Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

Cellular network perturbations by disease-associated variants

Jared A. Sewell and Juan I. Fuxman Bass

Abstract

Genetic and genome-wide association studies (GWAS) have identified a myriad of human disease-associated genomic variants. However, these studies do not reveal the mechanisms by which these variants perturb cellular networks, a necessary step to intervene and improve disease outcomes. This has been challenging because multiple variants are present in haplotype blocks, thereby confounding the identification of causal variants, and because most reside in noncoding regions. Here, we review recent advances in the identification of functional variants and gene-variant associations. In addition, we examine approaches used to study perturbations in protein–protein and protein–DNA interactions associated with disease, and discuss how these perturbations affect cellular networks.

Addresses

Department of Biology, Boston University, Boston, MA 02215, USA

Corresponding author: Fuxman Bass, Juan I (fuxman@bu.edu)

Current Opinion in Systems Biology 2017, **3**:60–66

This review comes from a themed issue on **Clinical and translational systems biology (2017)**

Edited by **Jesper Tegnér** and **David Gomez-Cabrero**

For a complete overview see the Issue and the Editorial

Available online 27 April 2017

http://dx.doi.org/10.1016/j.coisb.2017.04.009

Introduction

Genome-wide association studies (GWAS) and next generation sequencing have identified tens of thousands of human disease-associated variants and millions of single nucleotide polymorphisms [1–3]. Although these studies have been instrumental for predicting disease risk, they generally do not reveal the molecular mechanisms that affect cellular function, a necessary step towards developing therapeutics. Understanding the molecular mechanisms by which some genomic variants cause disease has been hampered by multiple challenges in identifying the genes affected by these genomic changes and in determining how risk alleles perturb cellular networks. For instance, given that numerous variants are in linkage disequilibrium, it is not straightforward to identify the causal- or disease-associated variant within a haplotype block [4]. In addition, most disease-associated variants reside in noncoding regions, which complicates the identification of the target gene affected [5]. Furthermore, it is challenging to determine how these variants perturb physical and functional interactions within cellular networks, and how these perturbations lead to disease.

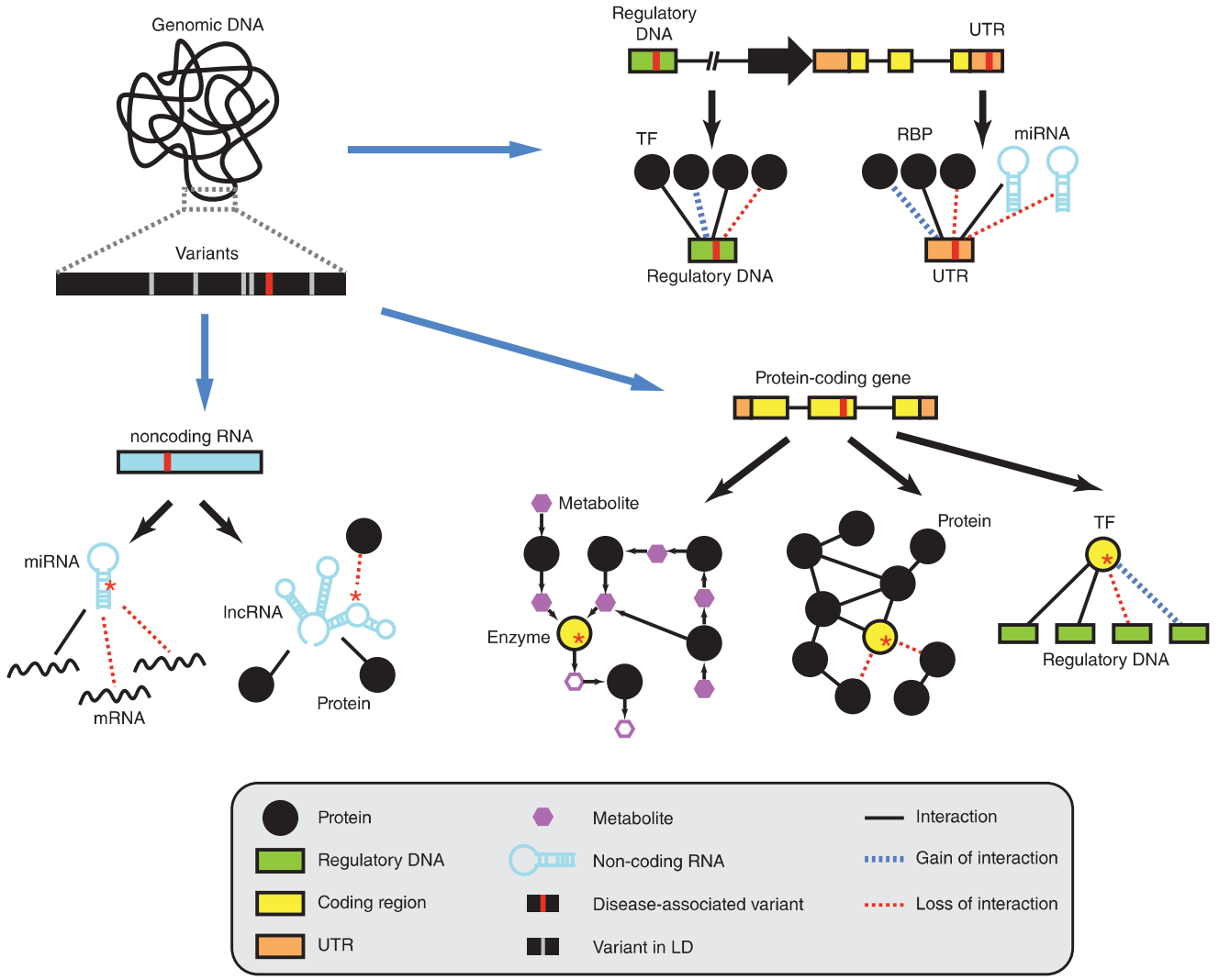
Systems and network mapping approaches are beginning to address some of the roadblocks that have limited the study of the functional consequences of genomic variants. These approaches have the potential to swiftly identify disease pathways and aid in the design of improved drug therapies. Here, we present an overview of recent advances in identifying gene-variant associations and how systems-based approaches are exploring the ways in which disease variants affect cellular networks.

Protein-coding variants

Disease-associated variants can affect gene function by disrupting or creating interactions with other molecules or by affecting enzymatic activity [6] (Figure 1). Alternatively, variants can affect transcriptional and/or post-transcriptional regulation by altering the binding of regulatory molecules such as transcription factors (TFs), RNA-binding proteins (RBPs) and micro RNAs (miRNAs) [7–10] (Figure 1).

Function-disrupting variants have been identified within protein-coding genes, miRNA and long noncoding RNAs (lncRNAs) [11,12]. Among these, protein-coding variants have been the most widely studied, as the genetic code and computational algorithms facilitate predictions of how protein function is likely to be affected. Indeed, most variants reported in the Human Gene Mutation Database (HGMD) and the Online Mendelian Inheritance in Man (OMIM) database are nonsense, missense, or frameshift mutations, or are sequence changes that disrupt splicing to affect protein structure [12,13]. These variants can alter protein function (e.g., increase, decrease, or change activity), and consequently lead to perturbations within the cellular network [6,14]. For instance, a recent study has shown that ~25% of disease-associated missense variants lead to protein inactivation as determined by increased binding to protein chaperones (suggesting decreased protein stability), or by the complete loss of protein–protein interactions (PPIs) in yeast two-hybrid assays [6]. Interestingly, ~30% of the missense variants result in edgetic perturbations (i.e., only affect a fraction

Figure 1



Outline of cellular network perturbations caused by disease-associated variants. Disease-associated variants (red bar) found in haplotype blocks with common variants (gray bars) must be identified prior to functional studies. Disease-associated variants within regulatory regions can alter gene expression by disrupting or creating interactions with TFs. Variants residing in untranslated regions can affect the binding of RBPs and miRNA. Protein-coding variants can affect multiple cellular networks including protein–protein, protein–DNA, protein–RNA and enzyme-metabolite networks leading to gain and loss of interactions. Variants affecting the sequence and/or structure or miRNAs and lncRNAs can also affect physical interactions with mRNAs and proteins, respectively. These perturbations in cellular networks underlie the basis of disease. Black solid edges represent physical interactions, red dotted edges represent lost interactions, blue dotted edges represent gained interactions.

of the PPIs in which the wild-type protein engages), suggesting partial loss of function [6]. For example, a study using affinity purification and mass spectrometry found that melanoma-associated mutations in CDK4 lead to loss of interactions with CDK inhibitors of the INK family, while other PPIs are retained or increased in affinity [15]. Importantly, different mutations within a gene that lead to different losses of PPIs often result in different disease outcomes, with disease severity correlating with the proportion of PPIs lost [6]. This shows that fine-mapping of PPIs with disease-associated variants may aid in disease stratification, one of the goals of personalized medicine.

PPI mapping methods reveal the interactions affected by disease variants but not the phenotypic consequences, which are often inferred based on the location of the variant within PPI network neighborhoods, connectivity, and protein and domain functions [16,17]. Surrogate genetics provides an alternative method to study the functional consequences of protein-coding disease-associated variants. In this case, human pathogenic variants are tested in yeast complementation assays to predict their role in disease severity [18]. However, this approach is only suitable for the small number of genes that have functional homologs in yeast, and cannot be applied to complex genetic diseases such

as those that involve multiple genes. A major obstacle in studying complex genetic diseases is determining how coding variants genetically interact with other variants. In this regard, recent high-throughput studies in yeast have explored epistatic effects and identified $\sim 10^6$ genetic interactions between 5416 genes and mapped genetic suppression [19,20]. In addition, synthetic lethal interactions in humans have been inferred from cancer genomic data from cell lines and clinical samples, and identified using gene perturbations [21,22]. Overall, these studies provide a framework to predict pathways and networks affected in human disease.

Protein-coding variants can also affect interactions with other cellular components. For instance, mutations in TFs not only affect interactions with other proteins such as cofactors and TFs, but also result in changes in DNA specificity and affinity, which can ultimately cause gain and/or loss of protein–DNA interactions (PDIs) simultaneously affecting many target genes (Figure 1) [6,7,23]. Mutations in Serine/arginine-rich splicing factor 2 (SRSF2) have been shown to misregulate splicing by affecting its affinity for RNA [24]. Overall, this highlights the importance of delineating reference physical interaction maps, or inferring them based on predicted interactions, including PPIs, PDIs, protein–RNA interactions and enzyme-metabolite interactions [25–27]. These networks constitute a blueprint for the study of perturbations caused by disease-associated variants and provide a framework for the mapping of cellular functions that may be affected in disease.

Regulatory variants

About 90% of the disease-associated variants identified in GWAS reside in intergenic or intronic regions [3,5,28]. Characterizing the functional effects of these noncoding variants presents additional challenges as compared to coding variants. Indeed, for noncoding variants identifying the target gene and the regulatory function affected is often much more difficult. For example, multiple regulatory functions can be affected by noncoding variants including the binding of TFs [7,8,29], chromatin structure or accessibility [28], DNA methylation [30,31], and 5′ and 3′UTR function [9,32] (Figure 1).

One challenge encountered when studying noncoding variants is determining which of the tens of thousands of variants identified in GWAS are regulatory, given that many have only modest or no effect on gene expression [33,34]. Early studies have focused on mutations residing mostly within promoter regions as these regions are easier to define and then test in reporter assays. However, the number of variants reported so far, and the fact that most are located several kilobases from transcription start sites render this approach impractical. Multiple studies have prioritized candidate causal noncoding variants by determining whether they reside within regulatory regions based on epigenomic marks such as chromatin accessibility, histone modifications, and regional genomic features [28,35–37]. In addition, recent studies using massively parallel reporter assays (MPRAs) have increased the throughput of functional testing of noncoding variants [38,39]. This method can evaluate thousands of variants in a single experiment by taking advantage of barcode next generation sequencing technology, while having a similar or higher sensitivity and reproducibility than low-throughput reporter assays [34]. Using MPRAs, researchers have identified 842 eQTL variants with differential allele expression and multiple functional variants associated with red blood cell traits [34,40]. However, given that variants are evaluated outside of their genomic context using plasmid constructs that are transfected into cell lines, many regulatory variants may be missed by this approach.

Another challenge is to identify the gene whose expression is affected by a noncoding variant as many of these variants reside hundreds of kilobases away from their target gene [28]. For instance, a recent study has shown that the rs1421085 variant associated with obesity (located within the *FTO* gene) increases the activity of an enhancer that controls the expression of *IRX3* and *IRX5*, located ~ 1 Mb away [41]. Many technologies based on crosslinking and ligation of spatially closed genomic regions, such as Hi-C [42,43] and chromatin conformation capture by paired-end tag sequencing (ChIA-Pet) [44], have emerged to identify target genes affected by noncoding variants. For example, a recent study examining the chromatin state effects of variants associated with autoimmune diseases found that these variants alter gene expression by disrupting the physical interactions between enhancers and promoters [45]. Computational methods have also been used to identify gene-disease and gene-variant associations. For instance, gene-disease associations have been predicted based on genome-scale shared-function networks [46] and on genomic features [47]. Bayesian methods have also been employed using GWAS data to identify disease-associated genes by creating tissue specific network maps [48]. Additionally, gene-variant associations have been determined by integrating genomic distance and joint expression/activity between genes and regulatory elements in given tissues [49].

Variants that affect transcriptional regulation frequently alter TF binding by either disrupting or creating DNA-binding sites [7,50]. Studies examining these types of variants have traditionally been low-throughput: first predicting differential TF binding using motif analyses, and then evaluating altered binding in functional assays such as chromatin immunoprecipitation (ChIP), electrophoretic mobility shift assays (EMSA) and reporter

assays. Therefore, higher throughput approaches are needed to test thousands of noncoding variants. Although great progress has been made in identifying *in vivo* TF binding by ChIP-seq, these studies have been limited to ~20% of human TFs, most tested in a single cell type or condition [35]. Furthermore, ChIP-seq cannot be applied to the *ab initio* discovery of TF binding differences as each TF needs to be tested in the appropriate tissue/condition with different allele variants. Alternatively, TF binding differences between alleles have also been predicted computationally using motif-based algorithms such as motifbreakR [51] and CIS-BP [52], or those that also integrate epigenomic data such as HaploReg [53], RegulomeDB [54], and RiVIERA [55]. However, motif predictions often result in false positives and false negatives and, thus require extensive experimental validation. Moreover, it is likely that only 10–20% of noncoding variants directly affect a TF binding site, suggesting that most variants either affect the binding of an uncharacterized TF or affect TF binding to nearby regions by altering DNA structure [56]. In addition, noncoding variants may affect DNA methylation which can alter TF binding [30,57], further complicating computational analyses. Thus, efforts have been made to infer differential TF binding from experimentally-derived allele-specific DNase I footprinting data across multiple individuals and cell types [50].

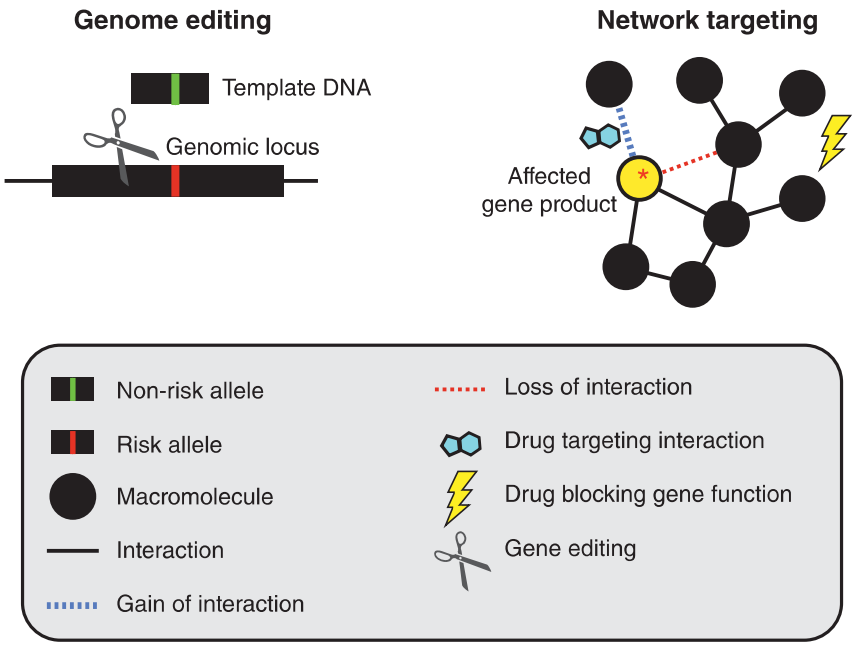
TF-variant-disease associations have also been determined by integrating allele- or genotype-specific TF binding between individuals and allele-specific expression [58]. However, these studies require ChIP data from many individuals and can be performed with only one TF at a time. A recent study has used enhanced yeast one-hybrid assays to increase the screening throughput of TF binding differences between noncoding alleles [7]. This study identified differential PDIs for 109 disease-associated variants showing that these genomic alterations not only lead to loss of TF interactions, but in ~40% of cases promote new interactions, suggesting that gain of PDIs may be a more prevalent disease mechanism than previously thought.

Beyond affecting transcriptional regulation, noncoding variants can also impact post-transcriptional and translational regulation. Recent studies have identified multiple disease-associated variants within miRNA gene regions and within miRNA-binding sites in target mRNAs [11,59]. Differential binding of miRNA and RBPs to 3′UTR variants have also been modeled *in silico* [9,10]. This highlights the importance of considering multiple levels of gene regulation when studying the functional consequence of noncoding variants.

Conclusions

GWAS constitute the first step towards therapeutic treatment of genetic diseases by identifying disease-associated genomic variants. The most conceptually straightforward approach for treatment would be to restore the non-risk allele using genome-editing technologies such as CRISPR/Cas9 or zinc finger- and transcription activator-like effector-nucleases (Figure 2).

Figure 2



Systems approaches can inform treatment strategies for disease-associated variants. After identifying the risk allele, gene-editing approaches can potentially be used to restore the non-risk allele. Alternatively, drugs can also be used to target interaction changes, or to affect epistatic genes to suppress disease phenotypes or to enhance cell death in cancer.

Although several studies have made strides toward this goal in *ex vivo* cells or in genetic diseases where restoration of function in a fraction of the cells is sufficient to alleviate disease [60,61], this approach has not yet been successfully applied in cases where most or all cells need to be engineered, such as in cancer. Further, genome editing has the risk of introducing new mutations due to off target effects. Determining how disease-associated variants affect cellular networks and mapping network perturbations are important steps towards designing alternative approaches to treat genetic disorders (Figure 2). For instance, gain of PPIs can potentially be blocked by small molecules, as in the case of AI-10-49 that inhibits binding of CBFβ-SMMHC to RUNX1 in acute myeloid leukemia [62].

Alternatively, neighborhood location in physical interaction networks and epistatic relationships can be leveraged to suppress a disease phenotype by targeting a different gene [19–21]. For example, targeting a suppressor of a gene associated with disease can potentially improve clinical manifestations in genetic disorders. Similarly, targeting genes that have synthetic lethal interactions with a gene affected by a somatic variant in cancer has the potential to reduce growth of malignant cells without affecting non-malignant cells [22]. As tools and technologies continue to develop, our understanding of the cellular networks and the effects that genetic variants have on macromolecular interactions will continue to improve. This will provide a global view of disease mechanisms, which is necessary for the improvement of therapeutics.

Acknowledgements

We thank Drs. Thomas Gilmore and Jessica Osterhout for critically reading and editing the manuscript. This work was supported by the National Institutes of Health to J.I.F.B. (R00 GM114296 from the NIGMS) and to J.A.S. (5T32HL007501-34 from the NHLBI).

References

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Consortium GTEx: **Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans.** *Science* 2015, **348**:648–660.

2. •• 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, *et al.*: **A global reference for human genetic variation.** *Nature* 2015, **526**:68–74.
This paper describes the genetic variation across 2504 individuals, including 88 million variants.

3. • MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, *et al.*: **The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog).** *Nucleic Acids Res* 2017, **45**: D896–D901.
This database contains a comprehensive manually curated collection of published GWAS SNPs and SNP-trait associations.

4. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, Adams DR, Altman RB, Antonarakis SE, Ashley EA, *et al.*: **Guidelines for investigating causality of sequence variants in human disease.** *Nature* 2014, **508**: 469–476.

5. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA: **Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.** *Proc Natl Acad Sci USA* 2009, **106**: 9362–9367.

6. •• Sahni N, Yi S, Taipale M, Fuxman Bass JI, Coulombe-Huntington J, Yang F, Peng J, Weile J, Karras GI, Wang Y, *et al.*: **Widespread macromolecular interaction perturbations in human genetic disorders.** *Cell* 2015, **161**:647–660.
This paper shows how missense mutations affect protein–protein and protein–DNA interactions resulting in partial or complete interaction changes.

7. • Fuxman Bass JI, Sahni N, Shrestha S, Garcia-Gonzalez A, Mori A, Bhat N, Yi S, Hill DE, Vidal M, Walhout AJ: **Human gene-centered transcription factor networks for enhancers and disease variants.** *Cell* 2015, **161**:661–673.
This paper shows how noncoding variants can lead to gain and loss of protein–DNA interactions.

8. Ma M, Ru Y, Chuang LS, Hsu NY, Shi LS, Hakenberg J, Cheng WY, Uzilov A, Ding W, Glicksberg BS, *et al.*: **Disease-associated variants in different categories of disease located in distinct regulatory elements.** *BMC Genomics* 2015, **16**(Suppl. 8):S3.

9. Mao F, Xiao L, Li X, Liang J, Teng H, Cai W, Sun ZS: **RBP-Var: a database of functional variants involved in regulation mediated by RNA-binding proteins.** *Nucleic Acids Res* 2016, **44**: D154–163.

10. Okada Y, Muramatsu T, Suita N, Kanai M, Kawakami E, Iotchkova V, Soranzo N, Inazawa J, Tanaka T: **Significant impact of miRNA-target gene networks on genetics of human complex traits.** *Sci Rep* 2016, **6**:22223.

11. Ghanbari M, Darweesh SK, de Looper HW, van Luijn MM, Hofman A, Ikram MA, Franco OH, Erkeland SJ, Dehghan A: **Genetic variants in microRNAs and their binding sites are associated with the risk of Parkinson disease.** *Hum Mutat* 2016, **37**:292–300.

12. • Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN: **The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine.** *Hum Genet* 2014, **133**:1–9.
This database contains literature-curated information for thousands of disease-associated variants.

13. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA: **Online Mendelian Inheritance in Man (OMIM), a knowledge-base of human genes and genetic disorders.** *Nucleic Acids Res* 2005, **33**:D514–D517.

14. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB, *et al.*: **A systematic survey of loss-of-function variants in human protein-coding genes.** *Science* 2012, **335**: 823–828.

15. Lambert JP, Ivosev G, Couzens AL, Larsen B, Taipale M, Lin ZY, Zhong Q, Lindquist S, Vidal M, Aebersold R, *et al.*: **Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition.** *Nat Methods* 2013, **10**:1239–1245.

16. Ghiassian SD, Menche J, Barabasi AL: **A DIseAse MOdule Detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome.** *PLoS Comput Biol* 2015, **11**:e1004120.

17. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabasi AL: **Disease networks. Uncovering disease-disease relationships through the incomplete interactome.** *Science* 2015, **347**:1257601.

18. Sun S, Yang F, Tan G, Costanzo M, Oughtred R, Hirschman J, Theesfeld CL, Bansal P, Sahni N, Yi S, *et al.*: **An extended set of yeast-based functional assays accurately identifies human disease mutations.** *Genome Res* 2016, **26**:670–680.

19. • Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, Wang W, Usaj M, Hanchard J, Lee SD, *et al.*: **A global genetic interaction network maps a wiring diagram of cellular function.** *Science* 2016, **353**.
This paper identifies positive and negative genetic interactions for nearly all pairs of yeast genes.

20. van Leeuwen J, Pons C, Mellor JC, Yamaguchi TN, Friesen H, Koschwanez J, Usaj MM, Pechlaner M, Takar M, Usaj M, *et al.*: **Exploring genetic suppression interactions on a global scale.** *Science* 2016, **354**.

21. Blomen VA, Majek P, Jae LT, Bigenzahn JW, Nieuwenhuis J, Staring J, Sacco R, van Diemen FR, Olk N, Stukalov A, *et al.*: **Gene essentiality and synthetic lethality in haploid human cells.** *Science* 2015, **350**:1092–1096.

22. Jerby-Arnon L, Pfetzer N, Waldman YY, McGarry L, James D, Shanks E, Seashore-Ludlow B, Weinstock A, Geiger T, Clemons PA, *et al.*: **Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality.** *Cell* 2014, **158**:1199–1209.

23. • Barrera LA, Vedenko A, Kurland JV, Rogers JM, Gisselbrecht SS, Rossin EJ, Woodard J, Mariani L, Kock KH, Inukai S, *et al.*: **Survey of variation in human transcription factors reveals prevalent DNA binding changes.** *Science* 2016, **351**:1450–1454.
This paper shows how missense variants in transcription factors affect their affinity and specificity for DNA.

24. Zhang J, Lieu YK, Ali AM, Penson A, Reggio KS, Rabadan R, Raza A, Mukherjee S, Manley JL: **Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities.** *Proc Natl Acad Sci USA* 2015, **112**:E4726–E4734.

25. Li X, Gianoulis TA, Yip KY, Gerstein M, Snyder M: **Extensive in vivo metabolite-protein interactions revealed by large-scale systematic analyses.** *Cell* 2010, **143**:639–650.

26. Rolland T, Tasan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R, *et al.*: **A proteome-scale map of the human interactome network.** *Cell* 2014, **159**:1212–1226.

27. Smola MJ, Calabrese JM, Weeks KM: **Detection of RNA-protein interactions in living cells with SHAPE.** *Biochemistry* 2015, **54**:6867–6875.

28. •• Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, *et al.*: **Systematic localization of common disease-associated variation in regulatory DNA.** *Science* 2012, **337**:1190–1195.
This paper shows that most noncoding variants reside in DNAse I hypersensitive regions active during development and affect TF binding sites.

29. Shi W, Fornes O, Mathelier A, Wasserman WW: **Evaluating the impact of single nucleotide variants on transcription factor binding.** *Nucleic Acids Res* 2016, **44**:10106–10116.

30. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, van Iterson M, van Dijk F, van Galen M, Bot J, *et al.*: **Disease variants alter transcription factor levels and methylation of their binding sites.** *Nat Genet* 2017, **49**:131–138.

31. Gutierrez-Arcelus M, Ongen H, Lappalainen T, Montgomery SB, Buil A, Yurovsky A, Bryois J, Padioleau I, Romano L, Planchon A, *et al.*: **Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing.** *PLoS Genet* 2015, **11**:e1004958.

32. Shalem O, Sharon E, Lubliner S, Regev I, Lotan-Pompan M, Yakhini Z, Segal E: **Systematic dissection of the sequence determinants of gene 3′ end mediated expression control.** *PLoS Genet* 2015, **11**:e1005147.

33. Corradin O, Saiakhova A, Akhtar-Zaidi B, Myeroff L, Willis J, Cowper-Sal lari R, Lupien M, Markowitz S, Scacheri PC: **Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits.** *Genome Res* 2014, **24**:1–13.

34. •• Tewhey R, Kotliar D, Park DS, Liu B, Winnicki S, Reilly SK, Andersen KG, Mikkelsen TS, Lander ES, Schaffner SF, *et al.*: **Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay.** *Cell* 2016, **165**:1519–1529.
This study shows how massively parallel reporter assays can be used to study expression changes caused by noncoding variants, and pinpoint causal alleles.

35. Consortium EP: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, **489**:57–74.

36. Lochovsky L, Zhang J, Fu Y, Khurana E, Gerstein M: **LARVA: an integrative framework for large-scale analysis of recurrent variants in noncoding annotations.** *Nucleic Acids Res* 2015, **43**:8123–8134.

37. Zhou J, Troyanskaya OG: **Predicting effects of noncoding variants with deep learning-based sequence model.** *Nat Methods* 2015, **12**:931–934.

38. Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, Feizi S, Gnirke A, Callan Jr CG, Kinney JB, *et al.*: **Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay.** *Nat Biotechnol* 2012, **30**:271–277.

39. Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, Lee C, Andrie JM, Lee SI, Cooper GM, *et al.*: **Massively parallel functional dissection of mammalian enhancers in vivo.** *Nat Biotechnol* 2012, **30**:265–270.

40. Ulirsch JC, Nandakumar SK, Wang L, Giani FC, Zhang X, Rogov P, Melnikov A, McDonel P, Do R, Mikkelsen TS, *et al.*: **Systematic functional dissection of common genetic variation affecting red blood cell traits.** *Cell* 2016, **165**:1530–1545.

41. Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C, Glunk V, Sousa IS, Beaudry JL, Puviindran V, *et al.*: **FTO obesity variant circuitry and adipocyte browning in humans.** *N Engl J Med* 2015, **373**:895–907.

42. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, *et al.*: **Comprehensive mapping of long-range interactions reveals folding principles of the human genome.** *Science* 2009, **326**:289–293.

43. Lu Y, Quan C, Chen H, Bo X, Zhang C: **3DSNP: a database for linking human noncoding SNPs to their three-dimensional interacting genes.** *Nucleic Acids Res* 2017, **45**:D643–D649.

44. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, *et al.*: **Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation.** *Cell* 2012, **148**:84–98.

45. • Corradin O, Cohen AJ, Luppino JM, Bayles IM, Schumacher FR, Scacheri PC: **Modeling disease risk through analysis of physical interactions between genetic variants within chromatin regulatory circuitry.** *Nat Genet* 2016, **48**:1313–1320.
This study identified physical chromatin interactions that are disrupted by risk SNPs and 'outside variants'.

46. Tasan M, Musso G, Hao T, Vidal M, MacRae CA, Roth FP: **Selecting causal genes from genome-wide association studies via functionally coherent subnetworks.** *Nat Methods* 2015, **12**:154–159.

47. Deo RC, Musso G, Tasan M, Tang P, Poon A, Yuan C, Felix JF, Vasan RS, Beroukhim R, De Marco T, *et al.*: **Prioritizing causal disease genes using unbiased genomic features.** *Genome Biol* 2014, **15**:534.

48. Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, Zhang R, Hartmann BM, Zaslavsky E, Sealfon SC, *et al.*: **Understanding multicellular function and disease with human tissue-specific networks.** *Nat Genet* 2015, **47**:569–576.

49. Marbach D, Lamparter D, Quon G, Kellis M, Kutalik Z, Bergmann S: **Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases.** *Nat Methods* 2016, **13**:366–370.

50. •• Maurano MT, Haugen E, Sandstrom R, Vierstra J, Shafer A, Kaul R, Stamatoyannopoulos JA: **Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo.** *Nat Genet* 2015, **47**:1393–1401.
This study identifies sequence variants that affect human TF occupancy based on DNAse I footprints and motif analyses.

51. Coetzee SG, Coetzee GA, Hazelett DJ: **motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites.** *Bioinformatics* 2015, **31**:3847–3849.

52. Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, *et al.*: **Determination and inference of eukaryotic transcription factor sequence specificity.** *Cell* 2014, **158**:1431–1443.

53. Ward LD, Kellis M: **HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease.** *Nucleic Acids Res* 2016, **44**:D877–D881.

54. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, *et al.*: **Annotation of functional variation in personal genomes using RegulomeDB.** *Genome Res* 2012, **22**:1790–1797.

55. Li Y, Kellis M: **Joint Bayesian inference of risk variants and tissue-specific epigenomic enrichments across multiple complex human diseases.** *Nucleic Acids Res* 2016, **44**:e144.

56. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, Shoresh N, Whitton H, Ryan RJ, Shishkin AA, *et al.*: **Genetic and epigenetic fine mapping of causal autoimmune disease variants.** *Nature* 2015, **518**:337–343.

57. Maurano MT, Wang H, John S, Shafer A, Canfield T, Lee K, Stamatoyannopoulos JA: **Role of DNA methylation in modulating transcription factor occupancy.** *Cell Rep* 2015, **12**:1184–1195.

58. Karczewski KJ, Dudley JT, Kukurba KR, Chen R, Butte AJ, Montgomery SB, Snyder M: **Systematic functional regulatory assessment of disease-associated variants.** *Proc Natl Acad Sci USA* 2013, **110**:9607–9612.

59. Ghanbari M, Ikram MA, de Looper HW, Hofman A, Erkeland SJ, Franco OH, Dehghan A: **Genome-wide identification of microRNA-related variants associated with risk of Alzheimer's disease.** *Sci Rep* 2016, **6**:28387.

60. Nelson CE, Hakim CH, Ousterout DG, Thakore PI, Moreb EA, Castellanos Rivera RM, Madhavan S, Pan X, Ran FA, Yan WX, *et al.*: **In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy.** *Science* 2016, **351**:403–407.

61. Yin H, Xue W, Chen S, Bogorad RL, Benedetti E, Grompe M, Koteliansky V, Sharp PA, Jacks T, Anderson DG: **Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype.** *Nat Biotechnol* 2014, **32**:551–553.

62. Illendula A, Pulikkan JA, Zong H, Grembecka J, Xue L, Sen S, Zhou Y, Boulton A, Kuntimaddi A, Gao Y, *et al.*: **Chemical biology. A small-molecule inhibitor of the aberrant transcription factor CBFbeta-SMMHC delays leukemia in mice.** *Science* 2015, **347**:779–784.