

The power of multiplexed functional analysis of genetic variants

Molly Gasperini¹, Lea Starita¹ & Jay Shendure^{1,2}

New technologies have recently enabled saturation mutagenesis and functional analysis of nearly all possible variants of regulatory elements or proteins of interest in single experiments. Here we discuss the past, present, and future of such multiplexed (functional) assays for variant effects (MAVEs). MAVEs provide detailed insight into sequence–function relationships, and they may prove critical for the prospective clinical interpretation of genetic variants.

Genome sequencing is now routine, and tools for annotating gene structures and regulatory elements are becoming increasingly mature. Yet in this time of genomic plenty, researchers remain poor at predicting genotype–phenotype relationships, that is the consequences of genetic variation. Which single-nucleotide changes will affect gene regulation? Which amino acid changes will affect protein function? Under what circumstances do the resulting biochemical phenotypes give rise to organismal phenotypes? For regulatory, protein-coding, and organismal phenotypes, what is the distribution of effect sizes within the space of all possible sequence variants? What risk does each confer for disease, and to what degree do they affect characteristics such as age of onset and severity of disease? Although methods for answering these questions by computational prediction have proliferated, their effectiveness is limited, and the conventional approach to confirm that an individual variant has a meaningful biochemical or organism-level effect is still to assay it in an *in vitro* system or model organism¹. This one-by-one, *post hoc* approach does not scale to the vast numbers of genetic variants that are being discovered each day by clinical exome and genome sequencing.

Within the past decade, various innovations have enabled the assignment of functional effects to hundreds to thousands of sequence variants in a single, highly multiplexed experiment; here we term these ‘multiplexed assays for variant effects’, or MAVEs. MAVE experiments have mapped sequence–function relationships with base-pair resolution for both proteins and regulatory elements in the form of deep mutational scans² and massively parallel reporter assays³, respectively. Deep mutational scans build on low-throughput predecessors such as alanine scanning, in which each variant is cloned and assayed individually⁴, and surveys of variant libraries by protein-display

methods^{5–7}, which suffer from reliance on capillary sequencing that allows function to be assigned to only a small number of winning variants. Massively parallel reporter assays build on methods such as saturation mutagenesis⁸ that are similarly limited with respect to scalability.

From a technical perspective, underlying the development of MAVEs are advances in massively parallel DNA synthesis⁹ and DNA sequencing¹⁰ that, respectively, enable the multiplex construction of genetic variant libraries and the multiplex quantification of functional consequences. Using oligonucleotide libraries generated by massively parallel DNA synthesis, we can now—within a single experiment—program every single-nucleotide change in a regulatory region or every possible amino acid change in a protein. Using next-generation DNA sequencing, it is now possible to track and quantify the functional effects of all of these variants within a single experiment. Even more recently, breakthroughs in genome engineering^{11–13} have enabled MAVE approaches for assaying the functional consequences of variants in their native genomic context¹⁴. With MAVE methods, the number of variants that can be functionally tested increases to hundreds or thousands per experiment.

Already, MAVE approaches have shown their utility for sequence–function analysis of diverse classes of sequence, including enhancers, promoters, mRNA untranslated regions, splice sites, and numerous kinds of proteins (**Fig. 1**). Though analogously multiplexed assays have been developed that validate thousands of putative regulatory elements at once¹⁵ or scan endogenous genomic space to dissect novel regulatory elements¹⁶, we focus here primarily on the dense dissection of sequences of interest, that is, measuring the effects of all possible nucleotide substitutions in a regulatory element or amino acid substitutions in a protein.

Although there is considerable variety in the details of each implementation, MAVE experiments share a basic framework, with key steps represented in **Figure 2**: (1) construction of a variant library (i.e., allelic series) of the sequence of interest, (2) delivery of this variant library to an *in vitro* or *in vivo* system,

¹Department of Genome Sciences, University of Washington, Seattle, Washington, USA. ²Howard Hughes Medical Institute, Seattle, Washington, USA. Correspondence should be addressed to J.S. (shendure@uw.edu).

Received 8 June; accepted 13 July; published online 1 September 2016;
doi:10.1038/nprot.2016.135

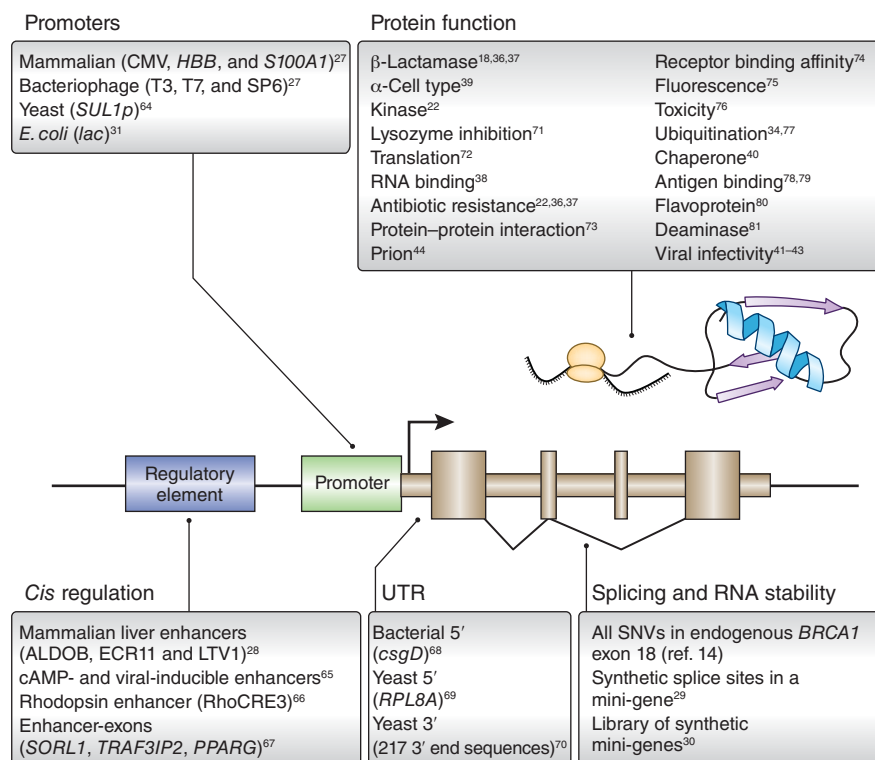


Figure 1 | Multiplexed assays for variant effects (MAVEs) throughout the central dogma. In the past decade, MAVEs have functionally tested at least tens to hundreds of thousands of variants across diverse sequences corresponding to all parts of the central dogma. Gene regulatory elements such as enhancers, promoters, and untranslated regions (UTRs) have been dissected, as have regions that affect splicing and RNA stability. All possible amino acid changes across proteins of diverse function have been characterized, yielding insights into biochemistry and disease.

(3) the functional assay (i.e., the stratification of variants by function), (4) sequencing to quantify each variant's representation in the context of the assay, and (5) calculation and calibration of functional scores for each variant. Ideally, this workflow results in sequence–function maps that capture the effect size of every possible variant at every position in the sequence of interest, with respect to the function assayed and potentially its correlates as well (e.g., clinical phenotype) (step 6 in Fig. 2). Below, we discuss each of these steps in greater detail.

Construction of a variant library

It is challenging to generate a variant library (also referred to as an allelic series) that uniformly represents all possible nucleotide or amino acid substitutions in an efficient and cost-effective manner. Classic methods such as error-prone PCR suffer from polymerase bias¹⁷. Doped oligonucleotides are limited in length (~200 bp), and although they can produce uniform libraries, methods that rely on individually synthesized oligonucleotides (i.e., one or two primers for each programmed mutation of a single base or codon) are costly and labor-intensive^{18–20}. However, despite these limitations, these remain viable options for constructing variant libraries of both regulatory and protein-coding sequences.

In 2004, Cleary *et al.*⁹ demonstrated that the products of microarray-based DNA synthesis can be used in a preparative rather than analytical fashion—that is, they can be released from the array surface and used as high-complexity oligonucleotide

libraries. Cost-effective, array-based DNA synthesis technologies are still improving in terms of the length and quality of their products. Despite this, the ability to program complex variant libraries from arrays has facilitated much of the MAVE work of the past decade. There are many ways by which array-derived libraries can be used to program variant libraries for MAVE experiments. For example, all possible alternative codons can be programmed on array-derived primers and incorporated into a coding sequence by sequential primer extensions from a wild-type sequence²¹ or by Gibson assembly²².

Delivery of the variant library

Variant libraries can be introduced to populations of cells via episomes or by insertion into the genome. If one is measuring regulatory activity via transcribed barcodes (Fig. 2, left), a high multiplicity of delivery per cell (i.e., multiple episomal or lentiviral reporters in a single cell) is permissible, if not desirable. However, in assays where the effect of a variant is assessed on the basis of the phenotype of its host cell (Fig. 2, right), the method must limit the number of alleles delivered per cell to one to avoid confounding the impact of any single variant. For bacteria- and yeast-based assays, alleles can be delivered by plasmid because the contribution of cotransformed cells is negligible. Delivery of a single allele to a mammalian cell is more challenging. Alleles can be randomly inserted in the genome by viral transduction^{23,24} at a multiplicity of infection of less than one, or by targeting only one locus for integration via integrase- or recombinase-mediated

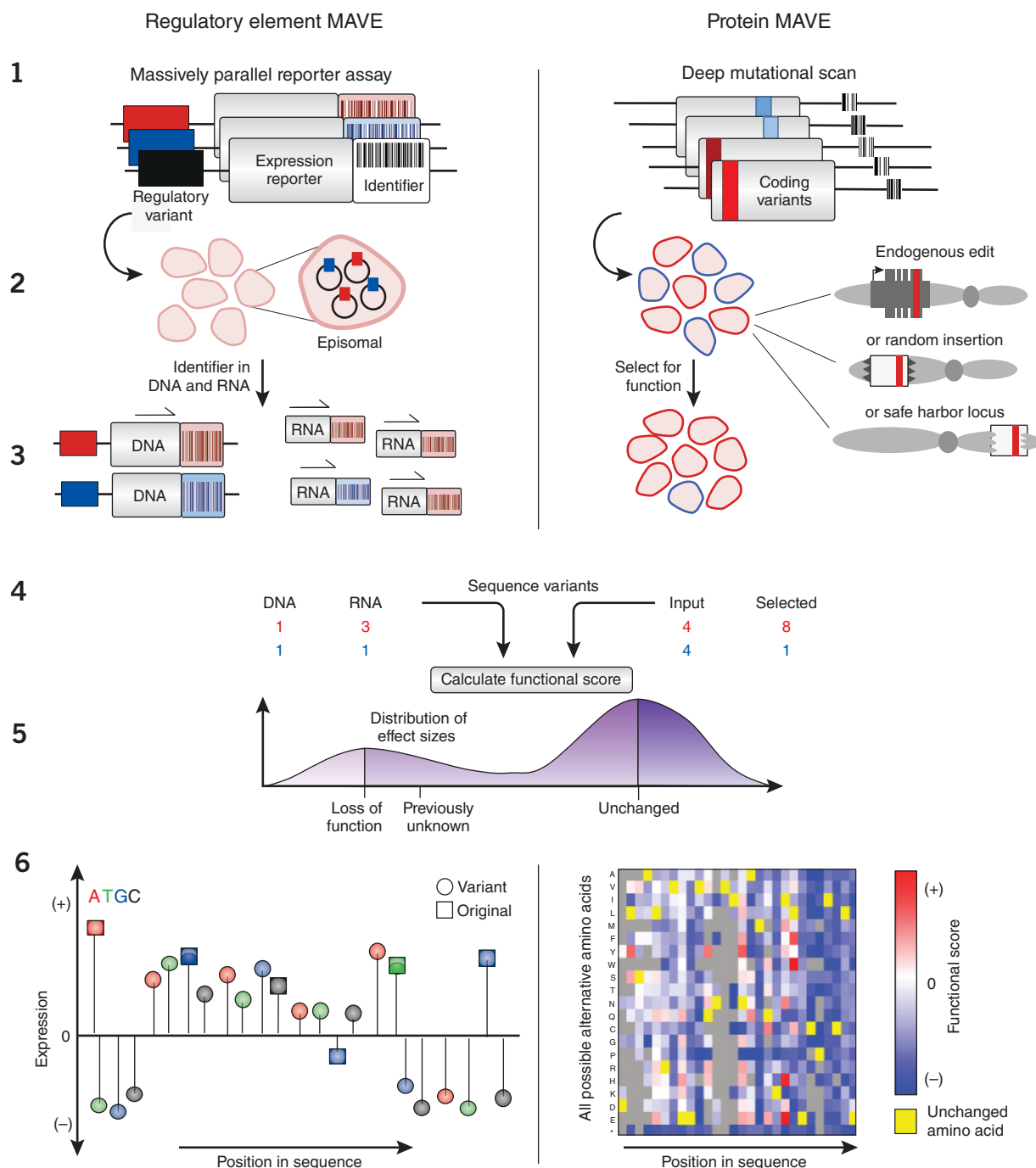


Figure 2 | The key steps of MAVE. Though diverse, all MAVE experiments rely on the same steps: (1) Construction of a variant library or allelic series of the sequence of interest. These variants might include all possible amino acid changes in a protein or all single-nucleotide changes in a regulatory element. (2) Delivery of this variant library to a model system. Variant libraries can be delivered episomally or via genomic integration by genome editing, by random insertion, or at a safe harbor locus. (3) A functional assay to stratify variants by function. Effects on RNA expression from variant regulatory sequences are measured by using sequencing to count transcripts under the influence of each variant. In protein MAVEs, assays are used that separate coding sequences for functional versus nonfunctional variants. (4) Sequencing to quantify each variant's representation in the context of the assay. For regulatory sequences, DNA and RNA that tag each variant can be sequenced to quantify effects on transcriptional output. Protein-coding variants (or variant-associated tags/barcodes) are sequenced before and after functional selection. (5) Calculation and calibration of functional scores for each variant. Sequencing read counts must be converted into a score for each variant. These scores range over a distribution of all possible effect sizes, and this distribution can be benchmarked by variants of known effect. (6) The genotype–phenotype relationship at every position in the interrogated sequence is represented in sequence–function maps.

insertion^{25,26}. We have also used CRISPR/Cas9 genome editing to introduce libraries of variants to their endogenous locus, such that ploidy limits the number of copies introduced per cell¹⁴.

Although currently somewhat inefficient, this approach has the major advantage of enabling variants to be assayed in their native genomic context.

Stratification of variants by function

The quality of MAVE measurements hinges on the ability of a functional assay to accurately stratify variants by their impact on the biochemical or cellular activity of interest. The design and validation of a well-performing functional assay is perhaps the most challenging aspect of implementing MAVEs. Broadly speaking, the design considerations are different for regulatory elements than for protein-coding sequences.

The impact of regulatory-sequence variation is most often stratified by changes in the transcriptional output of a reporter gene. To assess the effect of programmed regulatory variants on gene regulation, targeted RNA-seq can be used to count *cis*-linked reporter transcripts that contain a barcode uniquely paired with a specific variant²⁷. For example, for enhancer reporter assays, regulatory variants might increase or decrease transcriptional activation of associated barcodes, relative to the wild-type enhancer^{22,27,28}. Similar approaches have also been used to measure variant effects on splicing, wherein the regulatory variants might also serve as the barcode^{14,29,30}. Alternatively, a fluorescent reporter protein can be used as a proxy for RNA expression, with cells separated into brightness bins by fluorescence-activated cell sorting^{31,32}. In this case, the variants present in each bin can be quantified from the sorted DNA and inferred to have either increased or decreased expression of the reporter gene.

Designing suitable assays for protein MAVEs is more challenging. Protein MAVEs generally require delivery of one construct per cell, and protein function needs to be tied to either cellular growth or reporters that can be sorted by flow cytometry. Though each assay is highly specific to the protein of interest, recurring themes in the functional assays that are used for protein MAVE include protein display and capture^{33–35}, antibiotic resistance^{22,36,37}, cellular growth^{38–40}, viral infectivity^{41–43}, and protein- or antigen-binding affinity^{44,45}.

Looking to the future, we predict the development of new functional assays in two complementary directions. First, to enable the effective scaling of MAVEs to larger swaths of the proteome, we predict that MAVEs that measure generic protein properties such as stability or localization will be of great utility. Second, to accurately measure the effects of variants on genes associated with disease risk, there will be a strong incentive to develop multiplexing-compatible assays that specifically model the activities of a protein that are thought to be most relevant to its role in disease⁴⁶.

Sequencing to count variant frequency

After stratification in a functional assay, MAVEs rely on massively parallel DNA sequencing to provide a digital ‘count’ for each variant. Variants (or associated barcodes) are amplified from the functionally stratified DNA or RNA and sequenced to determine the frequency of each in each post-assay sample.

Provided that the assay results in changes in the representation of variants, the mutagenized region can be sequenced directly³³. However, if the variant is not part of the assay ‘output’, or if the mutagenized region is long and cannot be easily covered by cost-effective sequencing platforms, then each mutant can be tagged with a short barcode for the purposes of readout/quantification²⁷. Such barcodes can be linked to each

variant in the mutagenized regulatory element or protein via subassembly^{28,34,47} or long-read sequencing (L.M. Starita, M. Kircher, J. Underwood & J. Shendure, unpublished data). After barcodes have been linked to variants, only the barcodes need be sequenced to track or quantify the variants.

Calculating and calibrating functional scores for each variant

In order for MAVE experiments to be readily interpreted, the number of sequencing reads for each variant must be converted to a meaningful functional score. Multiple statistical models exist to convert read counts into scores^{48–51}, including simple but effective ratios of variant frequencies in RNA/DNA or selected/unselected populations.

A powerful and unique aspect of MAVE experiments, relative to conventional one-by-one functional assays, is that they result in a distribution of effect sizes for a large number of potential variants of a sequence of interest, all generated within a single experiment. To interpret this distribution, ideally one should compare a variant against benchmarks of known or expected effect. In protein MAVEs, stop codons represent the worst outcome, whereas synonymous changes are expected to have neutral effects. These can be used to validate and calibrate the distribution of observations, as well as to quantify uncertainty in measurements. For regulatory MAVEs, well-characterized regulatory motifs can be used as positive controls in noncoding MAVEs, whereas scrambled sequences with no expected regulatory effect can be used as negative controls⁵². The remaining variants of unknown effect are then compared to such controls. For disease-relevant sequences, previously observed pathogenic and benign substitutions can be used to calibrate MAVE scores for use in a clinical context^{45,53}. We predict that the increase in the number of available human genotype–phenotype data sets, such as that generated by the Exome Aggregation Consortium for estimating allele frequencies⁵⁴, and the expansion of ClinVar as a source for benign and pathogenic substitutions⁵⁵ will allow scientists to better calibrate and interpret the dense sets of experimental variant effects resulting from MAVEs of disease-relevant proteins and regulatory elements.

The future of MAVE

As the progress of human genetics is increasingly limited by the interpretation of genetic variants rather than by their ascertainment, we predict that the adoption and application of MAVE experiments will accelerate in the coming years. Ideally, by the 20th anniversary of *Nature Protocols*, sequence–function maps for thousands of proteins and regulatory elements^{56–58} will have been generated at single-residue resolution.

Such high-resolution sequence–function maps may provide the substrate for training more accurate computational models for directly predicting the impact of genetic variation on phenotypes. We are already observing this potential. For example, a model trained on the few hundred thousand splice sites in the human transcriptome⁵⁹ is less accurate than one trained on much larger data sets created via MAVE-like experiments designed to learn the rules of splice-site selection from millions of synthetic exons³⁰. By revealing relationships between regulatory sequence

composition and gene regulation, MAVEs will also shed light on the fine-scale ‘functional anatomy’ of the transcription factor binding motifs that comprise enhancers and promoters, and thereby allow scientists to build better algorithms to predict the effects of single-nucleotide variants on gene expression. For proteins, models trained on large numbers of MAVE experiments could also provide information about the relationship between primary sequence and protein properties such as stability, folding, and mutational tolerance.

Beyond advancing genomics in general, we also expect that MAVEs of disease-associated genes and regulatory elements will deliver experimental data needed to guide variant interpretation in the clinic. Genetic testing is identifying vast numbers of variants of unknown significance (VUSs). VUSs are often missense substitutions with unknown effects on the biochemistry of a disease-associated protein (though, importantly, splicing and other regulatory mutations can also lead to disease). To patients and physicians alike, ‘VUS’ is a confusing and unhelpful categorization. These variants are found at a high rate in panels of genes associated with cancer risk, where VUSs can outnumber interpretable ones by 95 to 1 (ref. 60). As a solution to the problem of VUS identification, the results of MAVEs have the potential to give rise to clinically calibrated scores for the impact of every possible amino acid change on the biochemistry of the tested protein-coding gene. Generating these variant interpretations prospectively will position those in the field to incorporate them into the routine practice of clinical genetics.

These high rates of VUSs come from surveying only coding sequences associated with disease, which represent about 1% of the genome. Scientists have just begun to understand how noncoding genetic variation in the other 99% of the genome influences disease. More than 24,000 loci, most of which occur in noncoding regions, have been associated with human phenotypes through genome-wide association studies⁵⁸. MAVEs can not only help determine which of these implicated sequences contribute to gene expression, but also test noncoding variants associated with disease^{61–63}. In recent regulatory MAVE studies, CRISPR-facilitated allelic replacement was used to functionally validate individual prioritized variants, as studying these hits in their native genomic context conferred biological advantages over episomal assays; in the genome, contextual transcription factors, polymerases and 3D architecture are maintained for promoters and distal regulatory elements⁶², and native splicing machinery is available^{14,63}. Saturation genome editing has already been applied to survey thousands of variants in the genome¹⁴, and we predict that as genome engineering protocols rapidly improve, the clear advantages of testing variants in their endogenous context will motivate the use of genome editing in the next decade’s MAVEs.

Additionally, we predict that genome editing may enable MAVEs of a given allelic series in the context of more than one cell line (i.e., reflecting the different genetic backgrounds of patients, which are known to influence the penetrance or variable expressivity of many genetic diseases even when the causal mutations are identical). One can also envision quantifying epistatic effects between variants or allelic series that are concurrently introduced to multiple endogenous locations

scattered across the genome. Such experiments have the potential to shed light on how epistatic effects influence phenotypic traits, which at present remains poorly understood.

Massively parallel functional dissection of proteins and regulatory elements generates empirical measurements of the consequences of thousands of variants per experiment. The distributions of effect sizes and sequence–function maps inform biology and may have a critical role in the clinical interpretation of genetic variation. We anticipate that as methods pioneered over the past 10 years are scaled up over the next decade, MAVEs may enable measurement of the functional consequences of millions to billions of genetic variants.

ACKNOWLEDGMENTS The authors thank the Shendure lab, and in particular R. Hause, for discussions. M.G. is a National Science Foundation Graduate Research Fellow. J.S. is an Investigator of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS M.G., L.S., and J.S. prepared the manuscript.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Botstein, D. & Shortle, D. Strategies and applications of *in vitro* mutagenesis. *Science* **229**, 1193–1201 (1985).
2. Fowler, D.M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **11**, 801–807 (2014).
3. Inoue, F. & Ahtuv, N. Decoding enhancers using massively parallel reporter assays. *Genomics* **106**, 159–164 (2015).
4. Cunningham, B. & Wells, J. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* **244**, 1081–1085 (1989).
5. Smith, G. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315–1317 (1985).
6. Boder, E.T. & Wittrup, K.D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557 (1997).
7. Amstutz, P. *et al.* In vitro selection for catalytic activity with ribosome display. *J. Am. Chem. Soc.* **124**, 9396–9403 (2002).
8. Myers, R., Tilly, K. & Maniatis, T. Fine structure genetic analysis of a beta-globin promoter. *Science* **232**, 613–618 (1986).
9. Cleary, M.A. *et al.* Production of complex nucleic acid libraries using highly parallel *in situ* oligonucleotide synthesis. *Nat. Methods* **1**, 241–248 (2004).
10. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26**, 1135–1145 (2008).
11. Ran, F.A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
12. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013).
13. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
14. Findlay, G.M., Boyle, E.A., Hause, R.J., Klein, J.C. & Shendure, J. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* **513**, 120–123 (2014).
15. White, M.A. Understanding how cis-regulatory function is encoded in DNA sequence using massively parallel reporter assays and designed sequences. *Genomics* **106**, 165–170 (2015).
16. Canver, M.C. *et al.* BCL11A enhancer engineering by Cas9-mediated *in situ* saturating mutagenesis. *Nature* **527**, 192–197 (2015).
17. Wong, T.S., Roccatano, D., Zacharias, M. & Schwaneberg, U. A statistical analysis of random mutagenesis methods used for directed protein evolution. *J. Mol. Biol.* **355**, 858–871 (2006).
18. Firnberg, E. & Ostermeier, M. PFunkel: efficient, expansive, user-defined mutagenesis. *PLoS One* **7**, e52031 (2012).
19. Jain, P.C. & Varadarajan, R. A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library. *Anal. Biochem.* **449**, 90–98 (2014).
20. McLaughlin, R.N. Jr., Poelwijk, F.J., Raman, A., Gosal, W.S. & Ranganathan, R. The spatial architecture of protein function and adaptation. *Nature* **491**, 138–142 (2012).

21. Kitzman, J.O., Starita, L.M., Lo, R.S., Fields, S. & Shendure, J. Massively parallel single-amino-acid mutagenesis. *Nat. Methods* **12**, 203–206 (2015).
22. Melnikov, A., Rogov, P., Wang, L., Gnirke, A. & Mikkelsen, T.S. Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Res.* **42**, e112 (2014).
23. Naldini, L. *et al.* In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272**, 263–267 (1996).
24. Deyle, D.R. & Russell, D.W. Adeno-associated virus vector integration. *Curr. Opin. Mol. Ther.* **11**, 442–447 (2009).
25. Craig, N.L. The mechanism of conservative site-specific recombination. *Annu. Rev. Genet.* **22**, 77–105 (1988).
26. Sauer, B. Site-specific recombination: developments and applications. *Curr. Opin. Biotechnol.* **5**, 521–527 (1994).
27. Patwardhan, R.P. *et al.* High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat. Biotechnol.* **27**, 1173–1175 (2009).
28. Patwardhan, R.P. *et al.* Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* **30**, 265–270 (2012).
29. Ke, S. *et al.* Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res.* **21**, 1360–1374 (2011).
30. Rosenberg, A.B., Patwardhan, R.P., Shendure, J. & Seelig, G. Learning the sequence determinants of alternative splicing from millions of random sequences. *Cell* **163**, 698–711 (2015).
31. Kinney, J.B., Murugan, A., Callan, C.G. Jr. & Cox, E.C. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc. Natl. Acad. Sci. USA* **107**, 9158–9163 (2010).
32. Sharon, E. *et al.* Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* **30**, 521–530 (2012).
33. Fowler, D.M. *et al.* High-resolution mapping of protein sequence–function relationships. *Nat. Methods* **7**, 741–746 (2010).
34. Starita, L.M. *et al.* Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis. *Proc. Natl. Acad. Sci. USA* **110**, E1263–E1272 (2013).
35. Whitehead, T.A. *et al.* Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol.* **30**, 543–548 (2012).
36. Stiffler, M.A., Hekstra, D.R. & Ranganathan, R. Evolvability as a function of purifying selection in TEM-1 β -lactamase. *Cell* **160**, 882–892 (2015).
37. Deng, Z. *et al.* Deep sequencing of systematic combinatorial libraries reveals β -lactamase sequence constraints at high resolution. *J. Mol. Biol.* **424**, 150–167 (2012).
38. Melamed, D., Young, D.L., Miller, C.R. & Fields, S. Combining natural sequence variation with high throughput mutational data to reveal protein interaction sites. *PLoS Genet.* **11**, e1004918 (2015).
39. Kim, I., Miller, C.R., Young, D.L. & Fields, S. High-throughput analysis of in vivo protein stability. *Mol. Cell. Proteomics* **12**, 3370–3378 (2013).
40. Hietpas, R.T., Jensen, J.D. & Bolon, D.N. Experimental illumination of a fitness landscape. *Proc. Natl. Acad. Sci. USA* **108**, 7896–7901 (2011).
41. Thyagarajan, B. & Bloom, J.D. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *eLife* **3**, 1–26 (2014).
42. Doud, M.B., Ashenberg, O. & Bloom, J.D. Site-specific amino acid preferences are mostly conserved in two closely related protein homologs. *Mol. Biol. Evol.* **32**, 2944–2960 (2015).
43. Bloom, J.D. An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Mol. Biol. Evol.* **31**, 1956–1978 (2014).
44. Doolan, K.M. & Colby, D.W. Conformation-dependent epitopes recognized by prion protein antibodies probed using mutational scanning and deep sequencing. *J. Mol. Biol.* **427**, 328–340 (2015).
45. Starita, L.M. *et al.* Massively parallel functional analysis of BRCA1 RING domain variants. *Genetics* **200**, 413–422 (2015).
46. Green, R.C. *et al.* ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* **15**, 565–574 (2013).
47. Hiatt, J.B., Patwardhan, R.P., Turner, E.H., Lee, C. & Shendure, J. Parallel, tag-directed assembly of locally derived short sequence reads. *Nat. Methods* **7**, 119–122 (2010).
48. Bloom, J.D. Software for the analysis and visualization of deep mutational scanning data. *BMC Bioinformatics* **16**, 168 (2015).
49. Fowler, D.M., Araya, C.L., Gerard, W. & Fields, S. Enrich: software for analysis of protein function by enrichment and depletion of variants. *Bioinformatics* **27**, 3430–3431 (2011).
50. Matuszewski, S., Hildebrandt, M.E., Ghenu, A.-H., Jensen, J.D. & Bank, C. A statistical guide to the design of deep mutational scanning experiments. Preprint at <http://biorxiv.org/content/early/2016/06/29/048892> (2016).
51. Ireland, W.T. & Kinney, J.B. Sort-Seq Tools: sequence-function relationship modeling for massively parallel assays. Preprint at <http://biorxiv.org/content/early/2016/05/21/054676> (2016).
52. White, M.A., Myers, C.A., Corbo, J.C. & Cohen, B.A. Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. *Proc. Natl. Acad. Sci. USA* **110**, 11952–11957 (2013).
53. Majithia, A.R. *et al.* Rare variants in PPARG with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes. *Proc. Natl. Acad. Sci. USA* **111**, 13127–13132 (2014).
54. Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
55. Landrum, M.J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2016).
56. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
57. Ardlie, K.G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
58. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014).
59. Xiong, H.Y. *et al.* The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806 (2015).
60. Maxwell, K.N. *et al.* Evaluation of ACMG-guideline-based variant classification of cancer susceptibility and non-cancer-associated genes in families affected by breast cancer. *Am. J. Hum. Genet.* **98**, 801–817 (2016).
61. Vockley, C.M. *et al.* Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. *Genome Res.* **25**, 1206–1214 (2015).
62. Tewhey, R. *et al.* Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell* **165**, 1519–1529 (2016).
63. Ulirsch, J.C. *et al.* Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell* **165**, 1530–1545 (2016).
64. Rich, M.S. *et al.* Comprehensive analysis of the SUL1 promoter of *Saccharomyces cerevisiae*. *Genetics* **203**, 191–202 (2016).
65. Melnikov, A. *et al.* Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* **30**, 271–277 (2012).
66. Kwasniewski, J.C., Mogno, I., Myers, C.A., Corbo, J.C. & Cohen, B.A. Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc. Natl. Acad. Sci. USA* **109**, 19498–19503 (2012).
67. Birnbaum, R.Y. *et al.* Systematic dissection of coding exons at single nucleotide resolution supports an additional role in cell-specific transcriptional regulation. *PLoS Genet.* **10**, e1004592 (2014).
68. Holmqvist, E., Reimegård, J. & Wagner, E.G.H. Massive functional mapping of a 5'-UTR by saturation mutagenesis, phenotypic sorting and deep sequencing. *Nucleic Acids Res.* **41**, e122 (2013).
69. Dvir, S. *et al.* Deciphering the rules by which 5'-UTR sequences affect protein expression in yeast. *Proc. Natl. Acad. Sci. USA* **110**, E2792–E2801 (2013).
70. Shalem, O. *et al.* Systematic dissection of the sequence determinants of gene 3' end mediated expression control. *PLoS Genet.* **11**, e1005147 (2015).
71. Procko, E. *et al.* Computational design of a protein-based enzyme inhibitor. *J. Mol. Biol.* **425**, 3563–3575 (2013).
72. Lind, P.A., Berg, O.G. & Andersson, D.I. Mutational robustness of ribosomal protein genes. *Science* **330**, 825–827 (2010).
73. Podgornaia, A.I. & Laub, M.T. Pervasive degeneracy and epistasis in a protein-protein interface. *Science* **347**, 673–677 (2015).
74. Pál, G., Kouadio, J.L.K., Artis, D.R., Kosiakoff, A.A. & Sidhu, S.S. Comprehensive and quantitative mapping of energy landscapes for protein-protein interactions by rapid combinatorial scanning. *J. Biol. Chem.* **281**, 22378–22385 (2006).
75. Sarkisyan, K.S. *et al.* Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397–401 (2016).
76. Adkar, B.V. *et al.* Protein model discrimination using mutational sensitivity derived from deep sequencing. *Structure* **20**, 371–381 (2012).
77. Roscoe, B.P., Thayer, K.M., Zeldovich, K.B., Fushman, D. & Bolon, D.N.A. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. *J. Mol. Biol.* **425**, 1363–1377 (2013).
78. Forsyth, C.M. *et al.* Deep mutational scanning of an antibody against epidermal growth factor receptor using mammalian cell display and massively parallel pyrosequencing. *MABS* **5**, 523–532 (2013).
79. Van Blarcom, T. *et al.* Precise and efficient antibody epitope determination through library design, yeast display and next-generation sequencing. *J. Mol. Biol.* **427**, 1513–1534 (2015).
80. Shin, H. *et al.* Exploring the functional residues in a flavin-binding fluorescent protein using deep mutational scanning. *PLoS One* **9**, e97817 (2014).
81. Gajula, K.S. *et al.* High-throughput mutagenesis reveals functional determinants for DNA targeting by activation-induced deaminase. *Nucleic Acids Res.* **42**, 9964–9975 (2014).