

Functional genomics

Functional genomics is a field of molecular biology where researchers attempt to understand the complex relationship between genotype and phenotype. While genetics and genomics cover the study of genes and all of an organism's genetic material, respectively, functional genomics aims to answer biological questions about how genes are activated and operate in a dynamic, context-dependent, and synergistic fashion, using a range of genomics and associated 'omics datasets. Since the sequencing of the entire human genome, the challenge has been to characterise the role of genes, understand how they are expressed, the function of proteins, and what impact genetic variation has on these processes. In this first of two policy briefings, we outline what functional genomics is and the approaches used in this field.

Summary

- Functional genomics describes an approach to investigating the activity of genes and gene products, how they are regulated and the consequences of variation on a genome-wide scale on an organism's biology.
- Using datasets from multiple high-throughput technologies (e.g. sequencing and mass spectrometry), researchers can gain information about different aspects of cellular, tissue, organ system and whole organism biology.
- Many large-scale, international projects and consortia exist that aim to comprehensively map genomic function in different cells, tissues, developmental stages and disease states.
- Findings from the field of functional genomics will inform all areas of genomic medicine by furthering our understanding of the impacts of genetic variation and how these relate to disease.

Functional genomics

Beyond genomics

The genome (i.e. DNA), whilst differing between individuals, is static and the same in every cell of a healthy human body. It is the expression of different genes, in different levels, that separates cell types from one another. Therefore, examining different biological processes that occur downstream of the DNA sequence is essential to knowing how genes give rise to biological function and how this process is regulated.

We do not fully understand the role or function of a large proportion of the genome. For example, we still do not fully understand the role of all 20,000 protein coding genes that make up around 1% of the human genome. The vast majority of the genome does not code for proteins but has important roles in regulating gene activity, e.g. turning gene expression on or off, and we are only just beginning to understand its importance.

Understanding function through multiomics

Functional genomics is not defined as a set of technologies or analyses but rather as an approach to investigate different aspects of biology to answer questions such as which genes are transcribed, how this process is regulated, how translation into proteins occurs, what functions the proteins have and their metabolic consequences. Any combination or number of 'omics datasets can be combined to understand genomic function. Due to recent technological advances including high throughput, next generation sequencing, improved protein and small molecule analysis techniques (e.g. mass spectrometry) and enhanced computing power, these investigations can be carried out on a large scale and in combination.

Different technologies can give us information about different processes in specific cells or tissues at a given time, including:

- The epigenome – modifications to DNA or RNA that control gene expression
- The transcriptome – all RNAs
- The proteome – all proteins
- The metabolome – all metabolites, which can include proteins and lipoproteins

Protein-DNA and protein-protein interactions can also be investigated.

Before the advent of next generation sequencing, the function of genes or regulatory sequences was determined one by one in animal or cell models, which is a laborious and slow process. For example, a gene of interest can be inactivated and the biological impacts to the animal or cell can be measured to determine the role of the gene.

Integrated 'omics experiments

Functional genomics investigations can include uncovering the natural variations in DNA, DNA modifications, RNA, proteins and metabolites over time, such as during an organism's development, or through space, such as comparing different organs or tissues. They can also investigate natural or experimental disruptions to these biological processes, which allows comparisons of diseased vs. healthy states, for example.

Functional genomics

Experimentally induced variations

Genome editing tools (e.g. CRISPR-Cas) can be used to systematically induce variations in genetic features in high-throughput studies to characterise the impact of changes to multiple genes, regulatory regions and/or non-coding regions in one experiment. These tools are highly specific and can be designed to target any region of the DNA, RNA and/or induce epigenetic modifications to study their impact on the biology of the cell.

Computational methods

Computational methods for predicting functional impacts of variations at the level of the DNA have become increasingly sophisticated due to the large and comprehensive data resources available.

Machine learning algorithms incorporated into bioinformatics tools have been developed to facilitate variant interpretation. These algorithms learn from labelled genomic data to predict functional impacts such as protein disruption (e.g. Polyphen, Mutation Taster and CADD) and phenotypic expression (e.g. Exomiser and eXtasy) that can be implicated in disease. Importantly, algorithms for predicting the functional consequences of variants that lie outside of protein-coding regions are evolving rapidly.

Functionally annotating the genome

There are many large-scale genome-wide catalogues publicly available as reference databases, with functional annotations, including initial insights on the consequences of variations for cell function and how they relate to phenotypic expression¹.

Functional differences between cells and tissues

Many projects aim to build functional maps across different body systems. Some of these employ a range of assays (e.g. ENCODE project and the GTEx consortium) whilst others focus on data derived from transcriptomics (e.g. FANTOM), epigenomics (e.g. The Roadmap Epigenomics Project) and proteomics (e.g. the Human Protein Atlas) investigations.

Functional changes over time

The HubMAP consortium and the Human Developmental Cell Atlas (a stream of the Human Cell Atlas) aim to deliver comprehensive, high-resolution reference maps of individual human cell types across diverse developmental stages using a range of 'omics technologies including transcriptomics.

Functional characteristics of disease

The Cancer Genome Atlas consortium uses integrated 'omics analysis of over 11,000 tumours from 33 most prevalent forms of cancer. Other projects have looked at complex diseases and the interaction with the microbiome (e.g. The Human Functional Genomics Project) and the epigenetic factors involved in blood disease (e.g. Blueprint).

The Functional Genomics Centre, a collaboration between Cancer Research UK and AstraZeneca, was set up in early 2019 with the aim to accelerate the development of novel treatments for cancer through pooling of expertise and resources. This research centre will use state of the art technologies including CRISPR to enable highly adaptable experiments.

Functional genomics

Limitations and challenges

- The use of cell lines and animal models in many multi-omic studies instead of primary cells/tissues (i.e. derived from patients) presents a major challenge in translating functional genomics findings for application in human disease.
- Where tissue/primary cell-based models have been used they are representative of relatively few individuals – i.e. different tissues from the same individual are used. Whilst data from these studies have cumulatively produced rich catalogues of functional elements across the genome in different cell and tissue types, they have not captured the variation in genome function between individuals. This would enable direct measurement of the genetic contribution to human diversity at the cellular level – e.g. why some people get disease and others do not.
- It is also not clear how accurately the microenvironment is currently modelled to correctly decipher variant impact *ex vivo*. For example, in cancer, tumour initiation, progression, and treatment are dependent on many factors including interactions between diseased and non-diseased cells and their impact on a systemic level. This can be difficult to simulate within model systems.
- Functional genomic resources provide insufficient population and demographic diversity. Typically, as in the GTEx project, the samples are predominantly from a European ancestry with a smaller sample from African ancestry with an age range spanning 20-70 but with an older age bias. The paucity of data from childhood and young adulthood limits the current applicability to the field of paediatric diseases and many common diseases that develop in early life¹.

The goal of functional genomics is to provide a comprehensive, annotated map of the downstream effects of all coding and non-coding parts of the genome. Due to material and resource constraints, and technological limitations it is currently not feasible to perform all types of functional genomics assays in every biological context and every biological sample of interest. However, the continued investigation into functional genomics is fundamental to achieving the goals of personalised medicine.

References

1. Ardlie KG and Guigo R. Data Resources for Human Functional Genomics. *Curr Opin Syst Biol*, 2017.1: p. 75-79

Authors: Dr Sarah Cook
Published: November 2020