematically test a library of bait proteins against a library of prey proteins to identify all possible interactions in a cell.

### AP/MS

Affinity purification and mass spectrometry (AP/MS) is able to identify proteins that interact with one another in complexes. Complexes of proteins are allowed to form around a particular “bait” protein. The bait protein is identified using an antibody or a recombinant tag which allows it to be extracted along with any proteins that have formed a complex with it. The proteins are then digested into short peptide fragments and mass spectrometry is used to identify the proteins based on the mass-to-charge ratios of those fragments.

## Loss-of-function techniques

### Mutagenesis

Gene function can be investigated by systematically “knocking out” genes one by one. This is done by either deletion or disruption of function (such as by insertional mutagenesis) and the resulting organisms are screened for phenotypes that provide clues to the function of the disrupted gene.

## RNAi

RNA interference (RNAi) methods can be used to transiently silence or knock down gene expression using ~20 base-pair double-stranded RNA typically delivered by transfection of synthetic ~20-mer short-interfering RNA molecules (siRNAs) or by virally encoded short-hairpin RNAs (shRNAs). RNAi screens, typically performed in cell culture-based assays or experimental organisms (such as *C. elegans*) can be used to systematically disrupt nearly every gene in a genome or subsets of genes (sub-genomes); possible functions of disrupted genes can be assigned based on observed phenotypes.

## Functional annotations for genes

### Genome annotation

Putative genes can be identified by scanning a genome for regions likely to encode proteins, based on characteristics such as long open reading frames, transcriptional initiation sequences, and polyadenylation sites. A sequence identified as a putative gene must be confirmed by further evidence, such as similarity to cDNA or EST sequences from the same organism, similarity of the predicted protein sequence to known proteins, association with promoter sequences, or evidence that mutating the sequence produces an observable phenotype.

### Rosetta stone approach

The Rosetta stone approach is a computation method of de novo protein function prediction, based on the hypothesis that some proteins involved in a given physiological process may exist as two separate genes in one organism and as a single gene in another. Genomes are scanned for sequences that are independent in one organism and in a single open reading frame in another. If two genes have fused, it is predicted that they have similar biological functions that make such coregulation advantageous.

## Functional genomics and bioinformatics

Because of the large quantity of data produced by these techniques and the desire to find biologically meaningful patterns, bioinformatics is crucial to analysis of functional genomics data. Examples of techniques in this class are data clustering or principal component analysis for unsupervised machine learning (class detection) as well as artificial neural networks or support vector machines for supervised machine learning (class prediction, classification). Functional enrichment analysis is used to determine the extent of over- or under-expression (positive- or negative- regulators in case of RNAi screens) of functional categories relative to a background sets. Gene ontology based enrichment analysis are provided by DAVID and Gene Set Enrichment Analysis (GSEA),<sup>[4]</sup> pathway based analysis by Ingenuity<sup>[5]</sup> and Pathway studio<sup>[6]</sup> and protein complex based analysis by COMPLEAT.<sup>[7]</sup>

## See also

- Systems biology
- Structural genomics
- Comparative genomics
- Pharmacogenomics
- MGED Society
- Epigenetics
- Bioinformatics
- Epistasis and functional genomics
- Protein function prediction

## References

1. ^ Gibson G, Muse SV. *A primer of genome science* (3rd ed.). Sunderland, MA: Sinauer Associates.
2. ^ Pevsner J (2009). *Bioinformatics and functional genomics* (2nd ed.). Hoboken, NJ: Wiley-Blackwell.
3. ^ Fields, S.; Song, O. (1989). "A novel genetic system to detect protein-protein interactions". *Nature* **340** (6230): 245–246. Bibcode:1989Natur.340..245F (<http://adsabs.harvard.edu/abs/1989Natur.340..245F>). doi:10.1038/340245a0 (<http://dx.doi.org/10.1038%2F340245a0>). PMID 2547163 (<https://www.ncbi.nlm.nih.gov/pubmed/2547163>).
4. ^ Subramanian A, Tamayo P, Mootha VK, *et al.* (2005). "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1239896>). *Proc. Natl. Acad. Sci. U.S.A.* **102** (43): 15545–50. Bibcode:2005PNAS..10215545S (<http://adsabs.harvard.edu/abs/2005PNAS..10215545S>). doi:10.1073/pnas.0506580102 (<http://dx.doi.org/10.1073%2Fpnas.0506580102>). PMC 1239896 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1239896>). PMID 16199517 (<https://www.ncbi.nlm.nih.gov/pubmed/16199517>).
5. ^ "Ingenuity Systems" (<http://www.ingenuity.com/>). Retrieved 2007-12-31.
6. ^ "Ariadne Genomics: Pathway Studio" (<http://www.ariadnegenomics.com/products/pathway-studio/>). Retrieved 2007-12-31.
7. ^ Vinayagam A, Hu Y, Kulkarni M, Roesel C, *et al.* (2013). "Protein Complex-Based Analysis Framework for High-Throughput Data Sets. 6, rs5 (2013)." (<http://www.flyrnai.org/compleat/>). *Sci. Signal.* **6** (r5). doi:10.1126/scisignal.2003629 (<http://dx.doi.org/10.1126%2Fscisignal.2003629>). PMID 23443684 (<https://www.ncbi.nlm.nih.gov/pubmed/23443684>).

## External links

- Functional genomics: An introduction to the EBI resources on Train OnLine (<http://www.ebi.ac.uk/training/online/course/functional-genomics-introduction-ebi-resources>)
- European Science Foundation Programme on Frontiers of Functional Genomics (<http://www.functionalgenomics.org.uk>)

- MUGEN NoE (<http://www.mugen-noe.org/>) — Integrated Functional Genomics in Mutant Mouse Models
- Nature insights: functional genomics (<http://www.nature.com/nature/insights/6788.html>)
- *Bioinformatics and functional genomics* (<http://www.bioinfbook.org/>) — Companion site for *Bioinformatics and functional genomics, 2nd ed.*
- ENCODE (<http://www.genome.gov/10005107>)
- 4th European Science Foundation Conference on Functional Genomics and Disease (<http://www.esffg2010.org/>)
- [1] ([http://www.ornl.gov/sci/techresources/Human\\_Genome/faq/compngen.shtml#function](http://www.ornl.gov/sci/techresources/Human_Genome/faq/compngen.shtml#function))

Retrieved from "[http://en.wikipedia.org/w/index.php?title=Functional\\_genomics&oldid=607264144](http://en.wikipedia.org/w/index.php?title=Functional_genomics&oldid=607264144)"

Categories: [Molecular biology](#) | [Genomics](#) | [Omics](#)

---

- This page was last modified on 6 May 2014, at 01:31.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.