Functional Assays I

Mike White
Bio5488
4/10/23
Lectures

- **MONDAY: Technologies**
  - Massively parallel reporter assays, deep mutational scans, pooled CRISPR screens, etc.

- **WEDNESDAY: Computational analysis**
  - Quantification, modeling, predicting variant effects
Today’s Learning Goals

Understand the technological basis of large scale functional assays:

• What are the main methods?

• What technical tricks make them work?

• What are the limitations of these technologies?
How would you do the following?

- Measure transcriptome for 1000 gene knockouts
- Test the cis-regulatory activity of all ATAC-seq peaks in a cell type
- Determine the functional impact of all possible single residue mutations in a disease-linked protein
- Mutagenize every base pair of a super enhancer at its native genomic locus (not on a plasmid)
- Measure drug sensitivity on knockdown of each of 500 tumor suppressor genes
- Identify which of the thousands of SNPs in human 3’ UTRs alter splicing efficiency
- Test the effect of knocking out TF binding sites in 1,000 ChIP-seq peaks
What do these experiments have in common?

They all require some perturbation or functional manipulation. Simply measuring naturally occurring molecules in the genome isn’t enough.
Why not just measure naturally occurring molecular species?

- Many important natural genetic variants are rare - it may be hard to find individuals with them.
- Making perturbations is a powerful way to test hypotheses.
- Deep learning: We often need bigger, more informative training sets to model cellular systems.

Example of need for functional testing:
Which sequences are enhancers?

To pick out genuine enhancers, epigenetic data are helpful but not definitive.
Characteristics of an Enhancer

Genome is packed with potential TF binding sites

“In a 1-kb segment of human DNA it is predicted that a new 7-8 bp protein-binding motif arises, by neutral evolution, on average every 60,000 years.”

In total, we defined 926,535 cCREs in human and 339,815 cCREs in mouse.
25-50% of Candidate CREs show activity in a functional assay

Key Technical Tricks of Functional Assays
Functional Assays: The View from 9,144 meters

• **We have a scale problem:** We need to characterize enormous numbers of non-coding elements and coding/non-coding genetic variants.

  • (ChIP-seq, RNA-seq are not enough - we need tests of *function* - hence, functional assays)

• **We have a technological opportunity:**

  • Cheap DNA sequencing **and** synthesis

  • Big catalogs of variants, ChIP-seq peaks, epigenomic data, etc.

• **Massively parallel functional assays solve the scale problem using cheap DNA synthesis & sequencing, and big data catalogs.**
High Throughput DNA Sequencing & Synthesis

Sequencing
2011: $10,000 per human genome
2021: < $1000 per human genome

Synthesis:
2011: $10/bp in 2011
2021: $0.09/bp in 2021

NHGRI: https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost
How do we leverage the power of cheap sequencing/synthesis to perform functional assays at scale?

Turn your assay into a sequencing assay! Pool samples and read output by sequencing.
Sequence Barcodes

- Minimal promoter
- Cis-regulatory element
- Reporter gene
- Unique DNA Barcode (9bp)
- mRNA
Sequence Barcodes

Dixit, et al., Cell 2016; 167(7):1853-1866.e17
Sort-seq

Genome-integrated reporter gene reading out some function

Flow sort into bins of fluorescence level

Sequence barcode in each bin

<table>
<thead>
<tr>
<th>Barcode</th>
<th>Bin 1</th>
<th>Bin 2</th>
<th>Bin 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC 1</td>
<td>10</td>
<td>234</td>
<td>57</td>
</tr>
<tr>
<td>BC 2</td>
<td>90</td>
<td>31</td>
<td>3</td>
</tr>
<tr>
<td>BC 3</td>
<td>30</td>
<td>1546</td>
<td>706</td>
</tr>
<tr>
<td>BC 4</td>
<td>25</td>
<td>984</td>
<td>33</td>
</tr>
</tbody>
</table>

Cell Count

GFP fluorescence
Landing Pads

B

1. Transfect cells with promoter library and CRE recombinase

2. Treat with Ganciclovir

Hong & Cohen, Genome Research (2022) 32:85-96
Summary of Key Tricks

• Turn your functional assay into a sequencing assay

• Leverage cheap DNA synthesis to build libraries of designed sequences

• Sequence barcodes

• Sort-seq

• Landing pads
Getting barcoded libraries into cells

- **Plasmids**
  - Easy to clone libraries, requires transfectable cell type, multiple library members per cell

- **Lentivirus/Adeno-associated virus**
  - genome integrated (lentivirus), genomic position effect can affect expression of construct, about one library member per cell (when applied at low MOI), construct size may be limited by what can be packaged in virus

- **Landing pads**:
  - Requires prior cell line engineering, library recombination efficiency into landing pad can be low, one library member per cell at a defined location.
Specific Technologies
We need a method to:

- Test the function of hundreds of thousands of putative enhancer sequences
- Test functional effect of tens of thousands of putative enhancer variants
Non-consensus GLI binding sites in Hedgehog target gene regulation.

Winklmayr M\textsuperscript{1}, Schmid C, Laner-Plamberger S, Kaser A, Aberger F, Eichberger T, Frischauf AM.
Classic Reporter Gene

Minimal promoter

Cis-regulatory element

Reporter gene

CRE

DsRed
The Barcode

Minimal promoter

Cis-regulatory element

CRE

DsRed

BC

Unique DNA Barcode (9bp)

Reporter gene

mRNA
Massively Parallel Reporter Assay (MPRA)

... pooled library of $10^5$ or more distinct reporters
MPRA Flavors

• Plasmid with upstream element and 3’ barcode

• Self-regulating element that serves as its own barcode (STARR-seq)

• Genome integrated reporters via lentivirus

• Sort-seq
Libraries by Custom Array Oligo Synthesis

Massively Parallel Reporter Assay (MPRA)

... pooled library of $10^5$ or more distinct reporters
Testing TF binding sites w. selected sequences & mutations

Proc Natl Acad Sci U S A. 2017 Feb 14;114(7):E1291-E1300.
Sheared Genomic DNA (STARR-seq)

STARR-seq
What are MPRA Limitations?

- Lack of native genomic context
- Multiple DNA copy number
- Mainly done in transfected cell lines which may not recapitulate enhancer biology
- Potential assay toxicity - plasmids, high reporter gene levels
- Compatibility with reporter/minimal promoter: minimal promoter used may lead to false positives, negatives
Deep Mutational Scans
Analysis of protein-coding genetic variation in 60,706 humans

Monkol Lek, Konrad J. Karczewski, Eric V. Minikel, Kaitlin E. Samocha, Eric Banks, Timothy Fennell, Anne H. O'Donnell-Luria, James S. Ware, Andrew J. Hill, Beryl B. Cummings, Taru Tukiainen, Daniel P. Birnbaum, Jack A. Kosmicki, Laramie E. Duncan, Karol Estrada, Fengmei Zhao, James Zou, Emma Pierce-Hoffman, Joanne Berghout, David N. Cooper, Nicole Deflaux, Mark DePristo, Ron Do, Jason Flannick, Exome Aggregation Consortium

Nature 536, 285–291 (2016) | Cite this article

319k Accesses | 6634 Citations | 1253 Altmetric | Metrics
"[With current cancer panels], the probability of detecting a VUS is higher than the probability of detecting a pathogenic variant."


Prevalence of pathogenic variants and variants of unknown significance in patients at high risk of breast cancer: A systematic review and meta-analysis of gene-panel data.

van Marcke C¹, Collard A², Vikkula M³, Duhoux FP⁴.
Commentary

Variant Interpretation: Functional Assays to the Rescue

Lea M. Starita 1, Nadav Ahituv 2, 3, Maitrey J. Dunham 1, Jacob O. Kitzman 4, 5, Frederick P. Roth 6, 7, 8, 9, Georg Seelig 10, 11, Jay Shendure 1, 12, Douglas M. Fowler 1, 13
Deep Mutational Scans

- Create a library of all possible protein aa variants
- Pick an assayable function
- Pick a cell type
- Integrate one variant per cell (usually)
- Read out assay by sequencing DNA of variants
Deep Mutational Scan of SARS-CoV2

Deep Mutational Scanning of SARS–CoV–2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding

Tyler N Starr 1, Allison J Greaney 2, Sarah K Hilton 3, Daniel Ellis 4, Katharine H D Crawford 2, Adam S Dingens 1, Mary Jane Navarro 5, John E Bowen 5, M Alejandra Tortorici 5, Alexandra C Walls 5, Neil P King 6, David Veesler 5, Jesse D Bloom 7

Affiliations + expand

PMID: 32841599    PMCID: PMC7418704    DOI: 10.1016/j.cell.2020.08.012
Test all possible ~3,800 RBD variants for effect on stability and ACE2 affinity:

- PCR-based codon mutagenesis
- Library of mutant variants
- Addition of barcode sequence
- Express in yeast display system
- PacBio sequencing links barcodes to mutant variants
- Illumina sequencing reads short barcodes
Measure ACE2 binding and RBD expression by sort-seq:

**EXPRESSION**

(A) unstained, stained, bin1, bin2, bin3, bin4

**BINDING**

(B) unmutated RBD, RBD mutant library

ACE2 RBD

yeast
Example 2: VAMP-Seq

Multiplex assessment of protein variant abundance by massively parallel sequencing


Premise: Protein stability is a good functional assay

VAMP-Seq
(Variant abundance by massively parallel sequencing)

1. Create variants by site-saturation mutagenesis (variable oligos with inverse PCR on a WT plasmid template)

2. Variants fused to GFP, sequence barcodes cloned in

3. Integrate library at single locus in HEK293 cells

4. Sort-seq: Sort cells on GFP levels, sequence sorted bins
Assessing VUS vs Pathogenic Variants
Pooled CRISPR Assays

Why?
What’s Needed for Pooled CRISPR?

• A library of designed guide RNAs
• A common readout of function (maybe reporter gene)
• Method to get Cas9 and 1 gRNA per cell. (Why 1)
• A way to link gRNA identity with phenotypic readout of the cell.
Big Picture Example: Genetic Screen for SHH Regulators

CRISPR Screens Uncover Genes that Regulate Target Cell Sensitivity to the Morphogen Sonic Hedgehog

Ganesh V Pusapati ¹, Jennifer H Kong ², Bhaven B Patel ², Arunkumar Krishnan ³, Andreas Sagner ⁴, Maia Kinnebrew ², James Briscoe ⁴, L Aravind ³, Rajat Rohatgi ⁵
Genetic Screen for SHH Regulators

Pool of 78,637 guides (+1000 control guides) targeting 19,674 genes (4 guides/gene)

NIH/3T3-CG cells (expressing Cas9 and GLI-GFP reporter)

- **NoSHH**: unsorted control, sorted
- **LoSHH**: unsorted control, sorted
- **HiSHH**: unsorted control, sorted

Cell number vs. GFP fluorescence
The Data
Single Cell Sequencing
Perturb-seq

Resource

Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens

Atray Dixit 1, 2, 9, Oren Parnas 1, 9, 10, Biyu Li 1, Jenny Chen 1, 2, Charles P. Fulco 1, 4, Livnat Jerby-Aronon 1, Nemanja D. Marjanovic 1, 3, Danielle Dionne 1, Tyler Burks 1, Raktima Raychowdhury 1, Britt Adamson 5, Thomas M. Norman 5, Eric S. Lander 1, 4, 6, Jonathan S. Weissman 5, 7, Nir Friedman 1, 8, Aviv Regev 1, 6, 7, 11

See also companion papers:
Cell, Volume 167, Issue 7, 15 December 2016, Pages 1867-1882.e21
Cell, Volume 167, Issue 7, 15 December 2016, Pages 1883-1896.e15
Aim: Identify factors in LPS immune response by genetic screen

Method: Combine pooled CRISPR library with single cell RNA-seq

What problems have to be solved to make the method work?
Pooled sgRNA libraries $\rightarrow$ Transduce cells $\rightarrow$ Selection $\rightarrow$ Barcoded droplet mRNA capture $\rightarrow$ Library construction $\rightarrow$ NGS Sequencing

$\lambda_{\text{MOI}}$ $\lambda_{\text{cells}}$ $\lambda_{\text{reads/cell}}$
<table>
<thead>
<tr>
<th>Cell type</th>
<th>sgRNA pool</th>
<th>Total cells</th>
<th>Time points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse BMDC</td>
<td>Transcription factors (67 guides)</td>
<td>70,000</td>
<td>0 and 3 hr post-LPS</td>
</tr>
<tr>
<td>Human K562</td>
<td>Transcription factors (46 guides)</td>
<td>104,000</td>
<td>7 and 13 days</td>
</tr>
<tr>
<td>Human K562</td>
<td>Cell cycle regulators (36 guides)</td>
<td>26,000</td>
<td>7 days</td>
</tr>
</tbody>
</table>
\[
\log(\text{Cells}) + 1 = Y = X \beta
\]

- **Y**: Expression matrix
- **X**: Design matrix
- **\beta**: Coefficient (regulatory) matrix

**Genes**
- Signature decomposition

**Covariates**
- Cell features
- Design of experiments

**Genes**
- Inference
- Interpretation
In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes

Xin Jin,1,2,3,4,* Sean K. Simmons,3,5,6 Amy Guo,3 Ashwin S. Shetty,2 Michelle Ko,2 Lan Nguyen,3,6 Vahbiz Jokhi,2 Elise Robinson,3,5,8 Paul Oyler,2 Nathan Curry,2 Giulio Deangeli,2 Simona Lodato,7 Joshua Z. Levin,3,5,6 Aviv Regev,3,6,9,10,† Feng Zhang,3,4,10,† and Paola Arlotta2,3,5,†

Abstract

The number of disease risk genes and loci identified through human genetic studies far outstrips our capacity to systematically study their functions. We applied a scalable genetic screening approach, in vivo Perturb-Seq, to functionally evaluate 35 autism spectrum disorder/neurodevelopmental delay (ASD/ND) de novo loss-of-function risk genes. Using CRISPR-Cas9, we introduced frameshift mutations in these risk genes in pools, within the developing mouse brain in utero, followed by single-cell RNA sequencing of perturbed cells in the postnatal brain. We identified cell type-specific and evolutionarily conserved gene modules from both neuronal and glial cell classes. Recurrent gene modules and cell types are affected across this cohort of perturbations, representing key cellular effects across sets of ASD/ND risk genes. In vivo Perturb-Seq allows us to investigate how diverse mutations affect cell types and states in the developing organism.
A

Lentiviral infection

Cas9^{+-}
E12.5 embryos

Ventricular zone

E12.5
E13.5
E14.5

Layer 2–3
Layer 4
Layer 5
Layer 6

Astrocytes
Oligodendrocytes

FACS enrichment for BFP^+

Single cell transcriptomics

Perturbation–cell type analysis

Neg. control
Risk gene 1
Risk gene 2
...
Risk gene 35

Cell type-specific effect of ASD risk gene perturbations
The Current Vision
Learning Objectives Again:

1. What are functional assays and why do we need them?
2. Technical strategies for building and assaying libraries
3. Understand key elements of experimental design

These assays are now widely used to test hypotheses/functionally validate results of genomic studies. They are accessible and affordable so consider adding them to your experimental repertoire!
Assigned Reading

Molecular Cell

Volume 71, Issue 6, 20 September 2018, Pages 1012-1026.e3

Article

Quantitative Activity Profile and Context Dependence of All Human 5’ Splice Sites

Mandy S. Wong, Justin B. Kinney, Adrian R. Krainer

Assigned Reading

• Don’t worry about the computational analysis or specific biological results.

• Don’t worry about memorizing details of the methods.

• Focus on assay design, library construction, strengths and weaknesses of the functional assay - the big themes we discussed.
As you read the paper ask these questions:

• What is the purpose of the functional assay(s)?

• How does the assay work?

• What are the limitations of the method?