MPRAbase A Massively Parallel Reporter Assay database

Jingjing Zhao[1,2,*], Fotis A. Baltoumas[3,*], Maxwell A. Konnaris[4,5], Ioannis Mouratidis[4], Zhe Liu[1,2,6], Jasmine Sims[1,2], Vikram Agarwal[7], Georgios A. Pavlopoulos[3], Ilias Georgakopoulos-Soares[4,†], Nadav Ahituv[1,2,†]

[1]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, California, USA.
[2]Institute for Human Genetics, University of California San Francisco, San Francisco, California, USA.
[3]Institute for Fundamental Biomedical Research, BSRC "Alexander Fleming", Vari, 16672, Greece.
[4]Institute for Personalized Medicine, Department of Biochemistry and Molecular Biology, The Pennsylvania State University College of Medicine, Hershey, PA, USA.
[5]Department of Statistics, Penn State University, State College, PA, USA.
[6]Department of Computer Science, City University of Hong Kong, Hong Kong, China.
[7]mRNA Center of Excellence, Sanofi Pasteur Inc., Waltham, MA, USA.

[†]Corresponding authors: izg5139@psu.edu, nadav.ahituv@ucsf.edu

Abstract

Massively parallel reporter assays (MPRAs) represent a set of high-throughput technologies that measure the functional effects of thousands of sequences/variants on gene regulatory activity. There are several different variations of MPRA technology and they are used for numerous applications, including regulatory element discovery, variant effect measurement, saturation mutagenesis, synthetic regulatory element generation or characterization of evolutionary gene regulatory differences. Despite their many designs and uses, there is no comprehensive database that incorporates the results of these experiments. To address this, we developed MPRAbase, a manually curated database that currently harbors 130 experiments, encompassing 17,718,677 elements tested across 35 cell types and 4 organisms. The MPRAbase web interface serves as a centralized user-friendly repository to examine online the activity of regulatory elements across cell types and organisms, and to download MPRA data for independent analysis.

Introduction

Since the initial sequencing of the human genome, millions of *cis*-regulatory elements with putative roles in transcriptional gene regulation have been identified (Encode Project Consortium 2012; Thurman et al. 2012). Following up on their annotation, a major challenge has been to functionally characterize these elements. Massively parallel reporter assays (MPRAs) were built on the framework of the classic reporter assay. In this framework, the assayed sequence is placed in front of a reporter gene for promoter assays and also a minimal promoter for enhancer assays (**Fig. 1**). If the sequence itself has regulatory activity, it will turn on the reporter gene. To overcome the one-by-one testing limitation of these classic reporter assays, MPRAs add a DNA barcode that is transcribed if the sequence has regulatory activity and can be measured via RNA-sequencing (RNA-seq), providing a way to examine the functional effects of thousands of sequences in parallel (Patwardhan et al. 2009; Inoue and Ahituv 2015; Patwardhan et al. 2012; Melnikov et al. 2012; Agarwal et al. 2025) (**Fig. 1**). In recent years, the rapidly declining cost of DNA synthesis and sequencing have led to the growing popularity in the use of MPRA experiments and rapid accumulation of MPRA data.

Since their invention more than a decade ago (Patwardhan et al. 2009), there has been a rapid emergence of novel variations of MPRA technology (**Fig. 1**) that differ by: 1) The positioning of the tested element and barcode relative to the reporter. For example, STARR-seq tests an element within a reporter's 3′ UTR (Arnold et al. 2013); 2) Library generation. Libraries can be generated by the synthesis of pre-defined oligonucleotide sequences (Smith et al. 2013; Patwardhan et al. 2009, 2012; Melnikov et al. 2012; Agarwal et al. 2025), input of natural sequences from whole-genome fragmentation (van Arensbergen et al. 2016, 2019; Liu et al. 2017; Kvon et al. 2014; Arnold et al. 2013), isolation of nucleosome-free regions (Murtha et al. 2014), or through the use of ATAC-seq (Wang et al. 2018); 3) The method of library delivery. MPRA libraries have been delivered to cells by transfection (Patwardhan et al. 2012; Melnikov et al. 2012; Kircher et al. 2019; Johnson et al. 2018; Liu et al. 2017), adeno-associated virus (AAV)-based MPRAs (Shen et al. 2016; Lambert et al. 2021; Chan et al. 2023) and lentivirus-based MPRAs (Inoue et al. 2017) that allow the integration of elements into the genome; 4) Computational processing tools, whereby the collected data are processed into activity scores, with appropriate processing pipelines chosen according to experimental design features (Lee et al. 2020; Gordon et al. 2020; Ashuach et al. 2019; Kim et al. 2021; Georgakopoulos-Soares et al. 2017). Experiments directly evaluating the aforementioned MPRA design choices have revealed a general consistency in measured element activities (Inoue et al. 2017; Klein et al. 2020).

Despite the exponential growth of published MPRA datasets, to date, there is no centralized repository that aggregates the results of such data. To address this shortcoming, we introduce MPRAbase (https://mprabase.ucsf.edu), a database that harbors 130 experiments, encompassing 17,718,677 sequences tested across 35 cell types and 4 organisms. In addition to storing published

data, MPRAbase provides a platform for users to deposit new MPRA data and rapidly disseminate it to the functional genomics community.

MPRAbase has processed high throughput experiments across 51 studies. For each study, we provide the PMID and a link to the original publication. We also provide the mean expression score of each sequence, along with the expression score of each sequence for every replicate in the same format. MPRAbase offers an advanced search option, in which the user can search based on the coordinates of interest, the technique used (MPRA/STARR-seq and their variations), organism, cell type, or motif. In addition, we provide an integration with the UCSC Genome Browser where available. MPRAbase also includes a filter for MPRAs carried out on synthetic sequences that do not exist in a specific genome. Another option includes searching MPRA datasets based on the article title, PubMed ID, or Gene Expression Omnibus (GEO) number. Finally, MPRAbase includes a separate tab for saturation mutagenesis MPRA that allows the selection of different regulatory elements, variants, and variant scores.

Results

Database overview

Our goal with MPRAbase was to provide a central repository for all published MPRA and STARR-seq experiments. MPRAbase provides a collection of MPRA and STARR-seq data from different experiments, organisms, and assays and presents them with a user-friendly web-based interface that allows users to easily download the data. The data provided in the database include the activity of sequences, measured as log2 RNA/DNA ratio, and is provided separately for experimental replicates and associated correlation plots, metadata, and statistics. MPRA experiments are divided into three types; i) standard MPRA, ii) synthetic MPRA, and iii) saturation MPRA experiments. Standard MPRAs are further subdivided into plasmid-based MPRAs, lentivirus-based MPRAs, and STARR-seq.

Collection of MPRA studies

For the development of our database, we scanned the literature using keywords and terms associated with MPRA experiments, resulting in the collection of 130 experiments. Studies were organized based on the organism of the assayed sequence, cell origin and type, and experimental assay. MPRA experiments across 4 organisms, 35 cell types/tissues, and 8 MPRA library techniques were downloaded, analyzed, and presented in MPRAbase (**Fig. 2; Supplementary Figure S1; Supplementary Table S1**). The size of the MPRA libraries varies between 98 and 16,092,560 sequences, with 8,384 being the median number of sequences per experiment tested. The total number of DNA sequences available across all the MPRA experiments in MPRAbase is currently 17,718,677.

Processing of MPRA data and expression quantification

In MPRA experiments, expression levels are generally quantified as the logarithmic RNA to DNA counts ratio with higher log-ratio reflecting increased *cis*-regulatory activity (Gordon et al. 2020). Data associated with each MPRA experiment across the collected studies were assembled and processed to provide the logarithmic ratio of RNA/DNA counts for each biological replicate and the mean log2 RNA/DNA expression levels across replicates.

MPRAbase also provides quality controls, including graphs for the correlation between biological replicates. The first type of plot compares the log2 RNA ratio between replicates, the second the logs DNA ratio between replicates, and the third the log2 RNA/DNA ratio correlations between biological replicates.

Library categories include plasmid-based MPRA experiments, lentivirus-based MPRA experiments (Gordon et al. 2020), and STARR-seq experiments (Muerdter et al. 2015). For lentiMPRA and plasmid-based MPRA experiments, the log2 RNA/DNA ratio is provided for the elements of the library design, whereas for STARR-seq experiments genome-wide log2

RNA/DNA ratios are quantified across the coordinates of the fragments derived from the experimental output data. The activity score for each element has been calculated as the log2 of the normalized count of RNA molecules from all barcodes corresponding to the element, divided by the normalized number of DNA molecules from all barcodes corresponding to the element. All coordinates provided in MPRAbase are 1-based on the website.

MPRAbase website and web-interface

The MPRAbase website contains interactive pie charts, tables, and drop-down menus that enable the selection of MPRA experiments based on the organism, cell type, and library strategy used (**Fig. 3A-E**). The user can select multiple combinations of the aforementioned groups, for which an interactive table is presented with the individual samples. MPRAbase also provides a search bar to search regions by chromosome coordinates (**Fig. 3F**). Therefore, MPRAbase provides an additional functionality, in which the user can examine *cis*-regulatory activity for particular loci of interest. A set of coordinates for a reference genome of interest can be inserted by the user for which all MPRA sequences across experiments and cell types will be returned. MPRAbase also allows search based on article title, PubMed ID, research group, or GEO number.

For each sample in the interactive table, the sample ID, organism, cell type, library strategy, Gene Expression Omnibus (GEO) number, and PubMed ID (PMID) of the experiment are provided (**Figure 3D-F**). The identifiers associated with GEO and PMID entries have embedded clickable hyperlinks that can take the user to the associated studies and raw sequencing experiment databases. The selected data can be downloaded for further processing by the user (**Figure 3G**). The format of the download is a zipped folder containing the sequence log2 RNA/DNA tables for the studies selected and a metadata file, which provides additional information for the selected studies. A selected study can be further examined at MPRAbase, including examination of the *cis*-regulatory activity of the elements and visualization of the correlation between replicates (**Figure 3H-I**).

MPRA is also carried out using synthetic sequences that may not exist in a certain genome. This is usually carried out to better understand the regulatory code, and to design regulatory elements that can drive tissue/cell type-specific expression, or sequences that respond to certain factors. To portray these in MPRAbase, we provide a specific tab called “Synthetic MPRA”. As synthetic sequences do not have genomic coordinates, we only provide the ability to download these datasets.

Saturation Mutagenesis

MPRAbase contains a separate tab for saturation mutagenesis MPRA (**Fig. 4**). In these MPRA experiments, a specific sequence is mutated, and the effect of these numerous mutations is tested in parallel using MPRA. As it measures variant effects across the same sequence, we provide the ability to select the different promoters/enhancers that were tested using this approach by the name

they were given in the experiment (**Fig. 4A-B**). Once the promoter/enhancer is selected, MPRAbase allows to select specific coordinates in the regulatory element for the variant scores, the number of unique tags, the log2 variant expression, and the p-value (**Fig. 4C-D**). Data is provided both as a table and also as a 'lollipop figure' that shows the different mutations and their effects (**Fig. 4C-D**).

Documentation and Help pages.

The website has a page in the 'About' tab which provides information about MPRA experiments to introduce potential users to the MPRA technology and its utility. In addition, a 'Help' tab is also provided, which provides explanations for the different functionalities of the database. This tab also provides information on how to submit MPRA datasets to the website along with a downloadable submission form.

Discussion

The recent advances in DNA synthesis and sequencing costs have enabled the widespread adoption of MPRA technologies. MPRAs and other related assays can provide insights into the roles of disease-associated variants, can be used to gain insights into *cis*-regulation, and have been implemented in studies of primate evolution, while they can also be used to examine synthetic sequences, with potential applications as therapeutic molecules (Whalen et al. 2023; Georgakopoulos-Soares et al. 2023, 2022; Arnold et al. 2013; Agarwal et al. 2025; Deng et al. 2024). Here, we generated MPRAbase that allows the user to view and analyze MPRA datasets in one location.

MPRAbase provides a curated database for MPRAs, consisting, at launch, of 130 experiments, for 36 cell types, across 4 organisms. The website is user-friendly and interactive, enabling users to select studies based on a list of criteria, including organism, cell type, and MPRA library type. For each study, we provide quality control metrics, enabling users to decide if the selected studies meet their quality requirements. We plan to have MPRAbase updated regularly to accommodate the increasing number of available MPRA experiments. With a continuously updated, comprehensive characterization of MPRA experiments across organisms and cell types, we believe MPRAbase will be a valuable resource to better understand gene regulatory grammar, illuminate the consequences of non-coding mutations, and be used to gain insights into evolutionary facets. We therefore anticipate this resource will have a broader impact on our broader understanding of genetics.

Methods

MPRA experiments in database. Publications with MPRA, STARR-seq, or other related assays were systematically collected. MPRA data were retrieved and manually curated. Curation included the collection of the PMID, organism name, cell type, experiment type, and library strategy used. Whenever available the GEO ID was also integrated. Correlation analyses were performed between replicates for each study for log2 RNA/DNA ratios. For human MPRAs, MPRA coordinates in hg18 (NCBI36) or hg19 (GRCh37) were revised to hg38 (GRCh38) using the UCSC liftOver tool (Kuhn et al. 2007); in mice, mm9 coordinates were converted to mm10.

Database implementation. MPRAbase contents are organized in a relational SQLite database (https://www.sqlite.org/). The user interface was implemented using R (https://www.r-project.org/) and the R/Shiny framework. Server-side operations are mainly handled by R. Data visualization and graphs are generated using the R/DT and R/plotly packages (Sievert 2020). MPRAbase is available online without fees for academic usage. The database is updated in regular 6-month intervals, as new MPRA studies become available, and have provided contact details in the Documentation for users who will be interested in having their datasets integrated.

Software and Data Availability

The code for MPRAbase can be found at GitHub (https://urldefense.com/v3/__https://github.com/Ahituv-lab/mprabase__;!!LQC6Cpwp!sEFajsZ1maorRtVEhT1eMfSkkuyc7Vqe0c_J_jz0rT95ctvuKwxgtfQsfisMW3gILnf5GFLavUU-HJWp--w$) and source code is also available as Supplemental Code. A stable version of the database has also been deposited in Zenodo (https://urldefense.com/v3/__https://doi.org/10.5281/zenodo.10920746__;!!LQC6Cpwp!sEFajsZ1maorRtVEhT1eMfSkkuyc7Vqe0c_J_jz0rT95ctvuKwxgtfQsfisMW3gILnf5GFLavUU-Pdiut6Q$).

Competing interest statement

V.A. is currently an employee of Sanofi Pasteur Inc., but pursued this work independently of the organization. N.A. is a co-founder and on the scientific advisory board of Regel Therapeutics and receives funding from BioMarin Pharmaceutical Incorporate.

Acknowledgments

Funding: National Human Genome Research Institute UM1HG009408 and UM1HG011966 (N.A.); National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM155468. (I.G.S.); Fondation Sante; Onasis Foundation; Hellenic Foundation for Research and Innovation (H.F.R.I) under the call 'Greece 2.0 - Basic Research Financing Action, sub-action II, Grant ID: 16718-PRPFOR; Program 'Greece 2.0, National Recovery and Resilience Plan', Grant ID: TAEDR-0539180.

Author Contributions: J.Z., V.A., I.G.S., and N.A. conceived of the study. J.Z.,. F.A.B.,. M.A.K,. V.A., I.M., Z.L., J.S., and I.G.S performed the computational analyses. J.Z.,. F.A.B.,. M.A.K,. and

I.G.S. generated the visualizations. G.A.P., I.G.S., and N.A. provided resources. I.G.S., and N.A. wrote the manuscript with input from all authors. V.A., G.A.P., I.G.S., and N.A. supervised the study.

Figure 1. Schematic illustration of MPRA experiments. Various MPRA uses are described in the top left. MPRA libraries are usually prepared via oligonucleotide synthesis, genome fragmentation or isolation of nucleosome-free or ATAC-seq regions. These sequences are then cloned into an MPRA vector along with a unique barcode, while in STARR-seq, the assayed sequence is used as the barcode. MPRAs are then delivered to the cells either via transfection or viral infection. DNA and RNA are then extracted and the barcodes are sequences and activity scores are provided.

Figure 2. Visualization of MPRAbase summary statistics. Pie chart displaying the breakdown of MPRAbase by different species, MPRA type, and cell type. The counts of experiments are listed in parentheses.

Figure 3. MPRA browser design. *(A)* Selection based on genome assembly or synthetic sequences with the option to search for specific loci of interest. ***(****B)* Selection based on species of interest or library type for different types of MPRAs to display by using a clickable box option. *(C)* Selection of cell types of interest to display by using a clickable box option. *(D)* Pie-chart displaying MPRA dataset distribution by species. ***(****E)* Pie chart displaying MPRA category breakdown to select from. ***(****F)* Table displaying the set of MPRA experiments in MPRAbase, with associated metadata including species, cell type, the library design used, the Gene Expression Omnibus (GEO) number of the experiment when available, and the PubMed ID of the relevant publication. ***(****G)* Datasets selected can be visualized or the associated tables can be downloaded for further research. *(H)* Scatter plots displaying pairwise correlations between replicates for log2 RNA/DNA ratios. The line of best fit is displayed in yellow. Both Pearson and Spearman correlations are shown. *(I)* Table displaying *cis*-regulatory element activity for each tile, for the selected experiment. *Cis*-regulatory activity is displayed separately for each replicate in each column and a column of the combined activity is also displayed.

Figure 4. Saturation mutagenesis MPRAs on MPRAbase. *(A)* Drag-bar for selecting variants based on their position or number of unique barcodes associated with a variant. ***(****B)* Drag-bar for selecting based on variant expression effects. *(C)* Table displaying individual mutations for the selected locus and their effect on expression. ***(****D)* Lollipop graph displaying individual mutations for the selected locus and their effect on expression. Mutations are colored by Reference and Alteration type.

References

Agarwal V, Inoue F, Schubach M, Penzar D, Martin BK, Dash PM, Keukeleire P, Zhang Z, Sohota A, Zhao J, et al. 2025. Massively parallel characterization of transcriptional regulatory elements. *Nature* **639**: 411–420.

Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. 2013. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**: 1074–1077.

Ashuach T, Fischer DS, Kreimer A, Ahituv N, Theis FJ, Yosef N. 2019. MPRAnalyze: statistical framework for massively parallel reporter assays. *Genome Biol* **20**: 183.

Chan Y-C, Kienle E, Oti M, Di Liddo A, Mendez-Lago M, Aschauer DF, Peter M, Pagani M, Arnold C, Vonderheit A, et al. 2023. An unbiased AAV-STARR-seq screen revealing the enhancer activity map of genomic regions in the mouse brain in vivo. *Sci Rep* **13**: 1–19.

Deng C, Whalen S, Steyert M, Ziffra R, Przytycki PF, Inoue F, Pereira DA, Capauto D, Norton S, Vaccarino FM, et al. 2024. Massively parallel characterization of regulatory elements

in the developing human cortex. *Science* **384**. https://pubmed.ncbi.nlm.nih.gov/38781390/ (Accessed July 17, 2024).

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

Georgakopoulos-Soares I, Deng C, Agarwal V, Chan CSY, Zhao J, Inoue F, Ahituv N. 2023. Transcription factor binding site orientation and order are major drivers of gene regulatory activity. *Nat Commun* **14**: 2333.

Georgakopoulos-Soares I, Jain N, Gray JM, Hemberg M. 2017. MPRAnator: a web-based tool for the design of massively parallel reporter assay experiments. *Bioinformatics* **33**: 137–138.

Georgakopoulos-Soares I, Victorino J, Parada GE, Agarwal V, Zhao J, Wong HY, Umar MI, Elor O, Muhwezi A, An J-Y, et al. 2022. High-throughput characterization of the role of non-B DNA motifs on promoter function. *Cell Genom* **2**. http://dx.doi.org/10.1016/j.xgen.2022.100111.

Gordon MG, Inoue F, Martin B, Schubach M, Agarwal V, Whalen S, Feng S, Zhao J, Ashuach T, Ziffra R, et al. 2020. lentiMPRA and MPRAflow for high-throughput functional characterization of gene regulatory elements. *Nat Protoc* **15**: 2387–2412.

Inoue F, Ahituv N. 2015. Decoding enhancers using massively parallel reporter assays. *Genomics* **106**: 159–164.

Inoue F, Kircher M, Martin B, Cooper GM, Witten DM, McManus MT, Ahituv N, Shendure J. 2017. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res* **27**: 38–52.

Johnson GD, Barrera A, McDowell IC, D'Ippolito AM, Majoros WH, Vockley CM, Wang X, Allen AS, Reddy TE. 2018. Human genome-wide measurement of drug-responsive regulatory activity. *Nat Commun* **9**: 5317.

Kim Y-S, Johnson GD, Seo J, Barrera A, Cowart TN, Majoros WH, Ochoa A, Allen AS, Reddy TE. 2021. Correcting signal biases and detecting regulatory elements in STARR-seq data. *Genome Res* **31**: 877–889.

Kircher M, Xiong C, Martin B, Schubach M, Inoue F, Bell RJA, Costello JF, Shendure J, Ahituv N. 2019. Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. *Nat Commun* **10**: 3583. doi: 10.1038/s41467-019-11526-w.

Klein JC, Agarwal V, Inoue F, Keith A, Martin B, Kircher M, Ahituv N, Shendure J. 2020. A systematic evaluation of the design and context dependencies of massively parallel reporter assays. *Nat Methods* **17**: 1083–1091.

Kuhn RM, Karolchik D, Zweig AS, Trumbower H, Thomas DJ, Thakkapallayil A, Sugnet CW, Stanke M, Smith KE, Siepel A, et al. 2007. The UCSC genome browser database: update 2007. *Nucleic Acids Res* **35**: D668-73.

Kvon EZ, Kazmar T, Stampfel G, Yáñez-Cuna JO, Pagani M, Schernhuber K, Dickson BJ, Stark A. 2014. Genome-scale functional characterization of Drosophila developmental

enhancers in vivo. *Nature* **512**: 91–95.

Lambert JT, Su-Feher L, Cichewicz K, Warren TL, Zdilar I, Wang Y, Lim KJ, Haigh JL, Morse SJ, Canales CP, et al. 2021. Parallel functional testing identifies enhancers active in early postnatal mouse brain. *Elife* **10**. http://dx.doi.org/10.7554/eLife.69479.

Lee D, Shi M, Moran J, Wall M, Zhang J, Liu J, Fitzgerald D, Kyono Y, Ma L, White KP, et al. 2020. STARRPeaker: uniform processing and accurate identification of STARR-seq active regions. *Genome Biol* **21**: 298.

Liu Y, Yu S, Dhiman VK, Brunetti T, Eckart H, White KP. 2017. Functional assessment of human enhancer activities using whole-genome STARR-sequencing. *Genome Biol* **18**: 219.

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

Muerdter F, Boryń ŁM, Arnold CD. 2015. STARR-seq — Principles and applications. *Genomics* **106**: 145–150. http://dx.doi.org/10.1016/j.ygeno.2015.06.001.

Murtha M, Tokcaer-Keskin Z, Tang Z, Strino F, Chen X, Wang Y, Xi X, Basilico C, Brown S, Bonneau R, et al. 2014. FIREWACh: high-throughput functional detection of transcriptional regulatory modules in mammalian cells. *Nat Methods* **11**: 559–565.

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

Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D, Shendure J. 2009. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat Biotechnol* **27**: 1173–1175.

Shen SQ, Myers CA, Hughes AEO, Byrne LC, Flannery JG, Corbo JC. 2016. Massively parallel cis-regulatory analysis in the mammalian central nervous system. *Genome Res* **26**: 238–255.

Sievert C. 2020. *Interactive Web-based Data Visualization with R, Plotly, and Shiny*. Chapman & Hall/CRC The R Series.

Smith RP, Taher L, Patwardhan RP, Kim MJ, Inoue F, Shendure J, Ovcharenko I, Ahituv N. 2013. Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nat Genet* **45**: 1021–1028.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al. 2012. The accessible chromatin landscape of the human genome. *Nature* **489**: 75–82.

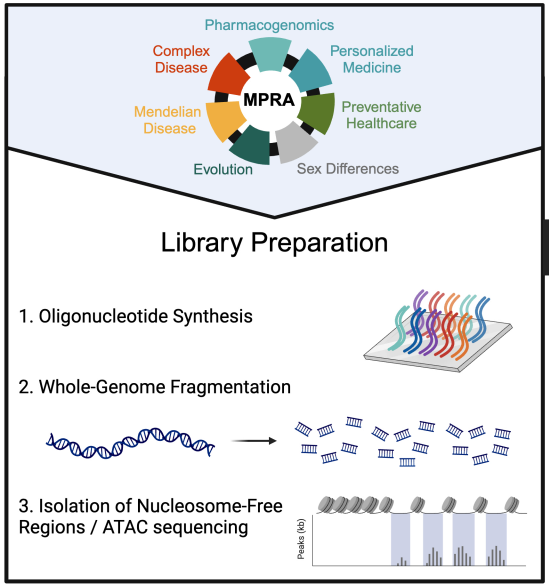
van Arensbergen J, FitzPatrick VD, de Haas M, Pagie L, Sluimer J, Bussemaker HJ, van Steensel B. 2016. Genome-wide mapping of autonomous promoter activity in human cells. *Nat Biotechnol*.

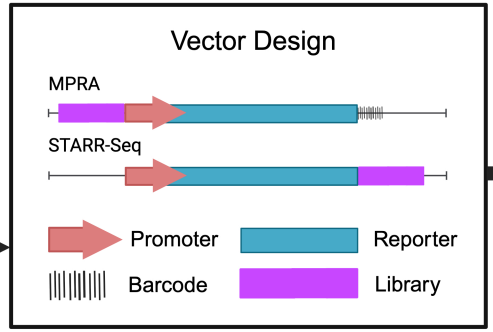
van Arensbergen J, Pagie L, FitzPatrick VD, de Haas M, Baltissen MP, Comoglio F, van der

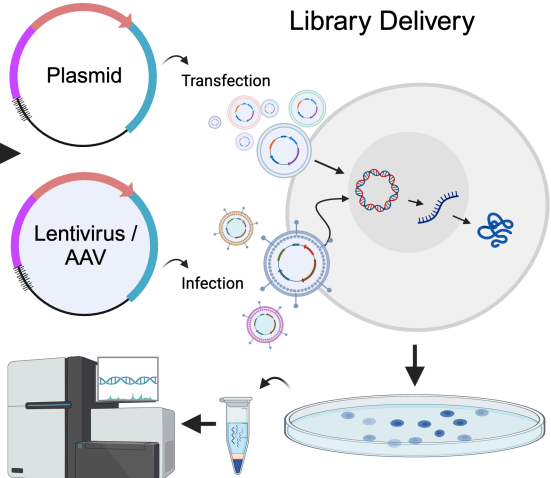
Weide RH, Teunissen H, Võsa U, Franke L, et al. 2019. High-throughput identification of human SNPs affecting regulatory element activity. *Nat Genet* **51**: 1160–1169.

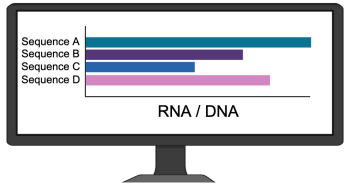
Wang X, He L, Goggin SM, Saadat A, Wang L, Sinnott-Armstrong N, Claussnitzer M, Kellis M. 2018. High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. *Nat Commun* **9**: 5380.

Whalen S, Inoue F, Ryu H, Fair T, Markenscoff-Papadimitriou E, Keough K, Kircher M, Martin B, Alvarado B, Elor O, et al. 2023. Machine learning dissection of human accelerated regions in primate neurodevelopment. *Neuron* **111**: 857-873.e8.





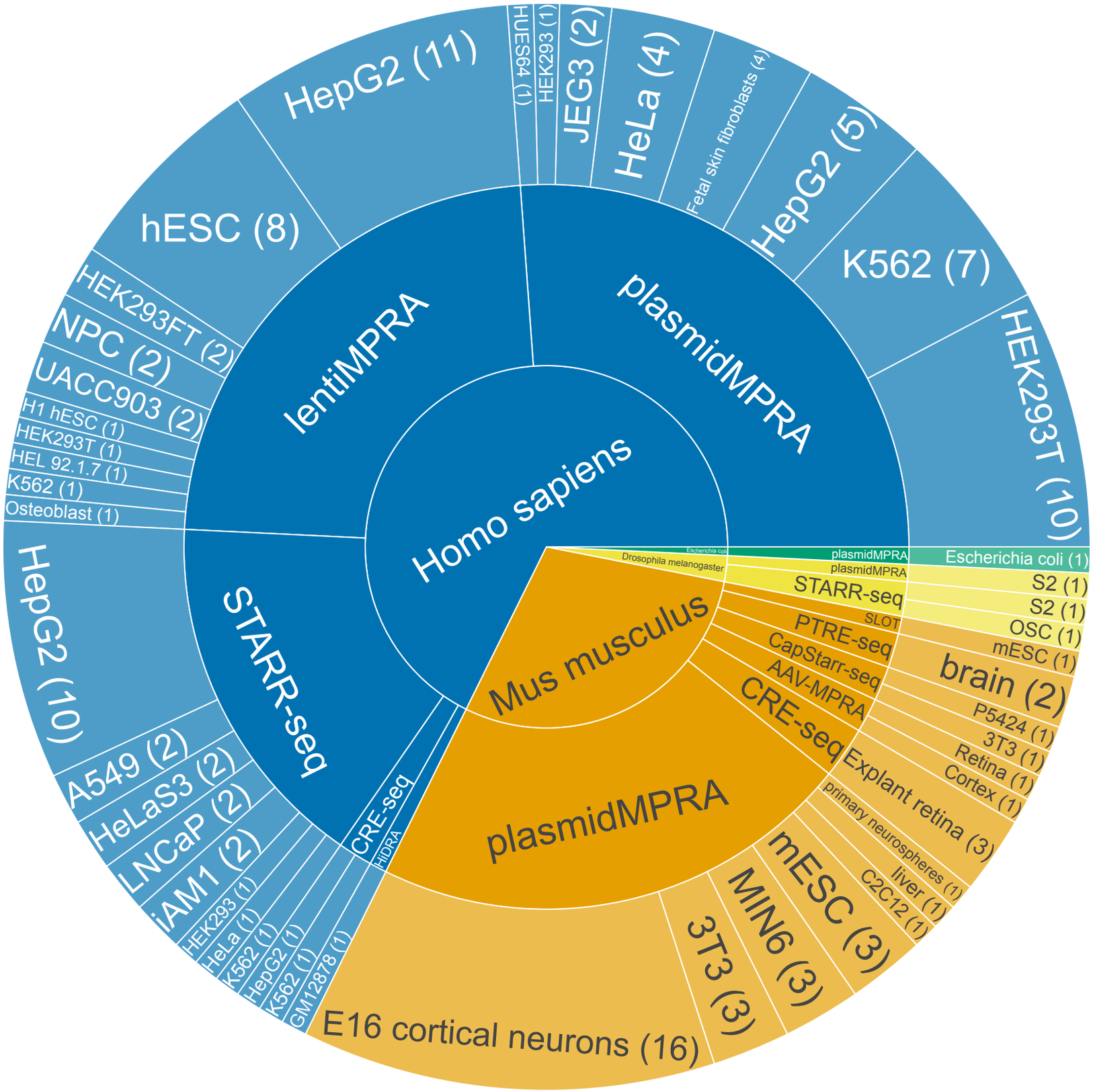




Computational Processing

Sequencing

DNA / RNA Extraction



A

Library Cis-regulatory Elements

1. Select library genome assembly:

✔ All | ✖ Human hg38 (GRCh38) | ✖ Human hg19 (GRCh37) | ✖ Human hg18 (NCBI36) | ✖ Mouse mm10 (GRCm38) | ✖ Mouse mm9 (MGSCv37) | ✖ D. melanogaster dm3 (BDGP R5)

2. Filter by cis-regulatory elements:

Start typing and select from list or enter a custom value...

Browse standard MPRA samples

B

Species & Library Strategies

Filter by species:

- ☐ Homo sapiens (67)
- ☐ Drosophila melanogaster (3)
- ☐ Mus musculus (34)
- ☐ Escherichia coli (1)

Filter by library strategy:

- ☐ lentiMPRA (19)
- ☐ STARR-seq (17)
- ☐ HiDRA (1)
- ☐ plasmidMPRA (56)
- ☐ CRE-seq (5)
- ☐ PTRE-seq (2)
- ☐ SLOT (1)
- ☐ CapStarr-seq (2)
- ☐ AAV-MPRA (2)

C

Cell Type

Filter by cell type:

- ☐ HepG2 (10)
- ☐ S2 (2)
- ☐ OSC (1)
- ☐ Hela (5)
- ☐ GM12878 (1)
- ☐ HEK293FT (2)
- ☐ UACC903 (2)
- ☐ E16 cortical neurons (16)
- ☐ 3T3 (4)
- ☐ H1 hESC (1)
- ☐ HEK293T (10)
- ☐ Escherichia coli (1)
- ☐ Explant retina (3)
- ☐ brain (2)
- ☐ Fetal skin fibroblasts (4)
- ☐ K562 (8)
- ☐ A549 (2)
- ☐ C2C12 (1)
- ☐ mESC (1)
- ☐ JEG3 (2)
- ☐ P5424 (1)
- ☐ HEL 92.1.7 (1)
- ☐ HEK293 (2)
- ☐ LNCaP (2)
- ☐ HeLaS3 (2)
- ☐ iAM1 (2)
- ☐ Cortex (1)
- ☐ Retina (1)
- ☐ primary neurospheres (1)
- ☐ NPC (2)
- ☐ hESC (8)
- ☐ Osteoblast (1)
- ☐ MIN6 (3)

D

Species / Cell Lines

E

Library Strategy

F

Show 10 entries | Search:

View	Sample ID	Sample Name	Species	Cell Type	Library Strategy	GEO Number	PMID
☐	DS0001-SID01	Inoue_Mutant_integrase_HepG2	*Homo sapiens*	HepG2	lentiMPRA	GSE83894	27831498
☐	DS0001-SID02	Inoue_Wild-type_integrase_HepG2	*Homo sapiens*	HepG2	lentiMPRA	GSE83894	27831498
☐	DS0002-SID01	Klein_STARR_Tiling_HepG2	*Homo sapiens*	HepG2	STARR-seq	GSE113978	30045748
☐	DS0002-SID02	Klein_STARR_Orthologs_HepG2	*Homo sapiens*	HepG2	STARR-seq	GSE113978	30045748
☐	DS0004-SID01	Stark_STARR-seq_S2	*Drosophila melanogaster*	S2	STARR-seq	GSE40739	23328393
☐	DS0004-SID02	Stark_STARR-seq_OSC	*Drosophila melanogaster*	OSC	STARR-seq	GSE40739	23328393
☐	DS0004-SID03	Stark_STARRseq_HeLa_BAC	*Homo sapiens*	Hela	STARR-seq	GSE40739	23328393
☐	DS0005-SID01	Wang_HiDRA_ GM12878	*Homo sapiens*	GM12878	HiDRA	GSE104001	30568279
☐	DS0007-SID01	Choi_HEK_Lib1	*Homo sapiens*	HEK293FT	lentiMPRA	GSE129250	32483191
☐	DS0007-SID02	Choi_HEK_Lib2	*Homo sapiens*	HEK293FT	lentiMPRA	GSE129250	32483191

Showing 1 to 10 of 105 entries | Previous | 1 | 2 | 3 | 4 | 5 | ... | 11 | Next

G

Selected datasets

Download Selected | Replicate Visualization

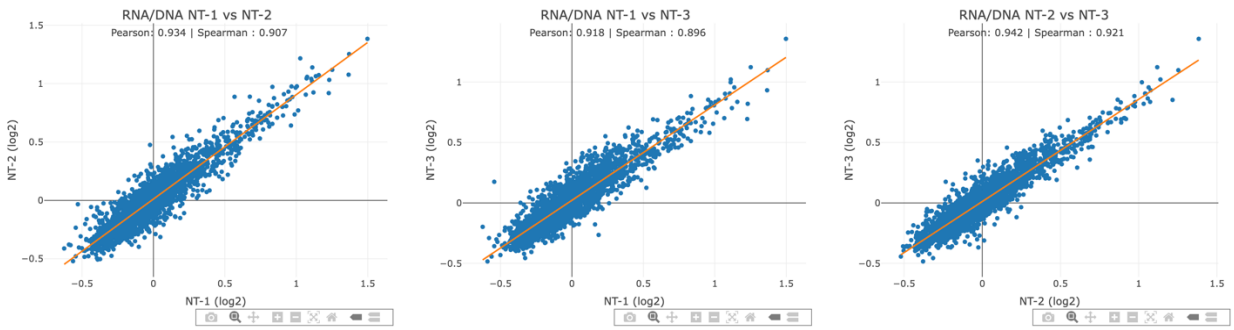
Select	Dataset ID	Dataset Name	Sample ID	PMID	GEO Number	Reference	Labs	Sample Name	Library Strategy	Species	Cell Type	Element Position	DNA/RNA replicates
☑	DS0001	Inoue_MPRA_HepG2	DS0001-SID01	27831498	GSE83894	A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity	Shendure,Ahituv	Inoue_Mutant_integrase_HepG2	lentiMPRA	*Homo sapiens*	HepG2	5'/3'	3

H

Library Viewer

RNA/DNA Correlations

RNA/DNA Replicate Distribution



I

Library Elements

Show 10 entries | Search:

Library	Library Element	Genome Assembly	Element coordinates	Score	NT-1	NT-2	NT-3		
Cloned_array_oligo_library	A:HNF4A-ChMod_chr10:11917871-11917984_[chr10:11917842-11918013]	hg19	chr10:11917871-11917984	0.985924917	0.9337856736	0.9713095383	0.9641061327	View	Open in UCSC
Cloned_array_oligo_library	A:HNF4A-ChMod_chr10:11917871-11917984_[chr10:11917842-11918013]	hg19	chr10:11917871-11917984	1.0589274795	0.9337856736	0.9713095383	0.9641061327	View	Open in UCSC
Cloned_array_oligo_library	A:HNF4A-ChMod_chr10:34165653-34165745_[chr10:34165613-34165784]	hg19	chr10:34165653-34165745	0.8999616967	0.9778621307	0.9803764001	1.1290221946	View	Open in UCSC
Cloned_array_oligo_library	A:HNF4A-ChMod_chr10:34165653-34165745_[chr10:34165613-34165784]	hg19	chr10:34165653-34165745	1.0603447826	0.9778621307	0.9803764001	1.1290221946	View	Open in UCSC
Cloned_array_oligo_library	A:HNF4A-ChMod_chr10:52009954-52010059_[chr10:52009921-52010092]	hg19	chr10:52009954-52010059	0.9077346708	0.8906822962	0.881488284	0.8690417856	View	Open in UCSC
Cloned_array_oligo_library	A:HNF4A-ChMod_chr10:52009954-52010059_[chr10:52009921-52010092]	hg19	chr10:52009954-52010059	0.9876427348	0.8906822962	0.881488284	0.8690417856	View	Open in UCSC
Cloned_array_oligo_library	A:HNF4A-ChMod_chr10:60767336-60767487_[chr10:60767326-60767497]	hg19	chr10:60767336-60767487	1.176927986	1.1143083074	1.1637132427	1.1470612608	View	Open in UCSC
Cloned_array_oligo_library	A:HNF4A-ChMod_chr10:60767336-60767487_[chr10:60767326-60767497]	hg19	chr10:60767336-60767487	1.4438039462	1.1143083074	1.1637132427	1.1470612608	View	Open in UCSC
Cloned_array_oligo_library	A:HNF4A-ChMod_chr10:60797400-60797480_[chr10:60797354-60797525]	hg19	chr10:60797400-60797480	0.7261392455	0.7421291999	0.7760066728	0.777035682	View	Open in UCSC
Cloned_array_oligo_library	A:HNF4A-ChMod_chr10:60797400-60797480_[chr10:60797354-60797525]	hg19	chr10:60797400-60797480	0.7886168916	0.7421291999	0.7760066728	0.777035682	View	Open in UCSC

MPRA Saturation – Mutagenesis Browser

A

Mutagenesis Study

Selected Study:

Homo sapiens (GRCh38/hg38)

Reference:
Kircher, M., Xiong, C., Martin, B. et al. Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. Nat Commun 10, 3583 (2019). (View on Pubmed)

Genomic Region

Selected Element:

BCL11A (chr:2)

Also include single base deletions

Genomic Position range:

60,494,940 – 60,495,539

60,494,940 60,495,060 60,495,180 60,495,299 60,495,419 60,495,539

Number of unique tags:

1 – 631,488

1 63,156 126,311 189,466 252,621 315,776 378,931 442,086 505,241 568,396 631,488

B

Variant Expression Effect

Log2 variant expression:

-1.65 – 1.49

-1.65 -1.34 -1.02 -0.71 -0.39 -0.08 0.23 0.55 0.86 1.18 1.49

P-value:

0 – 0.9984

0 0.0998 0.1997 0.2995 0.3994 0.4992 0.5991 0.6989 0.7988 0.8986 0.99845

View Data | Reset Form

C

Saturation Mutagenesis data for BCL11A

Data | Viewer

Copy | CSV | Excel | Search:

Chromosome	Position	Ref	Alt	Tags	DNA	RNA	Variant Expression Effect (log2)	P-value	Element
2	60,494,940	C	-	32	577	1345	-0.34	0.00546	BCL11A
2	60,494,940	C	A	146	2785	6772	-0.05	0.38889	BCL11A
2	60,494,940	C	G	60	975	2436	-0.13	0.13721	BCL11A
2	60,494,940	C	T	1084	8543	16057	-0.7	0	BCL11A
2	60,494,941	C	A	596	9425	23430	-0.08	0.00413	BCL11A
2	60,494,941	C	G	75	1506	3697	-0.06	0.47756	BCL11A
2	60,494,941	C	T	1130	19614	49138	-0.03	0.17139	BCL11A
2	60,494,942	T	A	191	3803	8753	-0.08	0.13585	BCL11A
2	60,494,942	T	C	342	6587	16703	0.05	0.21778	BCL11A
2	60,494,942	T	G	57	1317	3390	0.1	0.266	BCL11A

Showing 1 to 10 of 2,062 entries Previous 1 2 3 4 5 … 207 Next

Search parameters:

Chromosome	Gene	Position	Single base deletions	Variant Coeff.	No. of tags	P-value
2	BCL11A	60,494,940 - 60,495,539	true	-1.65 - 1.49	1 - 631,488	0.0000 - 0.9984

D

Saturation Mutagenesis data for BCL11A

Data | Viewer

Search parameters:

Chromosome	Gene	Position	Single base deletions	Variant Coeff.	No. of tags	P-value
2	BCL11A	60,494,940 - 60,495,539	true	-1.65 - 1.49	1 - 631,488	0.0000 - 0.9984

MPRAbase a Massively Parallel Reporter Assay database

Jingjing Zhao, Fotis A Baltoumas, Maxwell A Konnaris, et al.

Genome Res. published online April 22, 2025
Access the most recent version at doi:10.1101/gr.280387.124