MODELLING — OPINION

Cellular genomics for complex traits

Emmanouil T. Dermitzakis

Abstract | Recent developments in the collection and analysis of cellular multilayered data in large cohorts with extensive organismal phenotyping promise to reveal links between genetic variation and biological processes. The use of these cellular resources as models for human biology — known as 'cellular phenotyping' — is likely to transform our understanding of the genetic and long-term environmental influences on complex traits. I discuss the advantages and caveats of a deeper analysis of cellular phenotypes in large cohorts and assess the methodological advances, resource needs and prospects of this new approach.

Dissecting the aetiology of complex traits and disease is one of the essential steps required to obtain a better understanding of human phenotypic variation and to realize the potential for disease treatment[1–3]. There are two main obstacles towards fully characterizing the biological basis for variation in multifactorial phenotypes. First, current studies that comprehensively discover genetic variants associated with complex traits simply reveal statistical links between variation in a region of the human genome and the respective phenotype. This statistical link does not offer any direct biological insight; even when the gene is known, the biological effects that occur downstream of the genetic variant are not.

A second challenge is to measure how a phenotype is affected by variability in environmental exposures[1,2]. This is an arduous task because the environment is a combination of strong effects (for example, smoking) and subtle effects (for example, diet and atmospheric quality) that occur over the lifetime of an individual and that result from exposures that are notoriously difficult to record accurately through questionnaires. However, first principles suggest that all environmental exposures that have long-term effects on phenotypic variation are manifested in one of three ways: in our body structure (for example, obesity); in somatic mutations that may have a long term effect (for example, cancer); and in epigenetic marks in some of our cells. Any environmental exposure that does not leave a mark can be considered as having negligible effects on phenotypic variation. Therefore, the consequences of environmental effects can be measurable and quantifiable in human cells and tissues sampled from biopsies or autopsies without necessarily knowing the nature of the specific environmental cause. To gain a complete picture of the causes of our phenotypic variation, we need to document the consequences of genetic variants, both at the spatial level (among cells and tissues) and at the temporal level (over a period of time as individuals age)[3].

Here I argue that both challenges discussed above could be solved by developments in cellular phenotyping, which allows for an analysis of the molecular consequences of functional genetic variants as well as an assessment of environmental effects at the cellular level[1]. By cellular phenotyping, I mean the quantification of processes and functions that occur at the cellular level (for example, gene expression, metabolite production or mitotic rate) and represent, to a large extent, the state of the cell. These phenotypes can then be treated and analysed as conventional phenotypes, such as height or cholesterol levels. Many studies have addressed the genetics of cellular phenotypes, including gene expression[4,5], protein abundance[6] or response to challenges[7]. The task is to move towards using a multilayered approach in which we integrate multiple cellular phenotypes to understand the overall consequences of genetic and environmental variance.

In this article, I explain in detail the concept of cellular phenotyping and describe the available cellular resources, the type of phenotypes that can be measured in cells and the future requirements of this field. I then discuss the methodologies that have been described, as well as the upcoming approaches for the integrative analysis of cellular phenotypes and how these are likely to transform our understanding of biology. In particular, I highlight the challenges of coordinating the development of resources and analytical methods. Although this article is primarily focused on understanding human biology, a similar discussion is applicable to model organisms, such as mice, fruitflies or worms.

What is cellular phenotyping?

Cellular phenotyping is the attempt to quantify the output of processes and molecular interactions that occur in the cell. The main difference between such phenotypes and organismal phenotypes, such as height or disease status, is that the reference living unit is the cell as opposed to the organism. The phenotypes studied include gene expression levels, protein abundance, metabolite concentration within the cell or cell extract and pattern of cell division. There are also phenotypes that, although recorded in specific cells, are modified by other processes in the body or that are contributed by multiple tissues or organs. These phenotypes consist of metabolic products that can be quantified and analysed by tools developed by the emerging field of metabolomics. Such phenotypes are measured in blood or other body tissues and have more systemic properties; however, they are closely connected to cellular function and can be liberally categorized as cellular phenotypes. Although the basic components of this approach have been around for a decade or more, the integrated approach and the attempt to improve the degree of resources that are available have only recently been appreciated.

Cellular resources

For any phenotype, it is easy to pinpoint the affected tissue or organ, such as fat tissue in the case of body mass index (BMI) or the pancreas in type II diabetes. However, a correct assessment of the effect of genetic variants on phenotype crucially depends on identifying the location in the body that is affected by functional differences in the causal genetic variant. The link between the causal and affected tissues is not always obvious. For example, in obesity, the affected tissue is fat, but the causal tissue is, in some cases, the hypothalamus, in which known variants modify the control of appetite (for example, variants in melanocortin 4 receptor (*MC4R*))[8].

Our ability to identify the causal tissue in an unbiased way is crucial to our learning of the new biology of complex traits. A shift from a 'candidate tissue approach' — wherein we pretend that we know the causal tissue — to unbiased tissue studies will reveal new dimensions of biological effects. This is similar to the shift that took place in disease trait mapping by moving from candidate gene approaches to unbiased, genome-wide studies[9]. The problem of a candidate study is that only interrogating a small subset of tissues exaggerates the weight of a positive finding. It is not possible to be certain about whether the tissue that has been identified as causal is the only tissue in which the effect is present; this approach also makes it difficult to interpret negative findings. By contrast, by interrogating the molecular impact of variants that are associated with organismal traits in large collections of cell lines and tissues, we can detect where these variants are creating genotype-dependent functional profiles and can therefore identify tissues that are likely to be causal for a trait (BOX 1).

Unbiased tissue studies require extensive cellular resources with in-depth genetic and functional information, as described below.

Primary tissues. Primary tissues are sampled by autopsies and surgical biopsies. A large range of tissues can be sampled in this way; however, the stability of DNA, RNA and other molecules after sampling or after death can vary between cells. But if tissues are collected and have been handled and stored under optimal conditions, they integrate both genetic and environmental effects and are informative of the state of the organism at time of collection and of lifetime exposures. An approach of sampling extensive tissue samples from large number of individuals has been adopted by the US National Institutes of Health (NIH) Roadmap Genotype–Tissue Expression (GTEx) project.

Immortalized and primary cell lines. Cell lines can either be primary cultures (that is, they are isolated directly from the organism) or cells that have been transformed for long-term availability. The creation of a cell line allows the isolation of a specific cell type from a population of cell types in a tissue; however, a cell that has been isolated in this way also loses the effects of its endogenous environment. Although most of the epigenetic marks remain, even after transformation[10,11], some environmental information is erased. This especially applies to transient exposures that occur around the time of collection and to somatic mutations, which may be stochastically lost if they were only present in a subset of cells. Despite these caveats, cell lines are similar to primary tissues; they differ by being a sustainable source, capturing the effects of genetic variability well and allowing manipulation and studies of exposure to various agents (see below).

Induced pluripotent stem cells and derived cell types. Fully reprogrammed and pluripotent cells (induced pluripotent stem cells (iPSCs)) can be derived from various sources, such as skin cells or haematopoietic stem cells[12], and differentiated into other cell types. iPSCs that have been derived from a given individual should, in theory, carry no lifetime epigenetic information of that individual. They should, however, retain all of the germline and somatic changes of the population of cells from which they are derived. Ultimately, iPSCs only carry DNA information from their source cells, although it should also be noted that additional mutations may occur in the derivation process[13]. One of the unique properties of iPSCs is the accessibility to the differentiation process, which may allow the interrogation of variability in the stages of differentiation and the effect of genetic variation on them.

All of the above resources cover information ranging from purely genetic effects (in the case of iPSCs) to integrated genetic and environmental effects (in the case of tissues). Depending on the focus of the study and whether both genetics and environment are of interest, or whether only one of them is, different cellular resources could be used. Of course, culture cells (namely, cell lines and iPSC-derived cells) allow the manipulation and exposure of cells to stimuli, as I describe below (also see BOX 1).

Cellular phenotypes

Although the cell is a well-defined and small environment, it contains several different cellular phenotypes, each with different properties.

Nuclear-based phenotypes. These phenotypes usually refer to molecular interactions that take place on the genome or chromatin and are therefore genome-wide and relevant to all cell types and tissues. Nuclear phenotypes include levels of gene expression, epigenetic modifications and protein binding to DNA and/or chromatin. A number of studies on these phenotypes have discovered expression quantitative trait loci (eQTLs)[14] and chromatin and transcription factor QTLs[4,11,14–19] (FIG. 1a). These QTLs reflect the local effects of genetic variants in biochemical functions of the genome and are usually additive in nature. Studying a collection of such variants may result in a greater understanding of their global effects on cell function.

Box 1 | **Applications of cellular phenotyping**

Linking phenotypic variation to molecular and cellular phenotypes
Traditionally, it has been possible to measure phenotypes at the organismal level, but the molecular causes and effects of such phenotypes have remained elusive.

Functionally annotating disease variants
Although we are able to infer the involvement of variants in common disease statistically, the ability to integrate them with effects on cellular phenotypes brings direct biological effects into the foreground and highlights potential ways of intervening against the effects of pathogenic variants.

Pharmacogenomics and toxicity studies
The ability to assay cell lines and to expose them to certain agents and drugs allows the identification of the molecular toxic consequences of such exposures (which are probably not readily visible at the organismal level) and of the variants that modify these consequences.

Modelling the effects of the environment
By treating cells with natural environmental agents, we are able to measure the consequences of certain exposures and to obtain clues as to the tissue and manner by which they are having an effect on the organism.

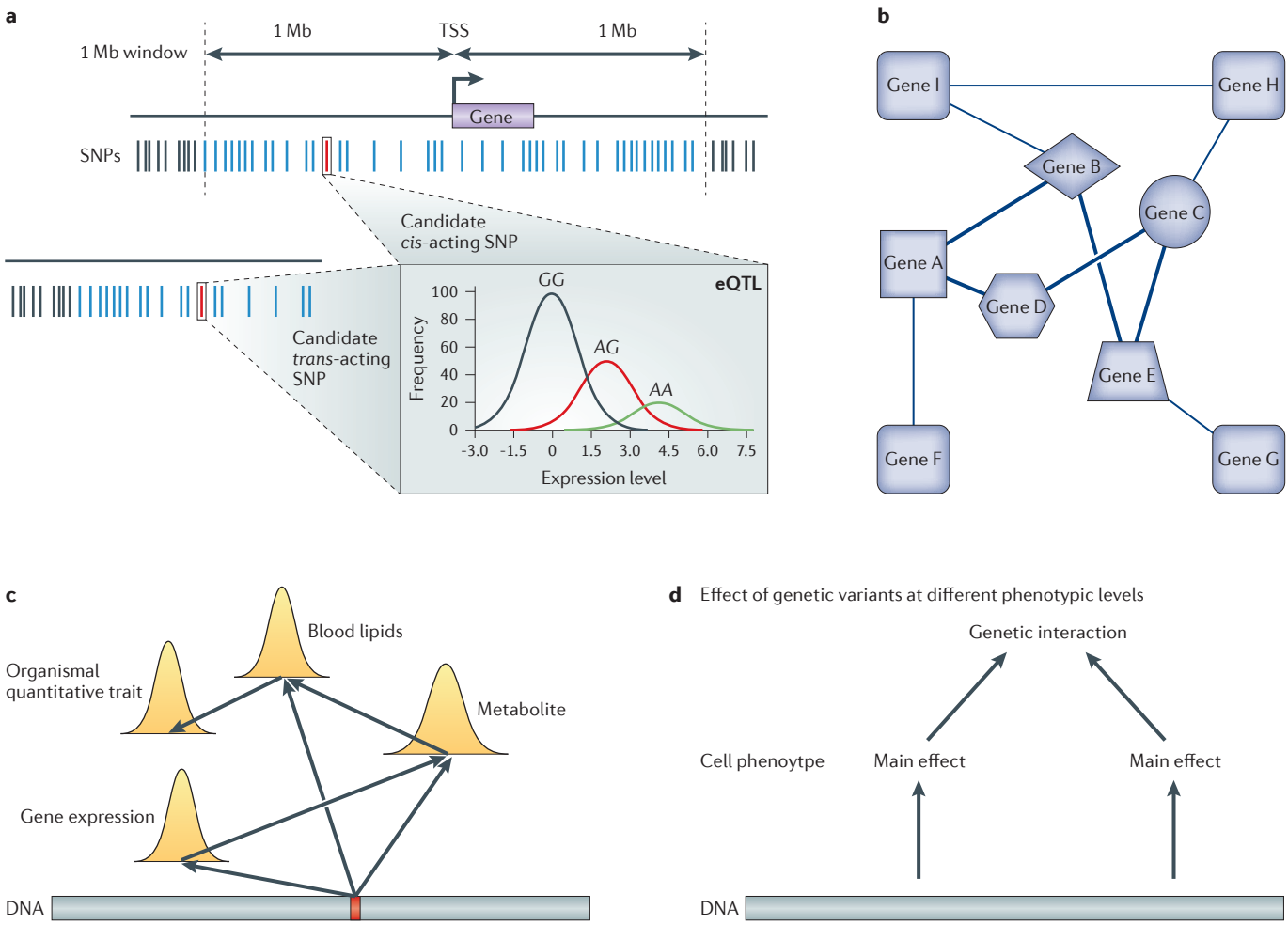

Figure 1 | **Analytical approaches made possible by using cellular phenotypes. a** | Identifying quantitative trait loci (QTLs) for molecular phenotypes: the link between a *cis*- and a *trans*-acting genetic variant and a molecular phenotype, such as gene expression. Detection of variants that act in *cis* begins by focusing on a small genomic window around the transcription start site (TSS). By contrast, detection of variants that act in *trans* involves testing the whole genome. Blue and red lines represent the SNPs to be tested in *cis* or in *trans*. The distribution shows the density plots of gene expression for the red SNPs that are stratified by genotypes *AA*, *AG* or *GG*. **b** | Inference of gene interaction networks. The aim is to infer the relationship and molecular interactions between genes either directly (in the form of protein–protein interactions) or indirectly (by modulation of one gene by another gene). These networks represent relationships between genes that usually lack directionality and can be represented by modules of genes with increased levels of interaction (as illustrated by the bold lines linking genes A–E) relative to other neighbouring genes with lower levels of interaction (as illustrated by the thin lines linking genes A–F). **c** | Inference of phenotype interaction networks and their integration with genetic information. The figure highlights the relationship between molecular phenotypes and how the variance of one phenotype has an impact on the variance of another. Directionality can be inferred when a genetic variant (that is unmodified in the lifetime of an individual) influences the phenotypic variance of one or more of them. **d** | Additive and interaction effects of genetic variants. A set of genetic variants that seem to have a linear (additive) effect on molecular phenotypes may appear to operate under epistatic interaction at an organismal phenotype level.

Cytoplasmic phenotypes. These phenotypes are defined with respect to the molecular composition of the cytoplasm or the interactions that take place in this cellular compartment. They are either quantities of molecules that stem from genomic phenotypes (for example, protein abundance[6]) or certain metabolites or by-products that result from signalling or biochemical cascades. Usually, the signalling or biochemical cascades are pathway-specific or linked to a certain cell type or tissue, or they are relevant to the function of only a few genes[20,21]. These properties make them more informative about specific processes in the cell and can represent the synergistic effect of many genomic phenotypes, such as gene expression and protein abundance.

Although the subcategorization described above is useful for defining a phenotype, it is desirable to obtain a combination of cytoplasmic phenotypes in a real experiment. The ability to integrate local effects on the genome with specific cellular processes allows us to understand how genome variation modulates function in a cell-type-dependent manner.

The phenotypic definitions given above can be classified further by the state of the cell when the measurement takes place.

Steady-state measurements. These are measurements that are taken under 'normal' phsyiological conditions, in which cells are simply dividing in culture or as part of a tissue biopsy. One caveat with such a definition is that 'normal' is a perceived state, and it is frequently unknown whether cells are under any type of stress or exposure to an agent. These issues are more

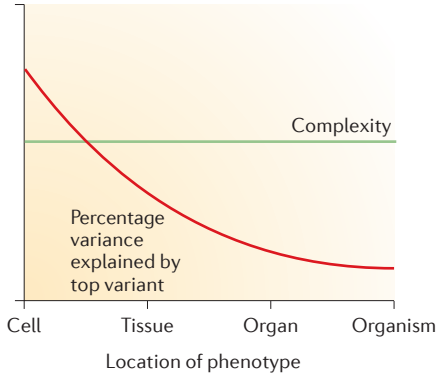

Figure 2 | **Challenges of dissecting genetic variance.** Expectation of phenotypic variance explained by each variant as we move away from the primary molecular effect of DNA variant. Even for cellular phenotypes, we may be able to detect the strongest variant fairly easily; however, the dissection of most of the genetic variance is as difficult a problem as the dissection of organismal phenotypes.

relevant in tissue samples, in which we have less control over environmental factors. However, any stress or exposure is usually random with respect to the variables of interest (for example, genotype) and are more likely to reduce statistical power than to generate false-positive signals. A well-known example of an exposure is the effect of copy number of the Epstein–Barr virus (EBV) in lymphoblastoid cell lines (LCLs): the copy number of EBV is likely to have effects on phenotype, such as on gene expression but, as far as we know, these effects are not genotype-dependent[7].

Measurements in challenged cells. Cells, particularly those in cell culture, can be challenged with various agents, such as pathogens, small molecules, drugs, toxic agents or even simply endogenous agents (for example, gene overexpression). The response of the cell that is conditional on the genotype is highly informative of what happens *in vivo*[7,22,23]. Challenges allow certain obscure signals to rise up in effect size and to become more visible to genetic analysis. In addition, they allow a direct assessment of the effect of certain molecules in the cell and the dissection of the molecular interactions responsible for the cellular response to any particular agent.

The availability of assays that challenge cell function allows the simultaneous agnostic exploration of genetic effects in the cell with hypothesis-driven studies in which certain conditions are simulated in culture. In a study on LCLs, cells that were exposed to radiation led to the identification of genetic variants whose function was conditional on exposure to radiation, revealing effects that would remain unobserved if cells had been tested in the steady state[24]. In another study on LCLs, the authors looked at genetic predictors of drug response by using cell lines and gene expression data[23]. In addition, even if it is possible to observe the molecular effect of a variant (for example, level of gene expression), it is not possible to assess some of its phenotypic consequences unless a challenge is introduced.

Analysing multilayer cellular data

Challenges of data integration. The challenges and value of incorporating cellular phenotypic data in the analysis of complex traits lie in integrating these data to derive new biology insights and to unravel complexity. The problem of integrating multilayered data is older than the ambition to generate the systematic data sets described above. However, in most of these older cases, the solutions and methodologies relied on integrating sparse data sets and making inferences about the biological relationship between genes without bespoke experimental designs (FIG. 1b). Although this literature provided useful insight into certain patterns of biological information, it also generated many false-positive signals or signals that were difficult to validate or replicate in the relevant conditions of discovery.

A more realistic and more comprehensive approach towards integrating data is now popular in which specific hypotheses are tested in comprehensive and almost complete data sets, and this generates opportunity for replication in similar data settings. This has been facilitated by the development of tools for gathering transcriptomic, proteomic and metabolomic data in important sets of samples and cohorts[25,26]. But the problem of integrating multilayered data is not yet solved: converting the available data into a biological model by which information is transmitted between the different layers requires the realization of two important properties of phenotypes. The first is that there are no levels of low and high complexity, but there are simply levels of different nature of complexity, whether genetic or otherwise (FIG. 2). The data layers are interacting such that, for example, a gene expression phenotype affects disease risk (that is, it has a causal expression effect) and, in turn, the disease manifestation results in the differential expression of other genes (that is, it has reactive effects)[27,28]. Second, interactions between layers are not linear but are in the form of networks, where multiple parallel and independent effects may contribute to the variability of a cellular or organismal phenotype. These two reasons make the inference of the transmission of the signal complex, as simple pairwise correlations, even in cases in which directionality is known, are not sufficient to piece together a cascade of effects. Current methods make an attempt to integrate multiple phenotypes simultaneously, but because we do not know enough about the nature of the networks, we are far from being able to do so in a manner that reveals clear biological signals.

Emerging approaches for integrating data

One of the breakthroughs in the analysis of multilayered data has been the use of genetic information together with large collections of phenotypic data points. Because the genotype is invariable in the lifetime of an individual, it can be considered to be an anchor of directionality, thus allowing the study of genetic perturbations in human populations. This has opened the field to a new subdiscipline called 'systems genetics', which aims to integrate data at a systems level with genetic variation[29] (FIG. 1c). A number of studies[26,30] have demonstrated the value of such approaches either by integrating large collections of data sets or by simply correlating one or two phenotypic data sets with genome-wide genetic variation. From the initial attempts of such approaches, it has become apparent that methodologies in which the analysis steps are a black box are less useful and less informative for elucidating biological signals. On the contrary, stepwise methods in which the assessment of the transmission of the variance is performed from one layer to the next reveal tremendously interesting insight into how certain biological interactions modulate molecular, cellular and organismal phenotypes. For instance, studies have revealed that genetic variation has a substantial influence on both epigenetic variation and gene expression, providing insight not only into the molecular effects of genetic variants but also into informing about the specific molecular interactions that are involved in gene regulation[11,31]. In addition, the use of multiple sources of data from the same people — such as genotypes, adipose gene expression and disease risk in the deCODE cohorts from Iceland — provided new insight into causality effects of genetic variants and the biological processes that mediate lipid levels and cardiovascular disease[26].

Another key development has been the use of reference data sets to derive models of cellular phenotypes and their genetic basis and to integrate them with specific phenotypic attributes in non-reference cohorts. This is an important area as, for large cohorts, we will probably not be able to gather simultaneously deep cellular information as well as disease or complex trait phenotypic information. In this case, we can infer the relationship of two phenotypic attributes through their correlation with the same genetic variant. An example of this has been the integration of variants discovered in large case–control genome-wide association studies (GWASs) with eQTLs discovered in independent small sample size cohorts and samples, where it was reported that there is a high enrichment of GWAS variants in overlapping *cis*-eQTLs[32,33]. These studies provide mechanistic dimensions to the statistical signals and suggest that a large fraction of common disease variants are likely to be *cis*-regulatory.

Finally, analysis of multilayered data will facilitate the dissection of what we currently call genetic interactions. It is well-known that many of the biological effects of genetic variants are context-dependent. Although we know that context-dependent effects exist with respect to sex, tissue or phenotypic status, we still have not discovered conditional effects (interactions) among genetic variants. It is hypothesized that genetic interactions harbour some degree of genetic variability in complex traits, although this is hard to disentangle owing to the statistical burden of the number of tests that are needed if all genetic variants are tested. However, it is becoming apparent that genetic interactions that occur in one phenotypic layer can be separated into two or more linear effects in another layer (FIG. 1d). With the availability of cellular phenotypic data, these marginal effects will be detected, and the bottom-up approach from cellular to organismal phenotypic analysis will render such effects much easier to identify. As an example, it was recently reported[34] that there is a signal of purifying natural selection at haplotypes that carry rare (and probably deleterious) protein-coding variants together with the high-expressing allele of an eQTL. The variants involved are of two separate types, each of which has an independent effect on protein function and gene expression levels. Although this type of interaction has not yet been explicitly tested in disease cohorts, it indicates that knowledge of the underlying interaction framework of the two variants allows for the hypothesis-driven assessment of their joint and conditional effect.

Glossary

Additive
In the context of a genetic effect, the linear relationship between the replacement of an allele and its effect on the phenotype.

Expression quantitative trait loci
(eQTLs). Loci at which genetic allelic variation is associated with variation in gene expression.

Induced pluripotent stem cells
(iPSCs). Cells that are derived from somatic cells by 'reprogramming' or de-differentiation that is triggered by the transfection of pluripotency genes, which alters the somatic cells to a state that is similar to that of embryonic stem cells.

Marginal effects
Also known as main effects, these are the effects of a variable assuming no dependency or conditionality of other variables.

Metabolomics
The directed use of quantitative analytical methods for analysing the entire metabolic content of a cell or organism (that is, the metabolome) at a given time.

Purifying natural selection
Natural selection that results in the reduction or elimination of the frequency of alleles with negative fitness effects.

Conclusion

The use of cellular resources and cellular phenotyping promises to make major breakthroughs in understanding the biology of complex traits and disease. It also offers a realistic alternative to the use of model organisms for resolving biological problems, such as understanding the specific effects of human genetic variation. The transformation of human cellular resources into a model system for the study of human biology will bring unprecedented resolution but also comes with caveats and realizations: the human body and interactions in it are more complex than we had hoped for in the effort to understand complex traits and disease. Coordinated efforts are now either in progress or are needed to create the relevant resources for such biological exploration. Cohorts that are under development or cohorts with the capability of recalling study participants should integrate the sampling of biological material for cellular analysis. In addition, funding bodies, which have largely resisted the decision to fund resource building, should realize the value and increased effort that such resources need versus the more traditional collection of DNA samples. The data analysis tools also need further development to maintain the balance between complex computational processes and realistic biological hypotheses. We are already able to analyse multiple cell types in co-cultures of multiple cells[35] and to perform three-dimensional modelling of tissues. The speed at which comprehensive experimental procedures are integrated with realistic and therefore complex computational methods promises an exciting future in biology and medicine, in which synergy among disciplines will be the norm.

Emmanouil T. Dermitzakis is at the Department of Genetic Medicine and Development, University of Geneva Medical School, 1 rue Michel-Servet, 1211 Geneva, Switzerland; and the Institute of Genetics and Genomics in Geneva (iGE3), Geneva, Switzerland.
e-mail: emmanouil.dermitzakis@unige.ch

doi:10.1038/nrg3115
Published online 14 February 2012

1. Gibson, G. The environmental contribution to gene expression profiles. *Nature Rev. Genet.* **9**, 575–581 (2008).
2. Idaghdour, Y., Storey, J. D., Jadallah, S. J. & Gibson, G. A genome-wide gene expression signature of environmental geography in leukocytes of Moroccan Amazighs. *PLoS Genet.* **4**, e1000052 (2008).
3. Dermitzakis, E. T. From gene expression to disease risk. *Nature Genet.* **40**, 492–493 (2008).
4. Cheung, V. G. *et al.* Mapping determinants of human gene expression by regional and genome-wide association. *Nature* **437**, 1365–1369 (2005).
5. Stranger, B. E. *et al.* Genome-wide associations of gene expression variation in humans. *PLoS Genet.* **1**, e78 (2005).
6. Foss, E. J. *et al.* Genetic basis of proteome variation in yeast. *Nature Genet.* **39**, 1369–1375 (2007).
7. Choy, E. *et al.* Genetic analysis of human traits *in vitro*: drug response and gene expression in lymphoblastoid cell lines. *PLoS Genet.* **4**, e1000287 (2008).
8. Loos, R. J. *et al.* Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nature Genet.* **40**, 768–775 (2008).
9. Montgomery, S. B., Lappalainen, T., Gutierrez-Arcelus, M. & Dermitzakis, E. T. Rare and common regulatory variation in population-scale sequenced human genomes. *PLoS Genet.* **7**, e1002144 (2011).
10. Caliskan, M., Cusanovich, D. A., Ober, C. & Gilad, Y. The effects of EBV transformation on gene expression levels and methylation profiles. *Hum. Mol. Genet.* **20**, 1643–1652 (2011).
11. McDaniell, R. *et al.* Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* **328**, 235–239 (2010).
12. Yamanaka, S. Elite and stochastic models for induced pluripotent stem cell generation. *Nature* **460**, 49–52 (2009).
13. Gore, A. *et al.* Somatic coding mutations in human induced pluripotent stem cells. *Nature* **471**, 63–67 (2011).
14. Montgomery, S. B. & Dermitzakis, E. T. From expression QTLs to personalized transcriptomics. *Nature Rev. Genet.* **12**, 277–282 (2011).
15. Kasowski, M. *et al.* Variation in transcription factor binding among humans. *Science* **328**, 232–235 (2010).
16. Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773–777 (2010).

17. Pickrell, J. K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768–772 (2010).
18. Stranger, B. E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–853 (2007).
19. Gaulton, K. J. *et al.* A map of open chromatin in human pancreatic islets. *Nature Genet.* **42**, 255–259 (2010).
20. Putoux, A. *et al. KIF7* mutations cause fetal hydrolethalus and acrocallosal syndromes. *Nature Genet.* **43**, 601–606 (2011).
21. Kim, S. *et al.* Nde1-mediated inhibition of ciliogenesis affects cell cycle re-entry. *Nature Cell Biol.* **13**, 351–360 (2011).
22. Duan, S. *et al.* Expression and alternative splicing of folate pathway genes in HapMap lymphoblastoid cell lines. *Pharmacogenomics* **10**, 549–563 (2009).
23. Gamazon, E. R., Huang, R. S., Cox, N. J. & Dolan, M. E. Chemotherapeutic drug susceptibility associated SNPs are enriched in expression quantitative trait loci. *Proc. Natl Acad. Sci. USA* **107**, 9287–9292 (2010).
24. Smirnov, D. A., Morley, M., Shin, E., Spielman, R. S. & Cheung, V. G. Genetic analysis of radiation-induced changes in human gene expression. *Nature* **459**, 587–591 (2009).
25. Dimas, A. S. & Dermitzakis, E. T. Genetic variation of regulatory systems. *Curr. Opin. Genet. Dev.* **19**, 586–590 (2009).
26. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
27. Schadt, E. E. Molecular networks as sensors and drivers of common human diseases. *Nature* **461**, 218–223 (2009).
28. Schadt, E. E. *et al.* An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genet.* **37**, 710–717 (2005).
29. Mackay, T. F., Stone, E. A. & Ayroles, J. F. The genetics of quantitative traits: challenges and prospects. *Nature Rev. Genet.* **10**, 565–577 (2009).
30. Suhre, K. *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **477**, 54–60 (2011).
31. Bell, J. T. *et al.* DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* **12**, R10 (2011).
32. Nica, A. C. *et al.* Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* **6**, e1000895 (2010).
33. Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
34. Lappalainen, T., Montgomery, S. B., Nica, A. C. & Dermitzakis, E. T. Epistatic selection between coding and regulatory variation in human evolution and disease. *Am. J. Hum. Genet.* **89**, 459–463 (2011).
35. Bouzakri, K. *et al.* Bimodal effect on pancreatic β-cells of secretory products from normal or insulin-resistant human skeletal muscle. *Diabetes* **60**, 1111–1121 (2011).

Acknowledgements
The author would like to thank the Louis-Jeantet Foundation and the Swiss National Science Foundation for support.

Competing interests statement
The author declares no competing financial interests.

FURTHER INFORMATION
Author's homepage: http://funpopgen.unige.ch
***Nature Reviews Genetics* Series on Modelling:** http://www.nature.com/nrg/series/modelling
NIH Roadmap Genotype-Tissue Expression (GTEx) project: http://commonfund.nih.gov/GTEx
ALL LINKS ARE ACTIVE IN THE ONLINE PDF