

# Functional genomics bridges the gap between quantitative genetics and molecular biology

Tuuli Lappalainen<sup>1,2</sup>

<sup>1</sup>New York Genome Center, New York, New York 10013, USA; <sup>2</sup>Department of Systems Biology, Columbia University, New York, New York 10032, USA

Deep characterization of molecular function of genetic variants in the human genome is becoming increasingly important for understanding genetic associations to disease and for learning to read the regulatory code of the genome. In this paper, I discuss how recent advances in both quantitative genetics and molecular biology have contributed to understanding functional effects of genetic variants, lessons learned from eQTL studies, and future challenges in this field.

Most of human genetics research falls under two main questions: What are the genetic origins of variation in human disease and other traits? How does the blueprint of the human genome function to give rise to a living individual? These questions have different historical roots—in quantitative or medical genetics and molecular biology, respectively—as well as different molecular and statistical methods, and thus for decades they have been largely distinct areas of research. However, a question of increasing importance for understanding the human genome lies at their intersection: What are the functional effects of genetic variants across the human genome?

The study of the evolutionary origins of human genetic variation and its contribution to human disease and traits has its origins in quantitative, statistical, and population genetics. Advances in high-throughput genotyping and sequencing technologies during the past 10 years have led to tremendous progress in this field, with the HapMap and 1000 Genomes projects (The International HapMap Consortium et al. 2007; The 1000 Genomes Project Consortium 2012) creating the foundation for hundreds of genome-wide association studies (GWAS) and now also rare variant analyses in the context of both common and rare diseases (Bamshad et al. 2011; Lee et al. 2014). However, these maps of genetic associations to disease do not give us direct information of the function of these variants: how they perturb the biology of the genome, the cell, and eventually the organism to affect disease risk—or from a population genetics perspective, to affect different selective pressures. Without such understanding, the information from genetic association studies will yield little benefit to human health.

On the other side, understanding the mechanistic function of the human genome—as well as genomes of other species—has always been one of the fundamental questions of molecular biology. During the past five years, the approach has become genome-wide via the development of diverse high-throughput sequencing assays, applied to multiple cell types. Projects such as ENCODE (The ENCODE Project Consortium 2012), the Epigenomics Roadmap (Roadmap Epigenomics Consortium 2015), and FANTOM (The FANTOM Consortium and the RIKEN PMI and CLST (DGT) 2014) have produced large catalogs of functional elements in the genome—or more accurately, some genomes, since naturally,

there is no archetype of the human genome. These studies do not typically capture variation in genome function among individuals, and the contribution of genetic differences in variation between samples is often ignored in study design. Thus, while these resources are used to annotate the putative regulatory function of genetic variants, this is done via indirect inference rather than direct measurement of genetic contribution to human phenotype diversity at the cellular level.

The need to bridge conventional quantitative genetics and functional or molecular genetics has now become widely acknowledged (Fig. 1). The concept is not new—medical genetics has a long history of characterizing cellular effects of disease-causing mutations. However, the development of genome-wide methods now allows systematic high-throughput analysis, which is eventually more cost-efficient and informative of generalizable patterns than laborious locus-specific characterization. High-throughput analysis, with scalable and robust molecular assays, careful statistical analysis, and deep biological interpretation, are essential to achieve the future goal of being able to accurately read the genetic code, i.e., predict functional and phenotypic effects of genetic variants.

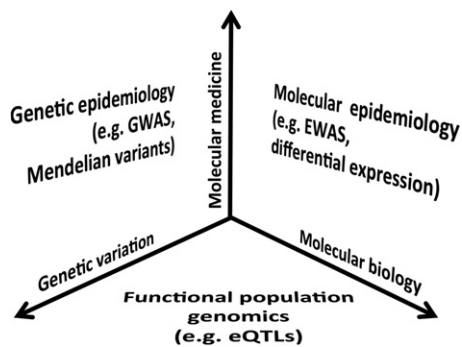
## Mapping regulatory variation by QTL approaches

Expression quantitative trait locus (eQTL) analysis has been the trailblazer in genome-wide functional population genomics. First applied in humans in the mid-2000s (Cheung et al. 2005; Stranger et al. 2007), associating genotypes to gene expression levels in population samples has become a mainstream approach to map variants that affect gene expression levels in *cis* (e.g., in Emilsson et al. 2008; Montgomery et al. 2010; Pickrell et al. 2010; Grundberg et al. 2012; Lappalainen et al. 2013; Battle et al. 2014; GTEx Consortium 2015); for a recent review, see Albert and Kruglyak (2015). Results from GWAS studies have been a major motivator for this work: 80% of genetic associations to common diseases are outside coding regions, which highlights the necessity of understanding regulatory variation (Farh et al. 2015). By now, eQTL studies have uncovered >10,000 genes with eQTLs, demonstrating that common regulatory variants are extremely widespread in the genome (Lappalainen et al. 2013; Battle et al. 2014; GTEx Consortium 2015). This has allowed us

**Corresponding author:** [tlappalainen@nygenome.org](mailto:tlappalainen@nygenome.org)

Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.190983.115>. Freely available online through the *Genome Research* Open Access option.

© 2015 Lappalainen This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.



**Figure 1.** Intersections of fields analyzing genetic variation, molecular biology, and medicine. GWAS and EWAS stand for genome-wide and epigenome-wide association studies, respectively, and eQTL is an abbreviation of expression quantitative trait loci.

to learn properties of proximal regulatory variants affecting gene expression in *cis*: They show widespread sharing across populations (Stranger et al. 2012), they are enriched for targets of positive natural selection (Fraser 2013; Grossman et al. 2013), and they are often located in promoter and enhancer regions but also, e.g., in 3' UTRs (Lappalainen et al. 2010; Gaffney et al. 2012; Battle et al. 2014; GTEx Consortium 2015).

Disease-associated variants are expected to impact cellular phenotypes that ultimately underlie the change in disease risk; indeed, several studies have shown an overrepresentation of eQTLs among GWAS loci (Nica et al. 2010; Nicolae et al. 2010). In hundreds of GWAS loci, eQTL associations have allowed the pinpointing of the GWAS variant to the likely target gene; annotation analyses sometimes indicate specific regulatory mechanisms, and tissue- or cell-type-specific eQTL data can point to tissue-specific mechanisms of disease etiology. For example, a genetic variant rs633185 with association to a QT interval is in high linkage disequilibrium with an eQTL that is particularly active in the heart but not in most other tissues (GTEx Consortium 2015). However, showing that the eQTL and GWAS association signals in the same locus are driven by the same causal variant, rather than randomly overlapping, is not trivial despite several proposed statistical methods (Nica et al. 2010; Giambartolomei et al. 2014). Even when such statistical evidence is solid, real proof of shared causality cannot be obtained without experimental perturbations in cell lines and/or model organisms. Furthermore, most GWAS hits in noncoding regions still remain unexplained by current eQTL catalogs, motivating further research—both more comprehensive eQTL analysis and other approaches (Farh et al. 2015).

A key feature of regulatory genetic effects is its context-specificity, i.e., varying effects of a given variant due to differences in the surrounding cellular or genomic environment. This is an area of intensive study, as many key questions are currently unknown: how widespread such variable effects are, what the key mechanisms are, and what consequences are at the level of the organism. Several studies have provided insight into how the effects of *cis*-regulatory variants can be modified by tissue-specificity, systemic effects such as sex, and cellular stimuli mimicking environmental effects (Dimas et al. 2012; Ye et al. 2014; GTEx Consortium 2015). The largest ongoing project in this domain is the Genotype Tissue Expression (GTEx) project (GTEx Consortium 2013, 2015), with analysis of genotype and RNA sequencing data, as well as other assays, eventually from over 30 tissues from 900 individuals. This

project is building a foundation of gene expression and eQTL variation across human tissues in the normal population and provides an unparalleled resource for the scientific community. However, the primary tissue samples in this and many other projects consist of multiple cell types, and further characterization of the architecture of regulatory variation in diverse, specific cell types will be important to capture the full biological complexity and avoid averaging out effects from rare cell types.

While samples from a few hundred individuals are sufficient for well-powered standard *cis*-eQTL analysis, further increase of sample sizes is essential for capturing other, more subtle genetic effects on gene expression. The most important gap in the current literature concerns *trans*-eQTLs associations to distal genes in the human genome. They are likely to explain a large proportion of heritable variation in gene expression and also act as modifiers of *cis*-eQTLs (Price et al. 2011; Grundberg et al. 2012; Buil et al. 2015). However, few studies have been large enough to capture them well (Westra et al. 2013; Battle et al. 2014), and characterizing their properties and mechanisms is an important topic for future research. Another controversial question in human genetics is epistasis or interaction between genetic variants in which combinations of variants either in *cis* and in *trans* may affect the trait outcome, and gene expression has been used as a model trait to detect such interactions (Brown et al. 2014; Hemani et al. 2014). However, pinpointing specific interactions has been challenging with the existing sample sizes and statistical methods, and the prevalence, mechanisms, and phenotypic importance of genetic epistasis remains currently unsolved. Finally, as larger and larger studies capture more of hereditary variation in gene expression, predictive imputation of gene expression levels in individuals, based on genotype data, is becoming possible, allowing association studies between disease phenotypes and predicted gene expression levels (Gamazon et al. 2015).

In addition to ongoing efforts in eQTL mapping, the same approach is increasingly being applied to other types of quantitative phenotypes of the cell, for example, to characterize genetic effects on chromatin state (Degner et al. 2012), methylation (Bell et al. 2011; Gutierrez-Arcelus et al. 2015), and transcript stability (Pai et al. 2012), as well as translation and protein levels (Battle et al. 2015). These cellular QTL studies are enabled by continuing development of scalable and affordable molecular assays that can be applied to hundreds of samples, ideally from multiple cell types and conditions. Other cellular QTLs have uncovered regulatory mechanisms of GWAS loci that are not captured by eQTL analysis, and thus QTL analysis of various cellular phenotypes is likely to continue to be one of the primary approaches for uncovering functional mechanisms of GWAS associations. Furthermore, integration of different QTL data provides extremely valuable information of causal mechanisms of genome function and gene regulation—for example, when and how genetic variants affecting epigenetic state lead to change in gene expression (Gutierrez-Arcelus et al. 2013; Pai et al. 2015).

## Functional effects of rare variants

One of the major caveats of the QTL approach is that, as an association analysis, it lacks statistical power to pinpoint effects of rare variants, which have become a major target in human genetics research. Currently, analysis methods for high-throughput analysis of cellular effects of rare variants are still under development (Li et al. 2014). Priors on the predicted functional effects can help, derived from annotation of the variants—such as whether a variant

introduces a premature stop codon, is in close proximity to an annotated splice site, or disrupts a transcription factor binding site. Analysis of allelic expression can be a powerful approach for detecting rare genetic effects on gene expression levels (Rivas et al. 2015). An essential component in this process is solid understanding of the normal spectrum of variation of the studied cellular trait in the population, which can be obtained from data collected in cellular QTL studies. However, given the difficulty of replicating the effects of rare variants, careful consideration is needed to distinguish effects that are beyond what is expected by chance. Sophisticated analysis of functional effects of rare variants requires increasing sample sizes, family-based data sets, and experimental approaches for validation via patient-derived iPSC cells and genome editing. Future advances in this area have the potential to contribute significantly to the understanding of causal molecular processes underlying Mendelian diseases and other phenotypes due to rare variants and to improve our understanding of selective forces that shape the spectrum of functional effects of genetic variants.

### Breaking the regulatory code with genetic perturbations

The importance of cellular QTL approaches is not only in filling in the functional gaps of the GWAS catalog. One of the ultimate goals of genomics is to learn to read the regulatory code and eventually predict regulatory changes caused by any genetic variants. Naturally occurring variation is still the world's largest mutagenesis "experiment," and systematic analysis of how genetic variants affect the cell is an important source of information for understanding the basic biology of the genome. Given how common eQTLs are, it is clear that the vast majority of them have no effect on organism-level phenotype, but yet they are informative of how genome perturbations affect gene expression—and the same applies to other cellular QTLs and their integrated analysis. While computational analyses of QTL data are starting to yield promising genome-wide results of sequence motifs, relevant annotations, and mechanisms of genome function (Gaffney et al. 2012; Lee et al. 2015; Pai et al. 2015), these analyses are complicated by the caveat of all association analyses, eQTLs included: They can capture only (common) variation observed in the study sample, and due to linkage disequilibrium, they uncover associated loci rather than the actual causal variants underlying the change in genome function. Emerging eQTL studies based on genome sequencing data have the opportunity for finding causal variants (Lappalainen et al. 2013), and analytical approaches developed for fine-mapping GWAS loci (Kichaev et al. 2014; Pickrell 2014) could be applied to eQTL loci as well. However, empirical validation of these methods is still lacking, and fundamentally, true evidence of causality in individual loci cannot be achieved by association analysis.

Experimental approaches for genome perturbation combined with functional readout are not bound to analysis of naturally existing variation in humans. They can circumvent caveats of linkage disequilibrium obscuring the identity of the causal variant and the bias toward capturing mainly existing common variants. Massively parallel reporter assays allow multiplexed analysis of sequences that control gene expression *in vitro*, with reporter bar codes that are analyzed by sequencing (Arnold et al. 2013; Kheradpour et al. 2013; Shalem et al. 2015). These assays have provided a wealth of evidence of the function of regulatory elements

of the genome, allow precise perturbation of the genetic code, and their high throughput yields comprehensive data for computational analysis of sequence motifs and their function. However, these approaches rely on artificial systems, with the elements being outside their native genomic context, and analysis *in vitro* may not always sufficiently recapitulate the complexity of the cellular environment *in vivo*.

The novel genome editing technology by CRISPR/Cas is opening a vast universe of new possibilities for analyzing how genetic variants affect phenotypes (Doudna and Charpentier 2014). Introducing variants in human cell lines and measuring the resulting cellular phenotypes in a high-throughput manner provides the possibility for experimental validation of cellular QTLs in their native genomic context, as well as testing cellular consequences of systematic high-throughput mutagenesis (Findlay et al. 2014). In addition to genome editing, CRISPR assays that allow targeted transcription regulation will be valuable tools for understanding causal networks of genome regulation (Konermann et al. 2015). Truly high-throughput applications of the CRISPR technology are currently limited to gene knock-out screens (Shalem et al. 2014), but this will likely change during the next few years as both molecular assays and analytical approaches develop. However, genome editing can be used to manipulate the human genome only in cell lines, and extrapolating that information to understand a complex living organism and its phenotypes is unlikely to be straightforward. Thus, observational data from human tissue samples as well as modified model organisms will continue to be important for interpreting and applying results from CRISPR assays.

### Functional genomics and human health

How has a decade of research into the cellular effects of genetic variants across the genome contributed to improving human health? Information of regulatory mechanisms behind genetic associations to disease can be informative of novel drug targets and other interventions, and this will hopefully be an increasingly fruitful approach in the near future. In the rare variant domain, knowing the functional effects of a variant causing (or protecting against) disease is important for knowing whether the treatment should, for example, block a truncated protein or boost gene expression or protein levels. Furthermore, understanding the range of functional variation observed in healthy individuals can be a powerful tool for understanding what type of manipulations of the functional landscape of the cell are likely to be well tolerated.

While exome and genome sequencing are rapidly becoming part of standard clinical practice, the same is not yet true for high-throughput assays in functional genomics such as RNA sequencing, epigenome analysis, and protein quantification. Yet, these assays can have significant clinical value. In addition to studies aiming to profile patients based on the transcriptome or the epigenome (Michels et al. 2013), these assays are also being pursued as personal biomarkers allowing a longitudinal monitoring of cellular state (Chen et al. 2012). Furthermore, geneticists are now painfully aware of how difficult the interpretation of an individual's genome is, but the epigenome and the transcriptome can provide a layer of information close to the genome that enables better interpretation of phenotypic effects of genetic variants. It is only now that the assays, analytical approaches, and general understanding of the spectrum of epigenome and transcriptome are starting to be advanced enough for clinical analysis. Although extensive benchmarking and standardization of bedside functional



genomics is still lacking, functional genomics assays hold substantial clinical potential for the future.

## Summary

Analysis of functional effects of genetic variants has become one of the fastest growing areas of human genetics—and rightly so, as it addresses some of the most burning questions in the quest toward understanding genome function as well as genetic background of phenotypic variation in humans. It brings together the formerly largely distinct fields of molecular biology and quantitative genetics, contributing to the development of both (cf. Fig. 1). Quantitative genetics is now reaching beyond disease associations as statistical constructs, toward real biological understanding. On the other hand, molecular biology has a lot to gain in understanding the range of population variation in genome function and using genetic effects as a causality anchor in cellular networks and disease etiology. The GWAS community has been exemplary in establishing commonly accepted gold standards for statistical analysis. While functional genomics data is more diverse in nature, the development toward similarly high standards must continue.

The future of this field looks bright: Increasing sample sizes allow deeper interrogation of more and more complex effects of genetic variants, and characterization of additional cell types and conditions with diverse assays will provide not only more comprehensive catalogs but also deeper mechanistic understanding beyond incremental increases. Genome editing technology will redefine the toolkit in unprecedented ways. Massive data sets often produced by consortium projects will continue to fuel research and provide the accessible, carefully curated data resources for discovery, both at the level of individual loci and, in particular, in genome-wide systems-level approaches. However, many novel biological phenomena, technologies, and statistical approaches will still be discovered and developed in individual laboratories in the future, via analysis of both humans and model organisms. Both detailed dissection of specific mechanistic components of genome function and systems-level approaches to link all the components back together are necessary. Finally, the application of technologies and results from functional genomics to improve drug development and interpretation of personal genomes has substantial potential to improve human health.

## Acknowledgments

I thank Ana Vasileva, Stephane Castel, Pejman Mohammadi, and Margot Bradt for helpful comments. The author is funded by National Institutes of Health (NIH) grant R01MH101814.

## References

- The 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**: 56–65.
- Albert FW, Kruglyak L. 2015. The role of regulatory variation in complex traits and disease. *Nat Rev Genet* **16**: 197–212.
- Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. 2013. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**: 1074–1077.
- Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J. 2011. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* **12**: 745–755.
- Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, Haudenschild CD, Beckman KB, Shi J, Mei R, et al. 2014. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res* **24**: 14–24.
- Battle A, Khan Z, Wang SH, Mitrano A, Ford MJ, Pritchard JK, Gilad Y. 2015. Genomic variation. Impact of regulatory variation from RNA to protein. *Science* **347**: 664–667.
- Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, Gilad Y, Pritchard JK. 2011. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol* **12**: R10.
- Brown AA, Buil A, Vinuela A, Lappalainen T, Zheng HF, Richards JB, Small KS, Spector TD, Dermitzakis ET, Durbin R. 2014. Genetic interactions affecting human gene expression identified by variance association mapping. *eLife* **3**: e01381.
- Buil A, Brown AA, Lappalainen T, Vinuela A, Davies MN, Zheng HF, Richards JB, Glass D, Small KS, Durbin R, et al. 2015. Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat Genet* **47**: 88–91.
- Chen R, Mias GI, Li-Pook-Tham J, Jiang L, Lam HY, Chen R, Miriami E, Karczewski KJ, Hariharan M, Dewey FE, et al. 2012. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* **148**: 1293–1307.
- Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT. 2005. Mapping determinants of human gene expression by regional and genome-wide association. *Nature* **437**: 1365–1369.
- Degner JF, Pai AA, Pique-Regi R, Veyrieras JB, Gaffney DJ, Pickrell JK, De Leon S, Michelini K, Lewellen N, Crawford GE, et al. 2012. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* **482**: 390–394.
- Dimas AS, Nica AC, Montgomery SB, Stranger BE, Raj T, Buil A, Giger T, Lappalainen T, Gutierrez-Arcelus M, Mu TC, et al. 2012. Sex-biased genetic effects on gene regulation in humans. *Genome Res* **22**: 2368–2375.
- Doudna JA, Charpentier E. 2014. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**: 1258096.
- Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, Carlson S, Helgason A, Walters GB, Gunnarsdottir S, et al. 2008. Genetics of gene expression and its effect on disease. *Nature* **452**: 423–428.
- The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.
- The FANTOM Consortium and the RIKEN PMI and CLST (DGT). 2014. A promoter-level mammalian expression atlas. *Nature* **507**: 462–470.
- Farh KK, Marson A, Zhu J, Kleinewieffeld M, Housley WJ, Beik S, Shores N, Whitton H, Ryan RJ, Shishkin AA, et al. 2015. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**: 337–343.
- Findlay GM, Boyle EA, Hause RJ, Klein JC, Shendure J. 2014. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* **513**: 120–123.
- Fraser HB. 2013. Gene expression drives local adaptation in humans. *Genome Res* **23**: 1089–1096.
- Gaffney DJ, Veyrieras JB, Degner JF, Pique-Regi R, Pai AA, Crawford GE, Stephens M, Gilad Y, Pritchard JK. 2012. Dissecting the regulatory architecture of gene expression QTLs. *Genome Biol* **13**: R7.
- Gamazon ER, Wheeler HE, Shah K, Mozaffari SV, Aquino-Michaels K, Carroll RJ, Elyer AE, Denny JC, Nicolae DL, Cox NJ, et al. 2015. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* doi: 10.1038/ng.3367.
- Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. 2014. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**: e1004383.
- Grossman SR, Andersen KG, Shlyakhter I, Tabrizi S, Winnicki S, Yen A, Park DJ, Griesemer D, Karlsson EK, Wong SH, et al. 2013. Identifying recent adaptations in large-scale genomic data. *Cell* **152**: 703–713.
- Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, Bell JT, Yang TP, Meduri E, Barrett A, et al. 2012. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* **44**: 1084–1089.
- GTEX Consortium. 2013. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**: 580–585.
- GTEx Consortium. 2015. The Genotype-Tissue Expression (GTEx) pilot analysis of multi-tissue gene regulation in humans. *Science* **348**: 648–660.
- Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A, Bryois J, Giger T, Romano L, Planchon A, et al. 2013. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *eLife* **2**: e00523.
- Gutierrez-Arcelus M, Ongen H, Lappalainen T, Montgomery SB, Buil A, Yurovsky A, Bryois J, Padioleau I, Romano L, Planchon A, et al. 2015. Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. *PLoS Genet* **11**: e1004958.
- Hemani G, Shakhbazov K, Westra HJ, Esko T, Henders AK, McRae AF, Yang J, Gibson G, Martin NG, Metspalu A, et al. 2014. Detection and replication of epistasis influencing transcription in humans. *Nature* **508**: 249–253.

- The International HapMap Consortium. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**: 851–861.
- Kheradpour P, Ernst J, Melnikov A, Rogov P, Wang L, Zhang X, Alston J, Mikkelsen TS, Kellis M. 2013. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res* **23**: 800–811.
- Kichaev G, Yang WY, Lindstrom S, Hormozdiari F, Eskin E, Price AL, Kraft P, Pasaniuc B. 2014. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet* **10**: e1004722.
- Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, et al. 2015. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**: 583–588.
- Lappalainen T, Salmela E, Andersen PM, Dahlman-Wright K, Sistonen P, Savontaus ML, Schreiber S, Lahermo P, Kere J. 2010. Genomic landscape of positive natural selection in Northern European populations. *Eur J Hum Genet* **18**: 471–478.
- Lappalainen T, Sammeth M, Friedlander MR, Hoen PA, Monlong J, Rivas MA, Gonzalez-Porta M, Kurbatova N, Griebel T, Ferreira PG, et al. 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**: 506–511.
- Lee S, Abecasis GR, Boehnke M, Lin X. 2014. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* **95**: 5–23.
- Lee D, Gorkin DU, Baker M, Strober BJ, Asoni AL, McCallion AS, Beer MA. 2015. A method to predict the impact of regulatory variants from DNA sequence. *Nat Genet* **47**: 955–961.
- Li X, Battle A, Karczewski KJ, Zappala Z, Knowles DA, Smith KS, Kukurba KR, Wu E, Simon N, Montgomery SB. 2014. Transcriptome sequencing of a large human family identifies the impact of rare noncoding variants. *Am J Hum Genet* **95**: 245–256.
- Michels KB, Binder AM, Dedeurwaerder S, Epstein CB, Grealis JM, Gut I, Houseman EA, Izzi B, Kelsey KT, Meissner A, et al. 2013. Recommendations for the design and analysis of epigenome-wide association studies. *Nat Methods* **10**: 949–955.
- Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermitzakis ET. 2010. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**: 773–777.
- Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, Dermitzakis ET. 2010. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet* **6**: e1000895.
- Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. 2010. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* **6**: e1000888.
- Pai AA, Cain CE, Mizrahi-Man O, De Leon S, Lewellen N, Veyrieras JB, Degner JF, Gaffney DJ, Pickrell JK, Stephens M, et al. 2012. The contribution of RNA decay quantitative trait loci to inter-individual variation in steady-state gene expression levels. *PLoS Genet* **8**: e1003000.
- Pai AA, Pritchard JK, Gilad Y. 2015. The genetic and mechanistic basis for variation in gene regulation. *PLoS Genet* **11**: e1004857.
- Pickrell JK. 2014. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet* **94**: 559–573.
- Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK. 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**: 768–772.
- Price AL, Helgason A, Thorleifsson G, McCarroll SA, Kong A, Stefansson K. 2011. Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *PLoS Genet* **7**: e1001317.
- Rivas MA, Pirinen M, Conrad DF, Lek M, Tsang EK, Karczewski KJ, Maller JB, Kukurba KR, DeLuca D, Fromer M, et al. 2015. Impact of predicted protein-truncating genetic variants on the human transcriptome. *Science* **348**: 666–669.
- Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**: 317–330.
- Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG, et al. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**: 84–87.
- Shalem O, Sharon E, Lubliner S, Regev I, Lotan-Pompan M, Yakhini Z, Segal E. 2015. Systematic dissection of the sequence determinants of gene 3' end mediated expression control. *PLoS Genet* **11**: e1005147.
- Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D, et al. 2007. Population genomics of human gene expression. *Nat Genet* **39**: 1217–1224.
- Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, Sekowska M, Smith GD, Evans D, Gutierrez-Arcelus M, et al. 2012. Patterns of cis regulatory variation in diverse human populations. *PLoS Genet* **8**: e1002639.
- Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE, et al. 2013. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**: 1238–1243.
- Ye CJ, Feng T, Kwon HK, Raj T, Wilson MT, Asinowski N, McCabe C, Lee MH, Frohlich I, Paik HI, et al. 2014. Intersection of population variation and autoimmunity genetics in human T cell activation. *Science* **345**: 1254665.