MaveDB: an open-source platform to distribute and interpret data from multiplexed assays of variant effect

Daniel Esposito¹, Jochen Weile²,³,⁴⁵†, Jay Shendure⁶,⁷, Lea M. Starita⁶,⁷, Anthony T. Papenfuss¹,⁹,¹⁰,¹¹,¹², Frederick P. Roth²,³,⁴,⁵,¹³*, Douglas M. Fowler⁶,¹³,¹⁴* and Alan F. Rubin¹,⁹,¹⁰*

Abstract

Multiplex assays of variant effect (MAVEs), such as deep mutational scans and massively parallel reporter assays, test thousands of sequence variants in a single experiment. Despite the importance of MAVE data for basic and clinical research, there is no standard resource for their discovery and distribution. Here, we present MaveDB (https://www.mavedb.org), a public repository for large-scale measurements of sequence variant impact, designed for interoperability with applications to interpret these datasets. We also describe the first such application, MaveVis, which retrieves, visualizes, and contextualizes variant effect maps. Together, the database and applications will empower the community to mine these powerful datasets.

Keywords: Deep mutational scanning, Massively parallel reporter assays, Large-scale mutagenesis, MAVE, Multiplexed assay of variant effect, Genome interpretation, Personalized medicine

Background

Experimentally interrogating the effects of genetic variation has helped reveal the mechanisms by which genes function and facilitate an understanding of the clinical consequences of human genetic variation. Multiplex assays of variant effect (MAVEs) leverage high-throughput DNA sequencing to greatly increase the scale at which variants can be experimentally investigated [1–3]. A MAVE yields a set of scores that describe the functional effect of thousands to tens of thousands of variants of a coding sequence, promoter, enhancer, or another genetic element relative to a reference sequence. MAVEs are being adopted rapidly for both basic research and clinical applications [4]. As a consequence, the total number of variants with functional data generated by MAVEs was predicted to surpass 200,000 by the end of 2018 [3], meaning there are now more MAVE-annotated variants in the literature than classified missense variants available in ClinVar [5].

These large-scale variant effect maps are yielding insights into protein function, structure, and evolution [6–10]: exploring gene regulation and promoter function [11–13]; improving computational variant effect prediction [14, 15]; and guiding variant interpretation in the clinic [16–21]. However, the impact of variant effect maps has been limited by shortcomings in data availability, dissemination, and discoverability. Nearly half of the publications describing large-scale variant effect mapping do not provide variant effect scores for all variants that were assayed (Additional file 1: Figure S1) [6–8, 11, 13, 16–169]. When variant effect scores are provided, they are typically accessible only as a supplementary table or via a bespoke web interface [13, 18–20, 70, 170] leading to a proliferation of inconsistent formats. Some publications, instead of including variant effect scores, deposit the associated high-throughput DNA sequencing data in the Sequence Read Archive or Gene Expression Omnibus [171, 172]. This raw data can be used to reconstruct variant effect scores, but accurately replicating the original analysis can be non-trivial. While databases
of variant effect information exist, they are typically designed for a specific application [173–175] or a specific group of target genes [176–178]. Larger and more general databases can sometimes contain variant effect data [179, 180], but these platforms were not developed with large-scale variant effect maps in mind, so valuable context for the variant effect scores and associated metadata may be lost. Furthermore, most existing resources lack support for noncoding targets entirely.

To overcome these challenges and facilitate future advances, we are establishing an open-source platform for MAVE resources. The foundation is MaveDB, a central repository that allows researchers to store and publish processed MAVE datasets, associated metadata, and linked raw data using a machine-readable, standardized, and searchable format. An easy-to-use web interface maximizes the impact and usefulness of researchers’ work by making the data readily accessible to the whole community, whether for clinical applications, meta-analysis, or reanalysis as computational techniques are refined.

This platform is designed to allow additional applications to communicate directly with MaveDB. The first of potentially many such applications, MaveVis, visualizes and provides context to protein variant effect maps by generating heatmaps and automatically integrating them with secondary structure, surface accessibility, interaction interfaces, and conservation data.

**Construction and content**

To capture the structure of real-world study designs, MaveDB is organized hierarchically into score sets, experiments, and experiment sets (Fig. 1). Score sets, the most basic unit of organization, contain the variant effect scores and additional metadata such as target sequence information and detailed methods. Each variant effect score is a numeric value. Optional data columns containing values related to each variant effect score such as variant counts and measures of uncertainty can also be included and named by the user.

Most experimental designs in MaveDB involve multiple score sets. For example, protein MAVEs commonly have one score set for nucleotide variants and another for amino acid variants [23]. Experiments with tiled designs [54, 92] can have score sets for each tile, and experiments with multiple distinct reference sequences [147] can have score sets for each reference sequence [26, 147]. In addition, we envision that reanalysis and renormalization of the existing datasets using updated methods will be commonplace [14, 15, 181, 182]. By grouping all analyses of a single raw dataset under one experiment, MaveDB ensures that the number of assays performed on each target sequence can be tracked accurately.

Each experiment describes one or more analyses arising from a single MAVE, including any technical and biological replicates. In addition to the links to score sets, experiments contain metadata including methodological details, links to raw data, and associated publications (Table 1), but no variant score information. Experiment sets contain one or more related experiments, for example, multiple MAVEs performed on the same target sequence under different conditions or multiple experiments from the same publication (Fig. 1).

MaveDB currently contains over one million variant effect scores across 45 unique targets. We welcome both new and previously described datasets from the community and have implemented a conversion tool, mavedb-convert, for datasets generated by Enrich [183], Enrich2 [182], and EMPIRIC [184] (see the “Availability of data and materials” section).

MaveDB is implemented in Python using the Django Python Web framework [185, 186]. The relational database backend is PostgreSQL [187]. The full database schema is shown in Additional file 1: Figure S2. Asynchronous tasks such as handling file uploads and sending emails are managed using RabbitMQ and Celery [188, 189]. Variant score and count data are stored using PostgreSQL JSONField objects, which offer additional flexibility for storing arbitrarily named data columns compared to a more traditional relational database design. Database accession numbers for publicly accessible entries are assigned in Universal Resource Name (URN) format [190].

Differences between each variant sequence and the target sequence are described using HGVS format [191]. MaveDB supports DNA and protein variant strings that describe substitutions or small indels in most sequence contexts, including splice site variants and multi-mutants.

Contributors are authenticated using their ORCID ID via the OAuth2 protocol [192, 193]. Consequently, an individual must have an ORCID ID to be named as a contributor to a MaveDB dataset. Users do not need to log in to browse or download publicly available data. MaveDB allows users to provide a private contact email address if they want to be contacted by administrators or receive alerts, but all other details are pulled from their public ORCID record.

Abstract and methods sections support Markdown blocks for formatted text with support for mathematical notation. Markdown blocks are rendered to HTML using Pandoc [195].

MaveVis is implemented using R [196] and Docker [197]. Surface accessibility and interface burial are calculated using FreeSasa [198]. Secondary structure is calculated using DSSP [199]. Conservation tracks are calculated using the AMAS algorithm [200], based on
Utility and discussion

Accessing datasets

MaveDB can be accessed through a standard web browser that allows users to explore by keyword, target gene, or organism. Alternatively, the advanced search function allows users to query all metadata fields, including the full text of methods and abstracts. Complete sets of variant effect scores and related values can be downloaded from any score set page in a comma-separated value format. These files can be parsed easily in most scientific programming environments or imported into spreadsheet applications.

multiple alignments computed using ClustalOmega [201] for the appropriate UniRef90 set of orthologous proteins with at least 90% sequence identity from UniProtKB [180].

**Fig. 1** Representative structures of MaveDB entries. Each panel depicts a single experiment set and its associated accession numbers. 

**A** A single target sequence analyzed using two different MAVEs. For example, MAVEs in phage display and yeast two-hybrid formats were performed on BRCA1 (experiment accessions ending in “-a” and “-b”) [16]. Each of these yielded two score sets, one for amino acid variants and one for nucleotide variants (score set accessions ending in “-1” and “-2”). 

**B** Two distinct target sequences analyzed using the same MAVE. For example, the VAMP-seq assay was performed on the PTEN and TPMT target sequences (experiment accessions ending in “-a” and “-b”) [18], yielding an amino acid score set for each target (score set accessions ending in “-1”).
The advanced search function is also accessible programmatically through the Representational State Transfer-Application Programming Interface (REST API). The API returns structured data, including full score sets and metadata, in JSON format, suitable for deserialization by most programming languages. Users of the R programming environment [196] can access MaveDB’s REST API using the rapimave library, which also includes a suite of exploration, searching, parsing, and filtering functions (see the “Availability of data and materials” section).

Creating new entries

Typically, a user starts by logging in using their ORCID ID and creating an experiment. The experiment can be added to an existing experiment set if desired; otherwise, a new one will be created automatically. The user provides a description of the assay used to generate the raw data, adds links to the raw data if available, and can then add contributors. After the experiment is created, the user creates one or more associated score sets. Here, the user enters the required information about the target such as its name and sequence and also describes the methods used to calculate the variant effect scores from the raw data. Variant effect scores and optional count files are then uploaded via the web interface and validated by the server.

Publishing datasets

When first created, score sets, experiments, and experiment sets are private and have temporary accession numbers. Private entries are only viewable by their contributors, and all values may be modified. Private entries can be accessed through the API by providing a contributor’s private access token generated on the contributor’s user profile page.

Completed private score sets can be published, making the score set publicly viewable. Publication creates a stable accession number and freezes the target sequence and variant effect score data, ensuring that all subsequent analyses based on the data are recomputable. Associated experiment and experiment sets are also published automatically if they are still private. Users may continue to edit some metadata such as the methods and descriptions after publication.

Published scores cannot be changed, but in case a correction is necessary, MaveDB allows score sets to be deprecated when creating a replacement. Users browsing MaveDB will only see the most recent version, but deprecated score sets will remain available by accession number to ensure that previous analyses are reproducible.

Contributor permissions

MaveDB supports three contributor roles: administrator, editor, and viewer. Administrators can add or remove contributors, modify entries, and publish score sets. Editors can modify entries but cannot affect the contributor list or make entries public. Viewers can see their private entries in the database but cannot change them.

All three roles appear in the contributor list with no visual distinction between them, and administrators can continue to change the contributor list for each score set or experiment after publication. Since score sets and experiments have independent contributor lists, MaveDB maintains clear attribution when datasets are reanalyzed.

Data licensing

Administrators may select one of several Creative Commons licenses for each score set [202–204], and additional licensing options may be added in response to user requests. The license information is included as score set metadata and as part of the header of each downloaded file. Administrators can relicense after publication, although users who download under a more permissive license would not be subject to a more restrictive license.

| Field name                | Experiment | Score set | Type        | Searchable | Link |
|---------------------------|------------|-----------|-------------|------------|------|
| Keyword                   | ✓          | ✓         | String      | ✓          | ✓    |
| Abstract                  | ✓          | ✓         | Markdown    | ✓          | ✓    |
| Method                    | ✓          | ✓         | Markdown    | ✓          | ✓    |
| Short description         | ✓          | ✓         | String      | ✓          | ✓    |
| Title                     | ✓          | ✓         | String      | ✓          | ✓    |
| PubMed ID                 | ✓          | ✓         | Accession   | ✓          | ✓    |
| DOI                       | ✓          | ✓         | Accession   | ✓          | ✓    |
| SRA accession             | ✓          |           | Accession   | ✓          | ✓    |
| RefSeq accession          | ✓          |           | Accession   | ✓          | ✓    |
| Ensembl accession         | ✓          |           | Accession   | ✓          | ✓    |
| UniProt accession         | ✓          |           | Accession   | ✓          | ✓    |
| Created by                | ✓          | ✓         | Contributor | ✓          | ✓    |
| Last modified by          | ✓          | ✓         | Contributor | ✓          | ✓    |
| Creation date             | ✓          | ✓         | DateTime    | ✓          | ✓    |
| Modification date         | ✓          | ✓         | DateTime    | ✓          | ✓    |
| Publication date          | ✓          | ✓         | DateTime    | ✓          | ✓    |
| License                   | ✓          |           | License     | ✓          | ✓    |
| Has replacement           | ✓          |           | Boolean     | ✓          | ✓    |

Table 1 MaveDB metadata fields
Visualizing variant effect maps

The MaveVis application allows users to quickly visualize score sets retrieved directly from MaveDB. One example of MaveVis output for a variant effect map of the protein SUMO1 [146] is shown in Fig. 2. Score sets are rendered as heatmaps with additional tracks representing integrated structural and conservation information from PDB [205] and UniprotKB [180]. The heatmap shows all possible amino acid changes at each protein sequence position, with colors reflecting the variant effect scores. The color scale is automatically calibrated based on the scores of reference and null alleles in the dataset or set manually by the user. Error bars are drawn directly on the heatmap fields to represent the measurement error provided in the score set, if present. Additional tracks show burial in protein interaction interfaces, residue-specific solvent accessibility, protein secondary structure, and sequence conservation.

MaveVis is hosted at http://varianteffect.org, a portal for applications built on MaveDB. Users can follow the MaveVis link on each MaveDB score set page or navigate directly to http://vis.varianteffect.org and search for datasets. Once a score set is selected, the corresponding UniProt accession from MaveDB is suggested when available. MaveVis automatically presents potentially relevant PDB structures for the selected protein that overlap with the score set target sequence, allowing users to select which structures to include in the visualization. The resulting plot can be downloaded in PNG, PDF, or SVG format.

In addition to the web interface, MaveVis also exists as an R package for local use (see the “Availability of data and materials” section). The R package provides direct access to both the visualization and underlying data integration functions, making it easy to automatically compile structural and conservation feature tables for individual proteins.

The MaveVis server automatically synchronizes with MaveDB at regular intervals via its API, caching any new score sets, automatically obtaining relevant PDB and UniProt data, and pre-calculating partial results for a more responsive user experience. MaveVis also exposes its own API, allowing it to be used within more complex workflows.

To facilitate communication between MaveVis and MaveDB, we developed an R package, hgvsParseR, to parse or assemble HGVS [191] strings that describe alleles (see the “Availability of data and materials” section). In addition to its utility for visualizing variant effect maps, we expect that this package will be generally useful for working with data from ClinVar [5], gnomAD [206], and other important sequence variation resources.

Conclusions

MaveDB is the foundation of an open-source platform for the collection, distribution, and analysis of variant effect maps. Designed to be flexible and extensible, the MaveDB repository can accommodate data from diverse target sequences and experimental methods as the field evolves. Using MaveDB to combine variant effect data with external contextual information, MaveVis is the first application built on this resource. We envision developing additional applications such as tertiary structure analysis, automatic imputation of missing values in variant effect maps [207], and a broadly applicable dashboard to assist dataset interpretation.

Fig. 2 Heatmap for the SUMO1 MAVE dataset rendered by MaveVis. The x-axis iterates over amino acid positions in the protein, while the y-axis lists all possible amino acid changes organized by their physicochemical properties. The heatmap color reflects the variant effect score, with blue being as deleterious as a full deletion (labeled “stop” in the color bar), white being equivalent to the reference allele (labeled “syn” in the color bar), and red representing a stronger phenotype than the reference residue at that position (labeled “hyper” in the color bar). Yellow cells indicate the reference amino acid at each position. Error bars represent standard error of the mean. The stacked bars above the heatmap represent the relative frequencies for each phenotype bin of the corresponding color at each position. Additional tracks show data integrated from other databases: orange heatmap tracks represent burial in protein interaction interfaces, the steel blue heatmap track represents solvent accessibility, the arrows and spirals correspond to secondary structure, and the yellow bar chart at the top shows sequence conservation.
MaveDB, MaveVis, and Enrich2 simplify, standardize, and democratize MAVE data analysis. These tools are the beginnings of a community-driven, open-source platform that allows researchers to explore these comprehensive datasets. The impact of each dataset will continue to increase as the number of assayed variants grows, contributing to a more complete understanding of genetic variation and sequence function.

Supplementary information
Supplementary information accompanies this paper at https://doi.org/10.1186/s13059-019-1845-6.

Additional file 1: Figure S1. Data unavailability is a persistent feature of the MAVE literature. We compiled a list of 159 publications that contained at least one deep mutational scanning or massively parallel reporter assay dataset and manually inspected the publication’s supplementary data and methods to determine whether counts or scores for the assayed variants were present. Refer to https://github.com/VariantEffect/MaveReferences for the full table. This figure was generated from release v0.1.1. Of the 159 total publications, 91 (57%) provide scores or counts. Figure S2. UML (Unified Markup Language) diagram of the complete MaveDB schema in PDF format. The diagram was generated using the Django Extensions package and visualized using Graphviz.

Additional file 2. Review history.

Acknowledgements
We would like to acknowledge Bernie Pope, Peter Georgeson, Nick Moore, Matthew Wakefield, and Dan Bolon for their helpful advice and guidance. We would also like to acknowledge the University of Washington Genome Sciences Information Technology Services team for their outstanding service and support of MaveDB during the development process.

Review history
Review history is available as Additional file 2.

Peer review information
Anahita Bishop was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Authors’ contributions
AFR, DMF, and FPR conceived the project. DE and AFR built the database and its web interface. JW built the MaveVis application. DE and JW built the APIs. All authors wrote the manuscript and approved the final version.

Funding
This work was supported by the National Institutes of Health (NIH; R01GM109110 and R01HG010461 to DMF) and the Brotman Baty Institute for Precision Medicine. DMF was a CIFAR Azrieli Global Scholar. ATP was supported by the Lorenzo and Pamela Galli Charitable Trust and by an Australian National Health and Medical Research Council (NHMRC) Program Grant (1054618) and NHMRC Senior Research Fellowship (1116955). The research benefited by support from the Victorian State Government Operational Infrastructure Support and Australian Government NHMRC Independent Research Institute Infrastructure Support. FPR and JW gratefully acknowledge the funding by the One Brave Idea Initiative, the National Human Genome Research Institute of the NIH Center of Excellence in Genomic Science Initiative (HG004233), the Canadian Excellence Research Chairs Program, and a Canadian Institutes of Health Research Foundation Grant.

Availability of data and materials
MaveDB is hosted at https://www.mavedb.org/.
MaveVis is hosted at https://vis.varianteffect.org/.
Source code for all websites, tools, and packages is available on GitHub at https://github.com/VariantEffect/

References
1. Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. Nat Methods. 2014;11:801–7.
2. Gasperini M, Sturita L, Shendure J. The power of multiplexed functional analysis of genetic variants. Nat Protoc. 2016;11:1782–7.
3. Welle J, Roth FP. Multiplexed assays of variant effects contribute to a growing genotype–phenotype atlas. Hum Genet. 2018. https://doi.org/10.1007/s00439-018-1916-x.
4. Sturita LM, Ahituv N, Dunham MJ, Kitzman JO, Roth FP, Seelig G, et al. Variant interpretation: functional assays to the rescue. Am J Hum Genet. 2017;101:315–25.
5. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014;42:980–5.
6. Stiffler MA, Hekstra DR, Ranganathan R. Evolvability as a function of purifying selection in TEM-1 β-lactamase. Cell. 2015;160:882–92.
7. Lee JM, Huddleston J, Doud NB, Hooper KA, Wu NC, Bedford T, et al. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. Proc Natl Acad Sci. 2018;201806133.
8. Cantor AJ, Shah NH, Kuriyan J. Deep mutational analysis reveals functional trade-offs in the sequences of E γFR autophosphorylation sites. Proc Natl Acad Sci. 2018;115:E7303–12.
9. Schmiedel JM, Lehner B. Determining protein structures using deep mutagenesis. Nat Genet. 2019;51:1177–86.
10. Rollins NJ, Brock KP, Poelwijk FJ, Stiffler MA, Gauthier NP, Sander C, et al. Inferring protein 3D structure from deep mutation scans. Nat Genet. 2019;51:1170–6.
11. Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, et al. Massively parallel functional dissection of mammalian enhancers in vivo. Nat Biotechnol. 2012;30:265–70.
12. Chatterjee S, Ahituv N. Gene regulatory elements, major drivers of human disease. Annu Rev Genomics Hum Genet. 2017;18:485–63.
116. Harris DT, Wang N, Riley TP, Anderson SD, Singh NK, Procko E, et al. Deep mutational scans as a guide to engineering high affinity T cell receptor interactions with peptide-bound major histocompatibility complex. J Biol Chem. 2016;291:24566–78.

117. Kowalsky CA, Whitehead TA. Determination of binding affinity upon mutation for type I dockerin–cohesin complexes from Clostridium thermocellum and Clostridium cellulolyticum using deep sequencing. Proteins Struct Funct Bioinforma. 2016;84:1914–28.

118. Ernest J, Melnikov A, Zhang X, Wang L, Rogov P, Mikkelsen TS, et al. Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. Nat Biotechnol. 2016;34:1180–90.

119. Brenan L, Andreev A, Cohen O, Pantel S, Kamburov A, Cacchiarelli D, et al. Phenotypic characterization of a comprehensive set of MAPK1/ERK2 missense mutants. Cell Rep. 2016;17:1171–83.

120. Qiu C, Ernine OC, Dave JM, Cui P, Jin H, Muthukrishnan N, et al. High-resolution phenotypic landscape of the RNA polymerase II trigger loop. PLOS Genet. 2016;12:e1006321.

121. Haddox HK, Dingens AS, Bloom JD. Experimental estimation of the effects of all amino-acid mutations to HIV’s envelope protein on viral replication in cell culture. PLOS Pathog. 2016;12:e1006114.

122. van Rosmak M, Janssen BMG, Hendrikse NM, van der Linden AJ, Pieters PA, Wanders D, et al. Affinity maturation of a cyclic peptide handle for therapeutic antibodies using deep mutational scanning. J Biol Chem. 2017;292:1479–89.

123. Fernandes D, Faust TB, Struhl NB, Smith C, Crosby DC, Nakamura RL, et al. Functional segregation of overlapping genes in HIV. Cell. 2016;167:1762–73 e12.

124. Keskin A, Akdoğan E, Dunn CD. Evidence for amino acid snorkeling from a high-resolution, in vivo analysis of Fis1 tall-anchor insertion at the mitochondrial outer membrane. Genetics. 2017;205:691–705.

125. Selcuk I, Groves B, Kuchina A, Rosenberg AB, Jojic N, Fields S, et al. Deep learning of the regulatory grammar of yeast 5′ untranslated regions. Genome Res. 2017;27:2015–24.

126. Spencer JM, Zhang X. Deep mutational scanning of S. pyogenes Cas9 reveals important functional domains. Sci Rep. 2017;7:16838.

127. Rabani M, Pieper L, Chew G-L, Schier AF. A massively parallel reporter assay of 3′ UTR sequences identifies in vivo rules for mRNA degradation. Mol Cell. 2017;68:1083–94 e5.

128. Ke S, Anquetil V, Zamalloa JR, Maity A, Yang A, Aitas MA, et al. Saturation mutagenesis reveals manifold determinants of exon definition. Genome Res. 2018;28:11–24.

129. Weile J, Sun S, Cote AG, Knapp J, Verby M, Mellor JC, et al. A framework for exhaustively mapping functional missense variants. Mol Syst Biol. 2017;13:3957.

130. Pleca C, Sidore AM, Lubock NB, Zhang D, Kosuri S. Multiplexed gene synthesis in emulsions for exploring protein functional landscapes. Science. 2018;360:943–7.

131. Vainberg Sluskin I, Weingarten-Gabbay S, Nir R, Weinberger A, Segal E. Unraveling the determinants of microRNA-mediated regulation using a massively parallel reporter assay. Nat Commun. 2018;9:529.

132. Staller MV, Holehouse AS, Swain-Lenz D, As KS, Pappu RV, Cohen BA. A high-throughput mutational scan of an intrinsically disordered acidic transcriptional activation domain. Cell Syst. 2018;6:444–55 e6.

133. Otsa G, Lehner B. The genetic landscape of a physical interaction. eLife. 2018;7:e32472.

134. Hartman EC, Jakobson CM, Favor AH, Lobba MJ, Álvarez-Benedicto E, Francis MB, et al. Quantitative characterization of all single amino acid variants of a viral capsid-based drug delivery vehicle. Nat Commun. 2018;9:1385.

135. Doud MB, Lee JM, Bloom JD. How single mutations affect viral escape from broad and narrow antibodies to H1 influenza hemagglutinin. Nat Commun. 2018;9:1386.

136. Mason DM, Weber CR, Parada C, Meng SM, Greiff V, Kelton WJ, et al. High-throughput antibody engineering in mammalian cells by CRISPR/Cas9-mediated homology-directed mutagenesis. Nucleic Acids Res. 2018;46:743–69.

137. Kotler E, Shani O, Goldfeld G, Lotan-Pompan M, Taricó O, Gershoni A, et al. A systematic p53 mutation library links differential functional impact to cancer mutation pattern and evolutionary conservation. Mol Cell. 2018;71:178–90 e8.

138. Salinas VH, Ranganathan R. Coevolution-based inference of amino acid interactions underlying protein function. eLife. 2018;7:e34300.

139. Mavor D, Barlow KA, Asarow D, Birman Y, Britain D, Chen W, et al. Extending chemical perturbations of the ubiquitin fitness landscape in a classroom setting reveals new constraints on sequence tolerance. Biol Open. 2018;7:bio036103.

140. Dorrity MW, Cuperus JT, Carlisle JA, Fields S, Quetsch C. Preferences in a trait decision determined by transcription factor variants. Proc Natl Acad Sci. 2018;115:E7997–8006.

141. Castaldi PJ, Guo F, Qiao D, Fu D, Naing ZZC, Li Y, et al. Identification of functional variants in the FAM13A chronic obstructive pulmonary disease

142. Ashenberg O, Padmakumar J, Doud MB, Bloom JD. Deep mutational scanning identifies sites in influenza nucleoprotein that affect viral inhibition by MXA. PLoS Pathog. 2017;13:e1006288.

143. Wrenbeck EE, Azouz LR, Whitehead TA. Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. Nat Commun. 2017;8:15669.

144. Wu NC, Xie J, Zheng T, Nycholat CM, Grande G, Paulson JC, et al. Diversity of functionally permissive sequences in the receptor-binding site of influenza hemagglutinin. Cell Host Microbe. 2017;21:742–53 e8.

145. Bhagavatula G, Rich MS, Young DL, Marin M, Fields S. A massively parallel fluorescence assay to characterize the effects of synonymous mutations on TP53 expression. Mol Cancer Res. 2017;15:301–7.

146. Bandaru P, Shah NH, Bhattacharyya M, Barton JP, Kondo Y, Cofsky JC, et al. Reconstruction of the Ras switching cycle through saturation mutagenesis. eLife. 2017;6:e27810.

147. Starr TN, Picton LK, Thornton JW. Alternative evolutionary histories in the sequence space of an ancient protein. Nature. 2017;549:409–13.

148. Woodsmith J, Apelt L, Casado-Medrano V, Oskan Z, Timmermann B, Steul B. Protein interaction perturbation profiling at amino-acid resolution. Nat Methods. 2017;14:213–23.

149. Staller MV, Holehouse AS, Swain-Lenz D, Das KS, Pappu RV, Cohen BA. A high-throughput mutational scan of an intrinsically disordered acidic transcriptional activation domain. Cell Syst. 2018;6:444–55 e6.

150. Ashenberg O, Padmakumar J, Doud MB, Bloom JD. Deep mutational scanning identifies sites in influenza nucleoprotein that affect viral inhibition by MXA. PLoS Pathog. 2017;13:e1006288.
218. Esposito D, Weile J, Shendure J, Starita LM, Papenfuss AT, Roth FP, Rubin AF. MaveReferences. GitHub. (2019). https://github.com/VariantEffect/MaveReferences
219. Esposito D, Weile J, Shendure J, Starita LM, Papenfuss AT, Roth FP, Rubin AF. MaveReferences. Zenodo. (2019). https://zenodo.org/record/3363411

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.