

6.874, 6.802, 20.390, 20.490, HST.506

Computational Systems Biology

Deep Learning in the Life Sciences

Lecture 18 – High-throughput Perturbations

MPRA, STARR-Seq, SHARPR, HiDRA, Perturb-Seq

Prof. Manolis Kellis

High-throughput expts: MPRA, SHARPR, HiDRA, Perturb-seq

1. High-throughput synthesis: Massively Parallel Reporter Assays (MPRA)

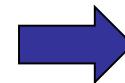
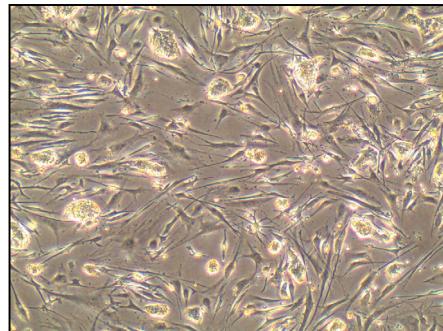
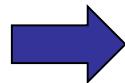
- MPRA technology: in vitro synthesis, reporter design, transfection
- 1 site X deep: single-base dissection, combinatorial changes, high-res map
- 2k sites x shallow: motif-guided perturbations, activators/repressors

2. Next-generation applications of MPRA + STARR-Seq, SHARPR, HiDRA

- 10k x deep: high-resolution dissection by tiling + deconvolution (Sharpr)
- 10M x deep: HiDRA, no synthesis (STARR), targeted capture (ATAC)

3. Endogenous genome editing: cutting and template-based repair

- Endonucleases: TALENs, ZNFs. Repair pathways. Diversity of outcomes.
- CRISPR-Cas9: origins, discovery, optimization, base-editing
- Next-generation: Perturb-seq, multiplexing, delivery, applications



Write DNA

Barcodes
Primer pools
Hybridization probes
RNAi libraries
Synthetic genes
Mutant libraries

...

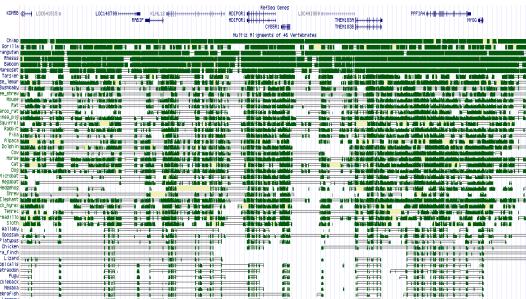
Read DNA

RNA-Seq
ChIP-Seq
Methyl-Seq
GRO-Seq
4C/5C/HiC

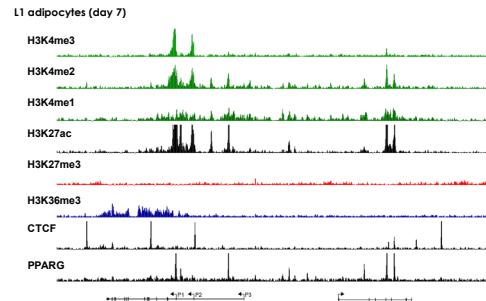
...

We can find regulatory elements ... but we don't know how to read them

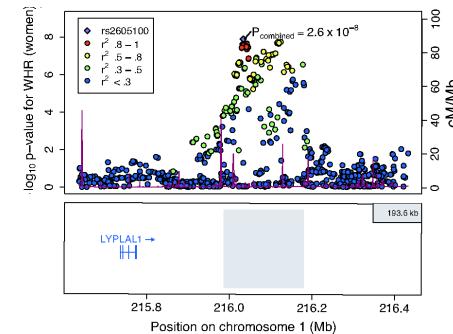
Comparative sequence analysis



Genome-wide chromatin/TF mapping



Genetics



100s of megabases of likely cis-regulatory elements

TATGGAACTGAAATGCCCACTGTCCAGAGACATCTTTCTTCAATTCTGGTCATAAACCTGGTTGAAAAAGCAAATTCAAGAATTCTCATCTAATAATGACAGAAAAAGAACATTCTGAATGAATTGTGGAAGTGTACAATTAAATTTCATTAAATTTCATTTACAATTTCATTTCAATTTCATTTCAATTAAATTTCCTCTTAATCCCATGCAAGGACACAGCAGTTAATAAAACTTAAAGAAATCTCTCAATTGCATACCATTCTTAATGCAACACTGCAACACTGGATTATACTGCTTCAAGGTTTATAATCCAGTTATTAGTATCTCAAACTTTAAATCAAAAGTCATGGCAGCTCTAAATATGCTTTCCTCAAGGTATCAACTCACCCCCACCTCACCCCCAAAGAGACTTGTGGTACGCCCTCTGGTGAGGCCCTTAAAGAGCAAAACAAATTGAGAATGTGTAGCCCTCCAGTTCCCAACCTTCATCCATTTTTTTTGCACTGCTGCCATGATCTGTACTTAAACAGGTATATTAGAAAATACCAAATTCTCTTAGAGAAAAAAATTGCACACAATCTCAACACTGAGAGGACTGCTCAAGGATGAAAGCAAGGTACTCTTAGGTTAGTTTTCTAAACCTATAGATGGCATTGGAAAGAAGTACACGCATAGGCTTCAAAAACCTGAACACCAGAAACCTAATTGCACTTTCTTTAGGGCACGGGTTTCCCTTACAGCCAACCTTAATGGGTCCAATTAGGCAACAATGAAAGGTTACAAAGTCTACAAAAGGTCCTTACAGCATTAAACAGGGTTCACTGTTTAACTCAAAATAGTTTACAGTGTGTTGTTGTT

Traditional regulatory element “bashing”

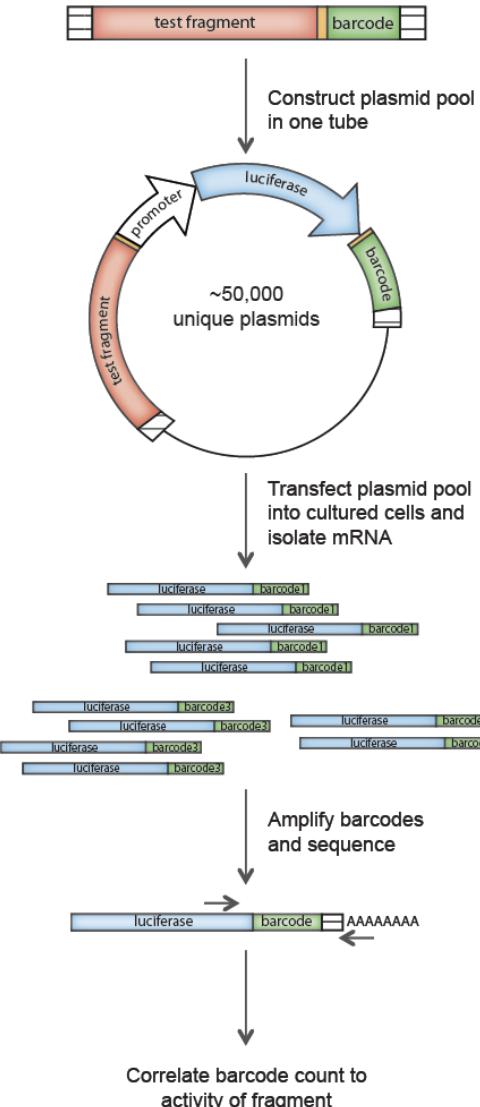


Bottlenecks:

1. Generating/cloning individual variants is tedious
2. Enzymatic/fluorescent reporters limit multiplexing

Massively Parallel Reporter Assays (MPRA)

50,000+ synthetic DNA fragments (200+ bp)

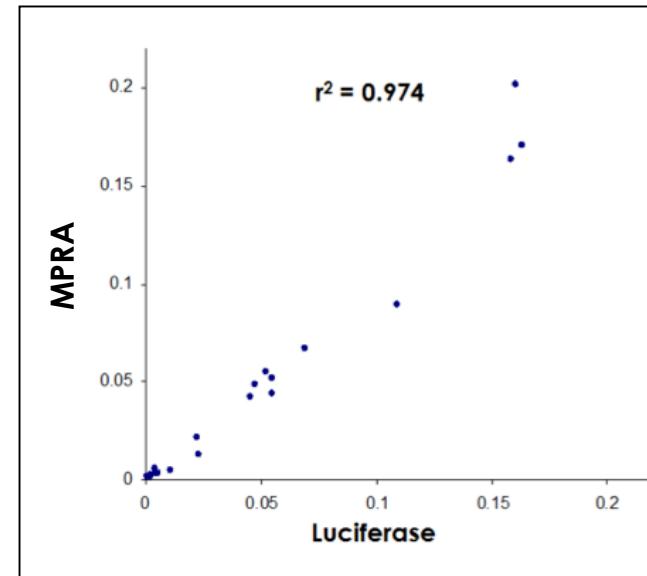


Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay

nature biotechnology

Alexandre Melnikov^{1,9}, Anand Murugan^{2,9}, Xiaolan Zhang^{1,9}, Tiberiu Tesileanu^{2,3}, Li Wang¹, Peter Rogov¹, Soheil Feizi^{1,4}, Andreas Gnirke¹, Curtis G Callan Jr^{2,3}, Justin B Kinney⁵, Manolis Kellis^{1,4}, Eric S Lander^{1,6,7} & Tarjei S Mikkelsen^{1,8}

- Flexible assay format:
Promoters, enhancers, silencers,
Insulators, RNA stability elements, ++
- Data is directly comparable to
traditional reporter assays:



- Throughput increased by
3 orders of magnitude

High-throughput expts: MPRA, SHARPR, HiDRA, Perturb-seq

1. High-throughput synthesis: Massively Parallel Reporter Assays (MPRA)

- MPRA technology: in vitro synthesis, reporter design, transfection
- 1 site X deep: single-base dissection, combinatorial changes, high-res map
- 2k sites x shallow: motif-guided perturbations, activators/repressors

2. Next-generation applications of MPRA + STARR-Seq, SHARPR, HiDRA

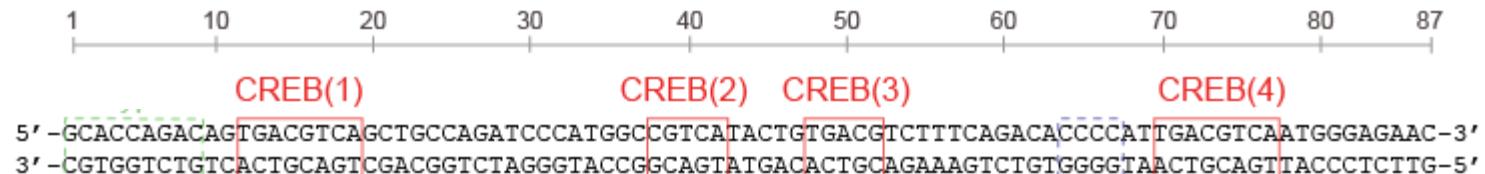
- 10k x deep: high-resolution dissection by tiling + deconvolution (Sharpr)
- 10M x deep: HiDRA, no synthesis (STARR), targeted capture (ATAC)

3. Endogenous genome editing: cutting and template-based repair

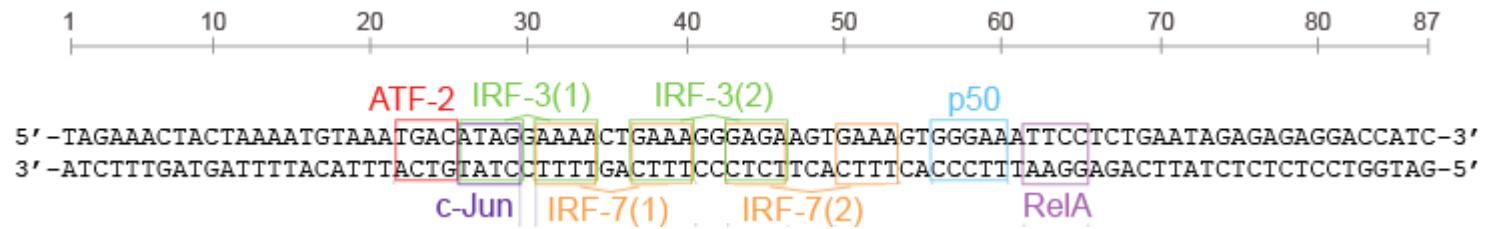
- Endonucleases: TALENs, ZNFs. Repair pathways. Diversity of outcomes.
- CRISPR-Cas9: origins, discovery, optimization, base-editing
- Next-generation: Perturb-seq, multiplexing, delivery, applications

MPRA Example 1: Inducible enhancers

Synthetic cAMP-responsive enhancer (Promega)

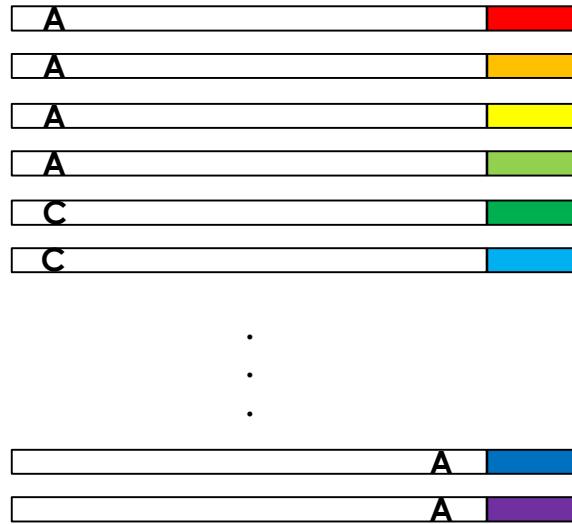


Virus-inducible enhancer of human interferon beta

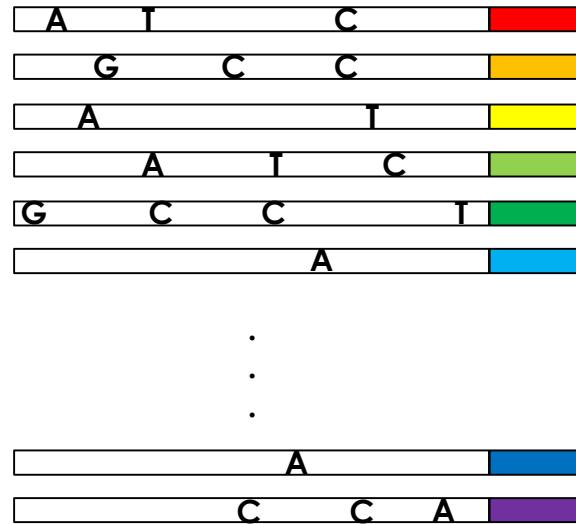


Two mutagenesis strategies

“Single-hit scanning”

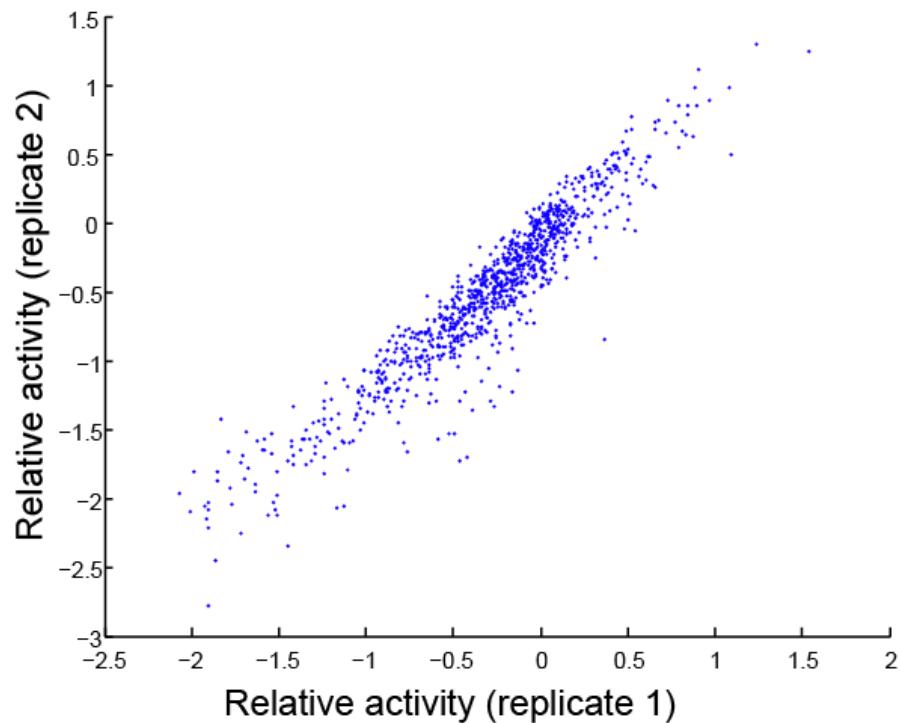


“Multi-hit sampling”

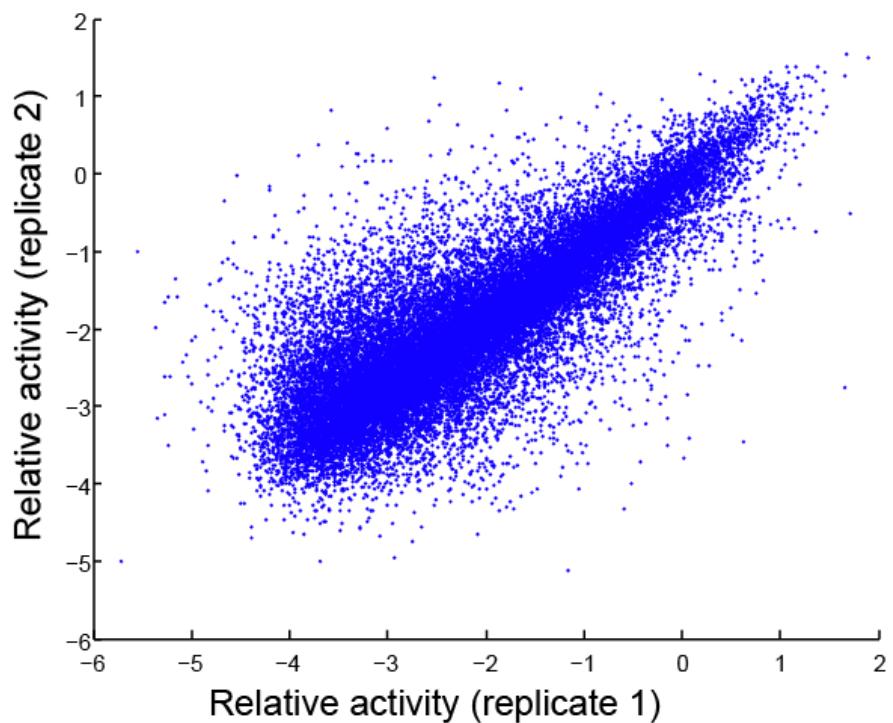


Two mutagenesis strategies

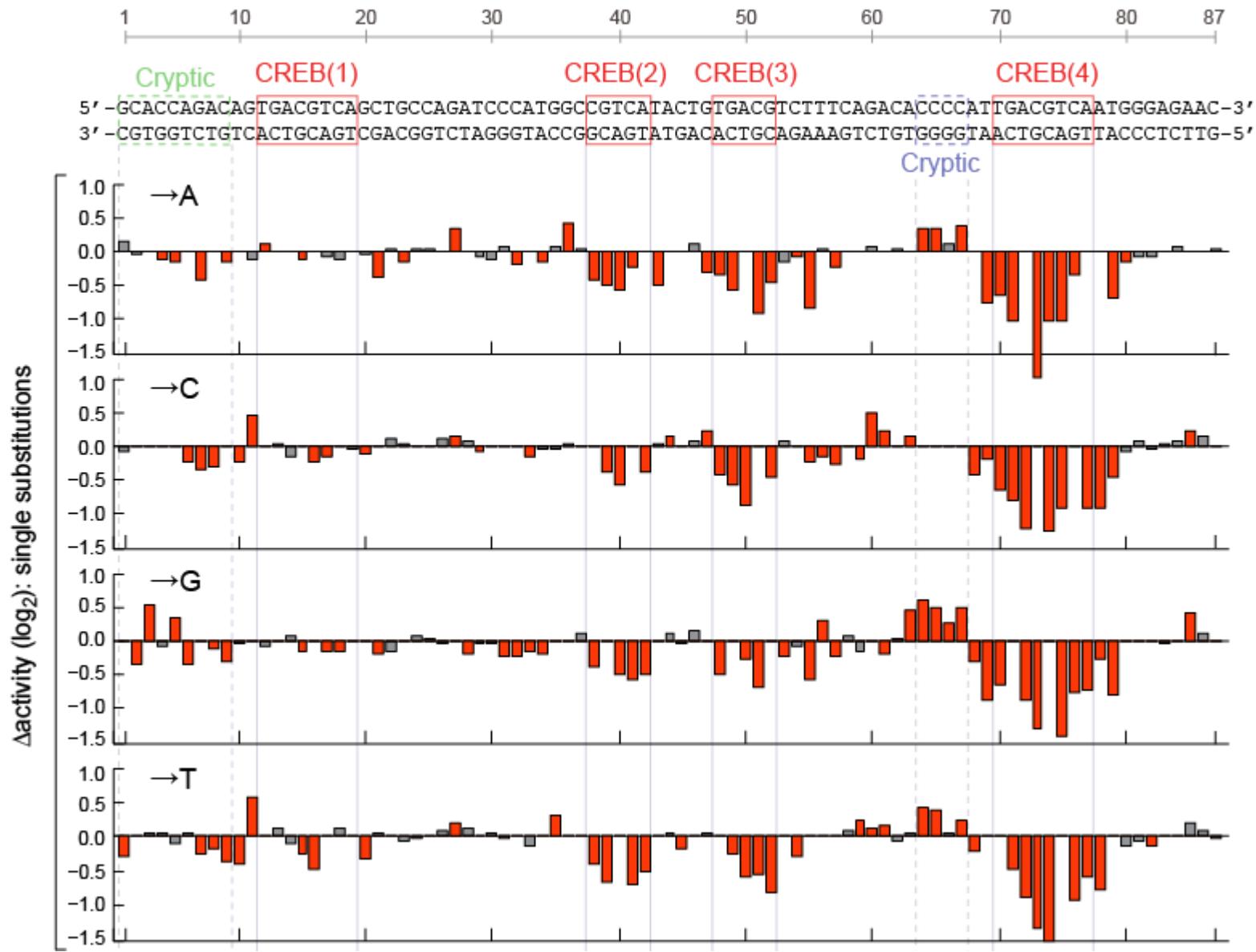
Single-hit (median of 13 tags), $r^2=0.89$



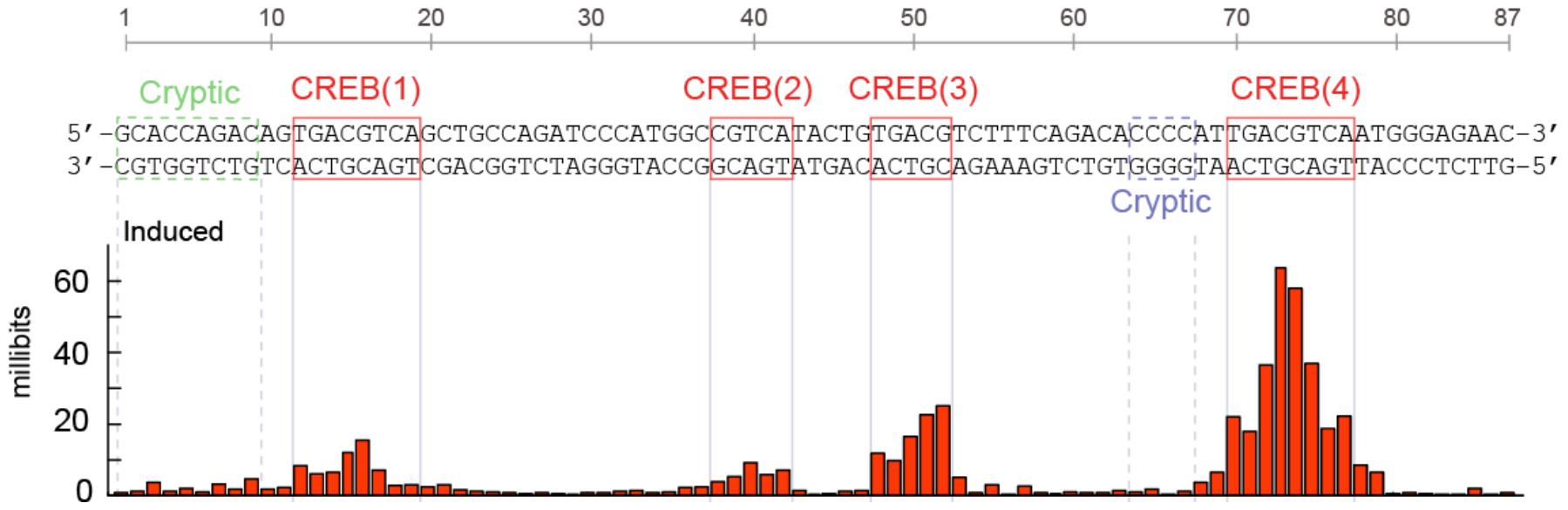
Multi-hit (single tags), $r^2=0.73$



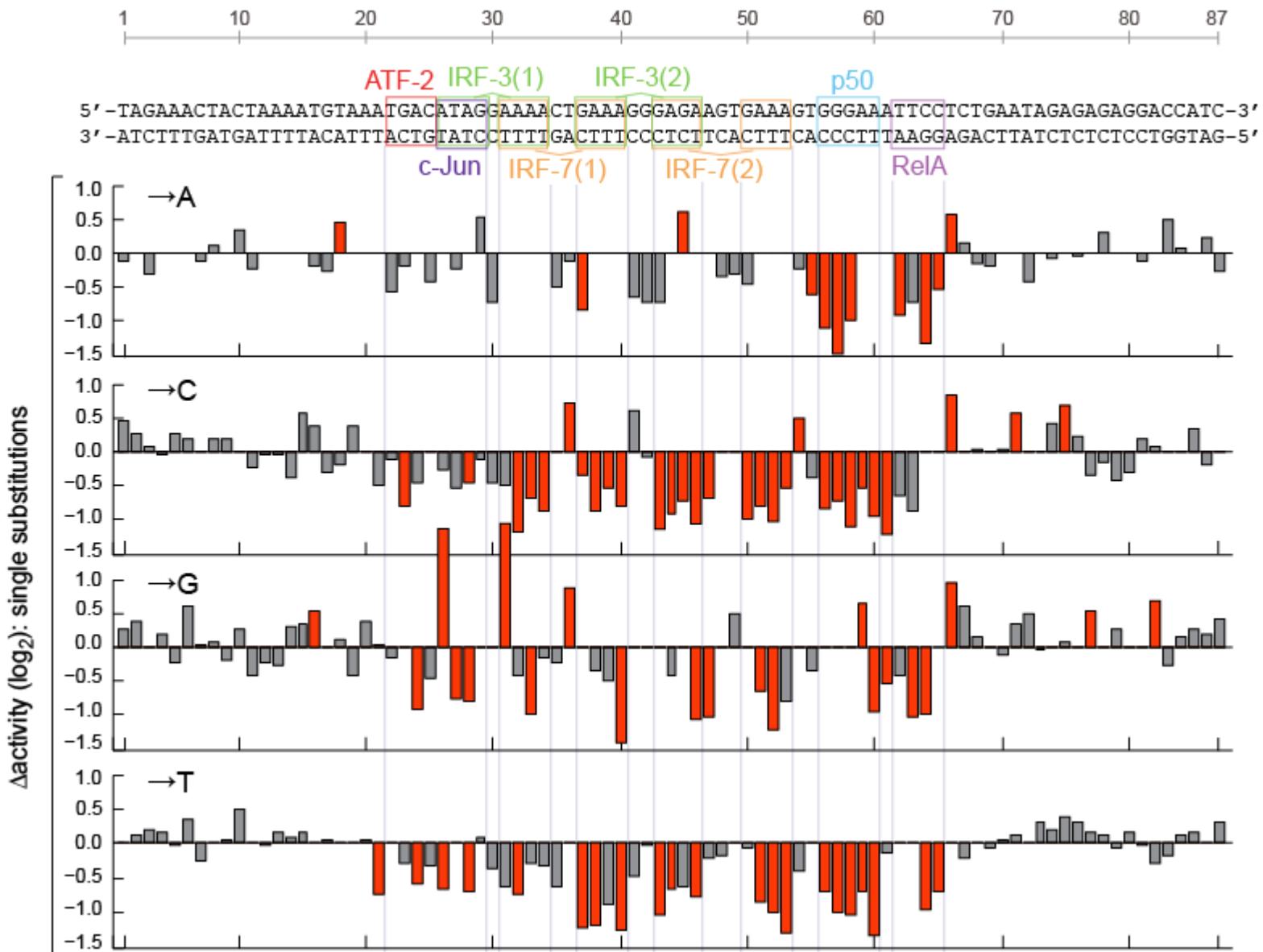
Single-hit scanning: Induced CRE



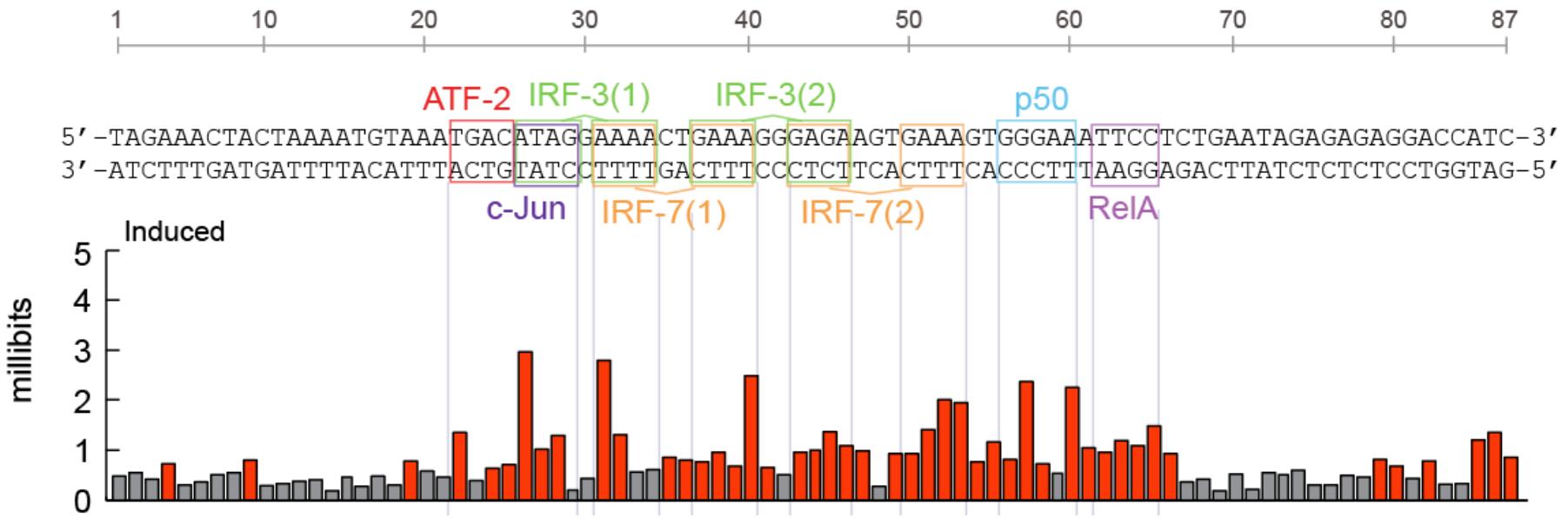
Multi-hit sampling: Induced CRE



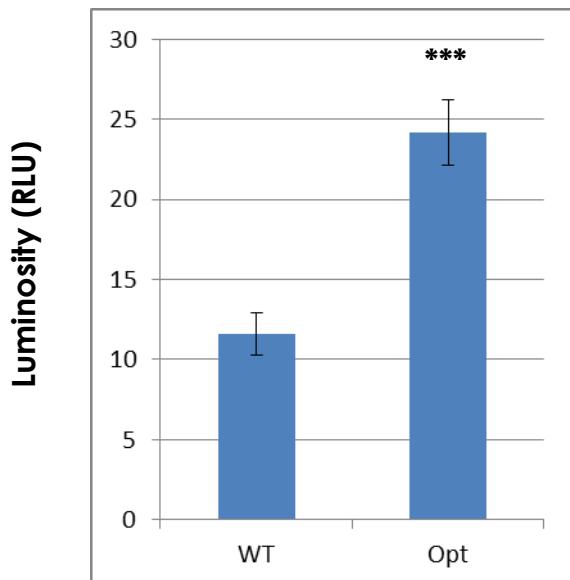
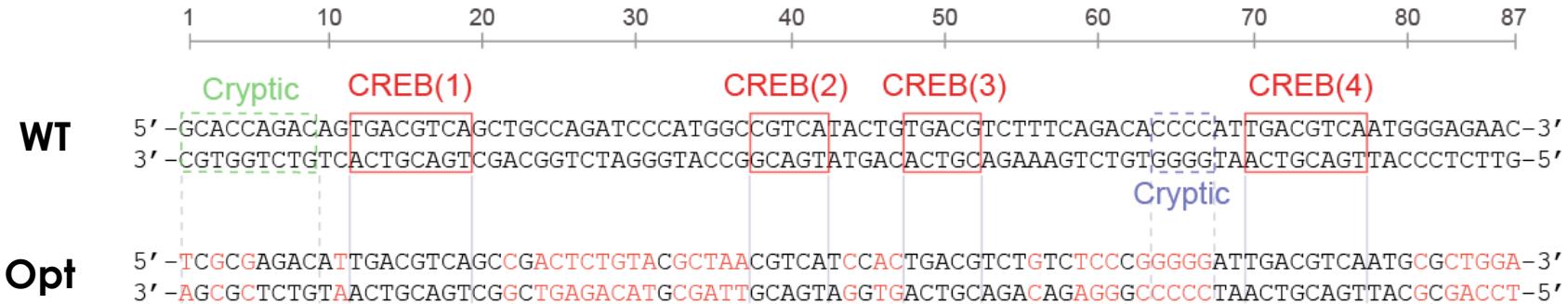
Single-hit scanning: Induced IFNB



Multi-hit sampling: Induced IFNB

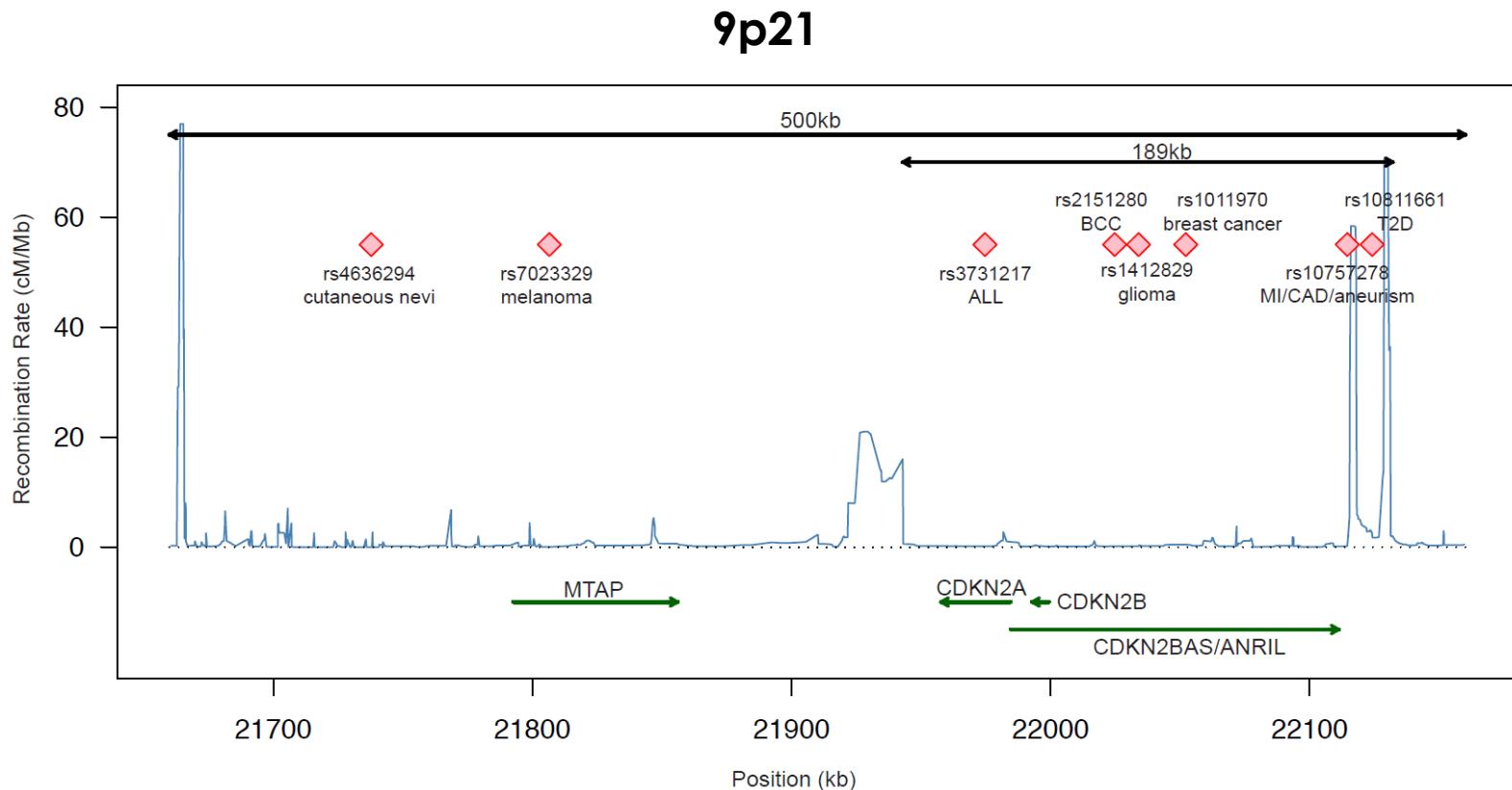


Expression optimization: CRE



Only 52 of 27,000 variants in the training set showed similar or higher increase

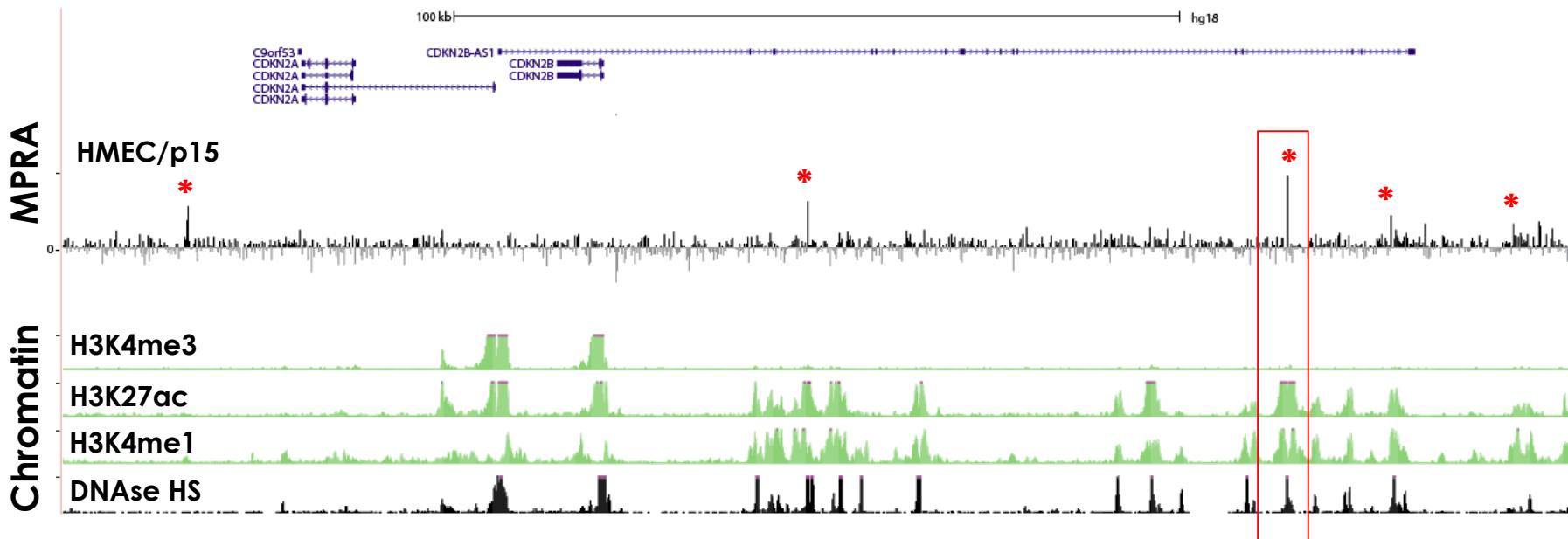
MPRA Example 2: Scanning for *cis*-regulatory elements



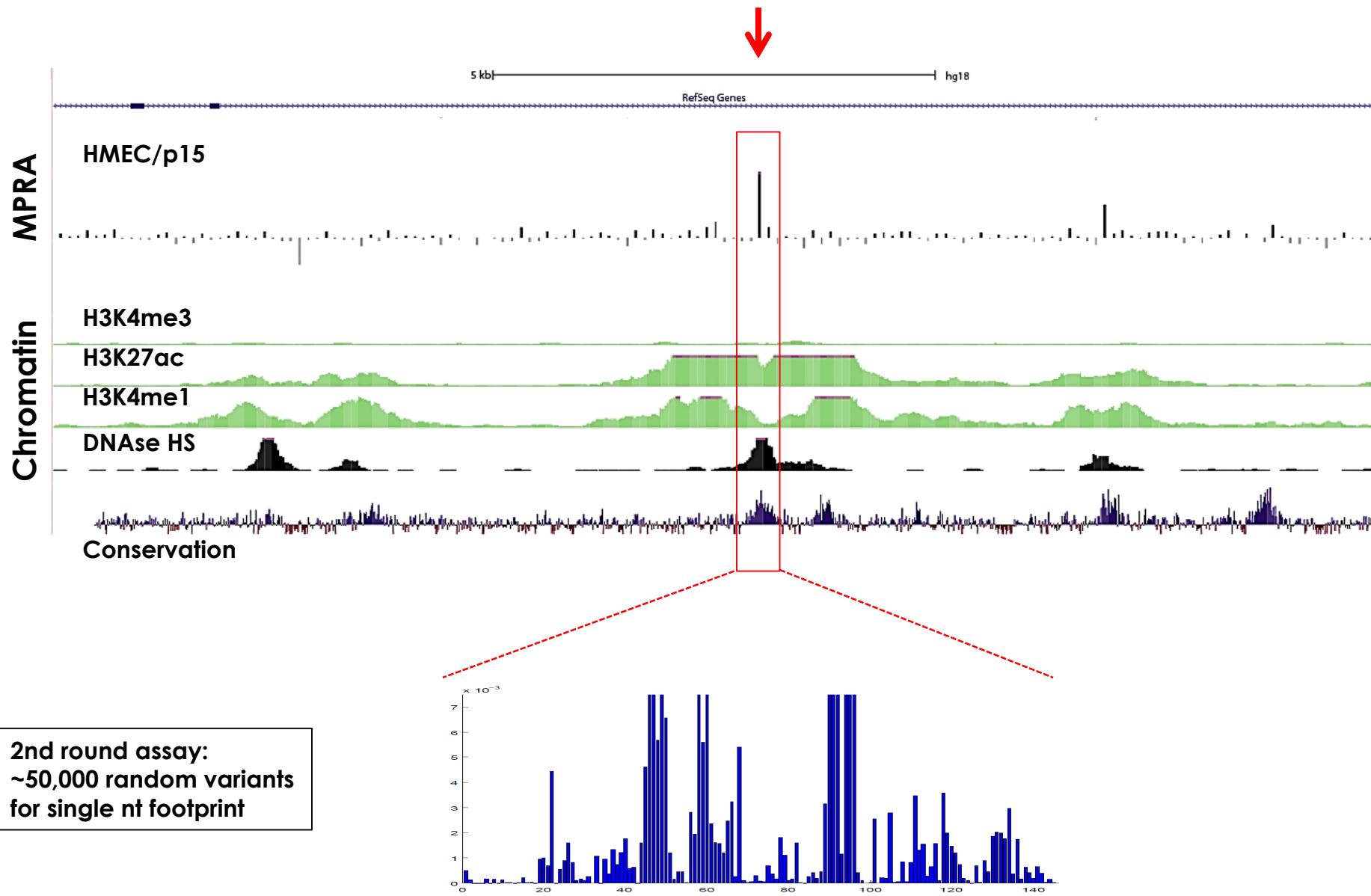
Experimental design

Tile across 210 kb region
at 100bp resolution
n= 2,897 distinct fragments,
27K oligos

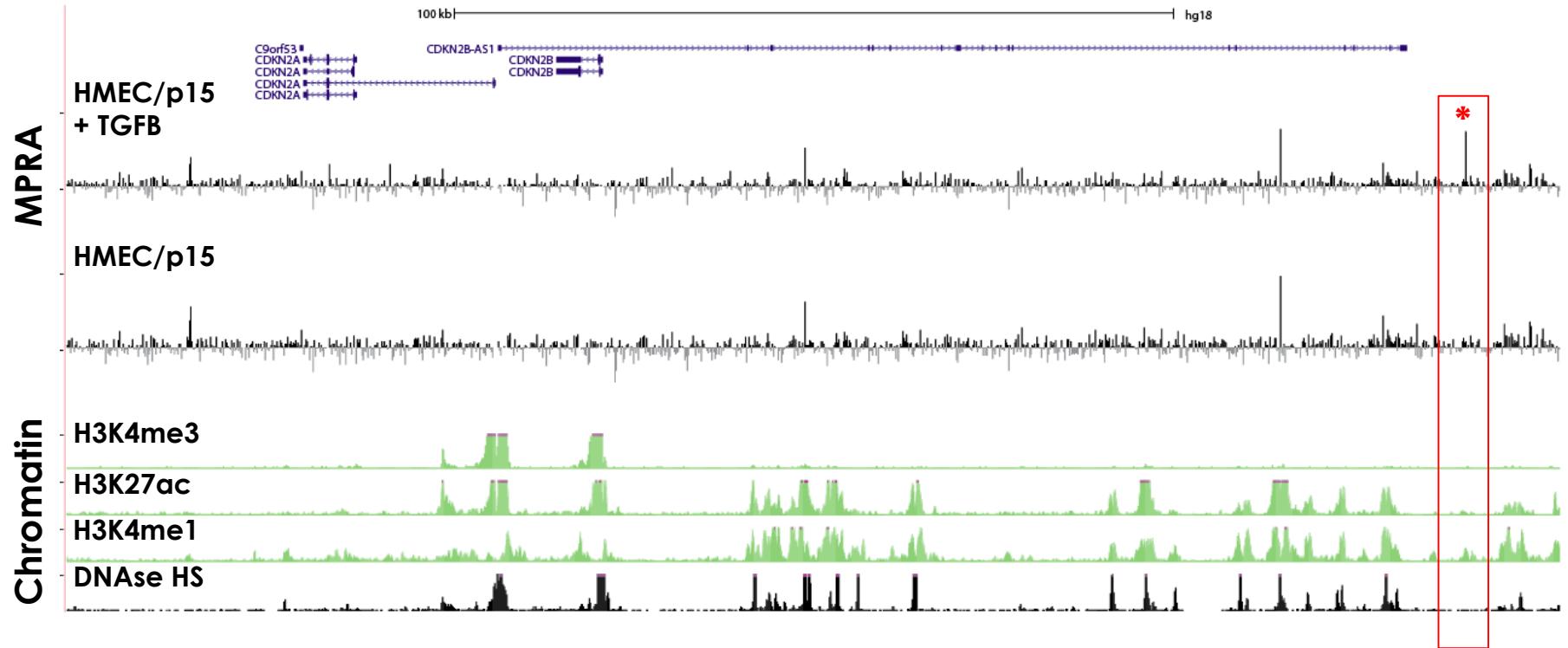
Putative enhancers across tiled 9p21 region



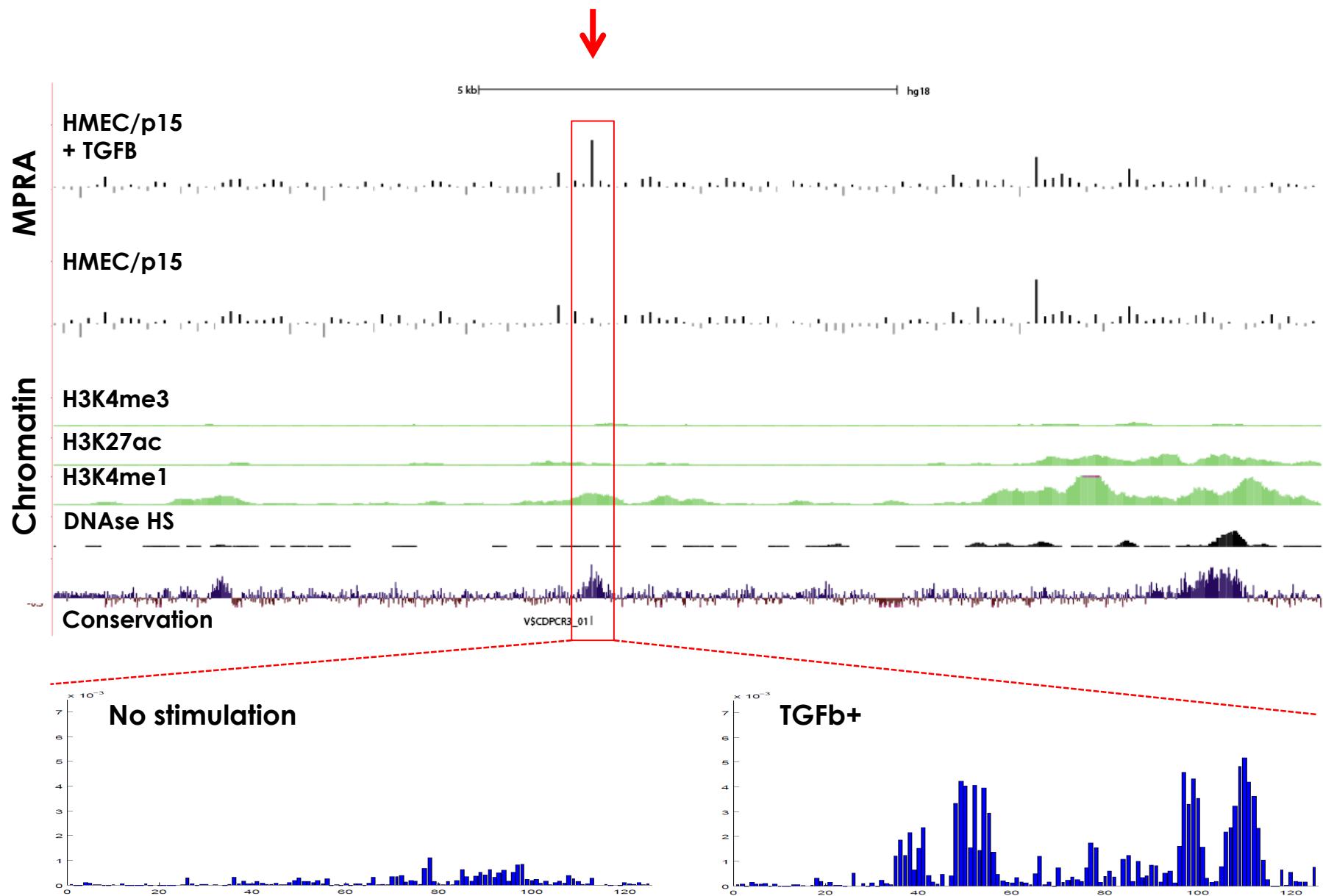
Zoom in and enhance!



Scanning for inducible *cis*-regulatory elements



TGFB-inducible enhancer



High-throughput expts: MPRA, SHARPR, HiDRA, Perturb-seq

1. High-throughput synthesis: Massively Parallel Reporter Assays (MPRA)

- MPRA technology: in vitro synthesis, reporter design, transfection
- 1 site X deep: single-base dissection, combinatorial changes, high-res map
- 2k sites x shallow: motif-guided perturbations, activators/repressors

2. Next-generation applications of MPRA + STARR-Seq, SHARPR, HiDRA

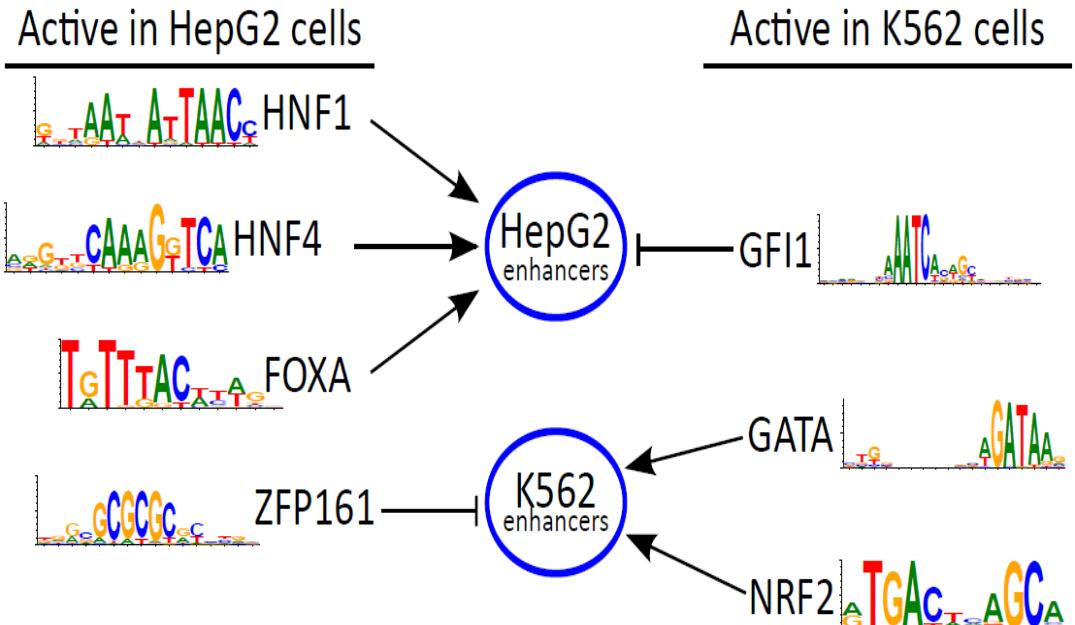
- 10k x deep: high-resolution dissection by tiling + deconvolution (Sharpr)
- 10M x deep: HiDRA, no synthesis (STARR), targeted capture (ATAC)

3. Endogenous genome editing: cutting and template-based repair

- Endonucleases: TALENs, ZNFs. Repair pathways. Diversity of outcomes.
- CRISPR-Cas9: origins, discovery, optimization, base-editing
- Next-generation: Perturb-seq, multiplexing, delivery, applications

Systematic motif disruption for 5 activators and 2 repressors in 2 human cell lines

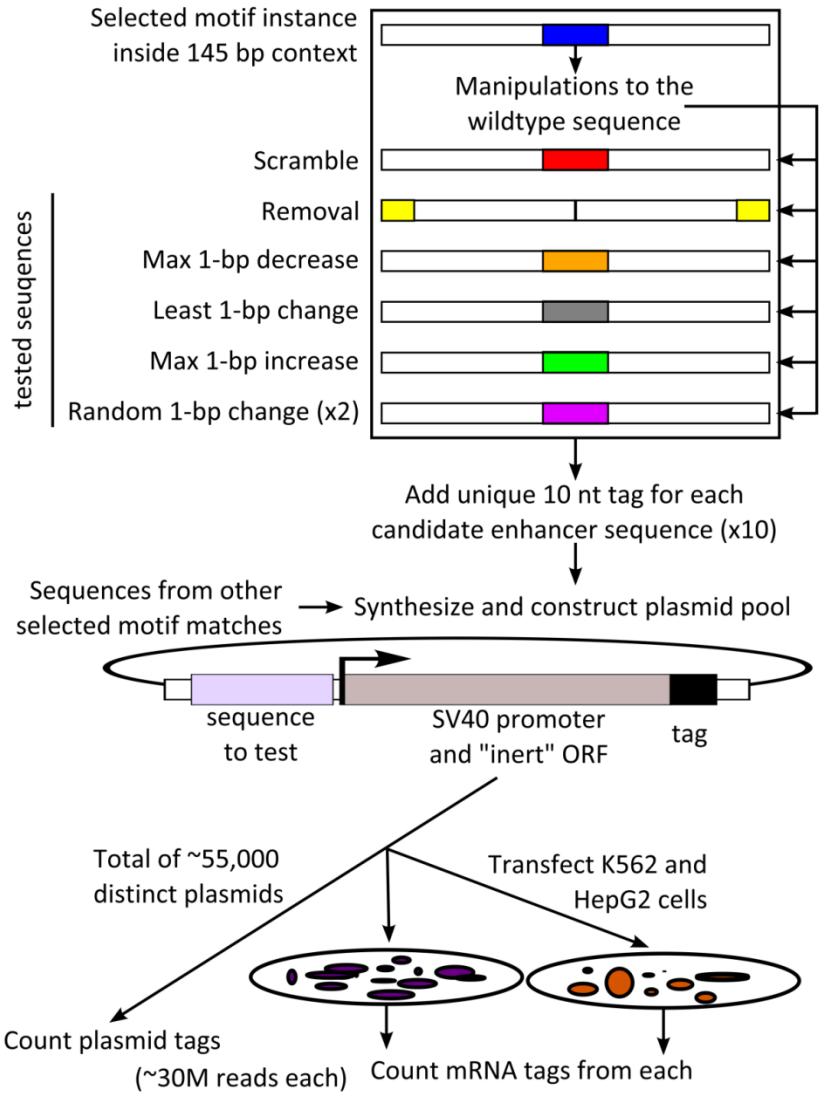
	Motif-motif similarity							Motif enrichment in enhancers		Factor expression	
	HNF1	HNF4	FOXA	GATA4	NRF2	ZFP161	GFI1	HepG2	K562	HepG2	K562
HNF1	1.0	0.4	0.4	0.4	0.4	0.1	0.4	1.5 2.3	0.0 0.4	1.0 1.0	0.8 0.5
HNF4	0.4	1.0	0.4	0.3	0.3	0.2	0.3	1.7 2.1	1.0 1.0	1.0 1.0	1.0 0.5
FOXA	0.4	0.4	1.0	0.3	0.5	0.1	0.4	1.4 1.7	1.0 1.0	1.0 1.0	2.2 2.1
GATA	0.4	0.3	0.3	1.0	0.3	0.1	0.5	1.0 1.0	2.1 2.8	0.1 0.3	0.4 0.4
NRF2	0.4	0.3	0.5	0.3	1.0	0.2	0.4	1.0 1.1	1.5 1.8	0.3 0.7	-0.1 -0.3
ZFP161	0.1	0.2	0.1	0.1	0.2	1.0	0.1	0.8 0.5	1.2 1.0	0.0 0.0	0.1 0.1
GFI1	0.4	0.3	0.4	0.5	0.4	0.1	1.0	1.0 1.0	0.6 0.5	0.4 0.3	1.3 1.1



Research

Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay

Pouya Kheradpour,^{1,2} Jason Ernst,^{1,2,5} Alexandre Melnikov,² Peter Rogov,² Li Wang,² Xiaolan Zhang,² Jessica Alston,^{2,3} Tarjei S. Mikkelsen,^{2,4} and Manolis Kellis^{1,2,6}



54000+ measurements (x2 cells, 2x repl)

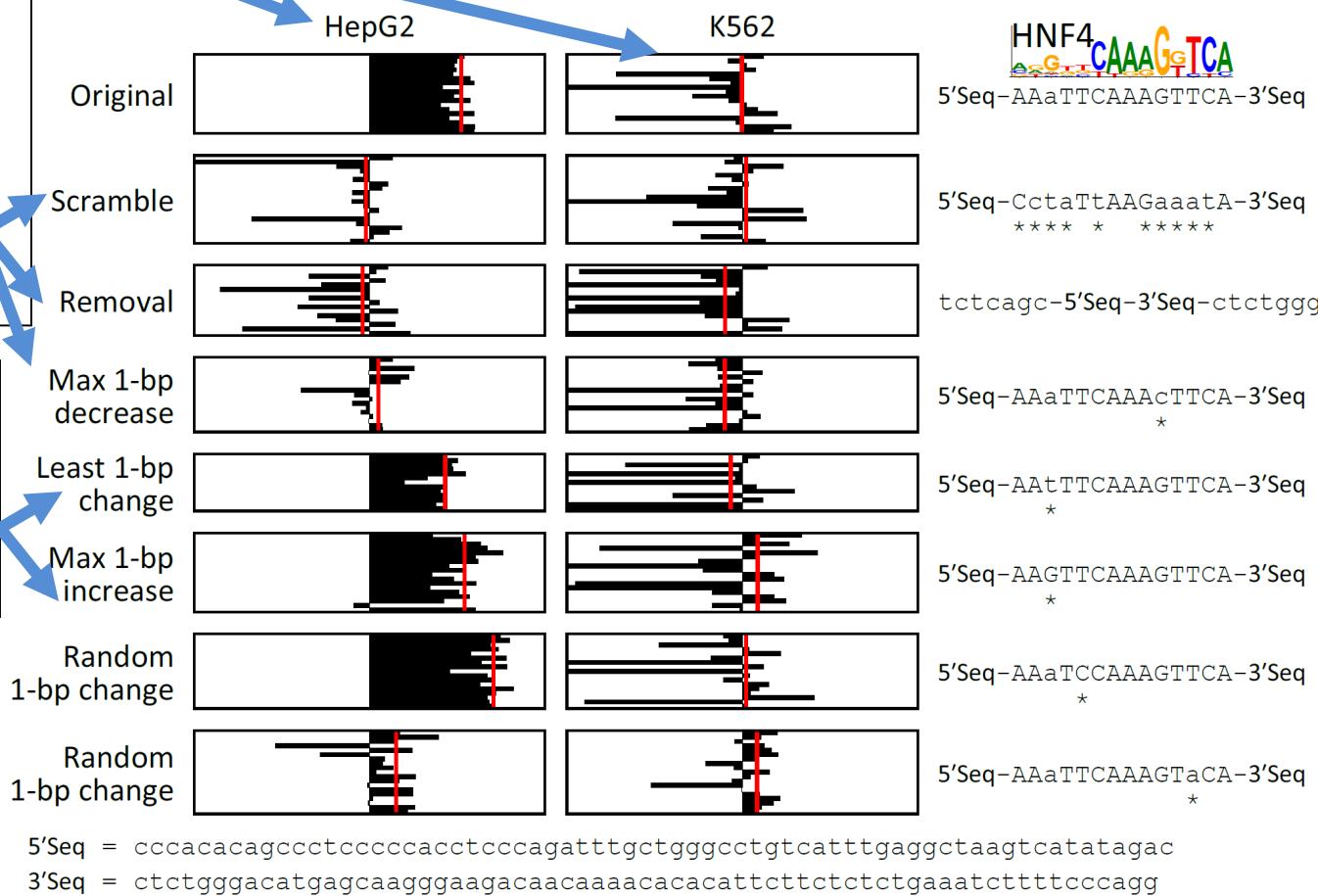
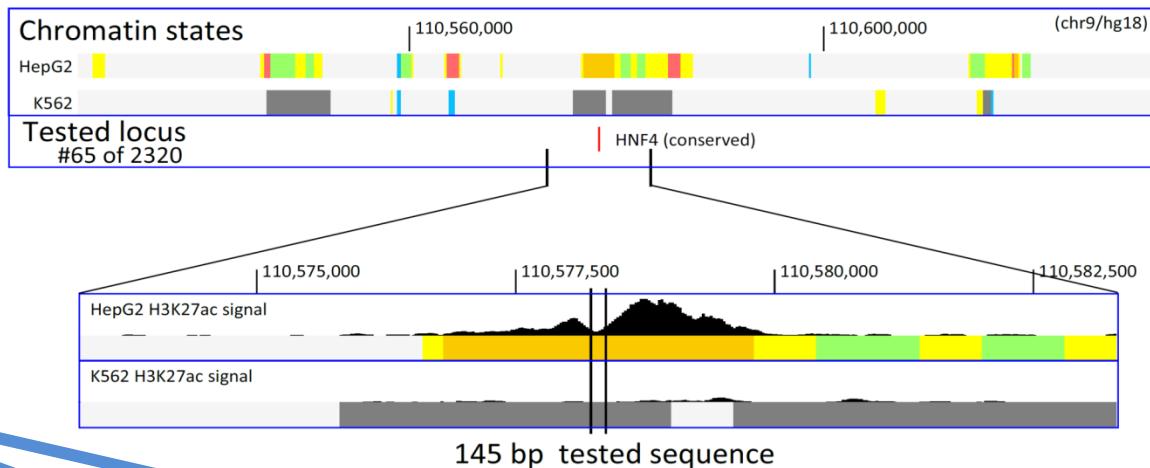
Example activator: conserved HNF4 motif match

WT expression
specific to HepG2

Motif match
disruptions reduce
expression to
background

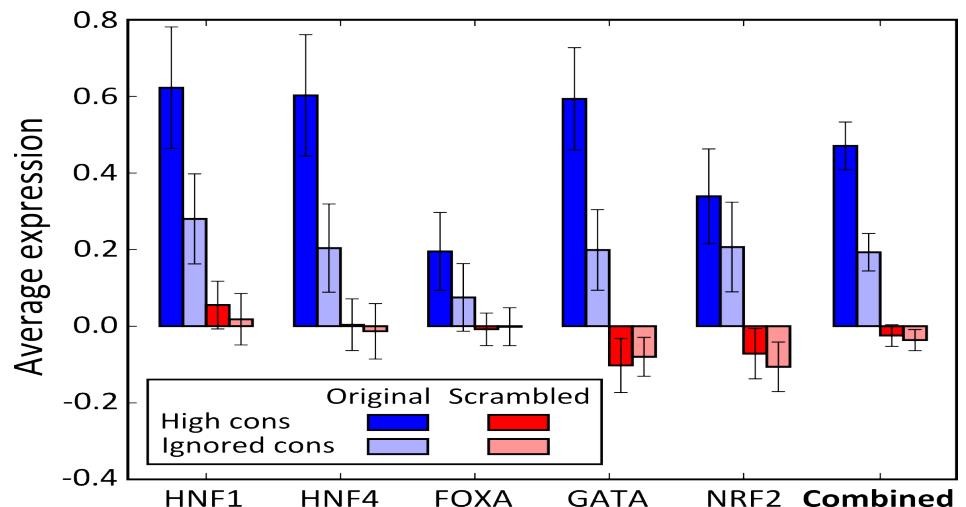
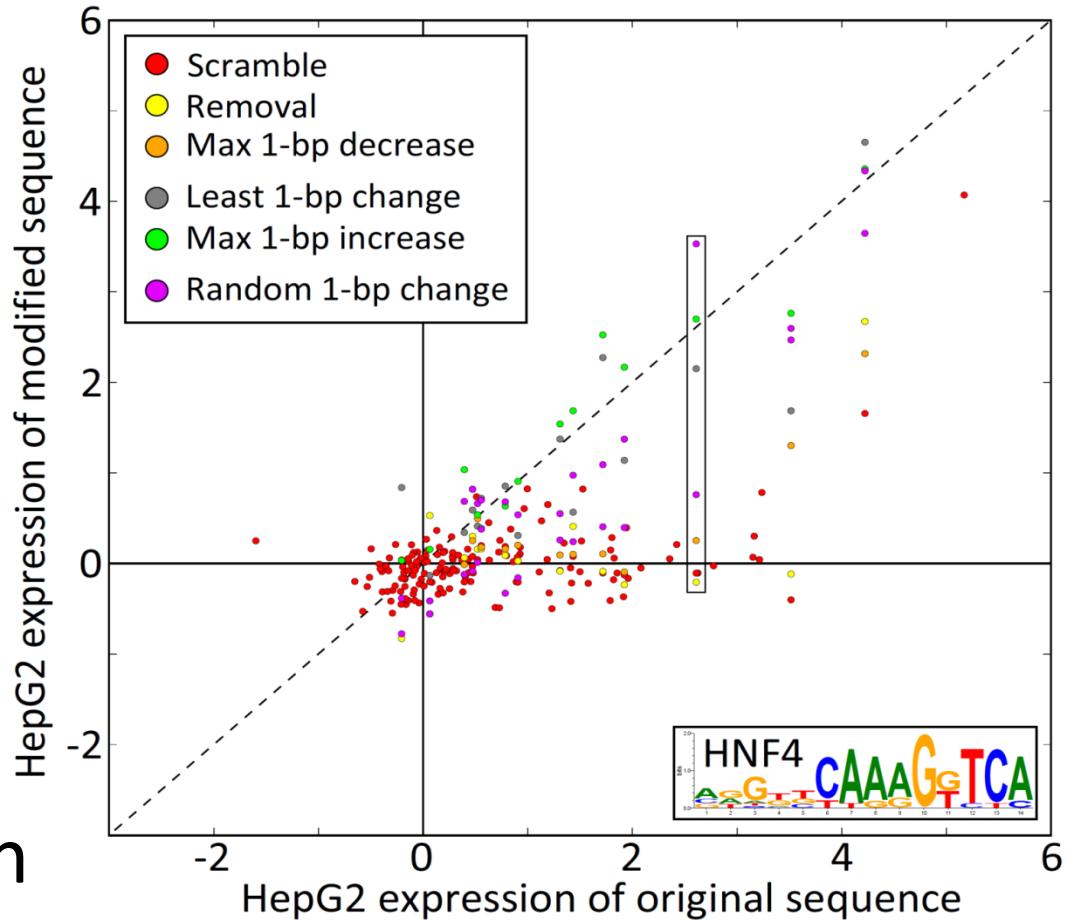
Non-disruptive
changes maintain
expression

Random changes
depend on effect
to motif match



Results hold across 2000+ enhancers

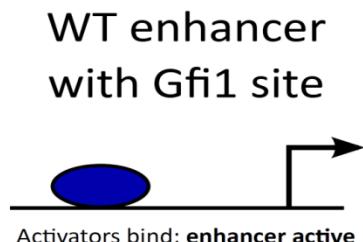
- Scramble abolishes reporter expression
- Neutral mutations show no change
- Increasing mutations show more expression
- However, only 40% show wild-type expression: context?



Repressors of HepG2 enhancer act in K562

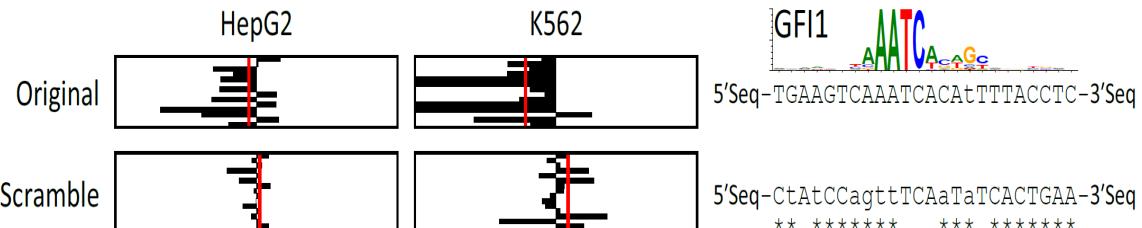
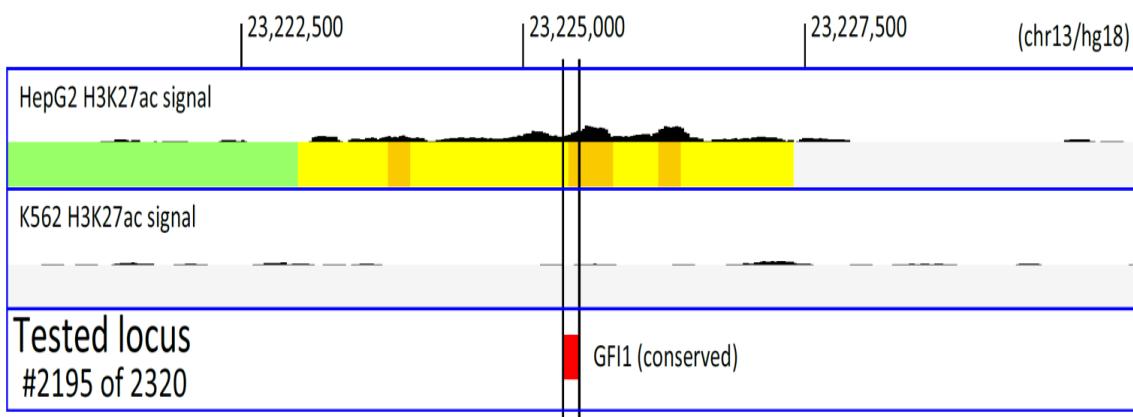
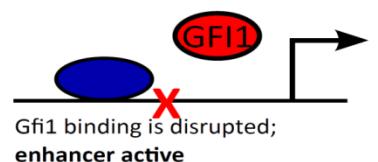
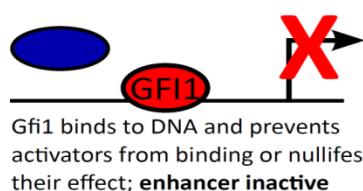
HepG2

Unknown activators expressed
Gfi1 not expressed



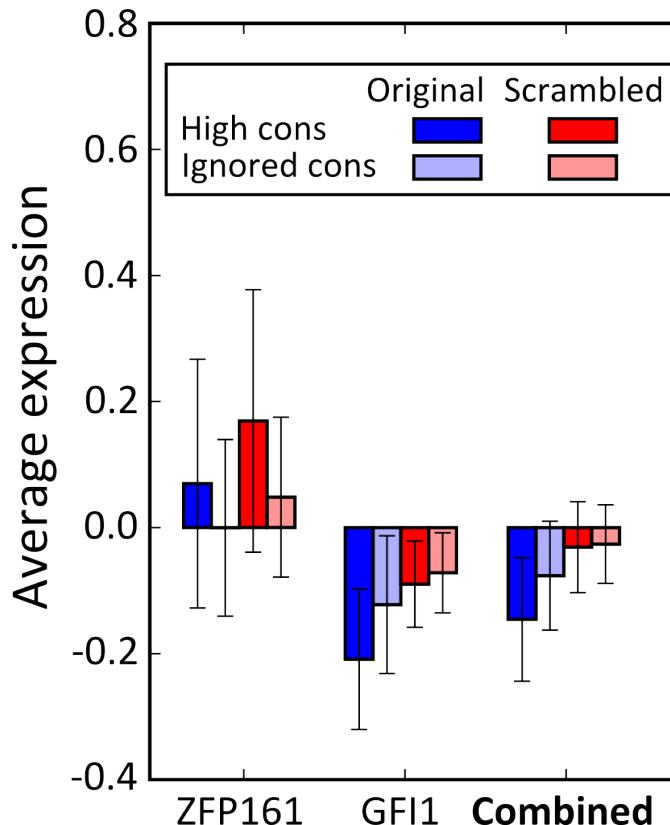
K562

Unknown activators expressed
Gfi1 expressed



5'Seq = actccatctaaaaaaaaaaaaataataataataaaaaaagaagtgttt

3'Seq = ccgcaaggaaattaagtgactttctaagtaatttagacagtagcagagtca



Research

Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay

Pouya Kheradpour,^{1,2} Jason Ernst,^{1,2,5} Alexandre Melnikov,² Peter Rogov,² Li Wang,² Xiaolan Zhang,² Jessica Alston,^{2,3} Tarjei S. Mikkelsen,^{2,4} and Manolis Kellis^{1,2,6}

**Repressor disruption
→ aberrant expression
in opposite cell types**

High-throughput expts: MPRA, SHARPR, HiDRA, Perturb-seq

1. High-throughput synthesis: Massively Parallel Reporter Assays (MPRA)

- MPRA technology: in vitro synthesis, reporter design, transfection
- 1 site X deep: single-base dissection, combinatorial changes, high-res map
- 2k sites x shallow: motif-guided perturbations, activators/repressors

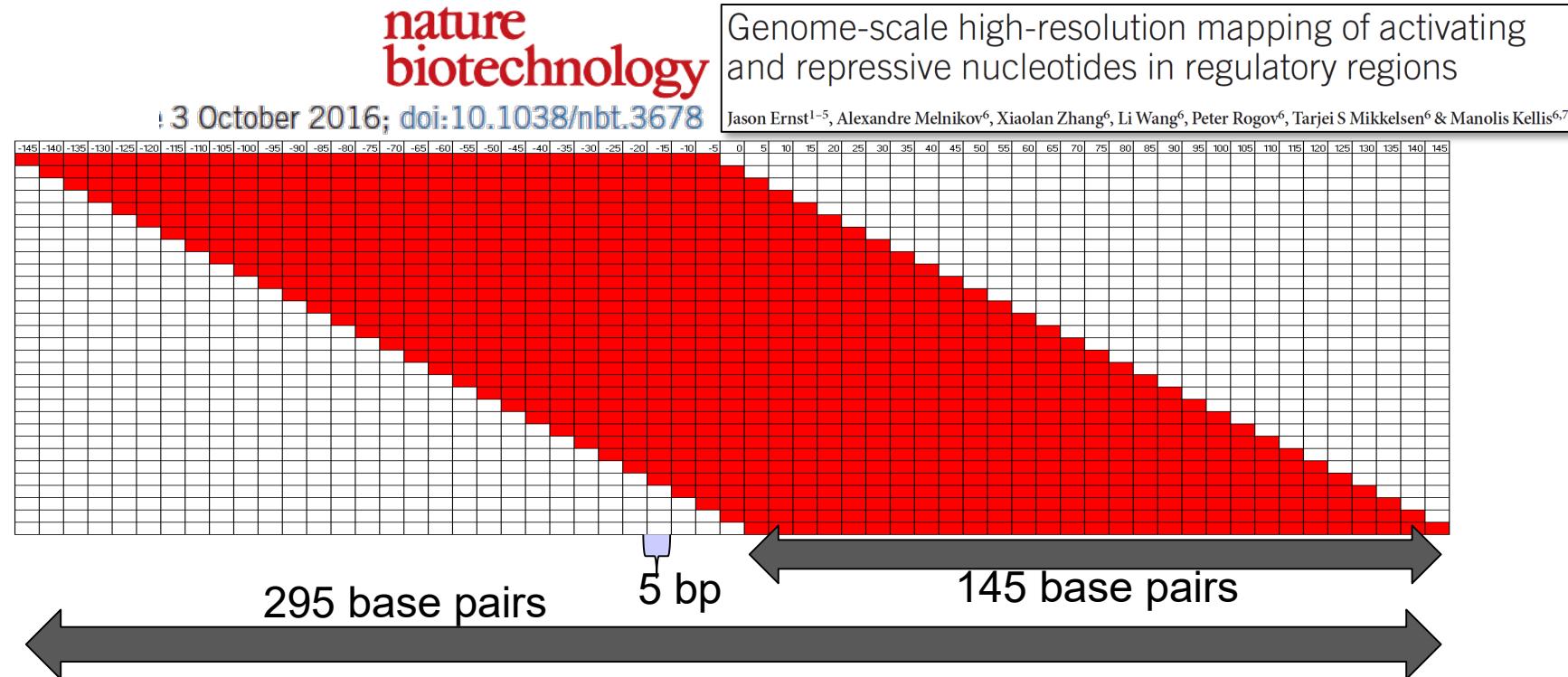
2. Next-generation applications of MPRA + STARR-Seq, SHARPR, HiDRA

- 10k x deep: high-resolution dissection by tiling + deconvolution (Sharpr)
- 10M x deep: HiDRA, no synthesis (STARR), targeted capture (ATAC)

3. Endogenous genome editing: cutting and template-based repair

- Endonucleases: TALENs, ZNFs. Repair pathways. Diversity of outcomes.
- CRISPR-Cas9: origins, discovery, optimization, base-editing
- Next-generation: Perturb-seq, multiplexing, delivery, applications

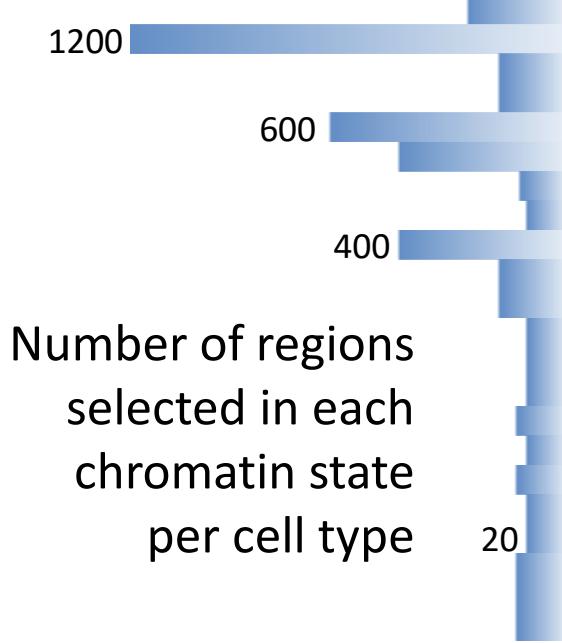
Experimental approach: Tiling Design



Advantages:

- 7860 unique regulatory elements can be tested in single experiment (assuming 243,660 total sequences)
- Coverage of 295 base pairs at each regulatory element tested
- Robustness to noise through overlaps
- Information to recover at high resolution regulatory bases
- Results not dependent on mutations

Selection of regions



State	H3K36me3	H4K20me1	Control	FAIRE	POL2	H3K27ac	H3K9ac	H3K4me3	H3K4me2	H3K4me1	DNase_UW	DNase_DUKE	CTCF	H3K27me3
1_Tss	1.8	6.4	3.0	33.1	98.5	96.1	99.0	99.9	95.9	43.0	85.9	92.2	36.1	4.9
2_TssF	21.4	23.9	0.7	4.4	55.1	72.4	86.5	97.8	99.4	73.3	24.4	15.4	3.9	1.7
3_PromF	3.3	6.2	0.2	0.6	7.5	2.3	8.2	32.2	89.1	43.8	4.8	4.2	1.7	2.2
4_PromP	0.1	7.1	0.5	7.7	38.4	2.2	37.3	88.0	96.7	61.5	51.3	86.9	31.7	42.1
5_Enh	11.4	7.4	4.0	49.5	68.9	98.4	72.7	33.9	97.1	98.2	89.0	67.3	17.2	0.1
6_EnhF	5.0	2.3	0.6	3.7	10.8	94.3	28.7	1.0	53.9	93.4	20.0	8.1	2.3	0.1
7_EnhWF	0.9	2.2	0.2	0.8	1.4	0.6	1.3	0.0	7.0	88.4	4.4	3.7	1.4	0.4
8_EnhW	0.2	8.3	0.7	16.6	18.9	10.8	6.3	0.6	61.6	93.6	62.6	59.9	17.0	0.7
9_DNaseU	0.4	1.4	0.1	2.8	3.0	1.6	0.5	0.1	2.7	9.1	95.1	13.6	0.3	0.3
10_DNaseD	1.1	3.5	0.2	0.3	1.6	0.5	0.7	0.1	0.3	3.1	2.5	69.5	0.9	0.3
11_FaireW	1.5	1.9	7.9	22.5	0.8	10.1	5.6	1.7	0.2	0.3	0.5	1.5	0.6	2.0
12_CtcfO	5.9	8.8	1.0	34.6	18.6	1.8	1.4	0.3	10.3	16.7	71.4	86.9	99.8	2.0
13_Ctcf	0.3	2.1	0.1	0.5	0.3	0.1	0.2	0.0	0.7	1.4	6.0	16.1	99.1	1.7
14_Gen5'	70.4	35.9	0.7	3.2	52.3	26.2	9.1	1.2	23.4	83.6	9.5	11.6	3.6	0.0
15_Elon	93.2	4.2	0.3	0.7	1.4	0.3	0.2	0.0	0.1	0.3	0.4	1.4	1.0	0.0
16_ElonW	0.0	0.4	0.0	0.3	0.8	0.1	0.0	0.0	0.0	0.1	0.1	0.2	0.0	0.0
17_Gen3'	93.1	17.6	0.4	1.2	83.5	1.2	0.6	0.1	0.1	0.7	2.3	7.5	2.8	0.0
18_Pol2	0.7	5.6	0.1	0.7	82.6	0.9	0.6	0.0	0.1	2.2	2.7	7.1	1.2	0.3
19_H4K20	3.4	85.7	0.3	1.0	2.1	0.2	0.3	0.0	0.2	2.1	0.4	1.4	0.5	0.6
20_ReprD	0.5	15.5	0.5	1.4	2.0	0.5	0.6	1.0	11.4	20.7	4.4	47.9	11.7	90.4
21_Repr	0.0	2.1	0.2	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	1.1	0.1	89.8
22_ReprW	0.0	0.2	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0
23_Low	0.1	0.0	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0
24_Quijes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25_Art	77.0	73.0	40.8	70.2	59.2	34.7	38.3	25.3	31.8	52.4	40.7	63.5	64.5	47.5

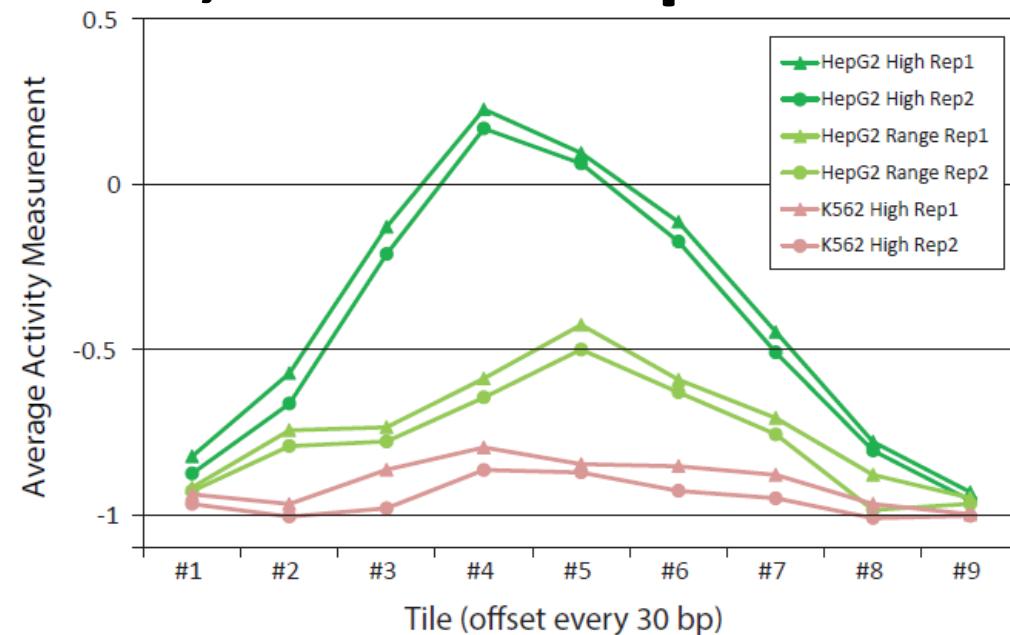
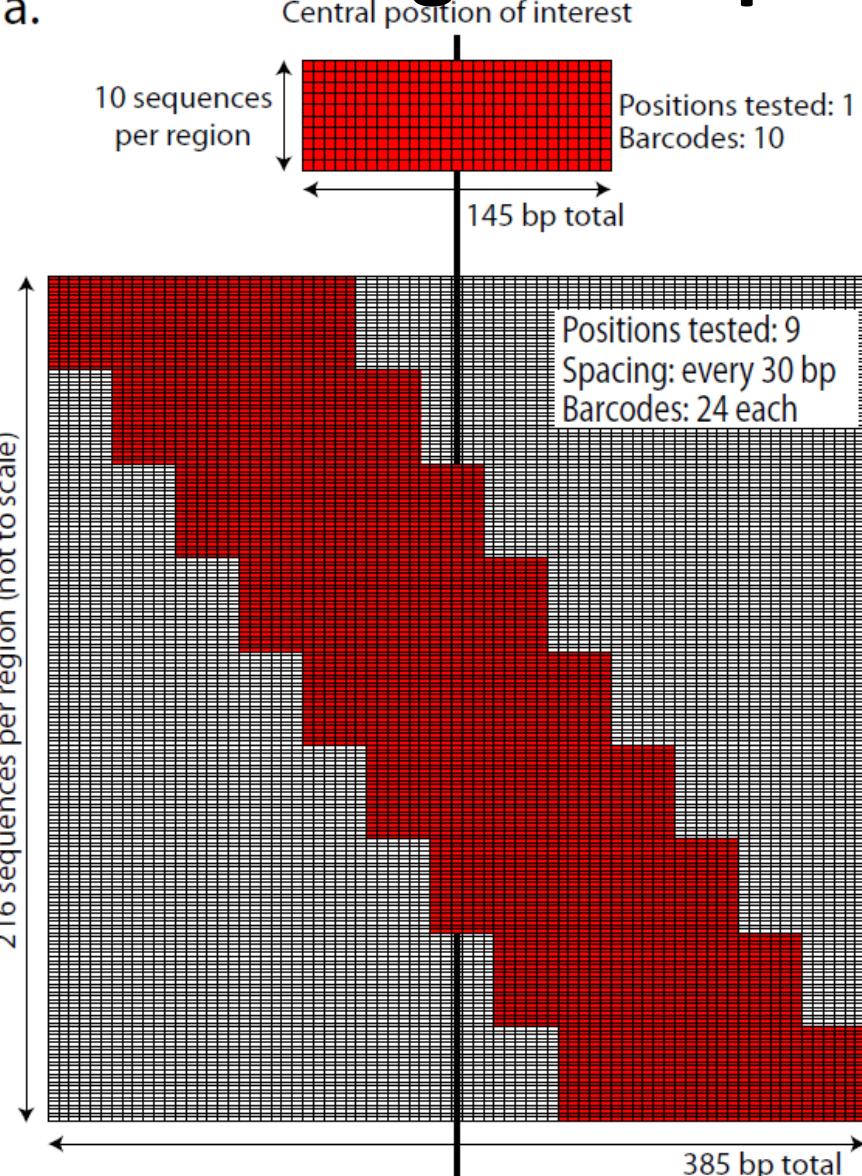
- 15,720 selections
- Centered on DNase peaks
- Split across four ENCODE cell types (K562, HepG2, H1hesc, Huvec)
- Split across chromatin states (ChromHMM states from Hoffman, Ernst, et al NAR 2012)

16 Experiments:

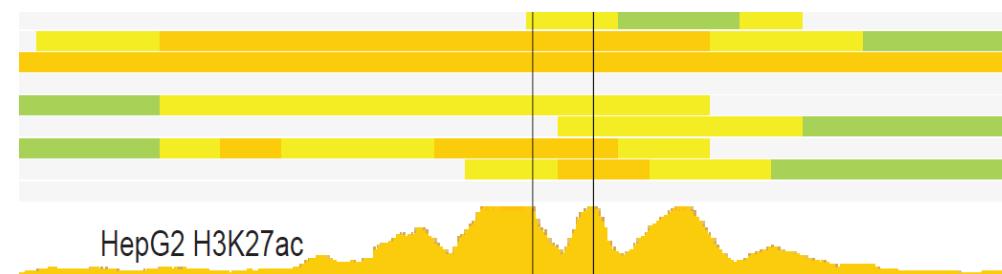
- 2 Cell Types (HepG2, K562)
 - x 2 Promoters (minimal, SV40)
 - x 2 sets of sequences
 - x 2 replicates

Pilot design: 30-bp offsets, each 20 replicates

a.



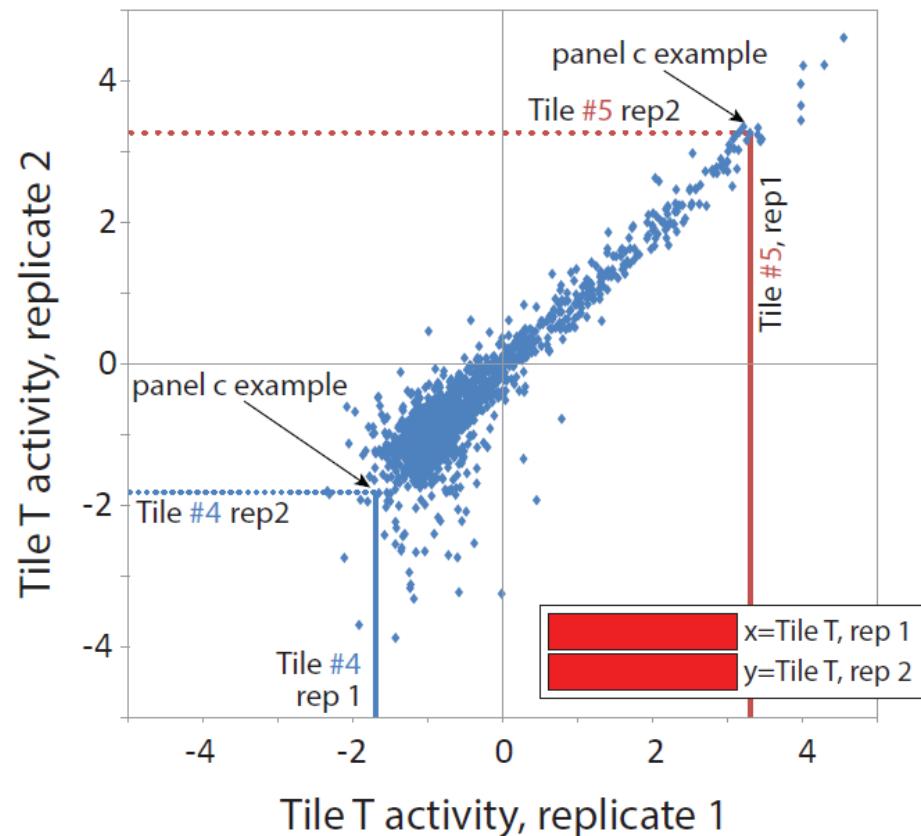
*Centers of selected regions
show strongest activity*



*Chromatin dips in matched cell show
strongest activity*

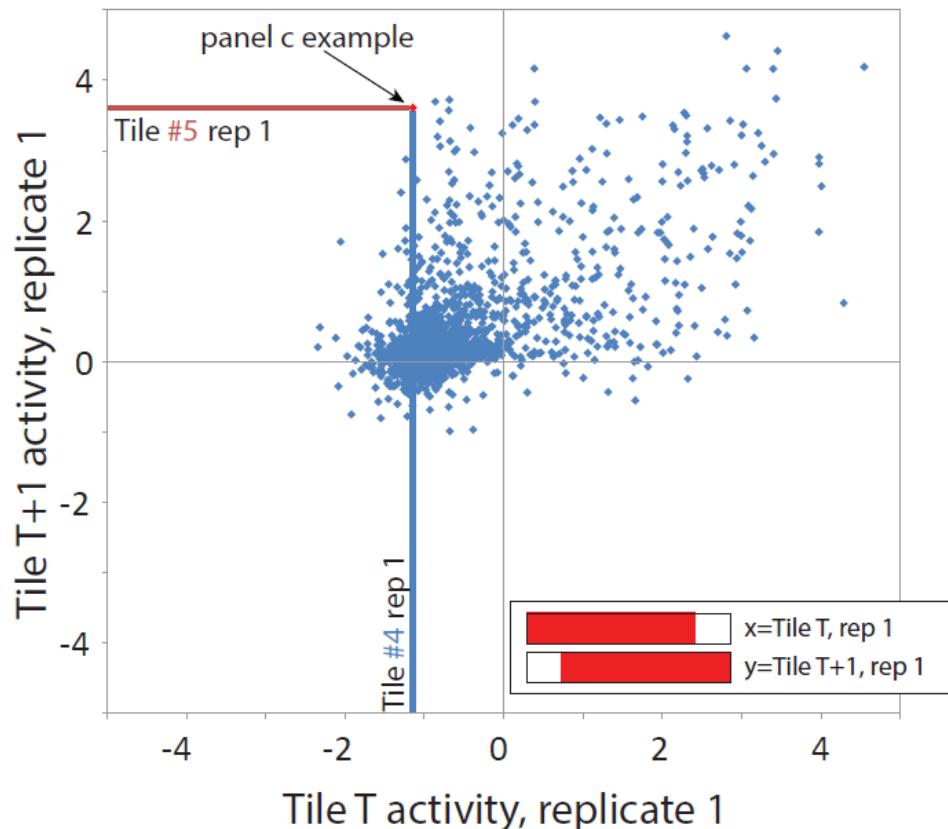
An offset of 30-bp can make a big difference

Comparison of replicates for the same tile



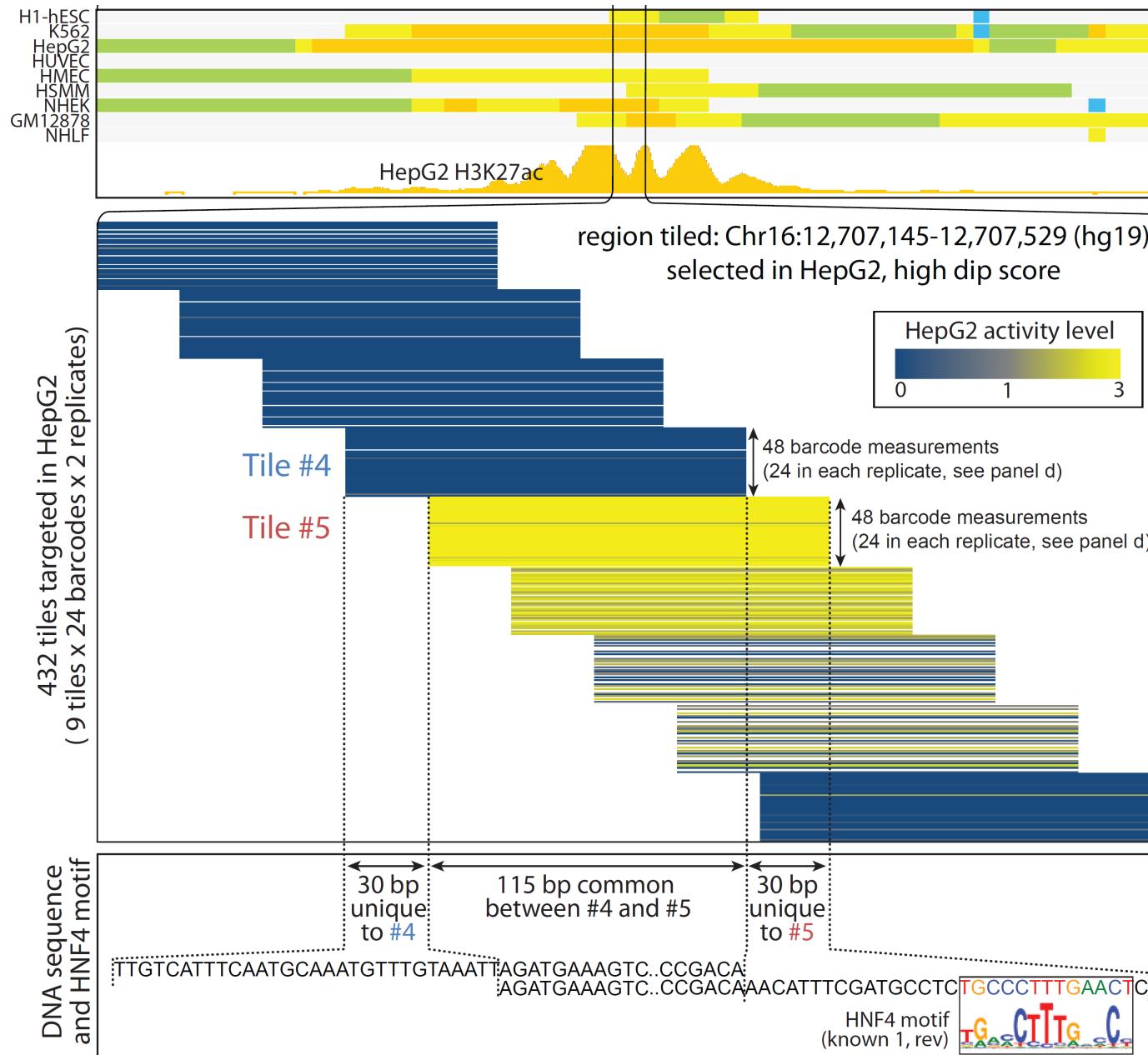
*Replicates of same tile
are highly consistent*

Comparison of consecutive tiles for same replicate



*Consecutive tiles
can differ greatly*

Consecutive tile diffs due to motif inclusion/exclusion



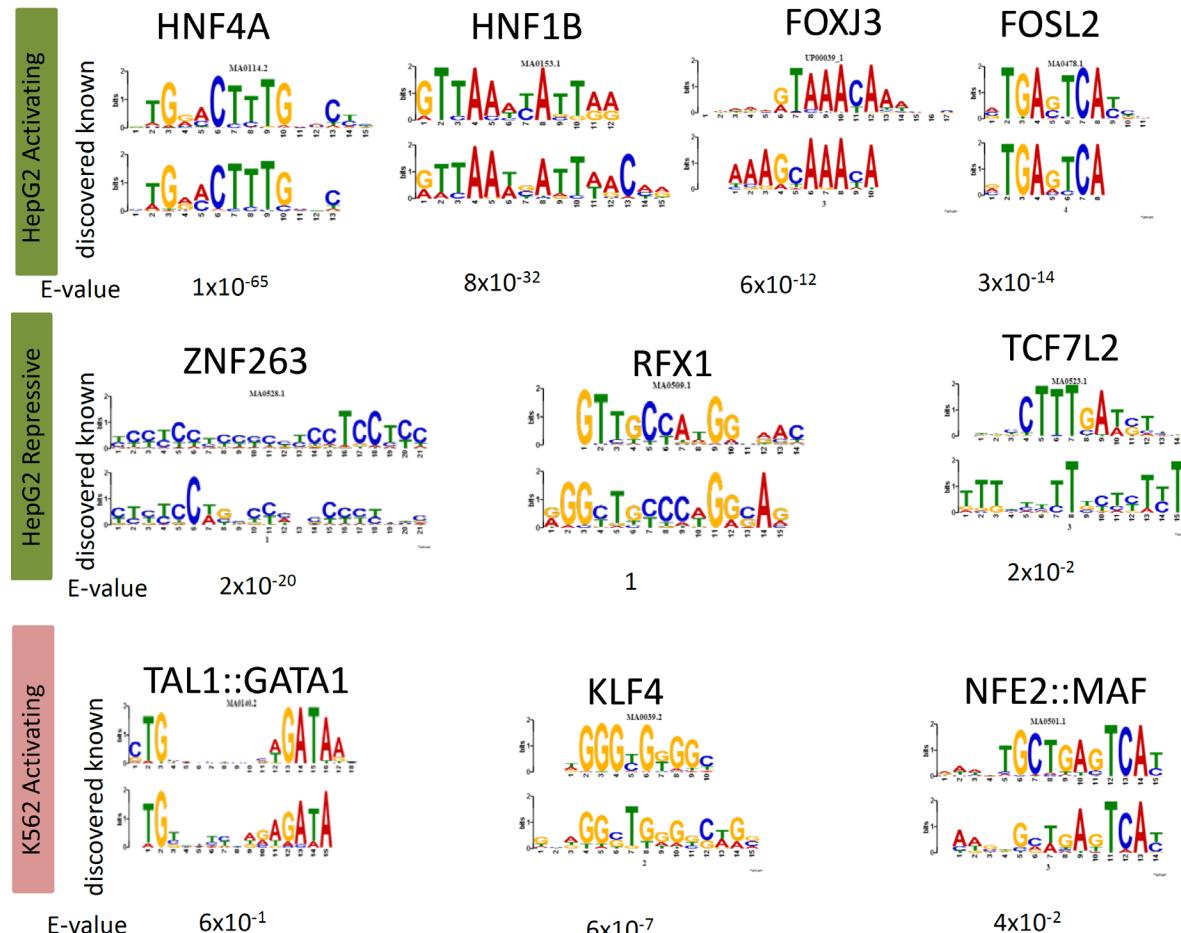
- Inclusion/exclusion of 30-bp intervals
- Akin to systematic disruption
- Increase resolution from tile (145bp) to offset (30bp)

Next:

- Use to discover motifs?
- Further increase resolution?

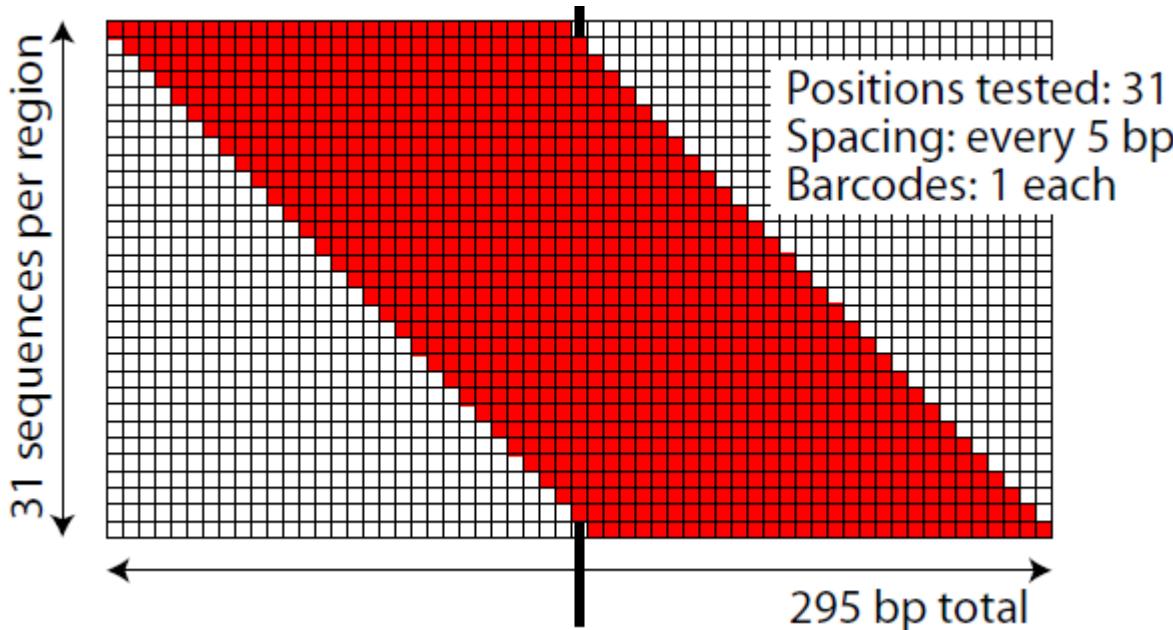
Consecutive tile differences allow motif discovery

	HepG2 Activating	HepG2 Repressing	K562 Activating	K562 Repressing
GATA_known14	1.0	1.0	6.4	1.0
LMO2_2	1.0	1.0	4.4	1.0
TAL1_disc1	1.0	1.0	3.0	1.0
CPHX_1	1.0	1.0	2.7	1.0
JDP2_2	1.0	1.0	2.4	1.0
NFE2L2_3	1.0	1.0	2.0	1.0
HNF4_known9	4.6	0.3	1.0	1.0
NR2F6_2	3.7	1.0	1.0	1.0
HNF1B_4	3.5	0.6	1.0	1.0
RXRA_known10	3.4	0.9	1.0	1.0
PPARA_4	3.1	1.0	1.0	1.0
HNF1A_4	2.8	1.0	1.0	1.0
HNF1_4	2.6	1.0	1.0	1.0
TCF7L2_disc1	2.4	1.0	1.0	1.0
AP1_known4	2.3	1.0	1.6	1.0
SMARC_disc1	2.2	1.0	1.1	1.0
CEBPB_known1	2.1	1.0	1.0	1.0
TLX2_2	2.0	1.0	1.0	1.0
FOXJ2_4	2.0	1.0	1.0	1.0
EGR1_disc2	2.0	1.0	1.0	1.0
RFX2_3	1.0	3.0	1.0	1.0
RFX5_known9	1.0	2.9	1.0	1.0

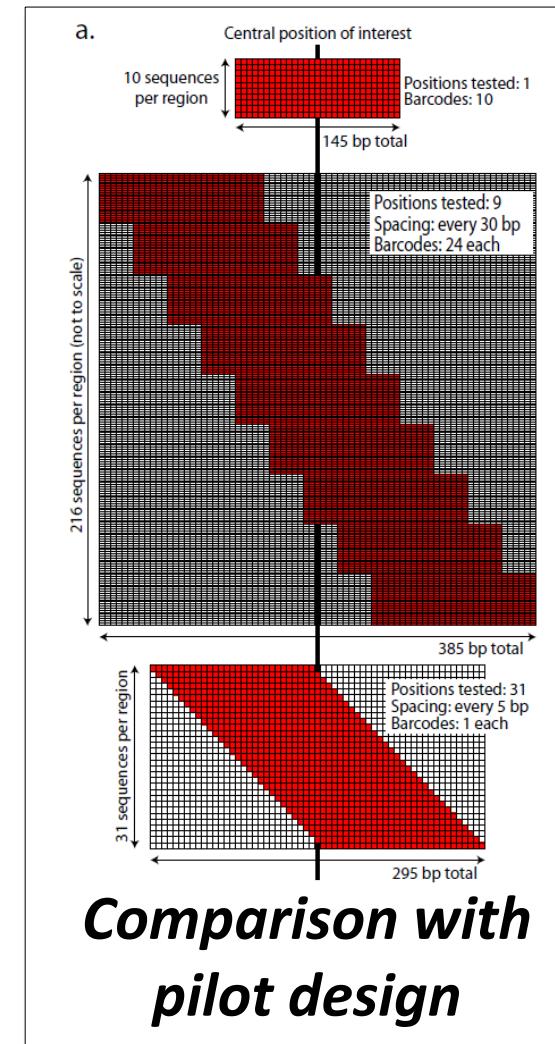


- Increased resolution allows testing of only 30-bp intervals
- De novo* discovered motifs match known motifs
- Discovery distinguishes activating vs. repressive factors

Scale-up: 5-bp tiles for nucleotide-level resolution

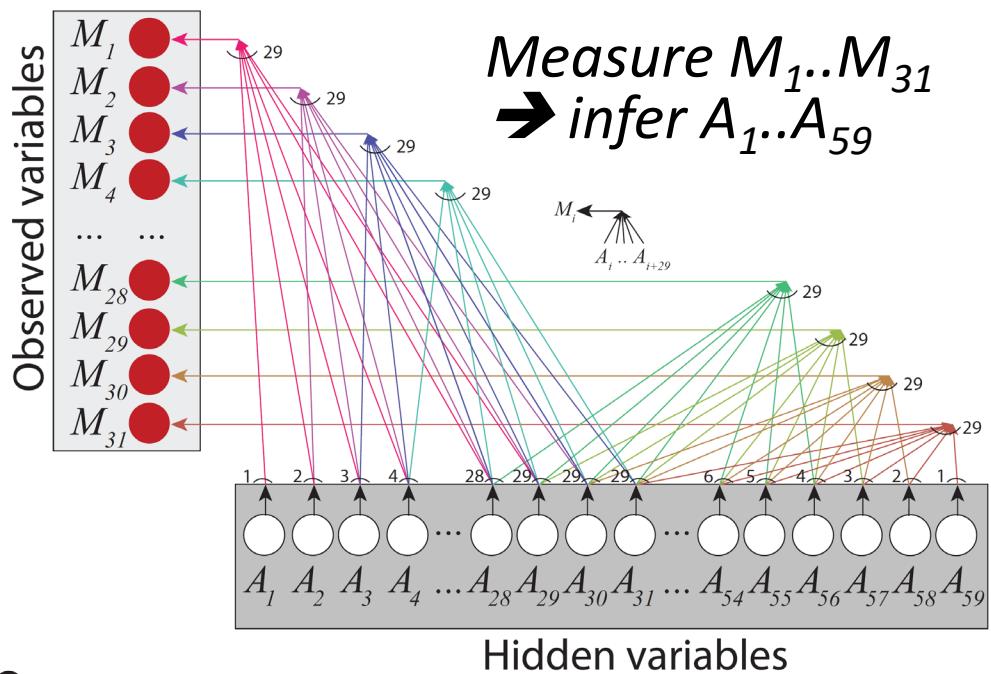
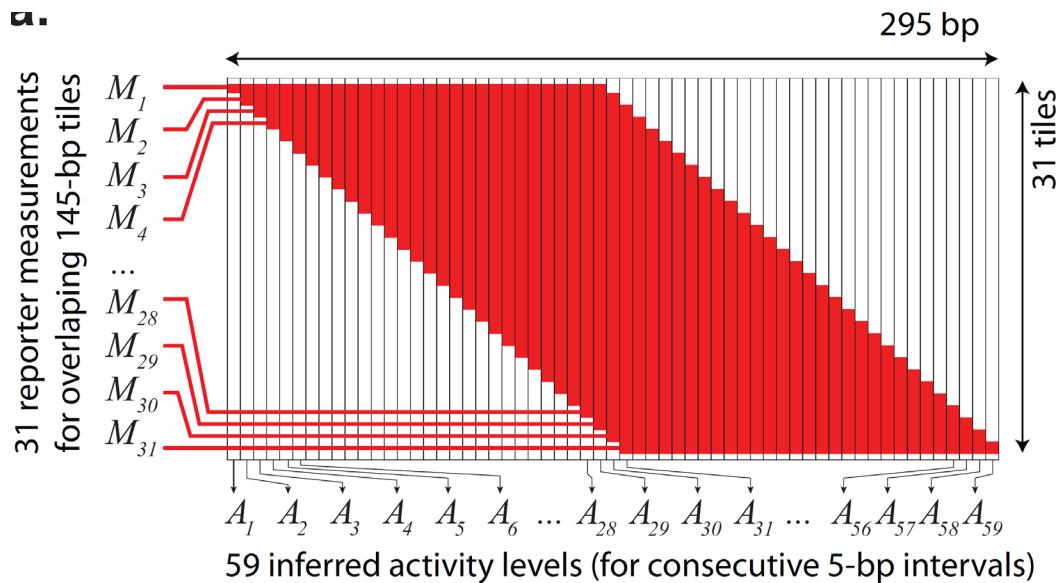


- *Tile 295-bp regions (vs. 385)*
- *Tile @ 5-bp offsets (vs @30)*
- *Center on DNase peaks (vs dips)*
- *Single barcode / offset (vs 24)*
- *244K spot array (vs 54K)*
- *25 chromatin states (vs. Enh)*
- *Both minP and SV40 promoters*



Computational model to infer per-nucleotide activity

a.



- Jointly estimate all elements jointly
- Adjust hidden variables to best match observed variables
- Hidden: 5-bp activity levels
- Observed: 145-bp reporter construct
- Interpolate to get single-nucleotide activity

Distributions used

Joint samples from multivariate normal $\begin{bmatrix} A_{r,t} \\ M_{r,t} \end{bmatrix} \sim N(\mu_{x_{r,t}}, \Sigma_{x_{r,t}})$

True activity normally distributed $A_{r,t,k} \sim N(\mu_{a_r}, \sigma_a^2) \quad k = 1, \dots, K$

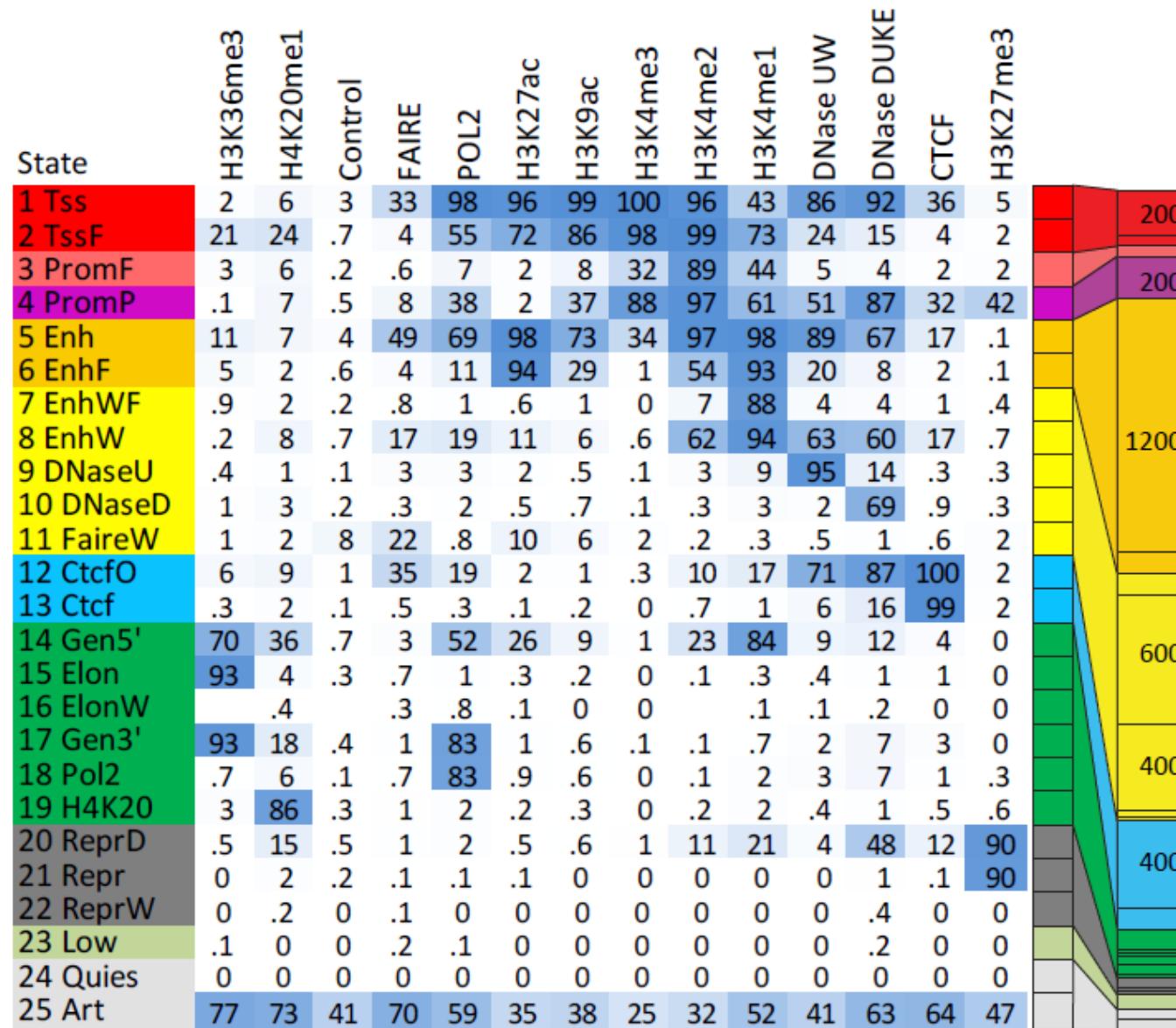
Observation is mean of true signals $\mu_{a_r} = \frac{1}{|\widehat{M}_r|} \sum_{m \in \widehat{M}_r} m$

Measurements normally distributed $M_{r,t,j} \sim N(\mu_{m_{r,t,j}}, \sigma_{m_r}^2)$

Mean set to empirical mean $\mu_{m_{r,t,j}} = \frac{1}{N} \sum_{i=0}^{N-1} A_{r,t,j+i}$

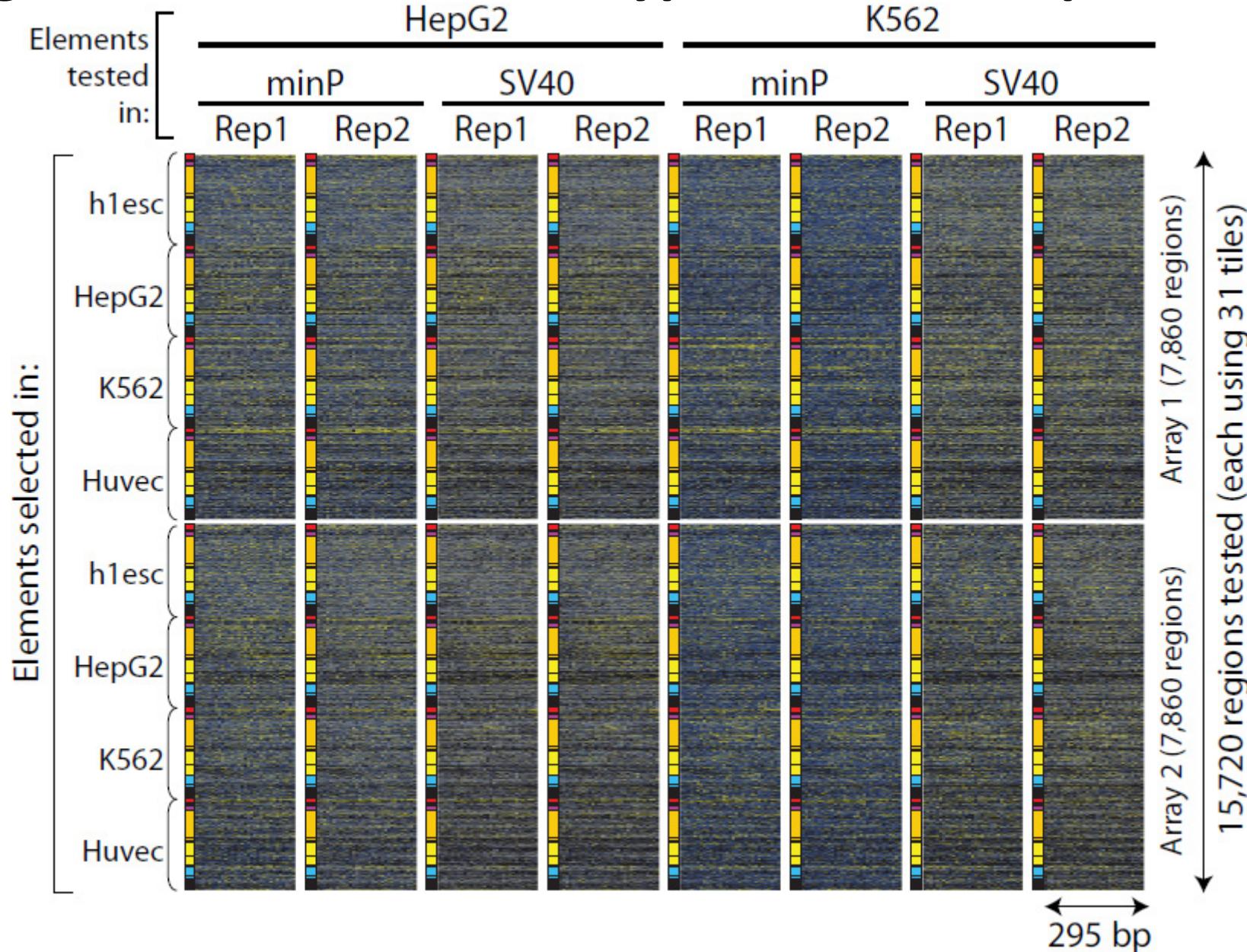
Variance set to empirical variance $\sigma_{m_r}^2 = \frac{1}{|\widehat{M}_r|} \sum_{m \in \widehat{M}_r} (m - \frac{1}{|\widehat{M}_r|} \sum_{m \in \widehat{M}_r} m)^2$

Chromatin state vs. reporter activity of DNase elmts



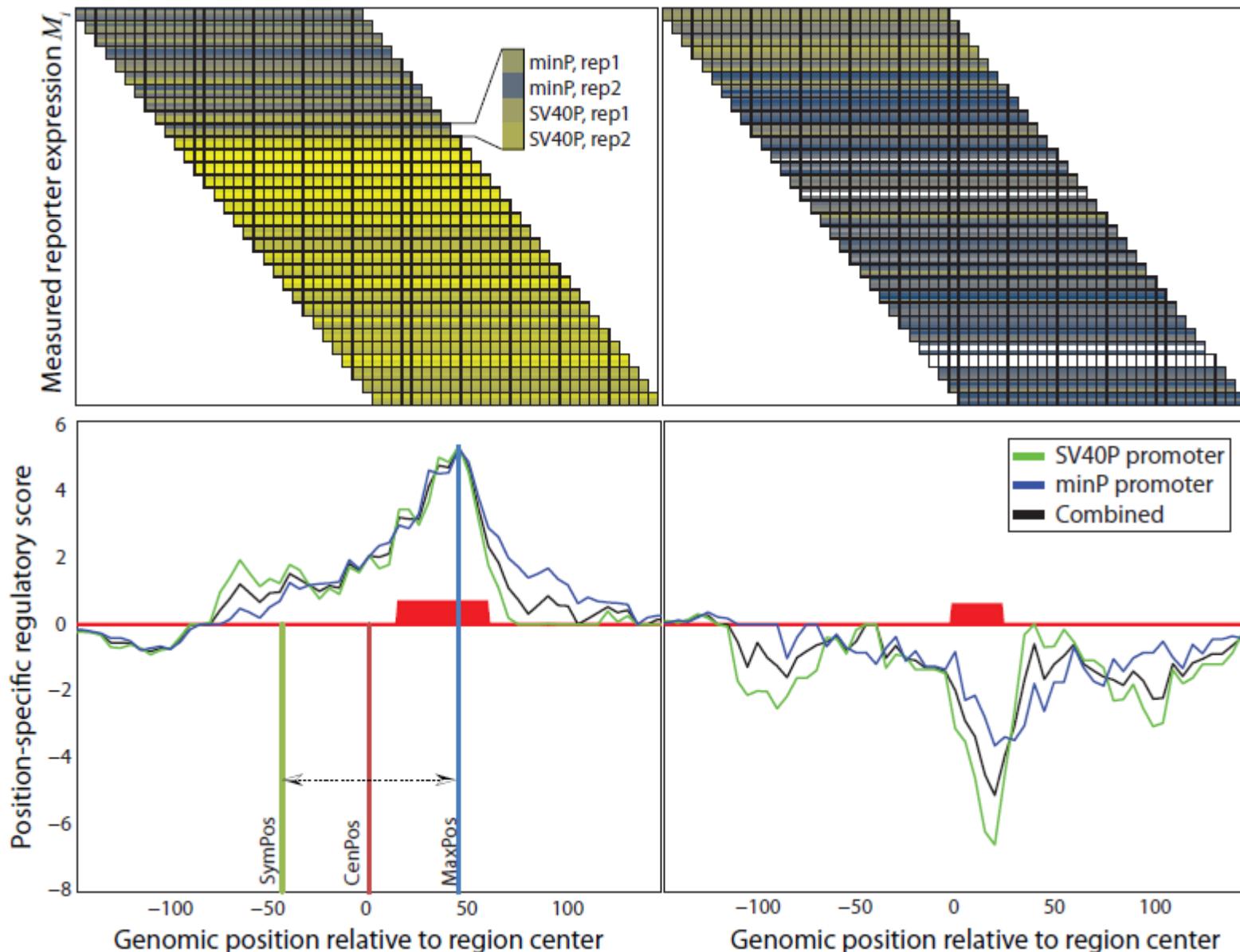
Select 15,720 DNase elements across all 25 chromatin states

Regions selected in 4 cell types, tiled in HepG2,K562



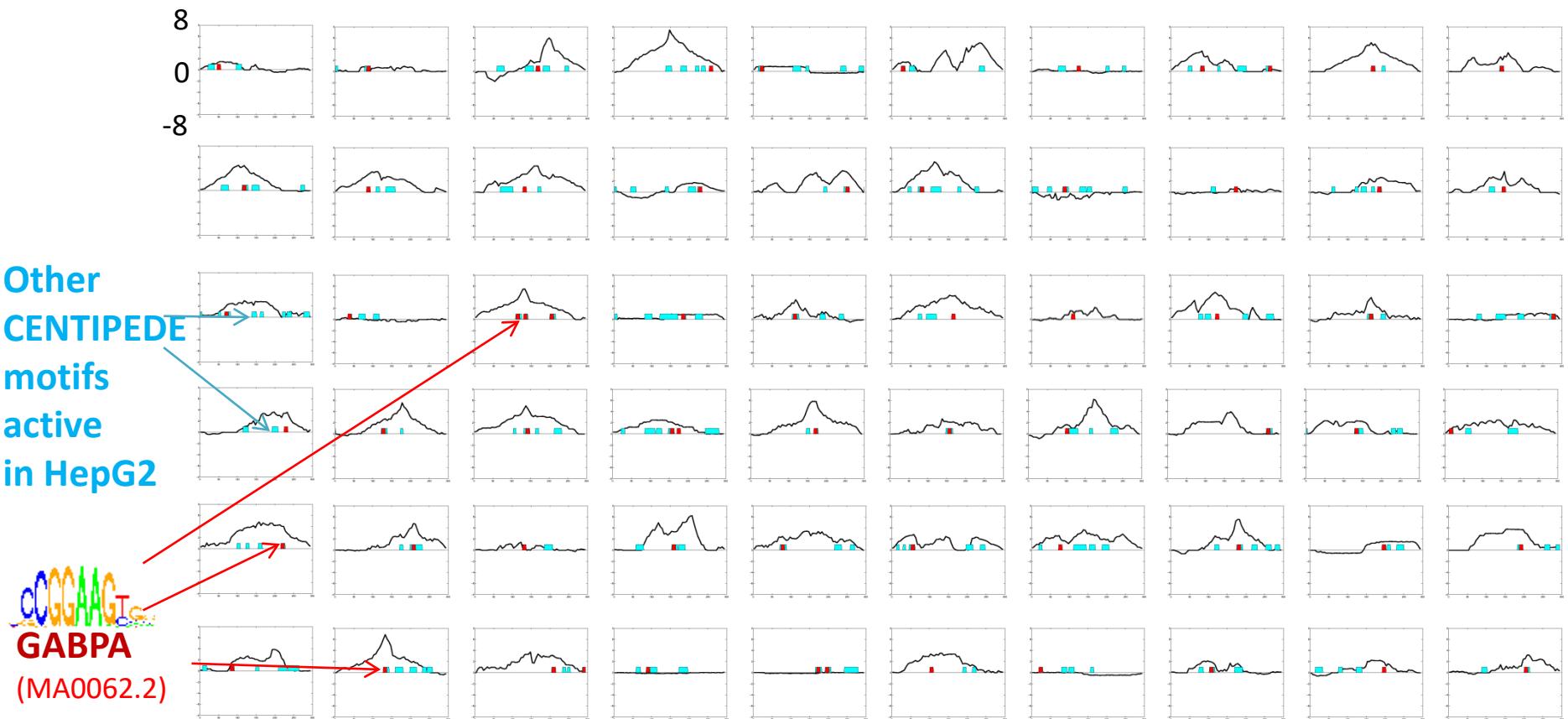
5,720 regions x 31 offsets x 2 promoters x 2 reps x 2 cell lines

Examples of tiling data deconvolution



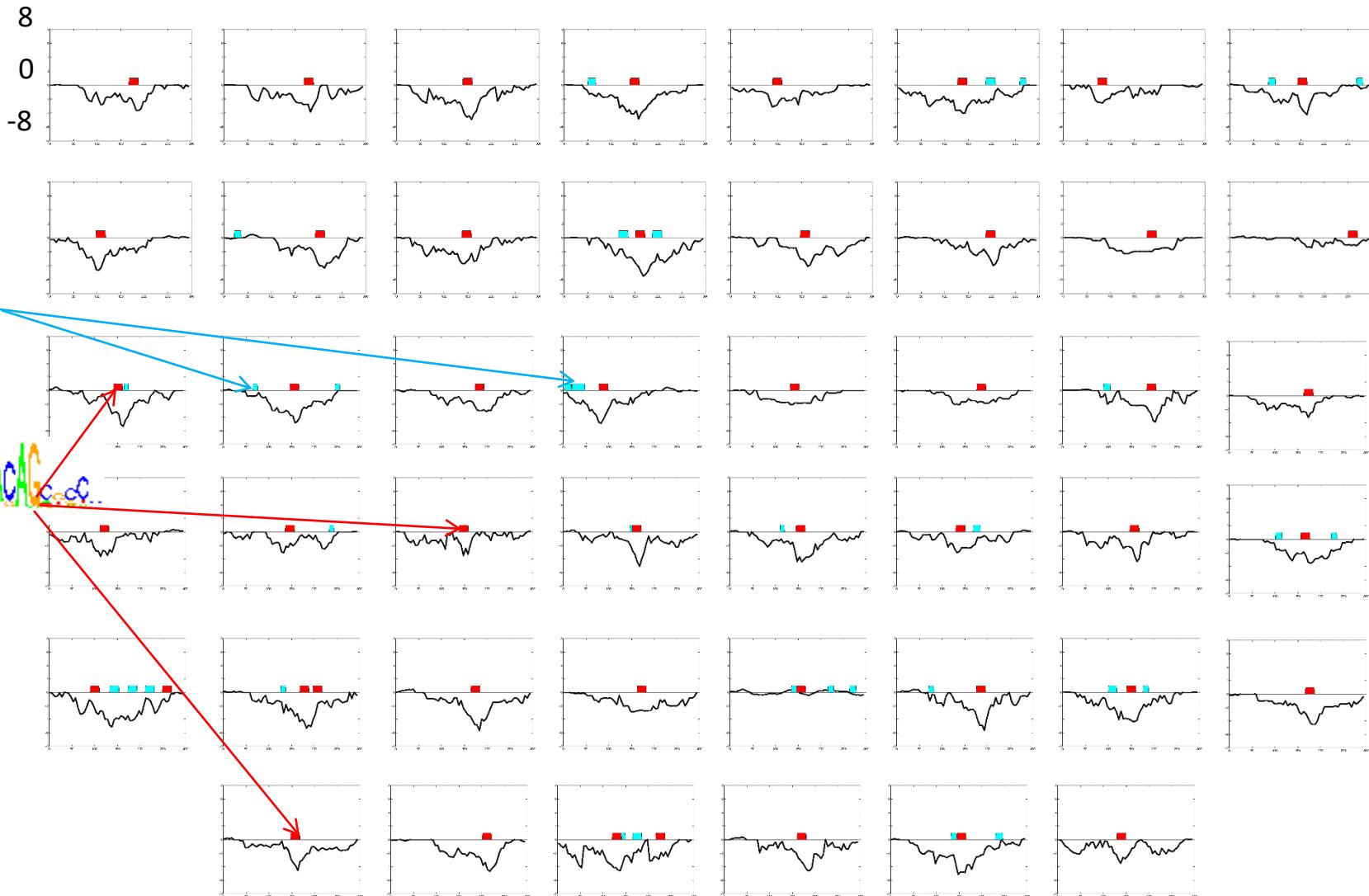
Detect activating/repressive elements at high resolution

Deconvolved regulatory signal vs. activator motif



60 sites containing GABPA HepG2 motifs predicted by CENTIPEDE

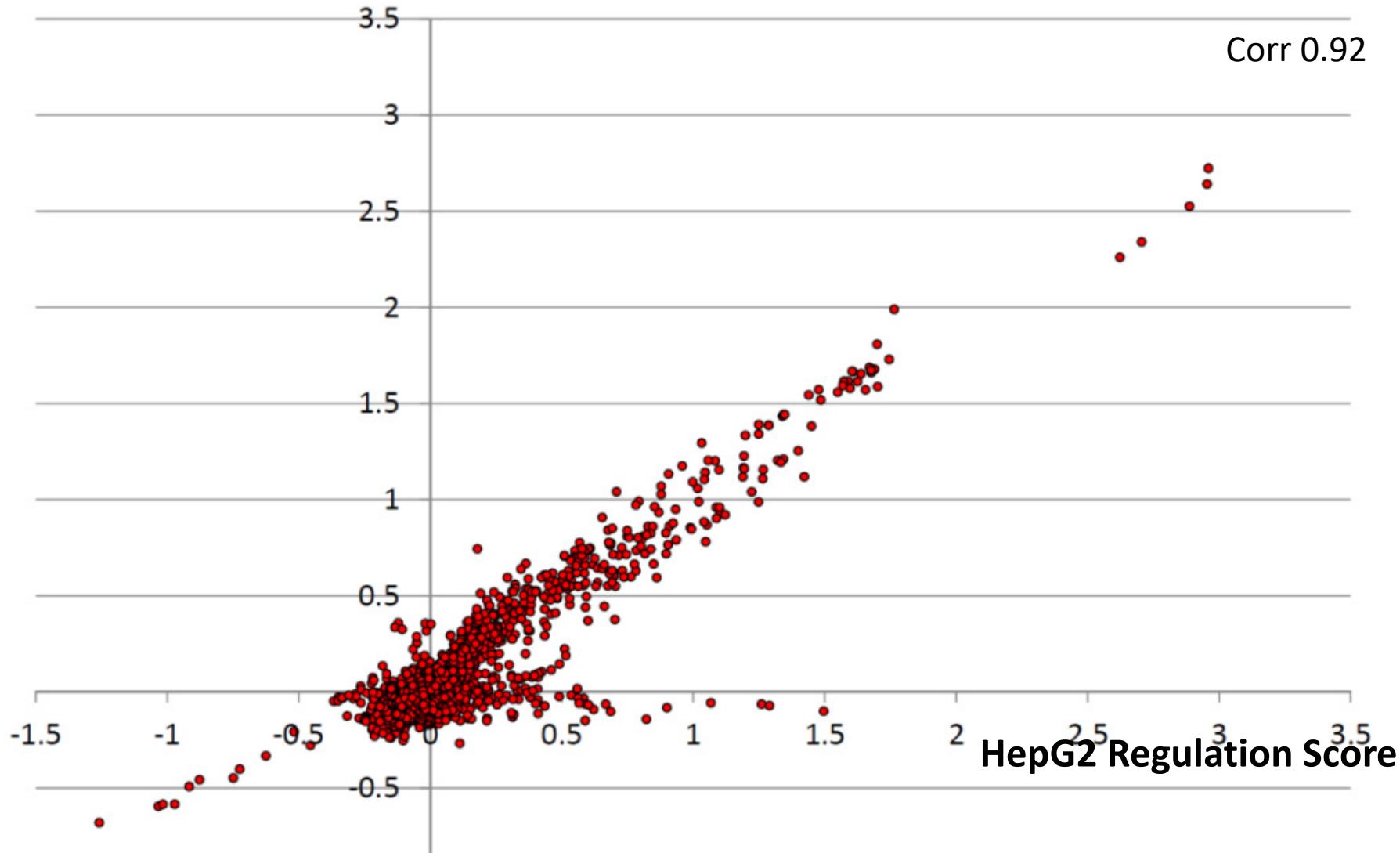
Deconvolved regulatory signal vs. repressor motif



46 sites containing NRSF HepG2 motifs predicted by CENTIPEDE

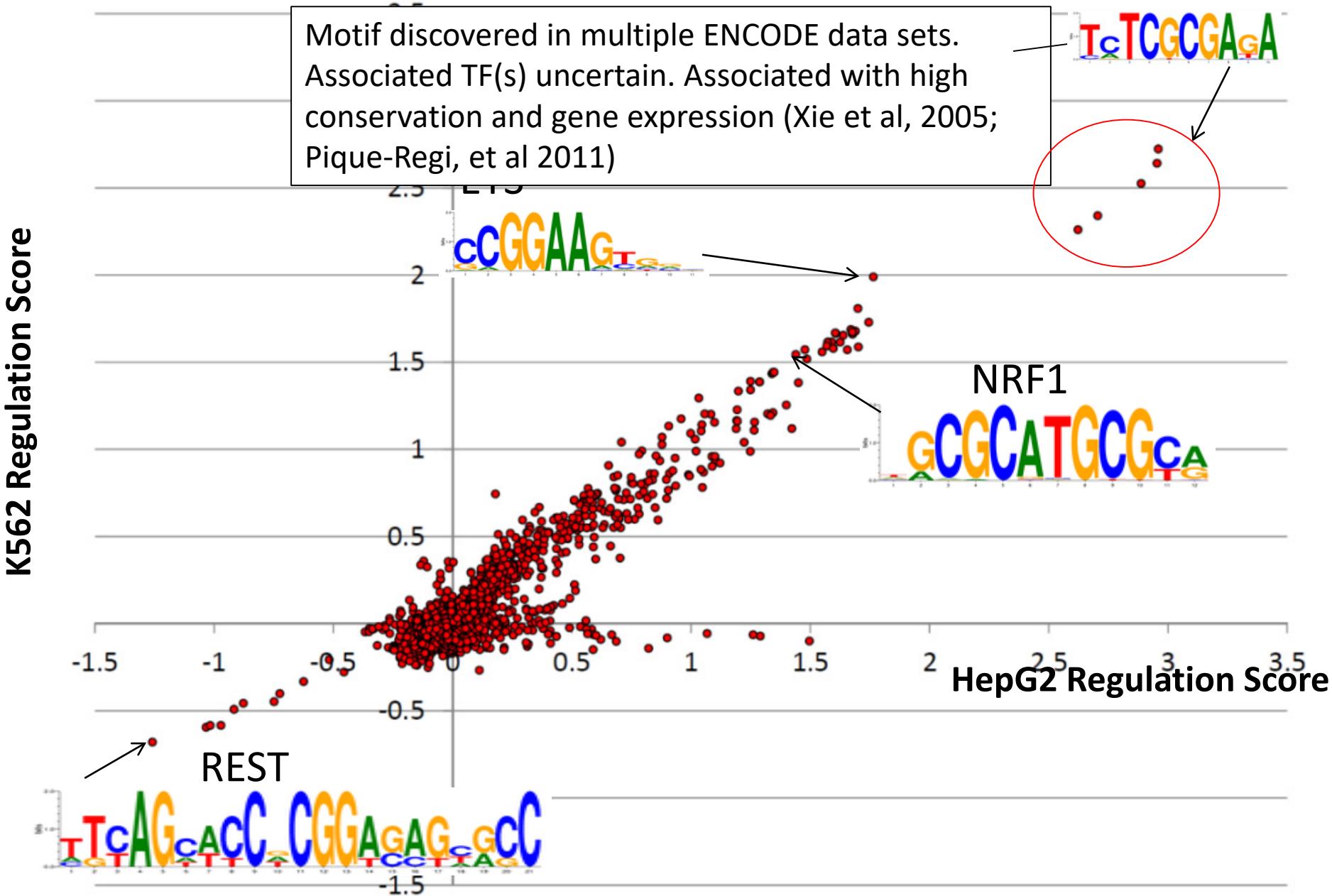
Aggregate Motif Score Highly Correlated between K562 and HepG2

K562 Regulation Score



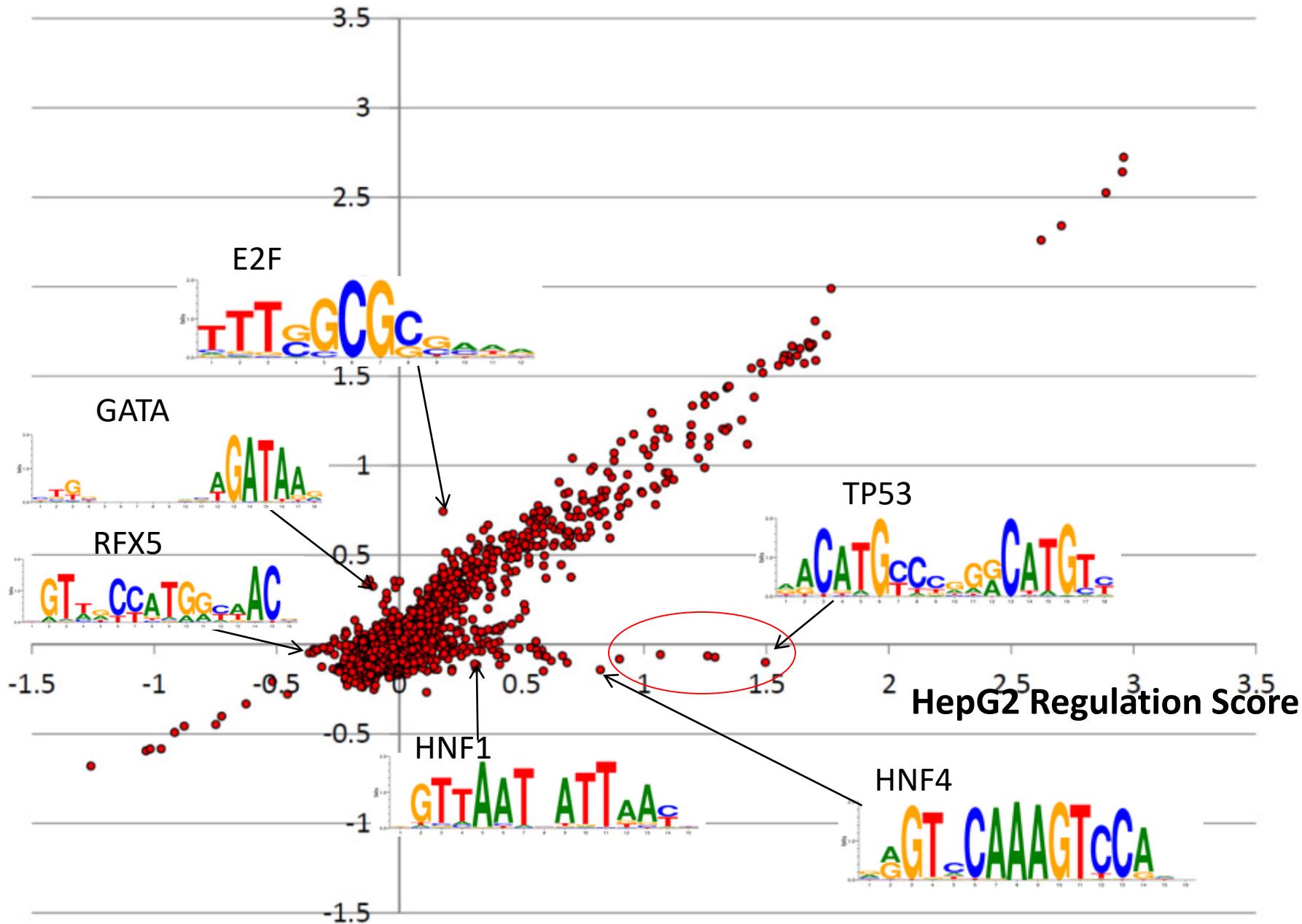
Comparing to ~1900 motifs - both known and discovered on ENCODE TF ChIP-seq data
(Kheradpour and Kellis, 2014) with ≥ 20 instances overlapping testing regions

Top Activating and Repressive Motifs Revealed

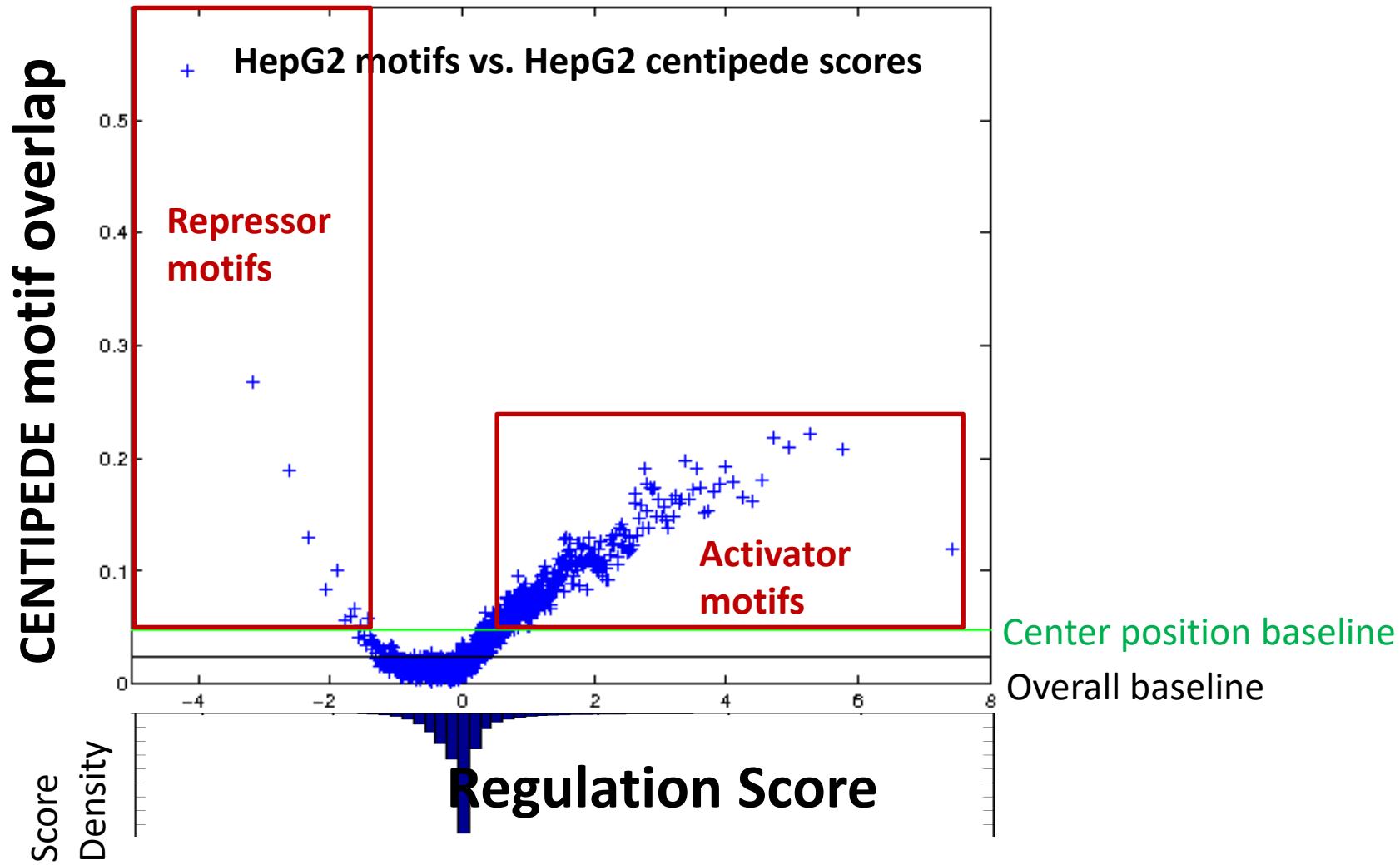


Cell Type Specific Motifs Revealed

K562 Regulation Score



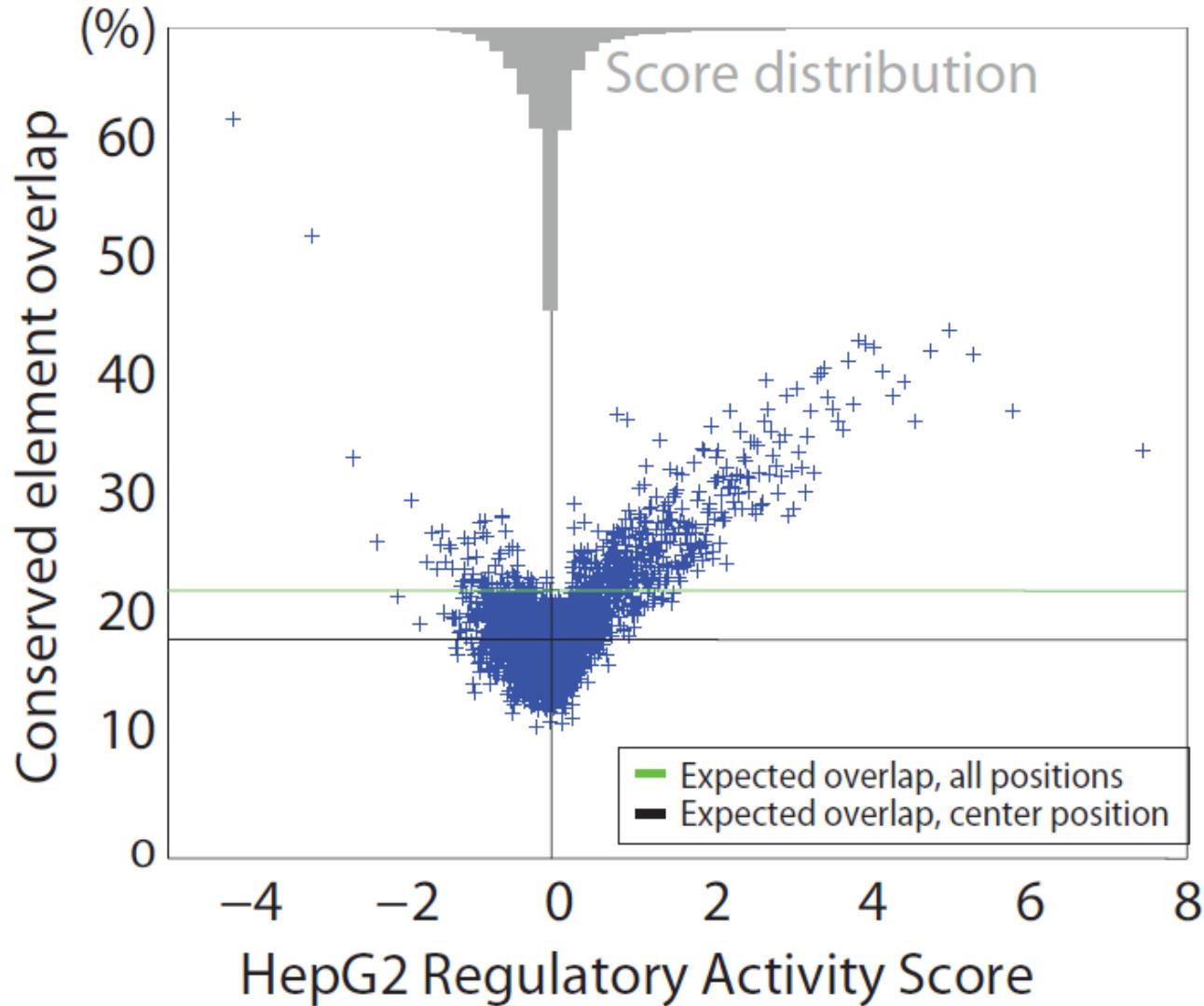
Inferred positions match regulatory motifs



44

*Predicted Activation and Repressive Bases Strongly Enrich
for Predicted Binding Sites in HepG2 + K562*

Active/repressed positions are evol. conserved

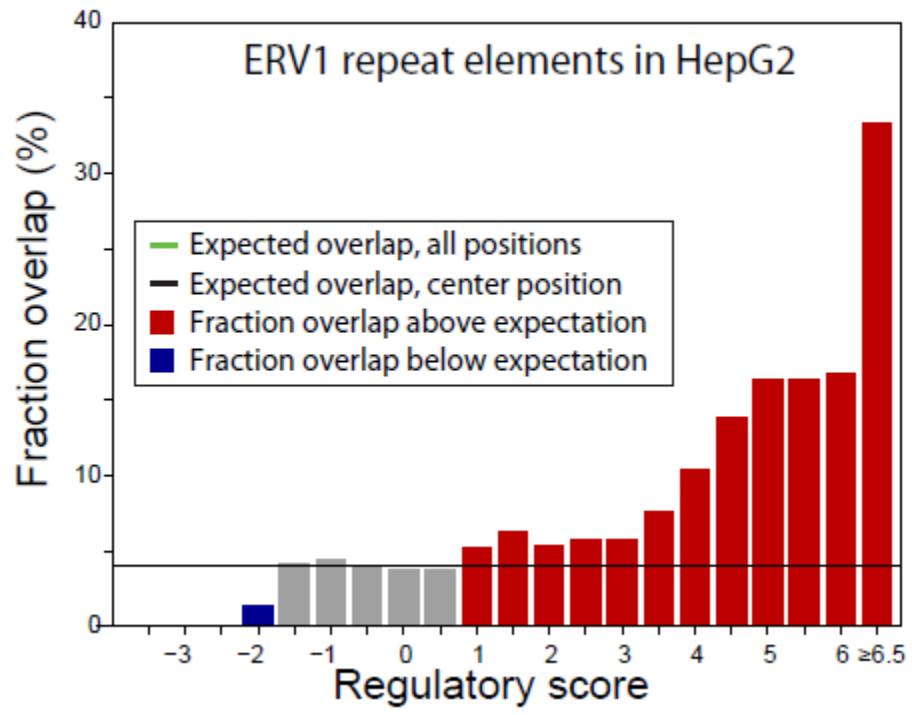


Strongest enrichment for repressive positions

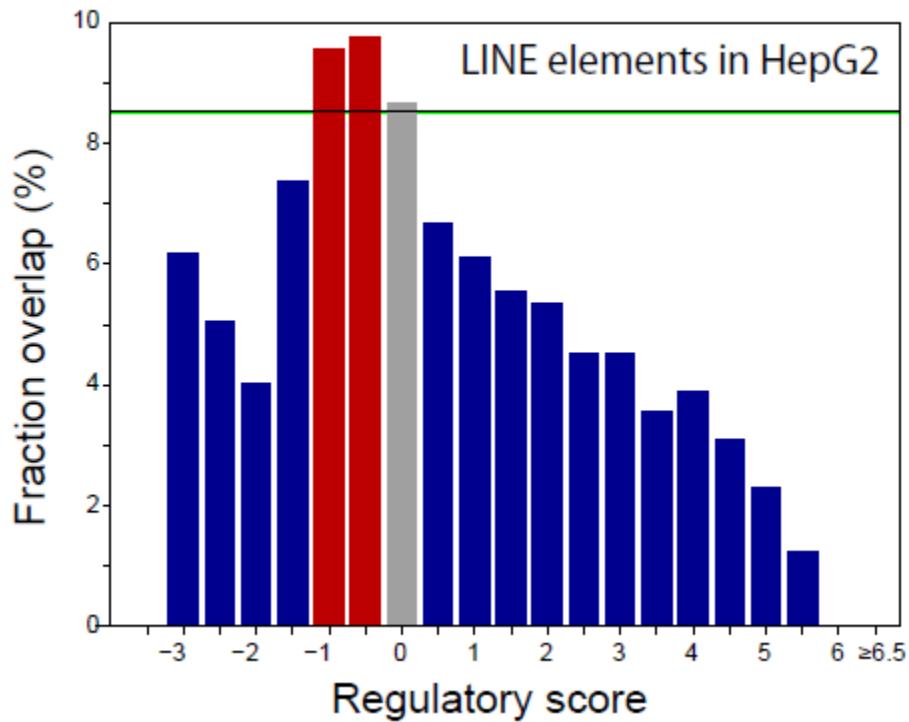
Slight depletion at strongest activating positions

ERV1 repeat elements can drive activity

a.

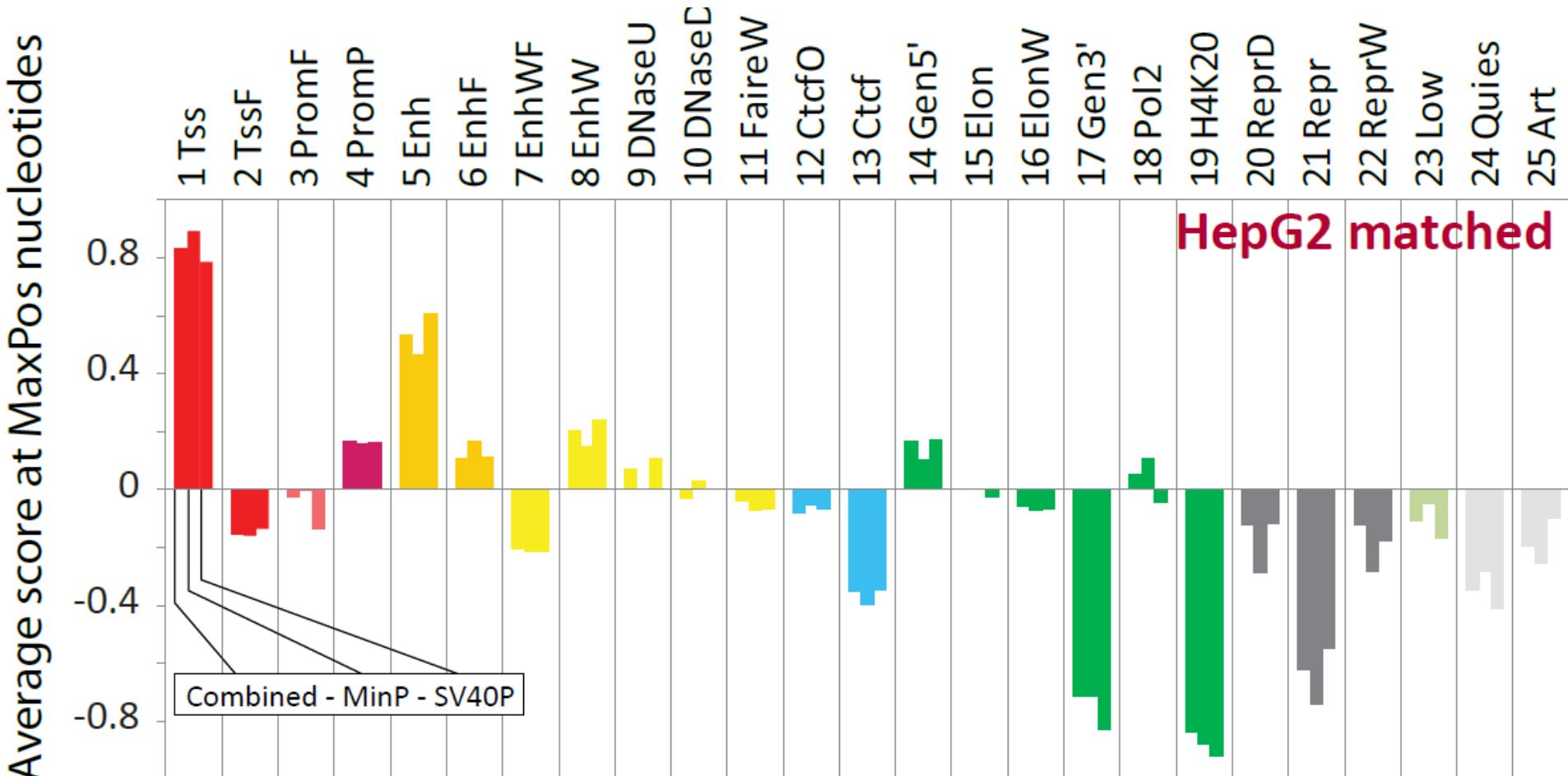


b.



*Strongest activating nucleotides match ERV1 repeats
(by contrast, LINE elements strongly depleted)
Enable rapid evolution of gene-regulatory networks*

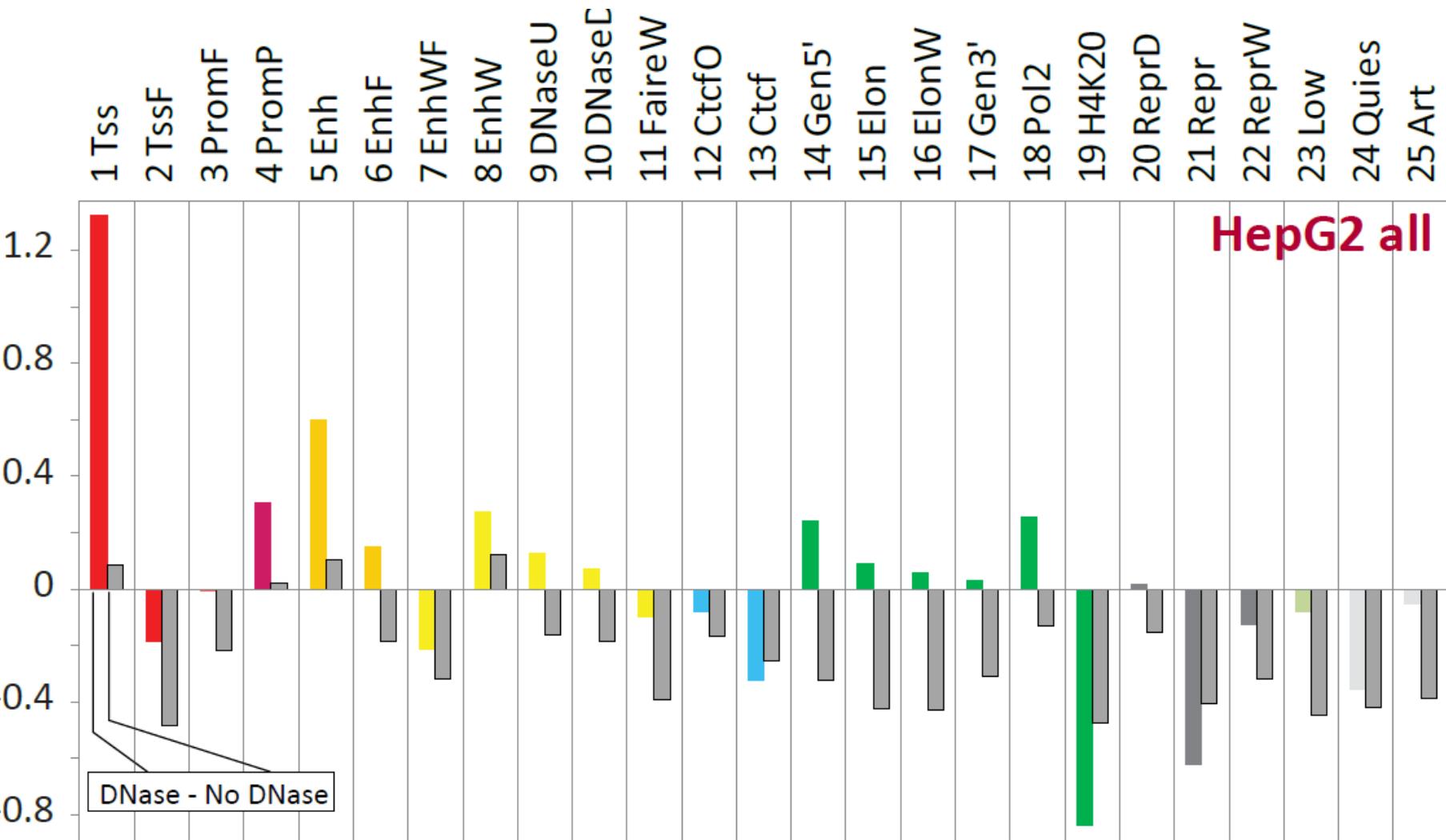
DNase elements in different chromatin states differ in their activity levels



Promoter, Enhancer regions highly activating.
ReprPC regions highly repressive

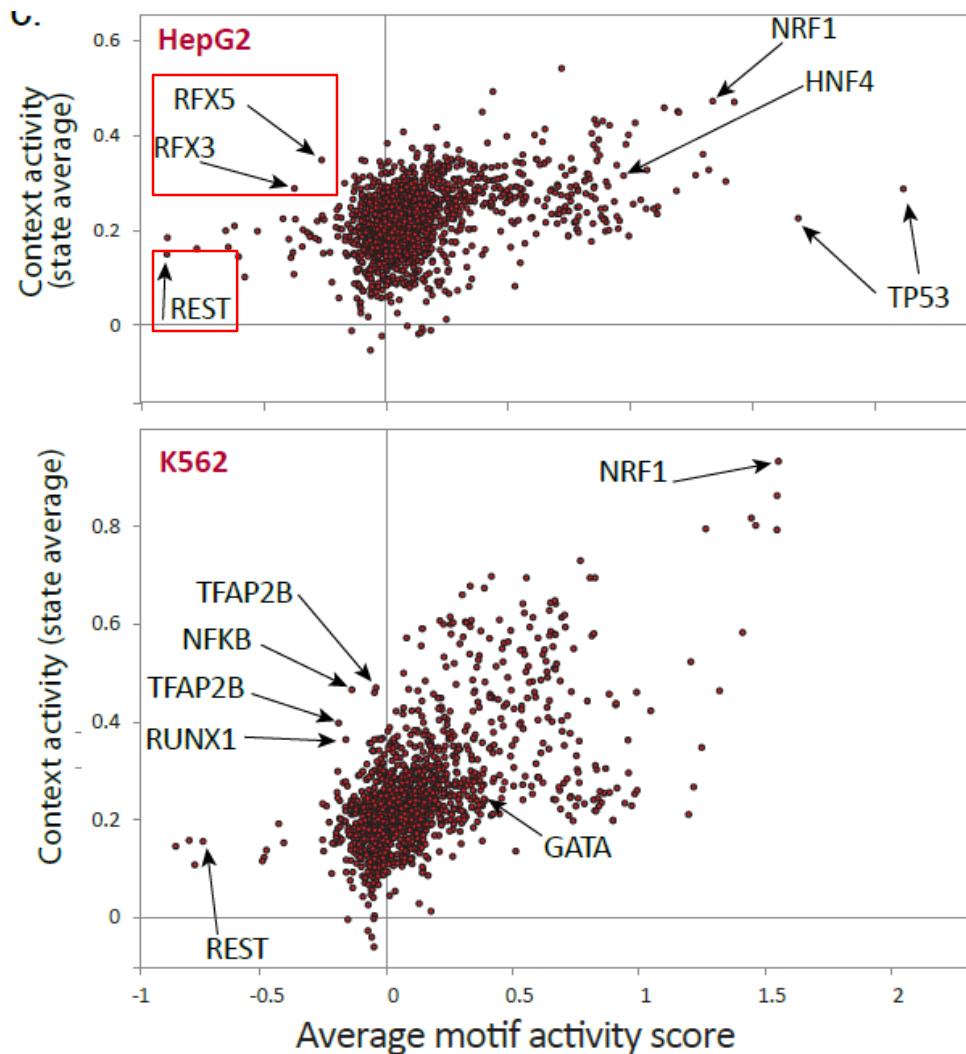
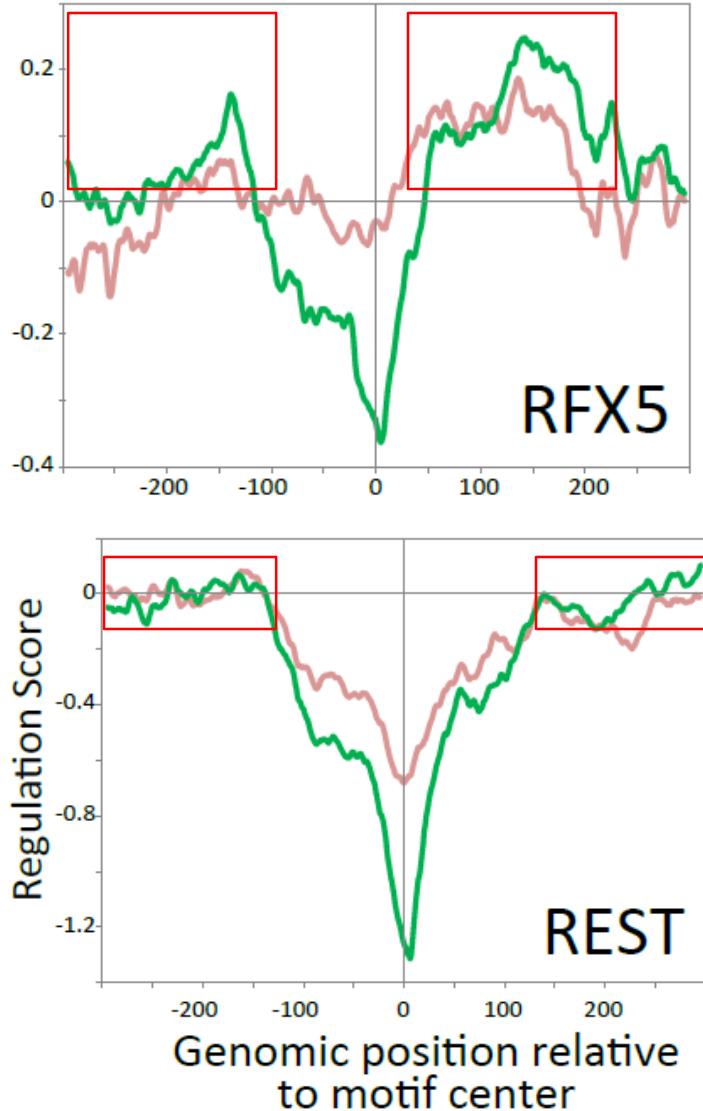
Accessible regions drive stronger activity

Average score at MaxPos nucleotides



For both activating and repressive positions

Discovery of repressors that act in active regions



- ***REST acts as a repressor in repressive regions (as expected)***
- ***But RFX5 acts as a repressor only in active regions (modulator?)***

High-throughput expts: MPRA, SHARPR, HiDRA, Perturb-seq

1. High-throughput synthesis: Massively Parallel Reporter Assays (MPRA)

- MPRA technology: in vitro synthesis, reporter design, transfection
- 1 site X deep: single-base dissection, combinatorial changes, high-res map
- 2k sites x shallow: motif-guided perturbations, activators/repressors

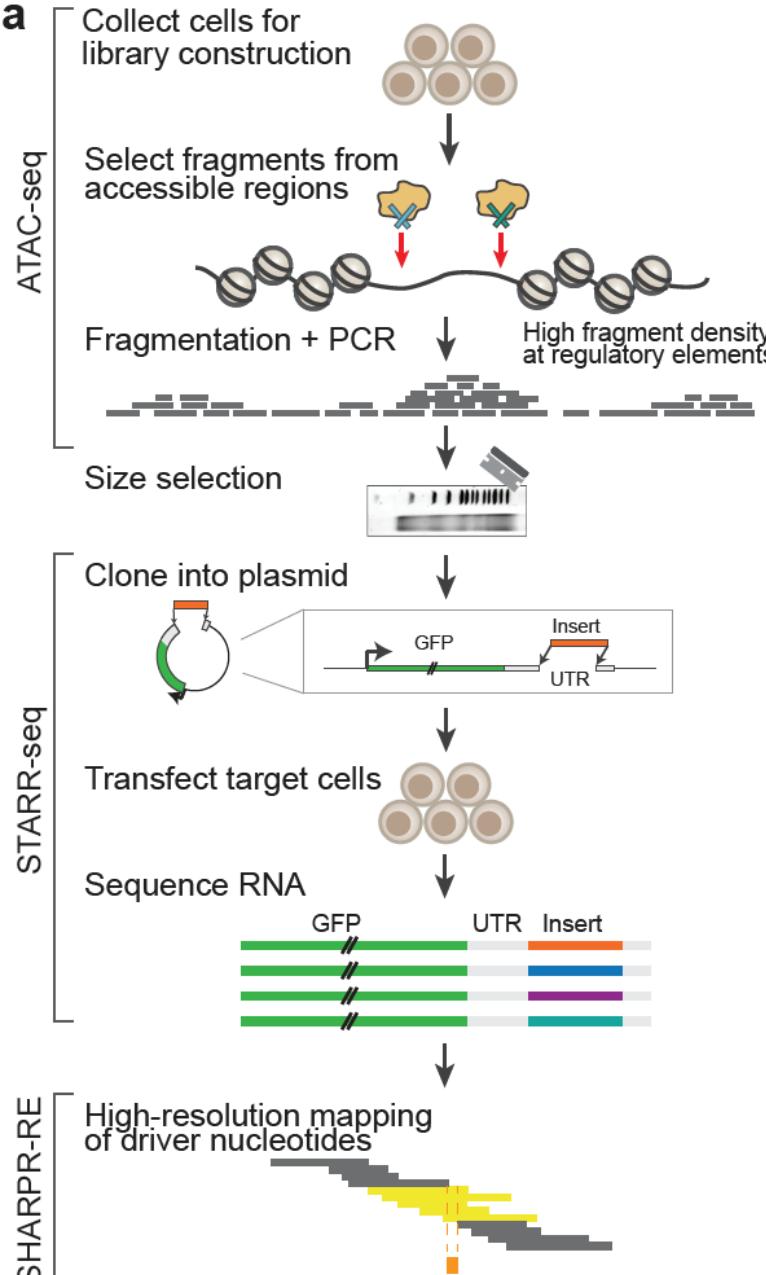
2. Next-generation applications of MPRA + STARR-Seq, SHARPR, HiDRA

- 10k x deep: high-resolution dissection by tiling + deconvolution (Sharpr)
- 10M x deep: HiDRA, no synthesis (STARR), targeted capture (ATAC)

3. Endogenous genome editing: cutting and template-based repair

- Endonucleases: TALENs, ZNFs. Repair pathways. Diversity of outcomes.
- CRISPR-Cas9: origins, discovery, optimization, base-editing
- Next-generation: Perturb-seq, multiplexing, delivery, applications

HiDRA: High-Definition Reporter Assay



Key features:

- No synthesis → 7M fragments tested in 1 expt
- No synthesis, size-selection → Test long fragments
- Select accessible DNA regions → High sensitivity
- 3'UTR integration → self-transcribing → No barcode
- Densely-overlapping fragments → Region tiling
- Unbiased, random starts/ends → Sharpr dissection

Putting it all together:

- Testing 7M fragments in 1 experiment
- High sensitivity, high specificity, quantitative assay
- High-res inference pinpoints driver nucleotides

ARTICLE

<https://doi.org/10.1038/s41467-018-07746-1>

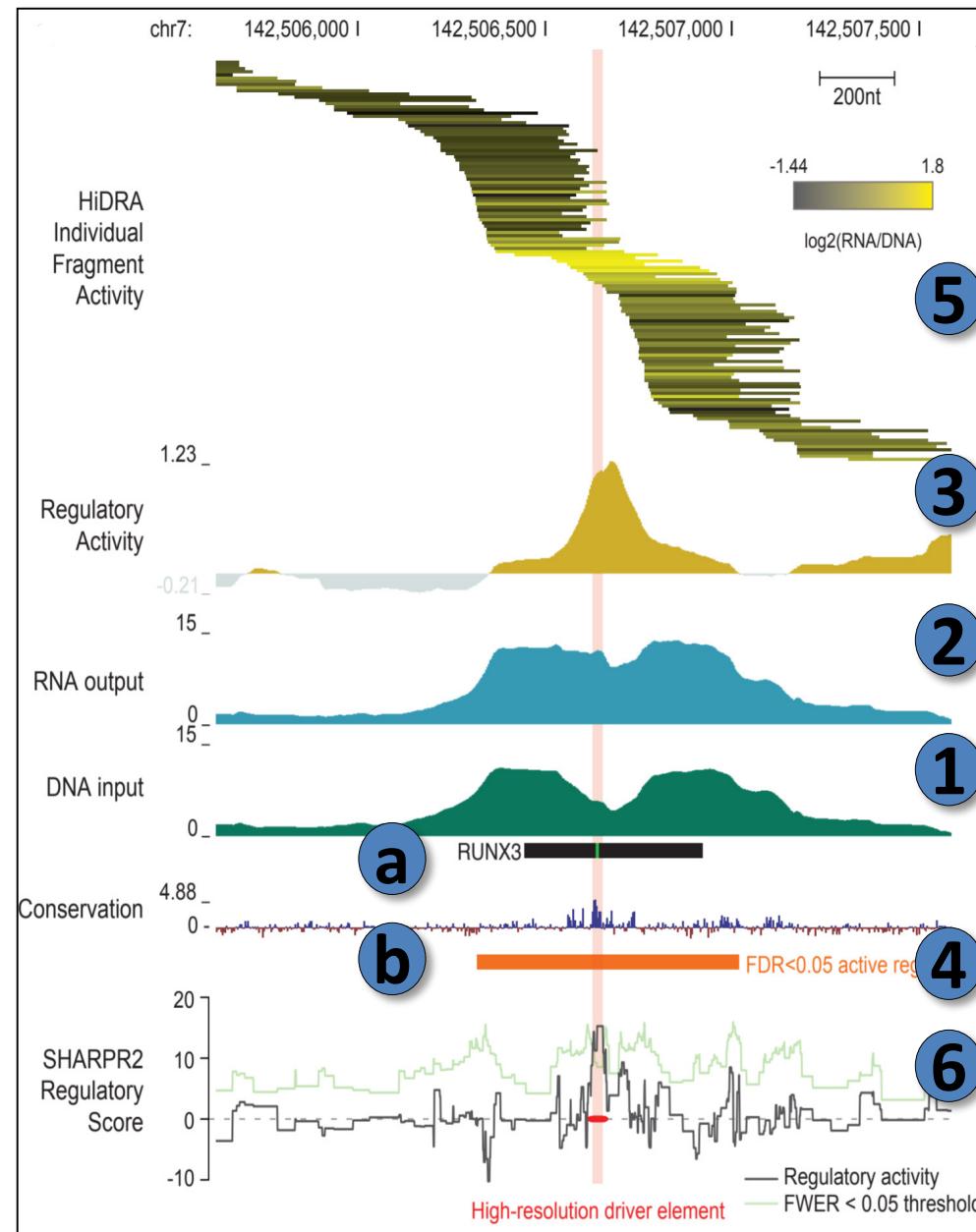
OPEN



High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human

Xinchen Wang 1,2,3,7, Liang He^{2,3}, Sarah M. Goggin², Alham Saadat², Li Wang², Nasa Sinnott-Armstrong², Melina Claussnitzer^{2,4,5,6} & Manolis Kellis 2,3

HiDRA data overview: DNA, RNA, Regulatory Activity, Sharpr2



1. Sequence DNA library

- Effectively a DNase/ATAC-Seq expt

2. Sequence RNA output

- How much expression does this drive

3. Take RNA/DNA ratio

- Measures regulatory activity

4. Pinpoint boundaries of active region

- FDR<0.05

5. Study activity of individual fragments

- Random start/end cuts (Transposase)

6. Infer high-resolution driver nucleotides

- Sharpr2 deconvolution algorithm
- Exploit diffs btw overlapping fragments

a. Compare with evolutionary conservation

- Capture evolutionarily-conserved nts

b. Compare with bound regulatory motifs

- Driver nucleotides are highly accurate

HiDRA input DNA library recapitulates DNase/ATAC-Seq

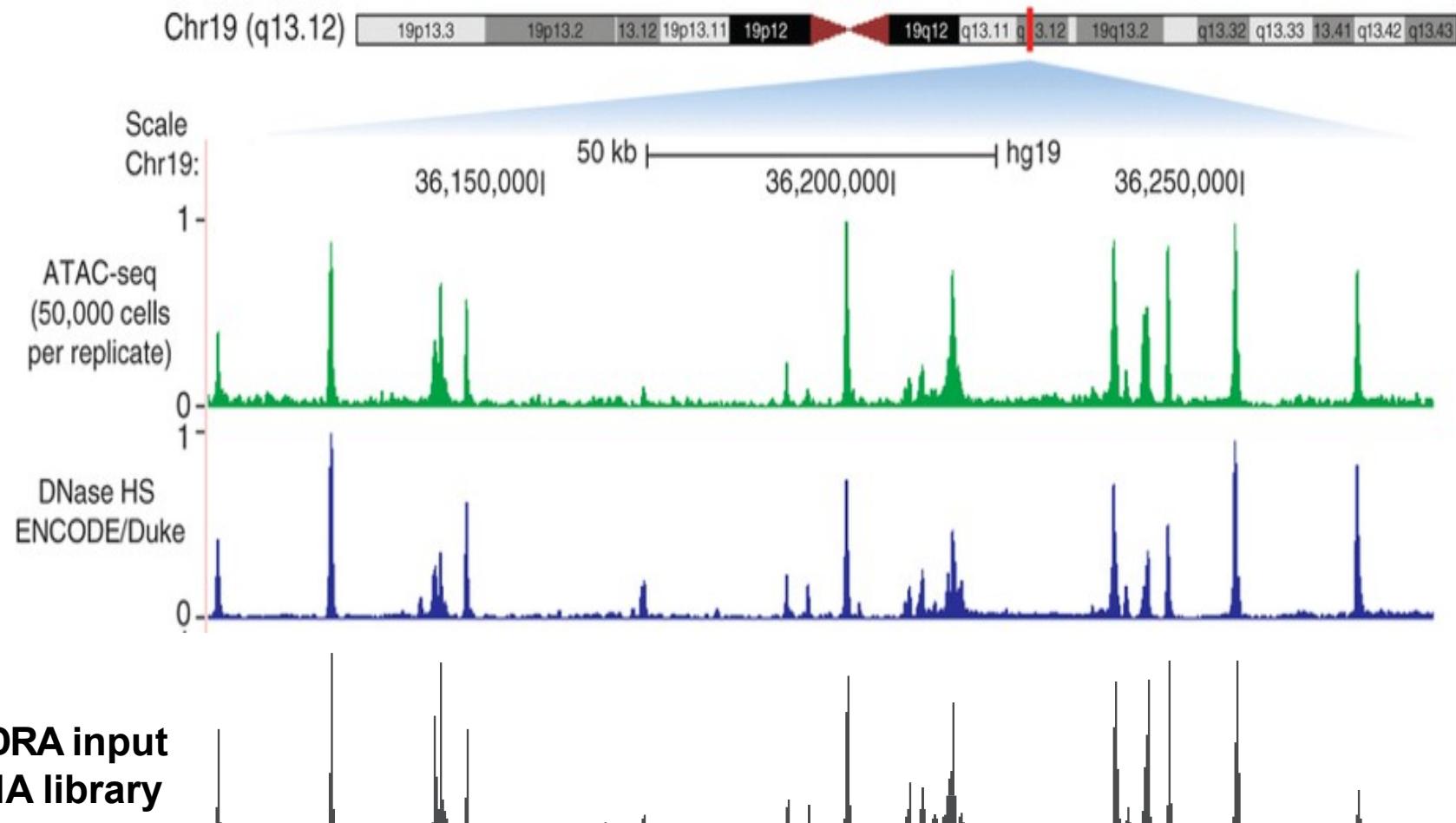
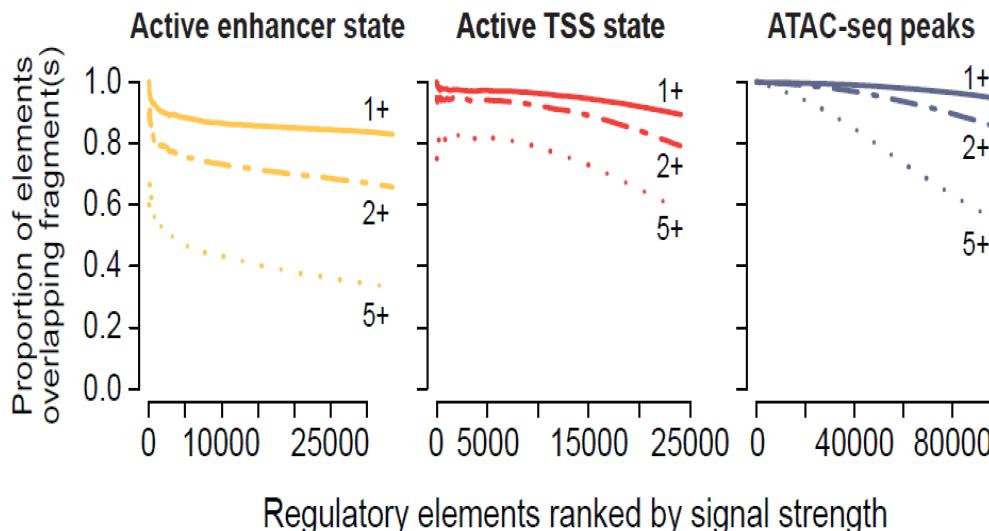


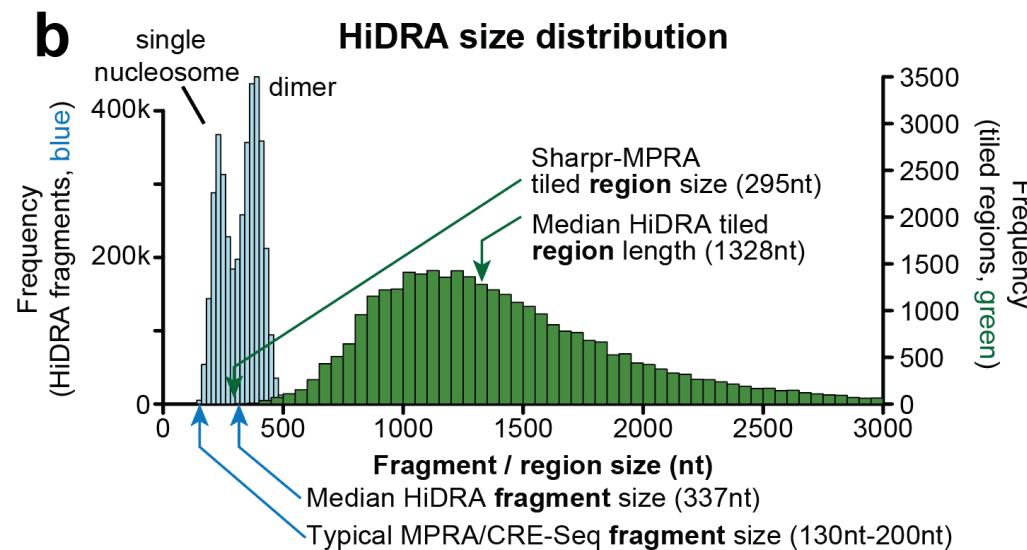
Fig 1c from Buenrostro et al. Nature Methods
2013

Preferential selection of putative regulatory elements

HiDRA input DNA library: long, active, densely-covered regions

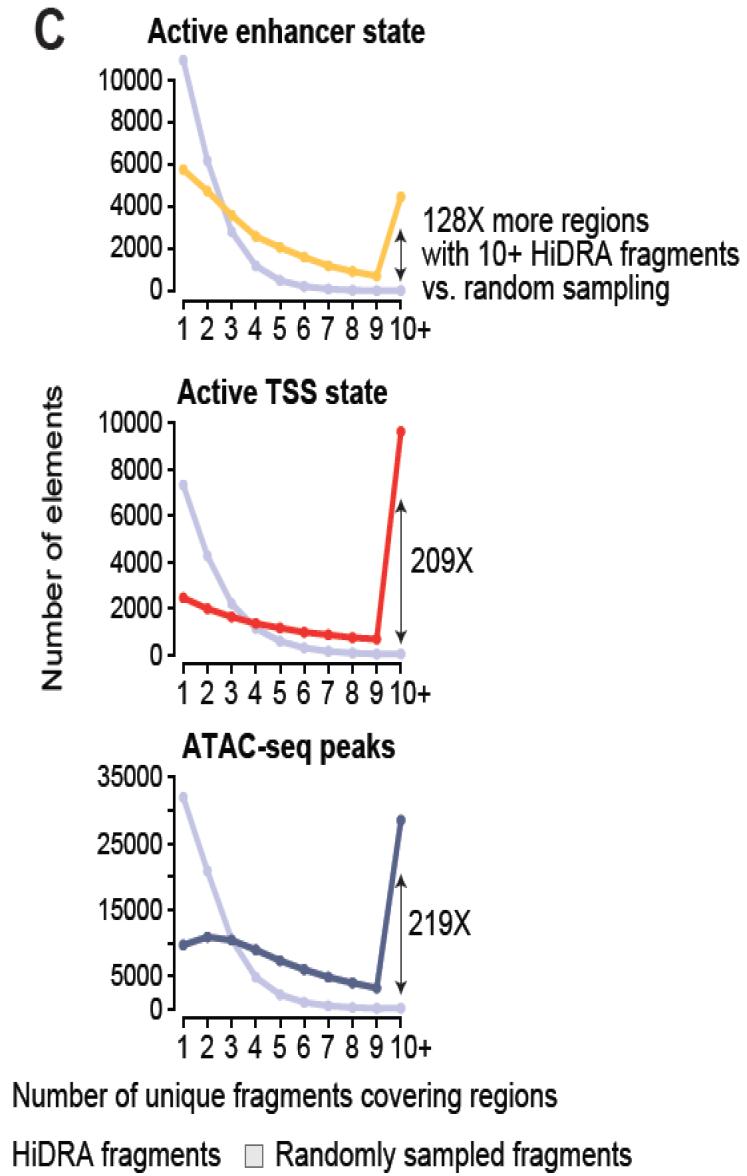


HiDRA DNA library captures more active elements



Fragments: 99% are 169-477 nt (median: 337nt)

Regions: 99% are 513-4,036 nt (median: 1,328nt)

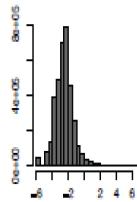


Up to 200-fold higher coverage for putative regulatory elements

HiDRA activity is highly reproducible

0.1 RPM cutoff

RNA 1



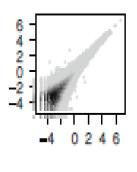
0.72
0.753

0.728
0.761

0.726
0.759

0.725
0.757

RNA 2

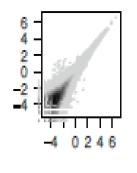


0.729
0.76

0.73
0.761

0.733
0.764

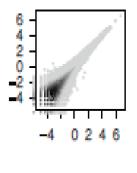
RNA 3



0.737
0.766

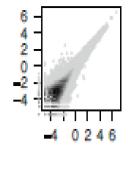
0.729
0.762

RNA 4



0.73
0.762

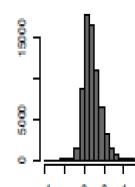
RNA 5



0.73
0.762

0.73
0.762

1.0 RPM cutoff

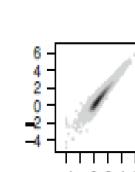


0.921
0.942

0.939
0.957

0.937
0.955

0.93
0.948

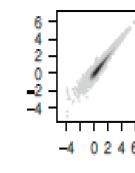


0.935
0.954

0.938
0.955

0.943
0.959

0.943
0.959

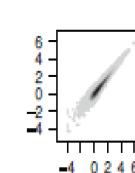


0.943
0.959

0.943
0.959

0.939
0.956

0.939
0.956



0.939
0.956

0.939
0.956

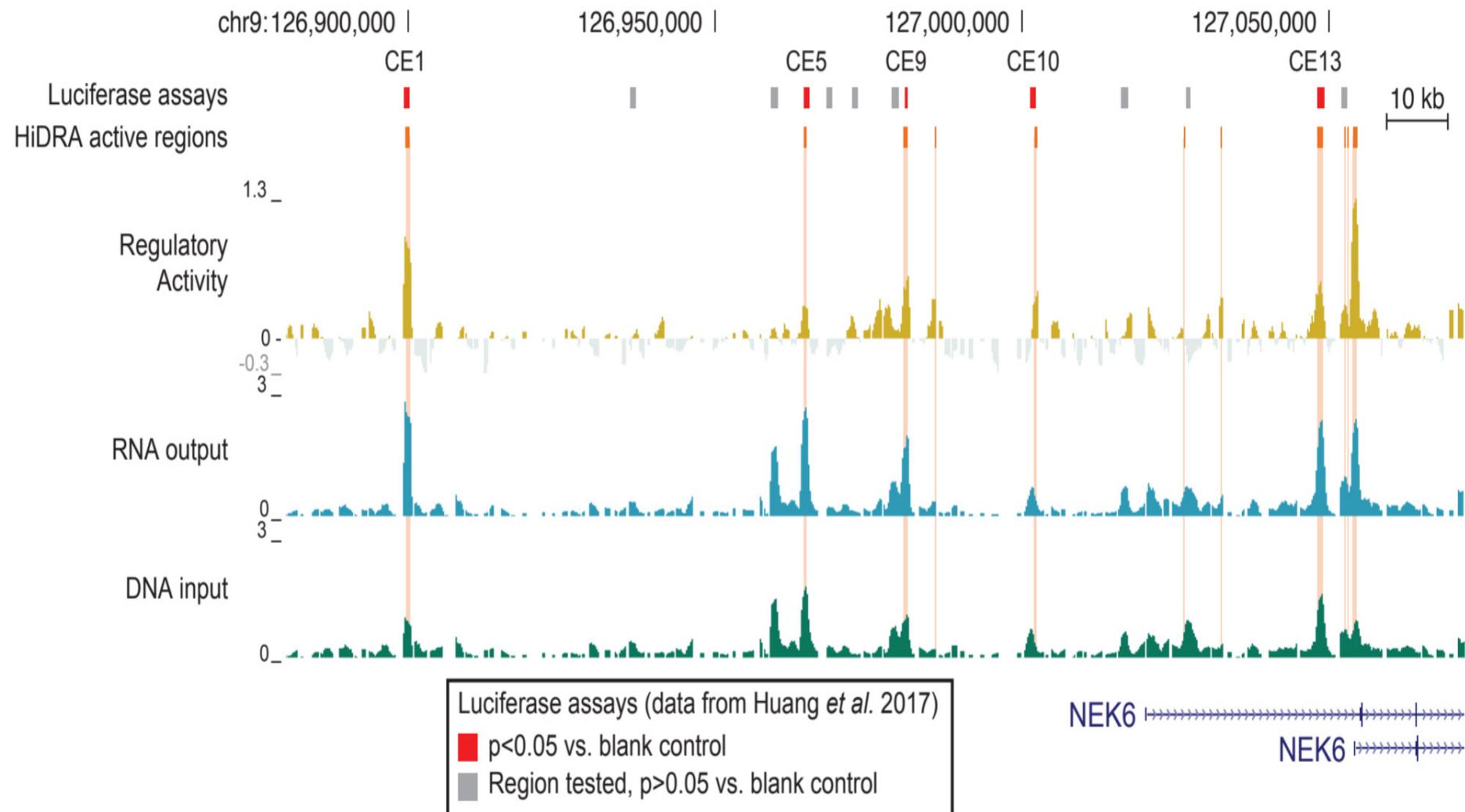
0.939
0.956

0.939
0.956

Avg Pearson Corr: 0.75 for low-activity fragments

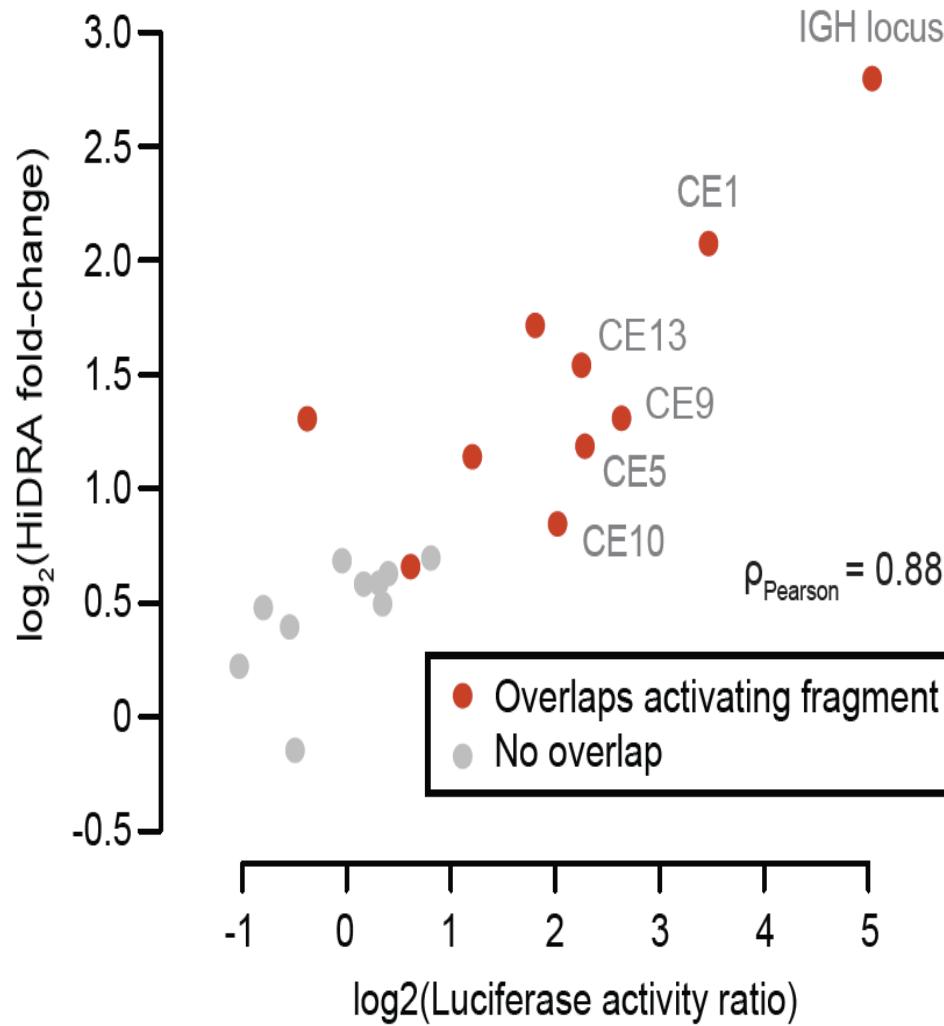
Avg Pearson Corr: 0.95 for high-activity fragments

HiDRA activity captures known enhancer regions



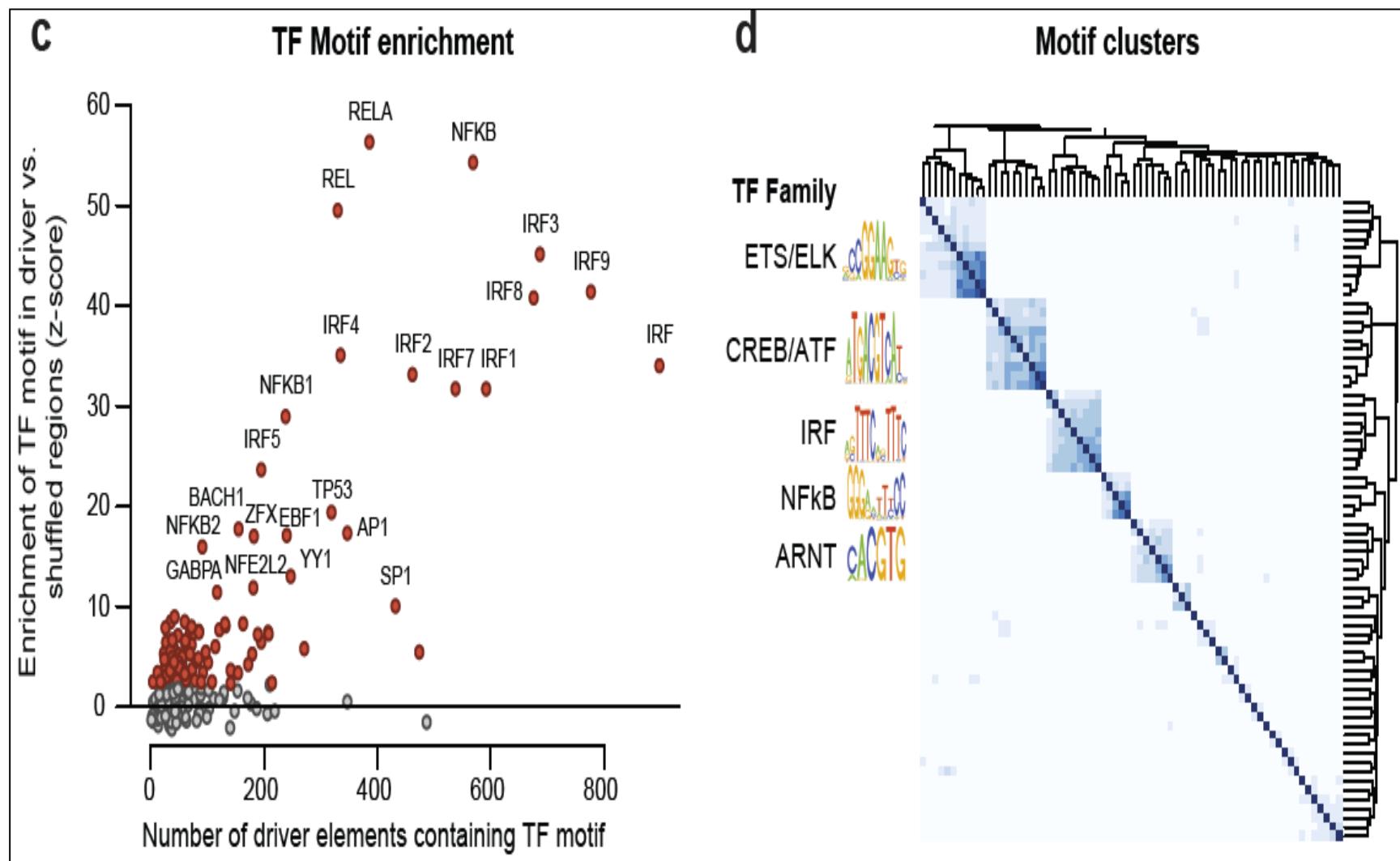
High sensitivity / high specificity vs. Luciferase assays

HiDRA provides a quantitative readout (vs. Luciferase)



Quantitative comparison with 20 luciferase experiments

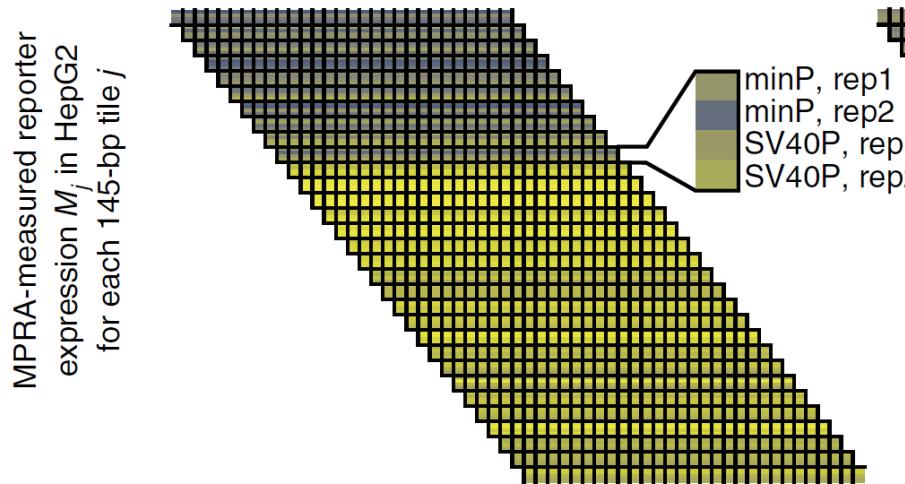
HiDRA captures motifs for known regulators



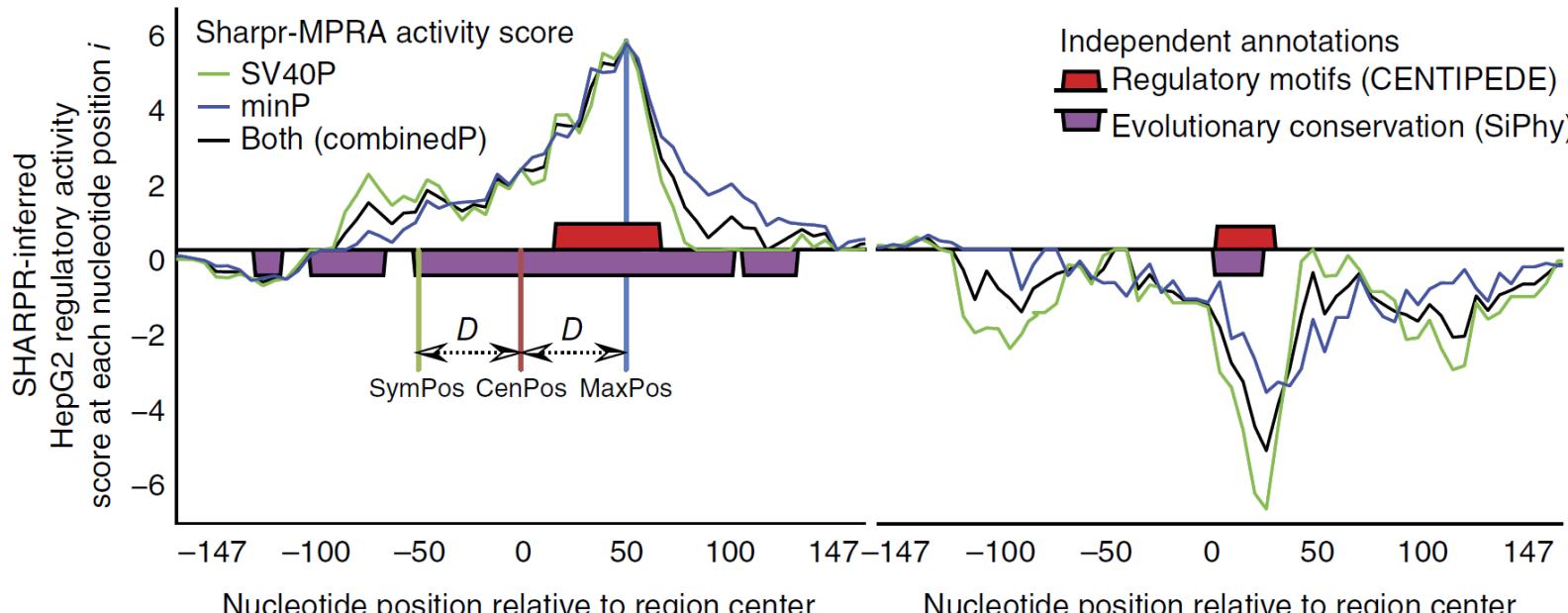
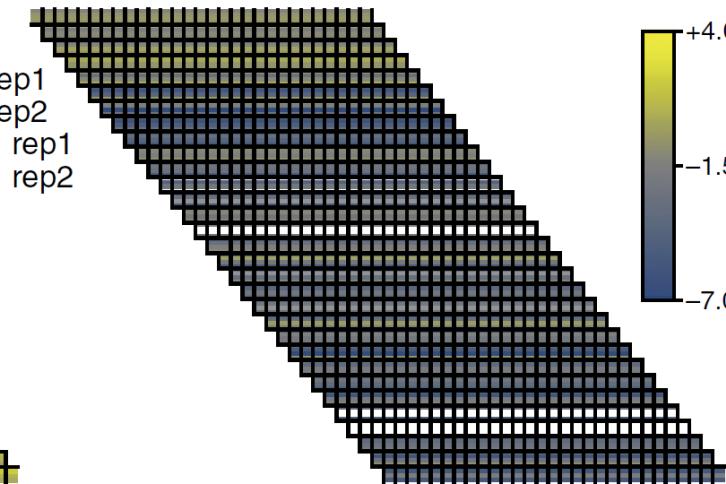
Capture known motifs

SHARPR: Exploit offsets for high-resolution driver inference

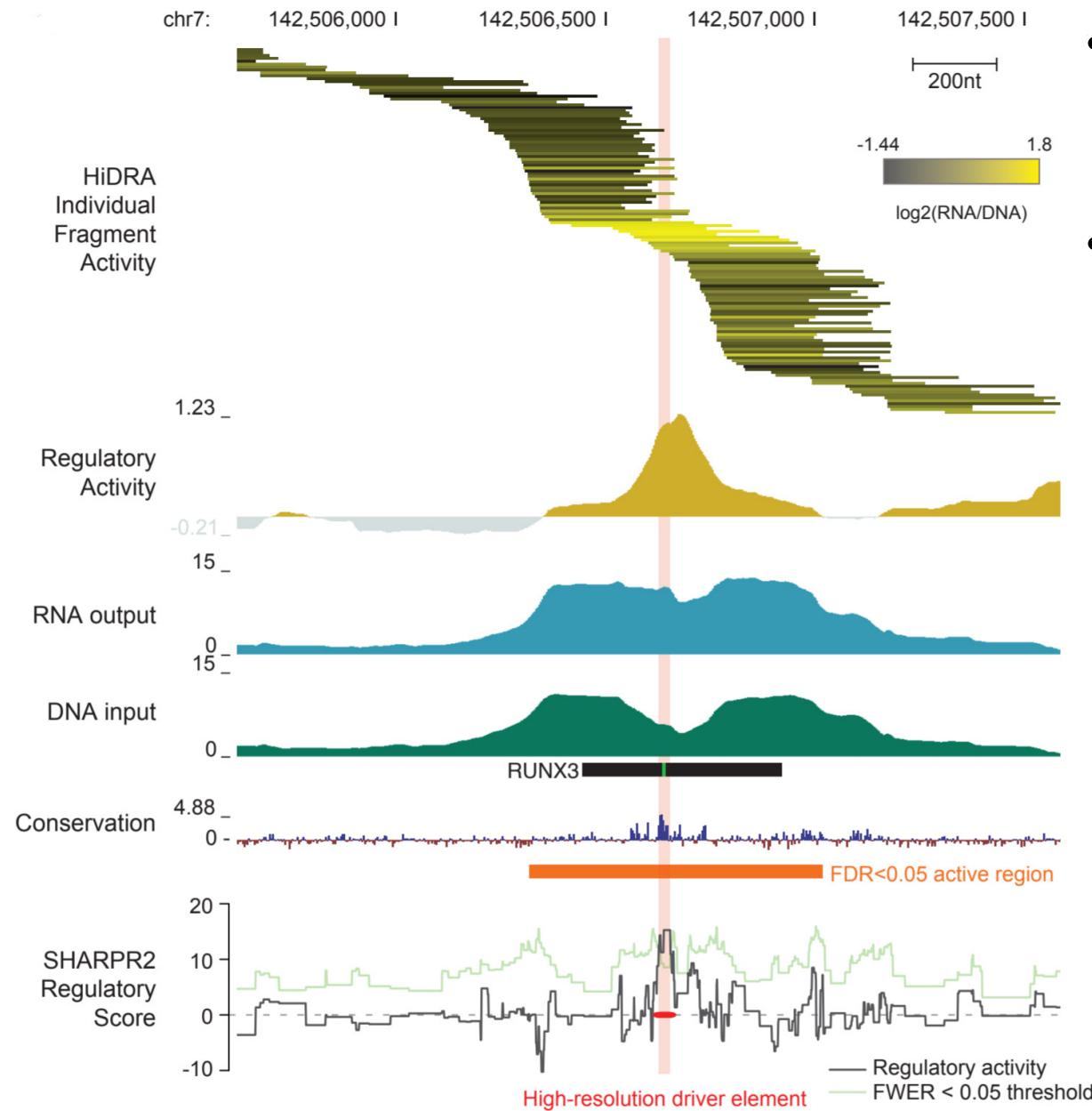
Region #2580 (chr16:69,564,569–69,564,863)
Selected in HepG2 state 5: Enh



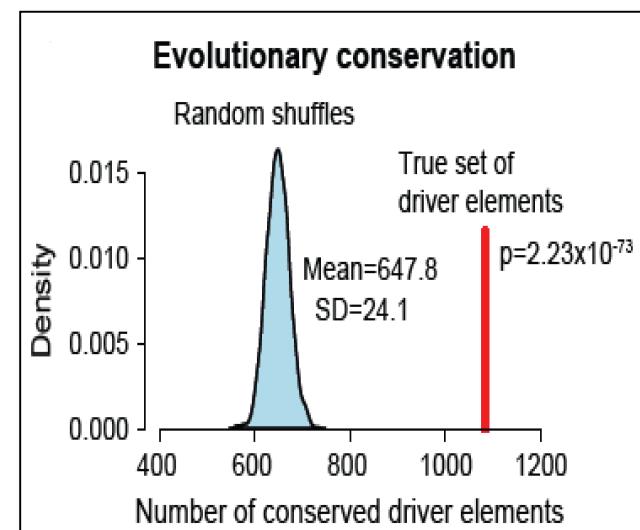
Region #11085 (chr12:103,218,269–103,218,563)
Selected in HepG2 state 9: DNaseU



SHARPR-RE algorithm infers high-resolution driver nucleotides

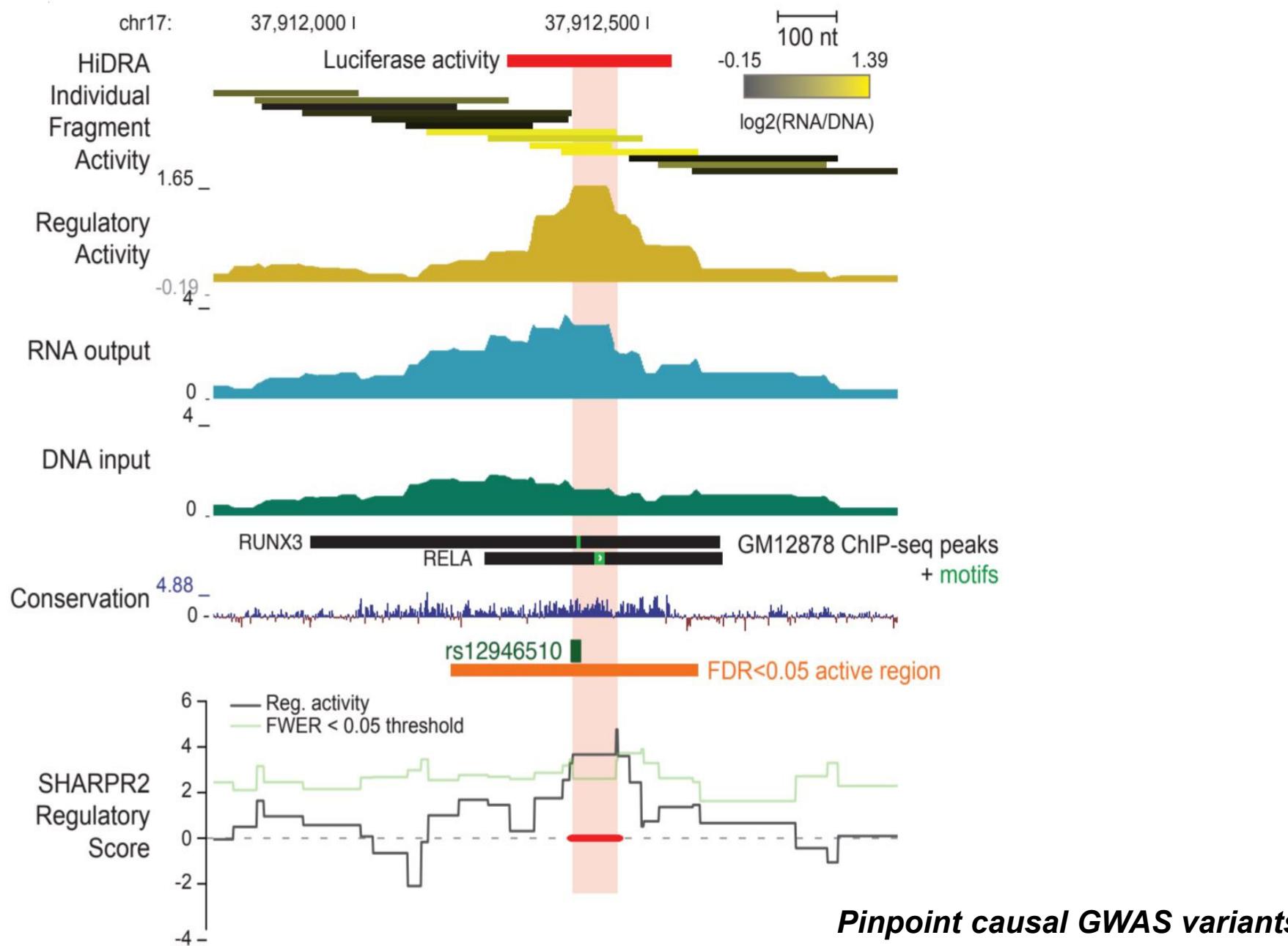


- Exploit differences between neighboring fragments
- Driver nucleotides match motifs, evolutionary conservation

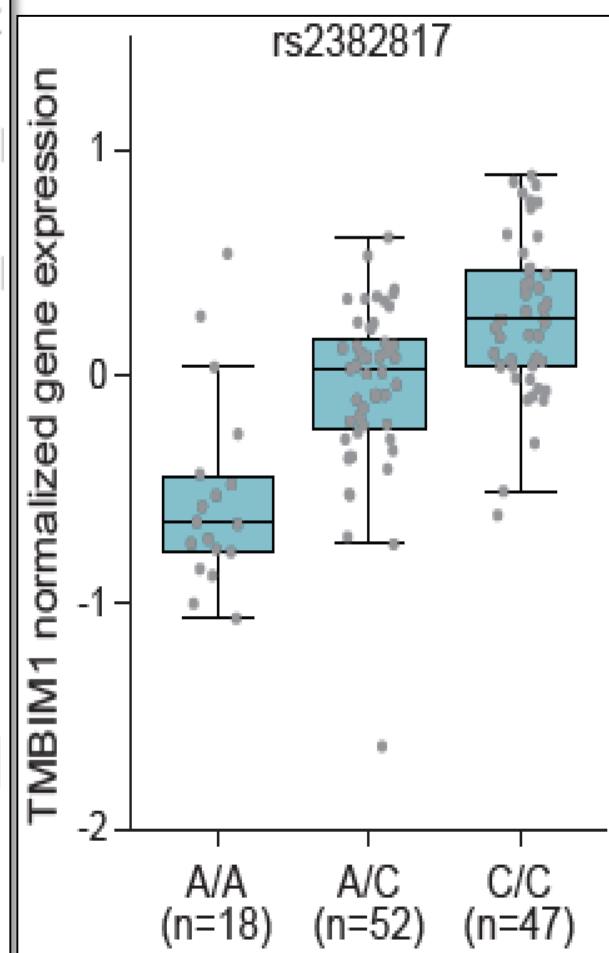
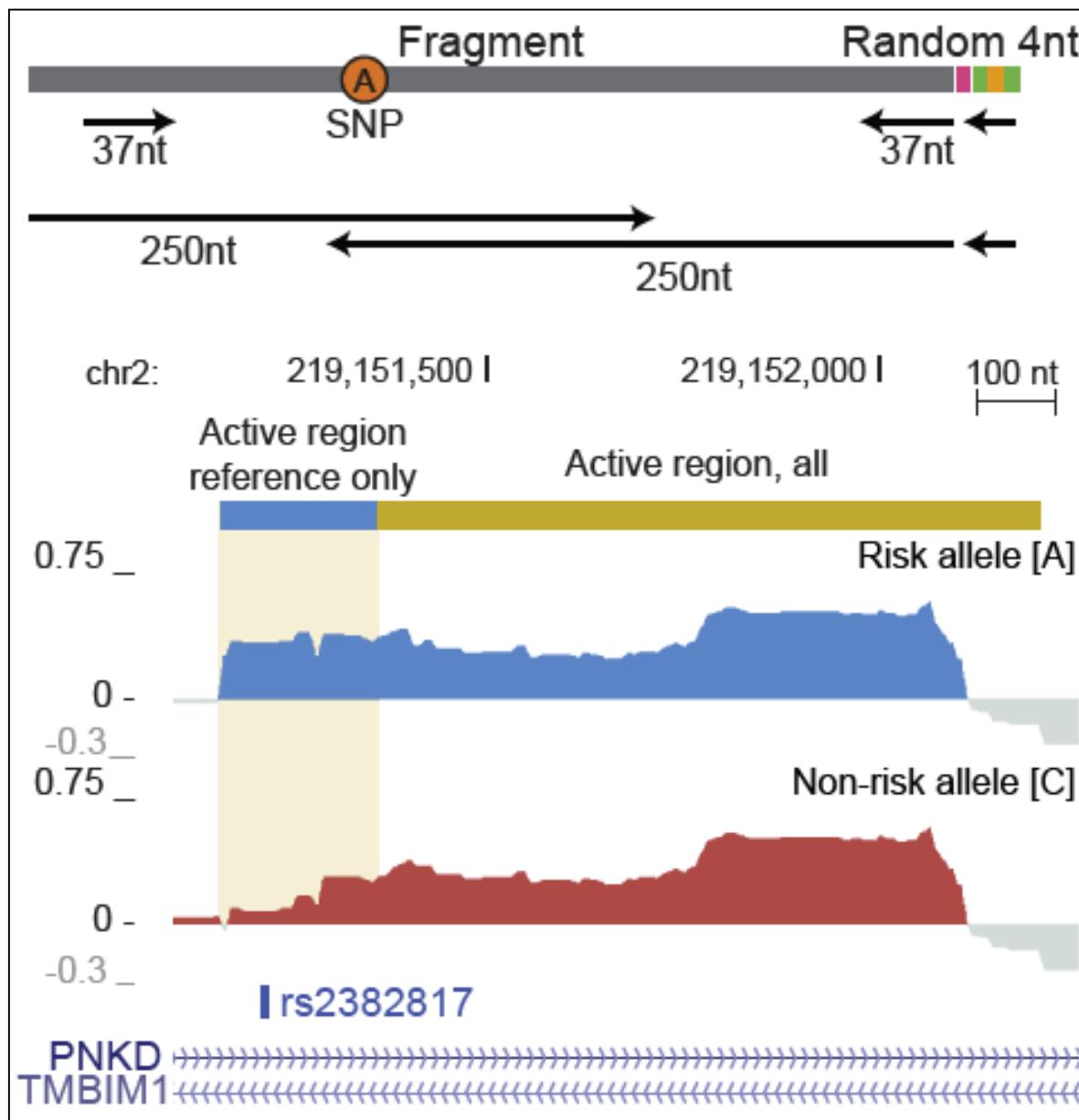


- Enrichment: $P < 10^{-73}$

HiDRA high-resolution drivers help dissect GWAS loci

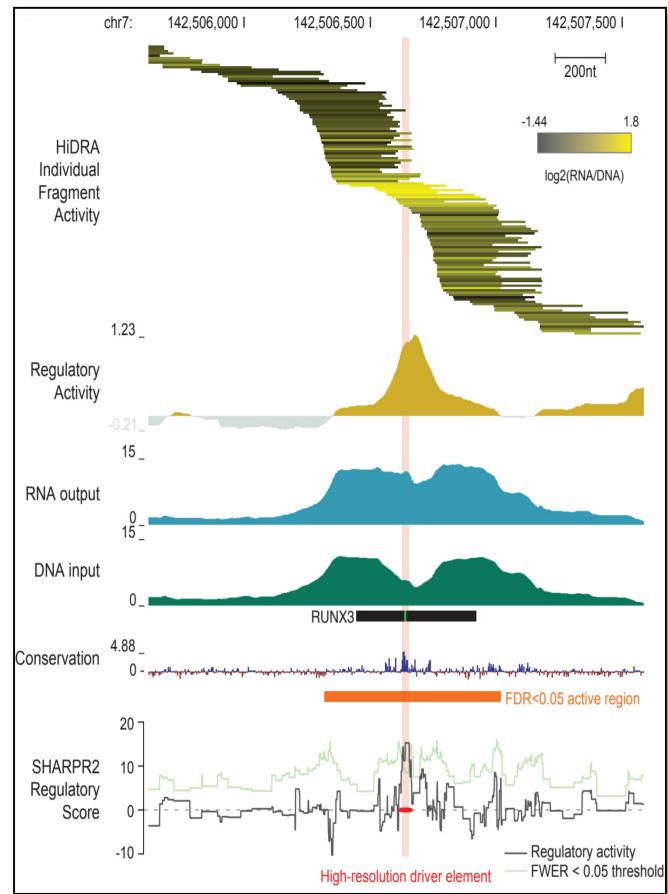


HiDRA activity differences between risk and non-risk alleles



**Allele-specific activity for
IBD-associated rs2382817**

HiDRA summary



- 7M fragments tested in one experiment
- Longer fragments (~350nt on average)
- High reproducibility, 0.95 for higher-activity elmt
- Up to 200-fold enrichment for regulatory regions
- High-resolution dissection of driver nts
- Captures known motifs, conserved nucleotides
- Pinpoints driver SNPs in GWAS loci
- Reveals diffs between risk and non-risk alleles
- **General tool for testing regulatory regions**

ARTICLE

<https://doi.org/10.1038/s41467-018-07746-1>

OPEN



High-resolution genome-wide functional dissection
of transcriptional regulatory regions and
nucleotides in human

Xinchen Wang 1,2,3,7, Liang He^{2,3}, Sarah M. Goggin², Alham Saadat², Li Wang², Nasa Sinnott-Armstrong², Melina Claussnitzer^{2,4,5,6} & Manolis Kellis 2,³

High-throughput expts: MPRA, SHARPR, HiDRA, Perturb-seq

1. High-throughput synthesis: Massively Parallel Reporter Assays (MPRA)

- MPRA technology: in vitro synthesis, reporter design, transfection
- 1 site X deep: single-base dissection, combinatorial changes, high-res map
- 2k sites x shallow: motif-guided perturbations, activators/repressors

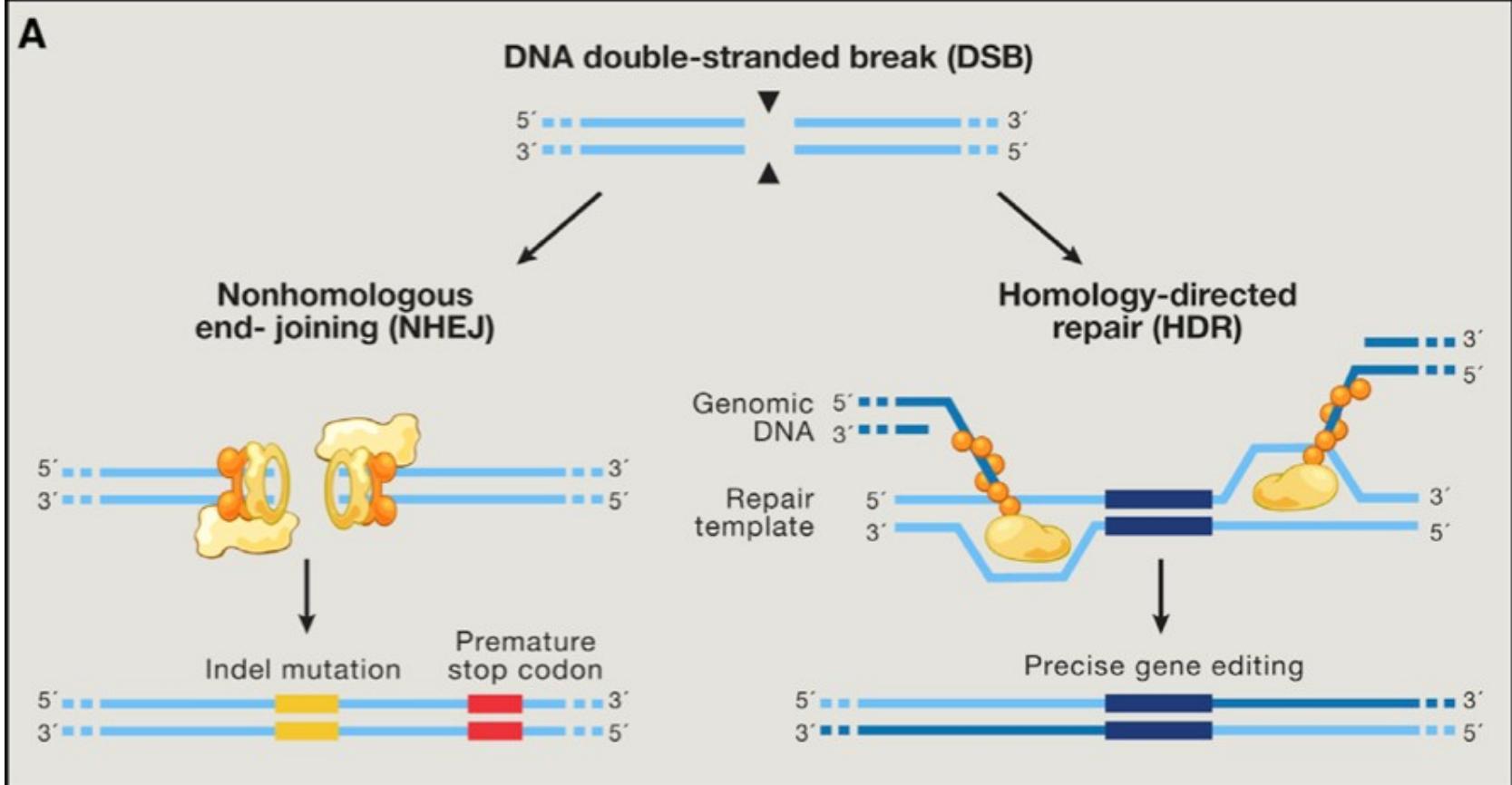
2. Next-generation applications of MPRA + STARR-Seq, SHARPR, HiDRA

- 10k x deep: high-resolution dissection by tiling + deconvolution (Sharpr)
- 10M x deep: HiDRA, no synthesis (STARR), targeted capture (ATAC)

3. Endogenous genome editing: cutting and template-based repair

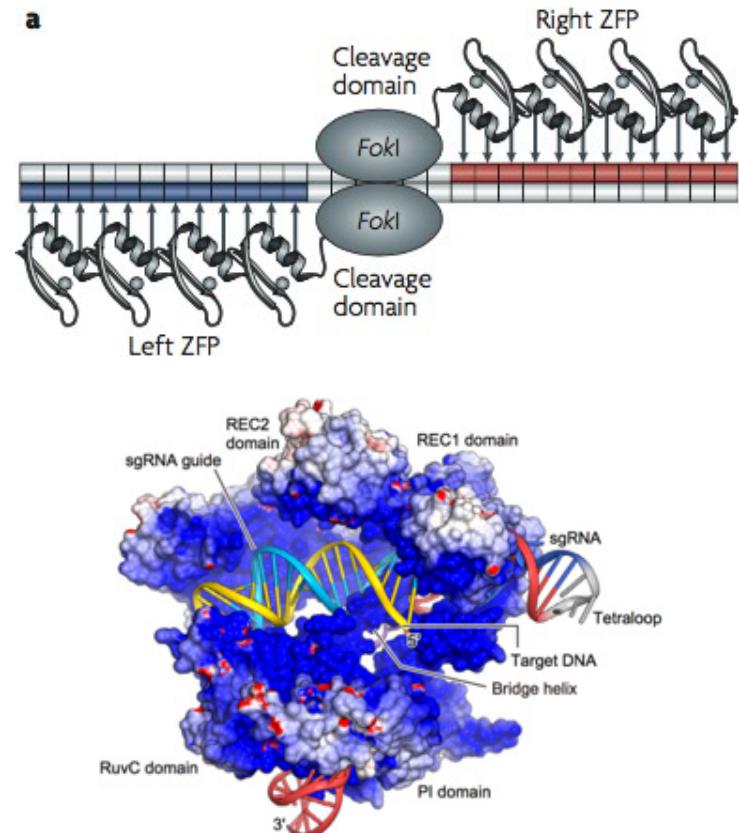
- Endonucleases: TALENs, ZNFs. Repair pathways. Diversity of outcomes.
- CRISPR-Cas9: origins, discovery, optimization, base-editing
- Next-generation: Perturb-seq, multiplexing, delivery, applications

How can we edit a genome?



How do we create artificial DNA DSBs?

- Using programmable DNA nucleases
- Zinc Fingers/TALENs: DNA binding provided by modular protein domains
- CRISPR: RNA-guided DNA cleavage



Customizing DNA Binding Proteins

Helix-turn-helix

Zinc finger

Leucine zipper

Winged helix

Winged helix turn helix

Helix loop helix

HMG-box

TAL effector

B3 domain

Immunoglobulin fold

High-throughput expts: MPRA, SHARPR, HiDRA, Perturb-seq

1. High-throughput synthesis: Massively Parallel Reporter Assays (MPRA)

- MPRA technology: in vitro synthesis, reporter design, transfection
- 1 site X deep: single-base dissection, combinatorial changes, high-res map
- 2k sites x shallow: motif-guided perturbations, activators/repressors

2. Next-generation applications of MPRA + STARR-Seq, SHARPR, HiDRA

- 10k x deep: high-resolution dissection by tiling + deconvolution (Sharpr)
- 10M x deep: HiDRA, no synthesis (STARR), targeted capture (ATAC)

3. Endogenous genome editing: cutting and template-based repair

- Endonucleases: **ZNFs**, TALENs. Repair pathways. Diversity of outcomes.
- CRISPR-Cas9: origins, discovery, optimization, base-editing
- Next-generation: Perturb-seq, multiplexing, delivery, applications

Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å

NIKOLA P. PAVLETICH AND CARL O. PABO

The zinc finger DNA-binding motif occurs in many proteins that regulate eukaryotic gene expression. The crystal structure of a complex containing the three zinc fingers from Zif268 (a mouse immediate early protein) and a consensus DNA-binding site has been determined at 2.1 angstroms resolution and refined to a crystallographic *R* factor of 18.2 percent. In this complex, the zinc fingers bind in the major groove of B-DNA and wrap partly around the double helix. Each finger has a similar relation to the DNA and makes its primary contacts in a three-base pair subsite. Residues from the amino-terminal portion of an α helix contact the bases, and most of the contacts are made with the guanine-rich strand of the DNA. This structure provides a framework for understanding how zinc fingers recognize DNA and suggests that this motif may provide a useful basis for the design of novel DNA-binding proteins.

THE ZINC FINGER IS ONE OF THE MAJOR STRUCTURAL motifs involved in eukaryotic protein-nucleic acid interactions. The fingers that were first discovered in the *Xenopus* transcription factor IIIA (TFIIIA) (1) contain a sequence motif of the form $X_3\text{-Cys}\text{-}X_{2,4}\text{-Cys}\text{-}X_{12}\text{-His}\text{-}X_{2,4}\text{-His}\text{-}X_4$ (where X is any amino acid), and hundreds of similar finger sequences have been reported (2). Only a few of the proteins that contain such fingers have been studied in detail, but it appears that many of these zinc finger domains are involved in DNA binding. Proteins with zinc finger domains are involved in many aspects of eukaryotic gene regulation. For example, such fingers occur in proteins induced by differentiation and growth signals [EGR1 (3, 4), EGR2 (5, 6)], in proto-oncogenes [GLI (7), Wilms' tumor gene (8)], in general transcription factors [Sp1 (9)], in *Drosophila* segmentation genes [*Hunchback* (10), *Krueppel* (11)], and in regulatory genes of lower eukaryotic organisms [ADRI (12), *BfA* (13)]. The term zinc finger

N. P. Pavletich and C. O. Pabo are in the Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21201. C. O. Pabo also is in the Howard Hughes Medical Institute, and may add his name as author after 1 July 1991. Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139.

10 MAY 1991

RESEARCH ARTICLES 809

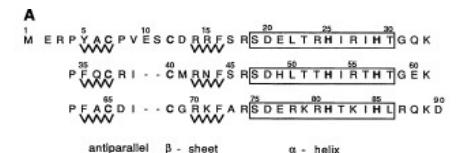
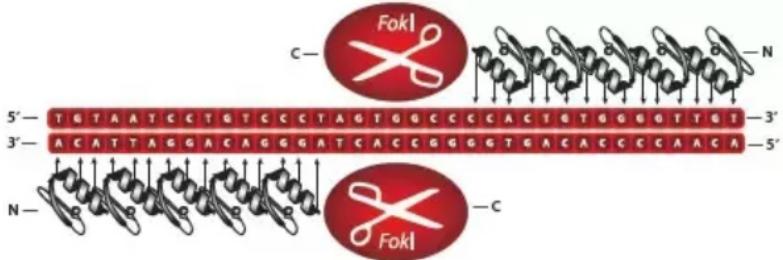


Fig. 1. Sequences of the Zif268 zinc finger domain and the DNA-binding site used in the cocrystallization. (A) The peptide used in the cocrystallization includes 89 residues from the Zif268 protein and the initiator methionine introduced in the cloning. Only residues 3 to 87 are present in the current model. (We presume that the terminal residues are disordered in the crystal.) The three zinc fingers are aligned to show the conserved residues and secondary structures. Helices are boxed, and the β sheets are indicated by zig-zag lines. The approximate positions for these regions of secondary structure could have been predicted from NMR studies of related zinc fingers; the precise positions were determined from our crystal structure. (B) DNA duplex used for cocrystallization. The subsites that the fingers bind to are either in shadowed or in bold letters. These alternate to highlight the 3-bp subsites recognized by the zinc fingers.

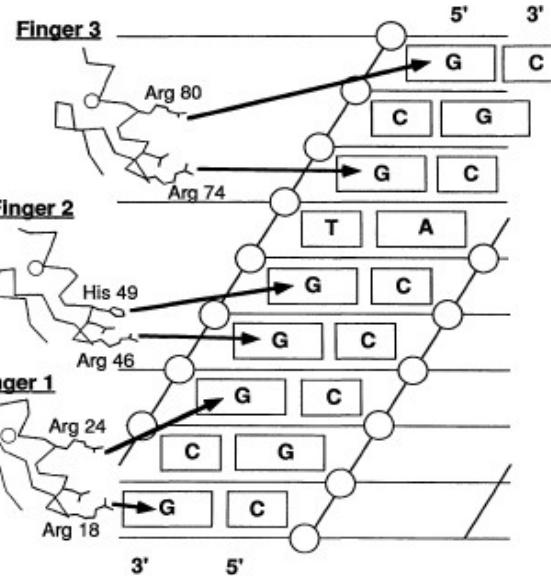
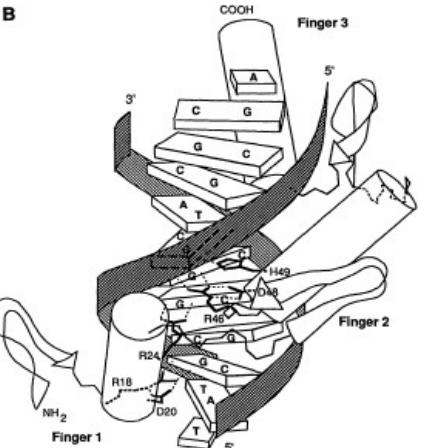


Fig. 5. Sketch summarizing all the base contacts made by the Zif268 peptide. The DNA is represented as a cylindrical projection.

ZNF-based genomic-editing strategy:

- Custom DNA endonucleases called zinc-finger nucleases (ZFNs)
- Zinc fingers are transcription factors
- Each finger module recognizes 3-4 bp of sequence
- By mixing and matching modules can target any sequence

High-throughput expts: MPRA, SHARPR, HiDRA, Perturb-seq

1. High-throughput synthesis: Massively Parallel Reporter Assays (MPRA)

- MPRA technology: in vitro synthesis, reporter design, transfection
- 1 site X deep: single-base dissection, combinatorial changes, high-res map
- 2k sites x shallow: motif-guided perturbations, activators/repressors

2. Next-generation applications of MPRA + STARR-Seq, SHARPR, HiDRA

- 10k x deep: high-resolution dissection by tiling + deconvolution (Sharpr)
- 10M x deep: HiDRA, no synthesis (STARR), targeted capture (ATAC)

3. Endogenous genome editing: cutting and template-based repair

- Endonucleases: ZNFs, **TALENs**. Repair pathways. Diversity of outcomes.
- CRISPR-Cas9: origins, discovery, optimization, base-editing
- Next-generation: Perturb-seq, multiplexing, delivery, applications

TALE: A New DNA Binding Protein

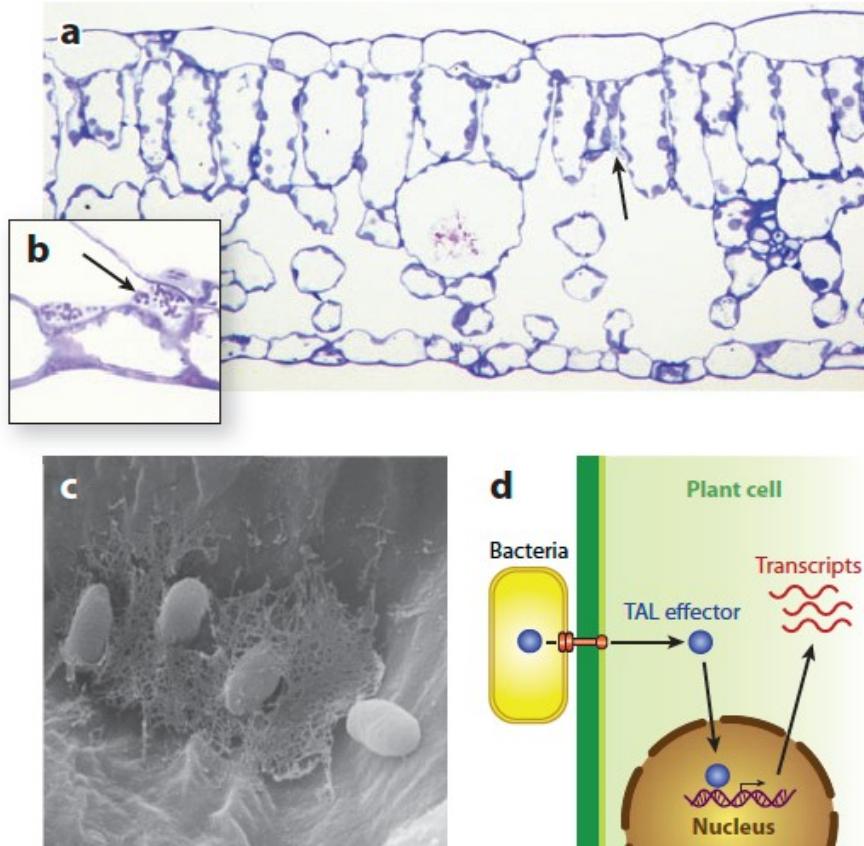
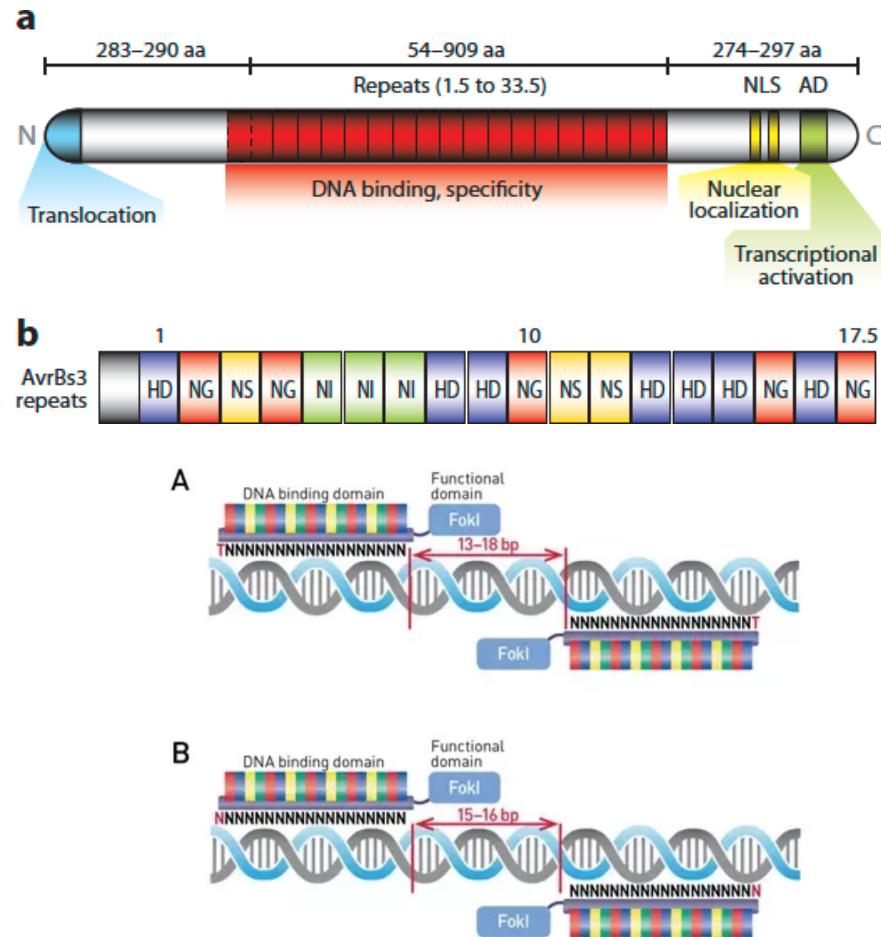


Figure 1

Xanthomonas pathovars inhabit the plant apoplast or xylem vessels and inject effector proteins into plant cells. (a) A microscopic image of a cross section through a pepper leaf infected by *X. campestris* pv. *vesicatoria* (Xcv). Note the large intercellular spaces and epidermal stomata. (b) Enlarged cross section of lower epidermis. (a, b) The arrows indicate bacterial microcolonies in the apoplast. (c) *Xcv* cells on pepper leaf cell surface inside the apoplast. The mesh corresponds to the exopolysaccharide xanthan. (d) *Xanthomonas* translocates a cocktail of effector proteins via a type III secretion system (red) into plant cells. Transcription activator-like (TAL) effectors localize to the plant cell nucleus, where they induce expression of specific target genes.



Transcription activator-like effector nucleases (TALENs):

- Dimeric transcription factor/nucleases
- Built from arrays of 33 to 35 amino acid modules
- Each module targets a single nucleotide
- By assembling arrays, can target nearly any sequence

High-throughput expts: MPRA, SHARPR, HiDRA, Perturb-seq

1. High-throughput synthesis: Massively Parallel Reporter Assays (MPRA)

- MPRA technology: in vitro synthesis, reporter design, transfection
- 1 site X deep: single-base dissection, combinatorial changes, high-res map
- 2k sites x shallow: motif-guided perturbations, activators/repressors

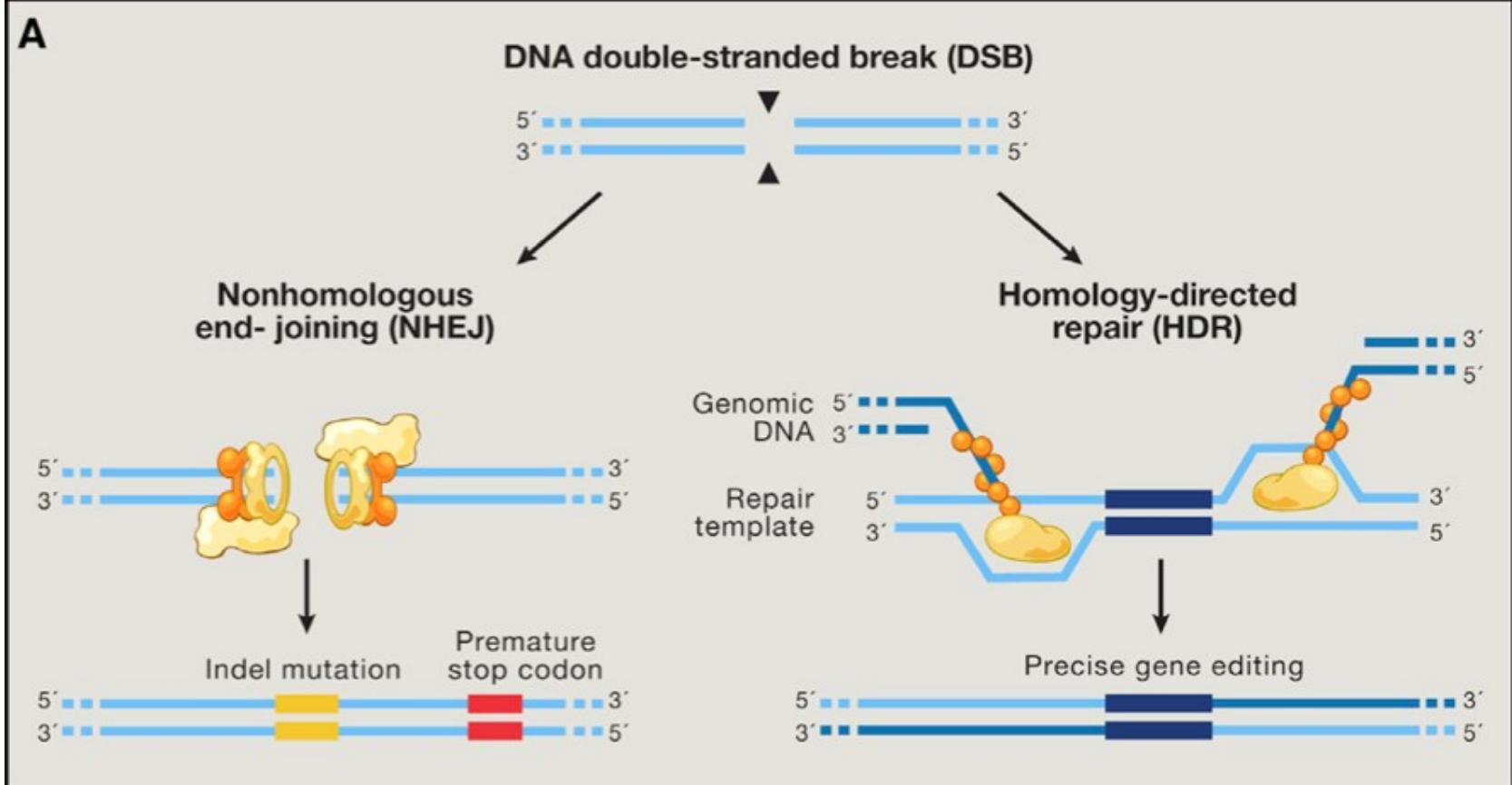
2. Next-generation applications of MPRA + STARR-Seq, SHARPR, HiDRA

- 10k x deep: high-resolution dissection by tiling + deconvolution (Sharpr)
- 10M x deep: HiDRA, no synthesis (STARR), targeted capture (ATAC)

3. Endogenous genome editing: cutting and template-based repair

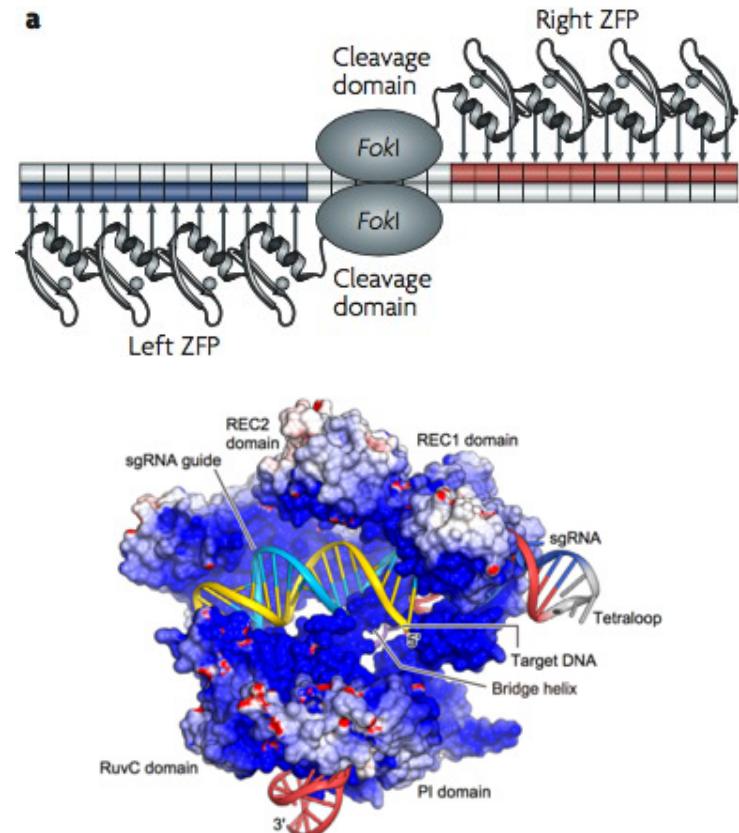
- Endonucleases: ZNFs, TALENs. **Repair pathways. Diversity of outcomes.**
- CRISPR-Cas9: origins, discovery, optimization, base-editing
- Next-generation: Perturb-seq, multiplexing, delivery, applications

How can we edit a genome?



How do we create artificial DNA DSBs?

- Using programmable DNA nucleases
- Zinc Fingers/TALENs: DNA binding provided by modular protein domains
- CRISPR: RNA-guided DNA cleavage



Beyond cutting DNA

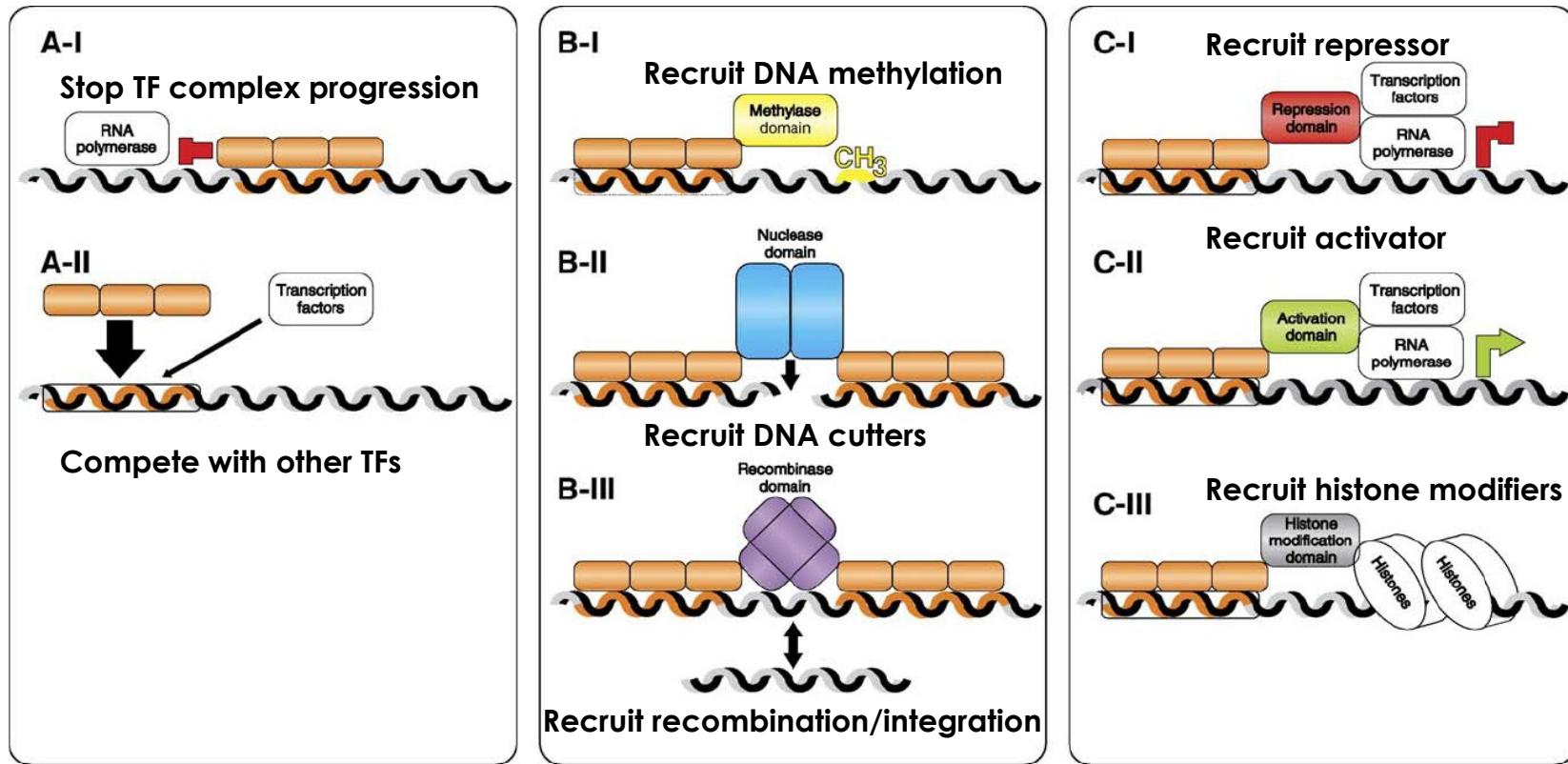


Fig. 2. Applications of designer Cys₂His₂ zinc fingers. The mechanism of action of artificial zinc finger factors depends on the DNA target site (grey and black DNA helix) and the effector domain (if any) attached to zinc fingers (3 orange rectangles). In each case DNA target bound by zinc fingers is marked in orange and framed, if it resides within promoter region. In absence of any effector domains (panel A) zinc fingers can be used as a natural barrier for the transcriptional complex progression (A-I) or to compete with other proteins (natural activators or repressors), which occupy the promoter region (A-II). Zinc fingers can be also used to create chimeric enzymes (panel B) capable of directly affecting DNA e.g. methylating (B-I), cleaving (B-II), mediating recombination/integration (B-III). In this mode of action the zinc finger binding site can be chosen in or outside the promoter region (as represented by dotted framing over binding site in B-I). Engineered zinc fingers can be used as recombinant transcription factors targeted to the promoter region of the gene of interest (panel C). In this scenario novel transcription factors interact with other elements of transcriptional machinery and depending on the character of their effector domain exert either inhibitory (C-I) or stimulatory (C-II) effect on transcription. Engineered chromatin remodeling factors (C-III) can mediate either activation or repression.

High-throughput expts: MPRA, SHARPR, HiDRA, Perturb-seq

1. High-throughput synthesis: Massively Parallel Reporter Assays (MPRA)

- MPRA technology: in vitro synthesis, reporter design, transfection
- 1 site X deep: single-base dissection, combinatorial changes, high-res map
- 2k sites x shallow: motif-guided perturbations, activators/repressors

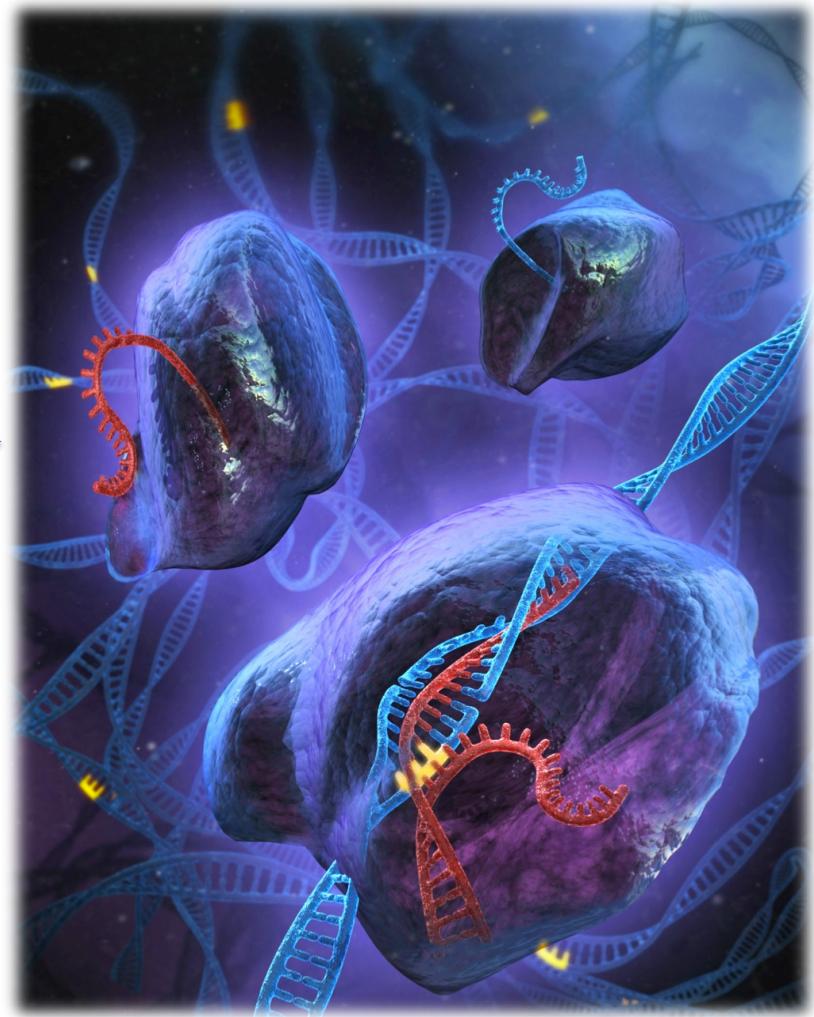
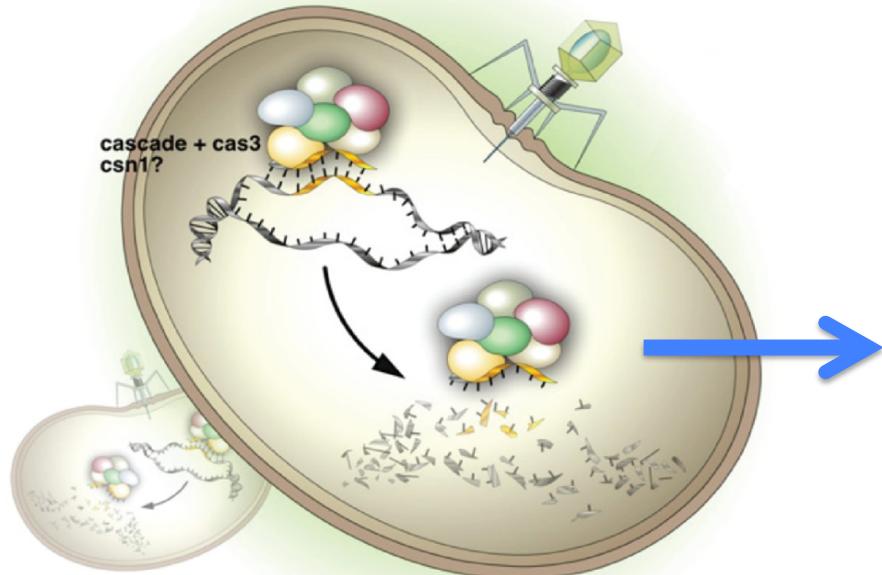
2. Next-generation applications of MPRA + STARR-Seq, SHARPR, HiDRA

- 10k x deep: high-resolution dissection by tiling + deconvolution (Sharpr)
- 10M x deep: HiDRA, no synthesis (STARR), targeted capture (ATAC)

3. Endogenous genome editing: cutting and template-based repair

- Endonucleases: ZNFs, TALENs. Repair pathways. Diversity of outcomes.
- CRISPR-Cas9: origins, discovery, optimization, base-editing
- Next-generation: Perturb-seq, multiplexing, delivery, applications

from yogurt to precise genome perturbation



programmable Genome Positioning System enables precise editing and modulation of the genome

1987

Ishino *et al.*, *J. Bacteriol.*
(discovery of CRISPR sequence)

2007

Barrangou *et al.*, *Science*
(elucidation of an α -bacterial phage function, type II CRISPR)

2008

Brouns *et al.*, *Science*
(elucidation of an α -bacterial phage function, type I CRISPR)

Marraffini *et al.*, *Science*
(elucidation of plasmid DNA-interference function, type III CRISPR)

2010

Garneau *et al.*, *Nature*
(type II CRISPR cuts phage DNA)

2011

Deltcheva *et al.*, *Nature*
(type II CRISPR needs 3 components:
CRISPR RNA, tracrRNA, and Cas9*)

Sapranauskas *et al.*, *Nucleic Acids Res.* (type II CRISPR from *S. thermophilus* can be transplanted into *E. coli*)

* Cas9 also called Csn1, Cas9, Csx12

Nucleotide Sequence of the *iap* Gene, Responsible for Alkaline Phosphatase Isozyme Conversion in *Escherichia coli*, and Identification of the Gene Product

YOSHIZUMI ISHINO, HIDEO SHINAGAWA, KOZO MAKINO, MITSUKO AMEMURA, AND ATSUO NAKATA*
Department of Experimental Chemotherapy, The Research Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565, Japan

CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes

Rodolphe Barrangou,¹ Christophe Fremaux,² Hélène Deveau,³ Melissa Richards,¹ Patrick Boyaval,² Sylvain Moineau,³ Dennis A. Romero,¹ Philippe Horvath^{2*}

Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes

Stan J. J. Brouns,^{1*} Matthijs M. Jore,^{2*} Magnus Lundgren,³ Edze R. Westra,¹ Rik J. H. Stilkhuis,¹ Ambrosius P. L. Snijders,² Mark J. Dickman,² Kira S. Makarova,³ Eugene V. Koonin,³ John van der Oost^{1†}

CRISPR Interference Limits Horizontal Gene Transfer in *Staphylococci* by Targeting DNA

Luciano A. Marraffini and Erik J. Sontheimer*

The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA

Josiane E. Garneau¹, Marie-Eve Dupuis¹, Manuela Villion¹, Dennis A. Romero², Rodolphe Barrangou², Patrick Boyaval³, Christophe Fremaux², Philippe Horvath³, Alfonso H. Magdaleno¹ & Sylvain Moineau¹

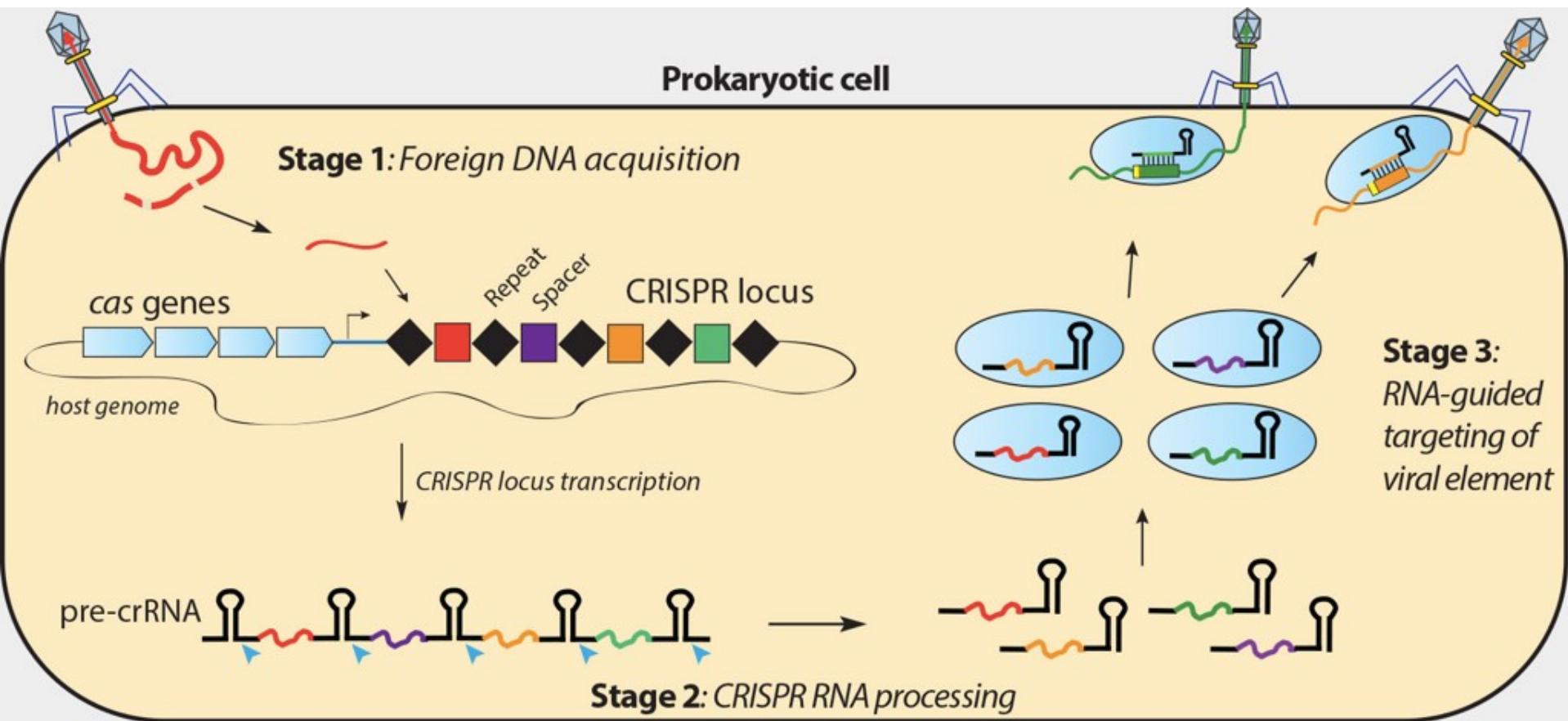
CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III

Eliza Deltcheva^{1,2}, Krzysztof Chylinski^{1,2*}, Cynthia M. Sharma^{3*}, Karine Gonzales², Yanjie Chao^{3,4}, Zaid A. Pirzada², Maria R. Eckert¹, Jörg Vogel^{1,4} & Emmanuelle Charpentier^{1,2}

The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*

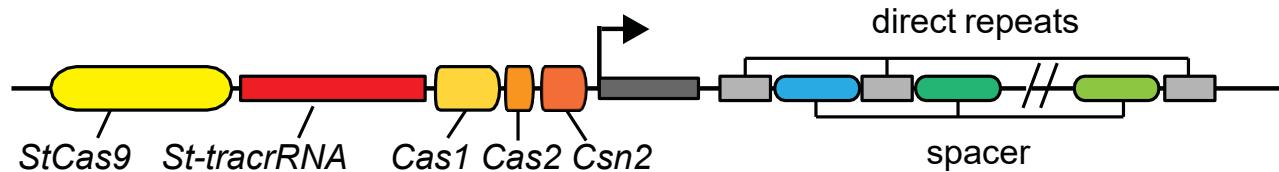
Rimantas Sapranauskas¹, Giedrius Gasiunas¹, Christophe Fremaux², Rodolphe Barrangou³, Philippe Horvath² and Virginijus Siksnys^{1,*}

CRISPR (clustered regularly interspaced short palindromic repeats) is a naturally occurring bacterial immune system

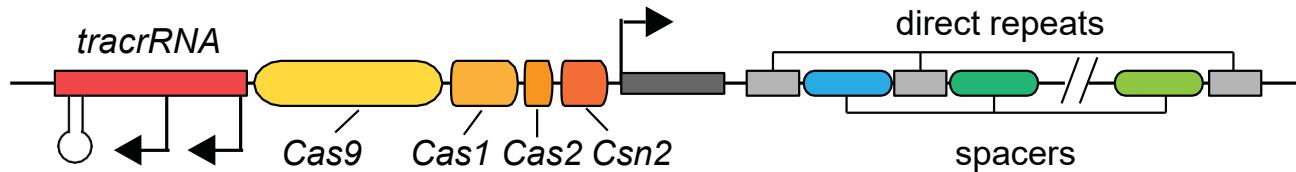


CRISPR-Cas9: an RNA-guided nuclease

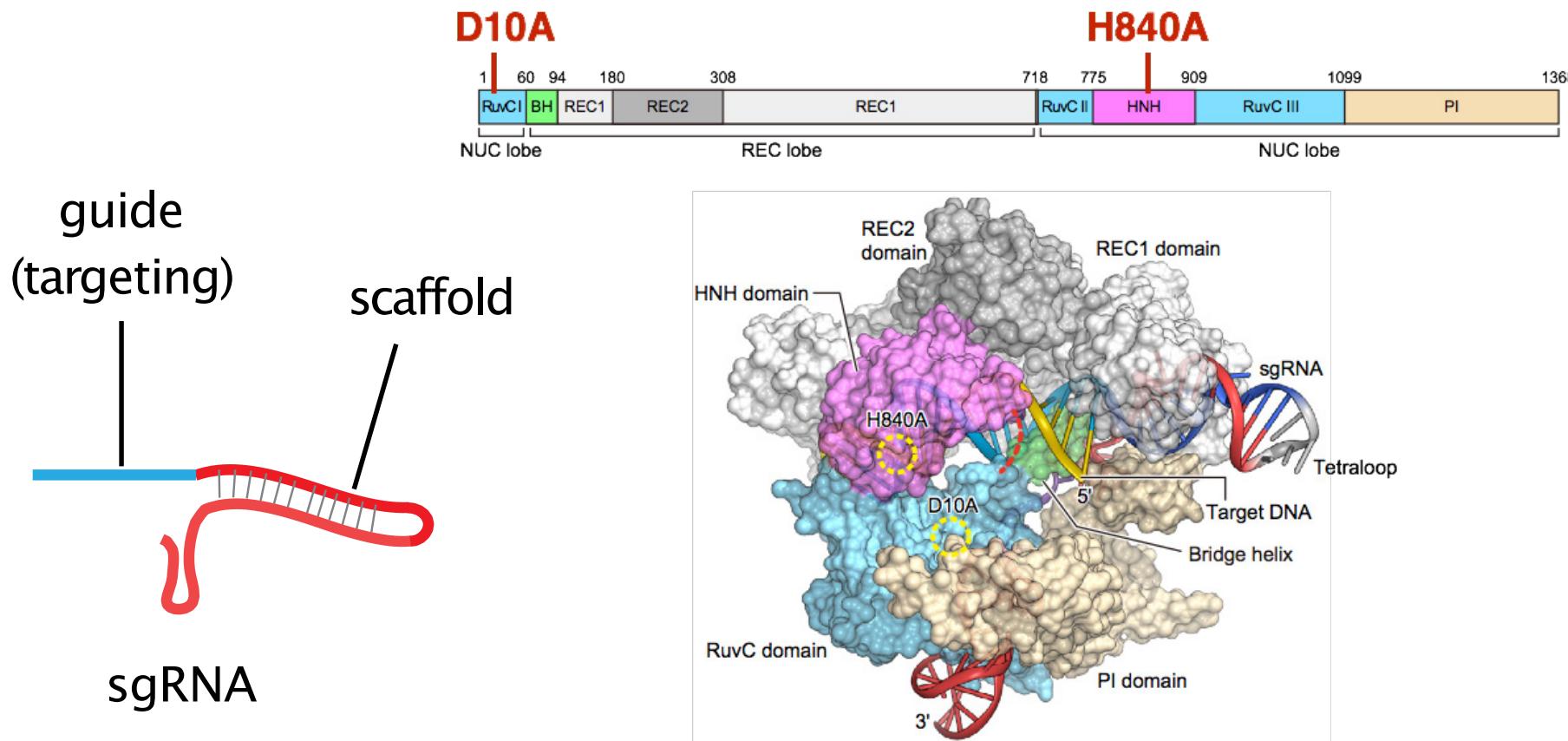
Streptococcus thermophilus LMD-9 CRISPR1



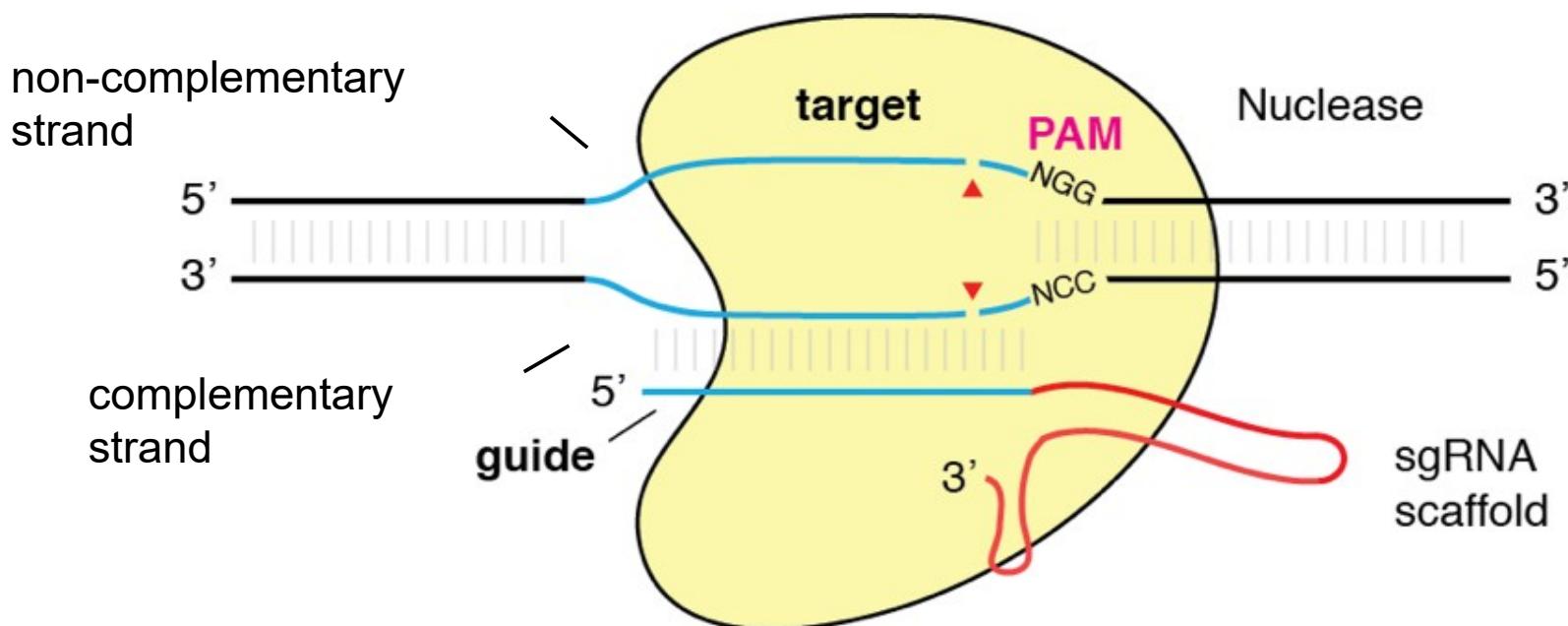
Streptococcus pyogenes SF370 type II CRISPR locus



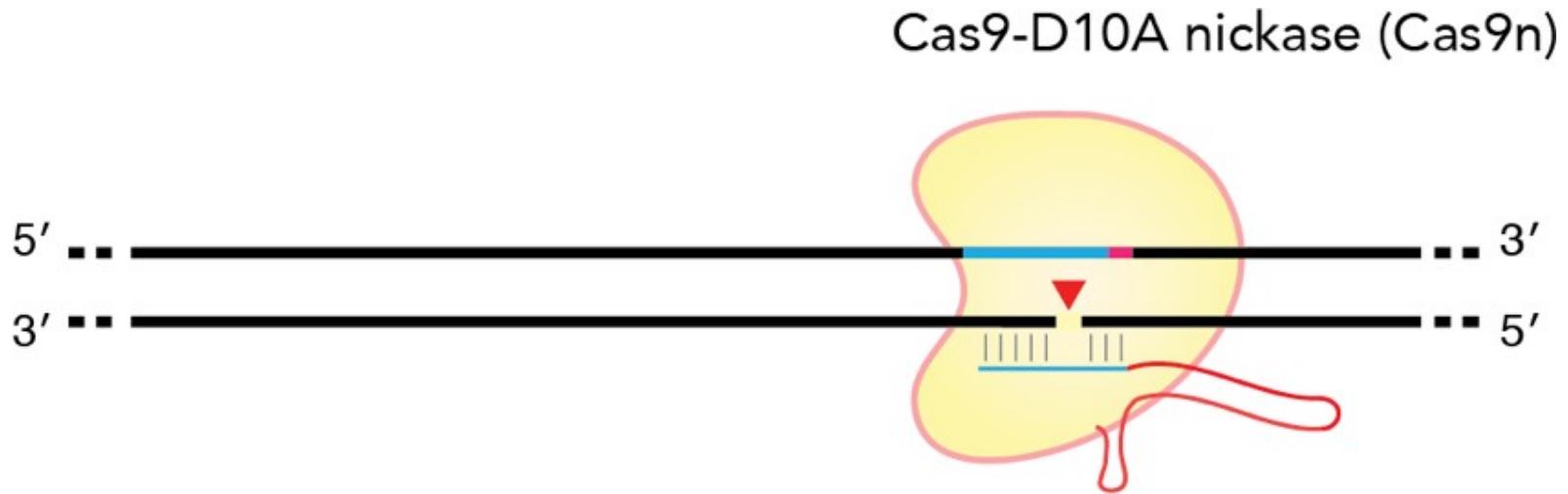
CRISPR-Cas9: an RNA-guided DNA endonuclease



Cas9 from *S. pyogenes* cleaves upstream of a NGG motif



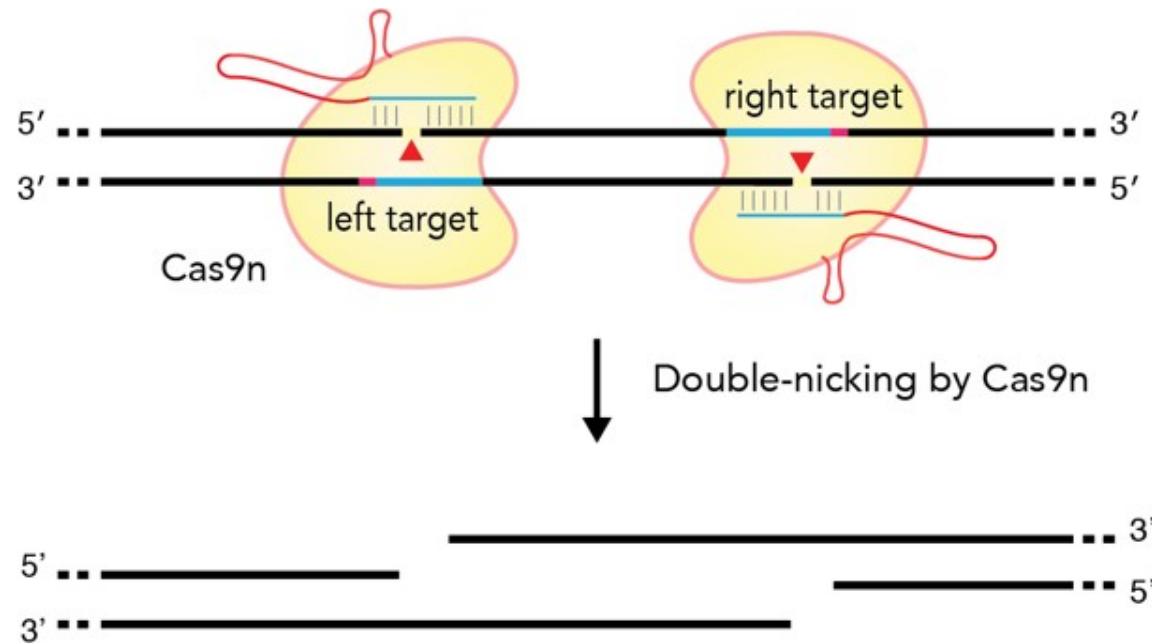
What is nickase and what does it do?



Cas9n **nicks** DNA instead of inducing double-stranded breaks

Single nicks typically does not result in indel mutations, but can **induce HDR (knockin)**, although with lower efficiency

What is nickase and what does it do?



Paired Cas9n can be used to introduce DSBs by **double-nicking** to improve **targeting specificity**.

For guide design purposes, we use the same principles

High-throughput expts: MPRA, SHARPR, HiDRA, Perturb-seq

1. High-throughput synthesis: Massively Parallel Reporter Assays (MPRA)

- MPRA technology: in vitro synthesis, reporter design, transfection
- 1 site X deep: single-base dissection, combinatorial changes, high-res map
- 2k sites x shallow: motif-guided perturbations, activators/repressors

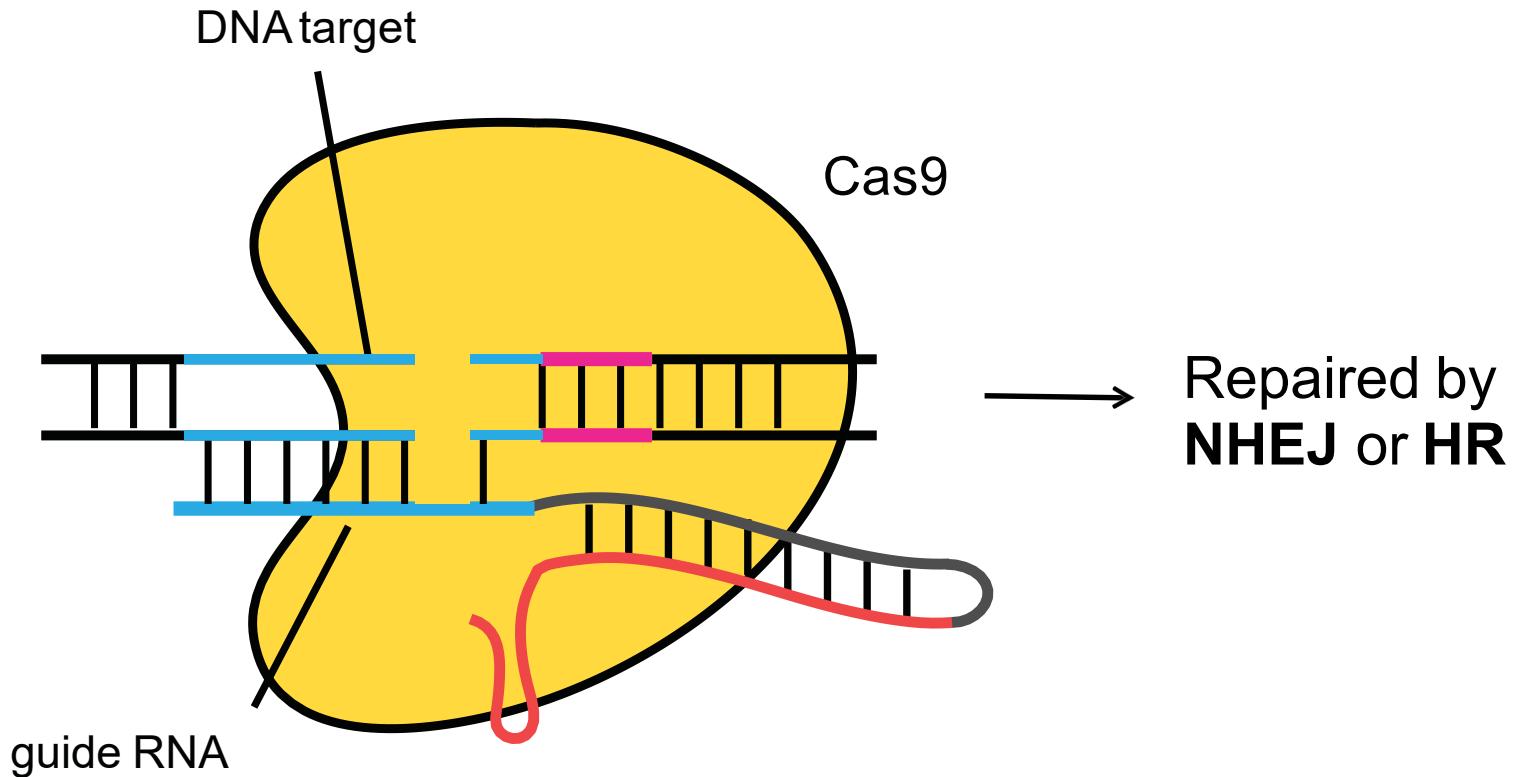
2. Next-generation applications of MPRA + STARR-Seq, SHARPR, HiDRA

- 10k x deep: high-resolution dissection by tiling + deconvolution (Sharpr)
- 10M x deep: HiDRA, no synthesis (STARR), targeted capture (ATAC)

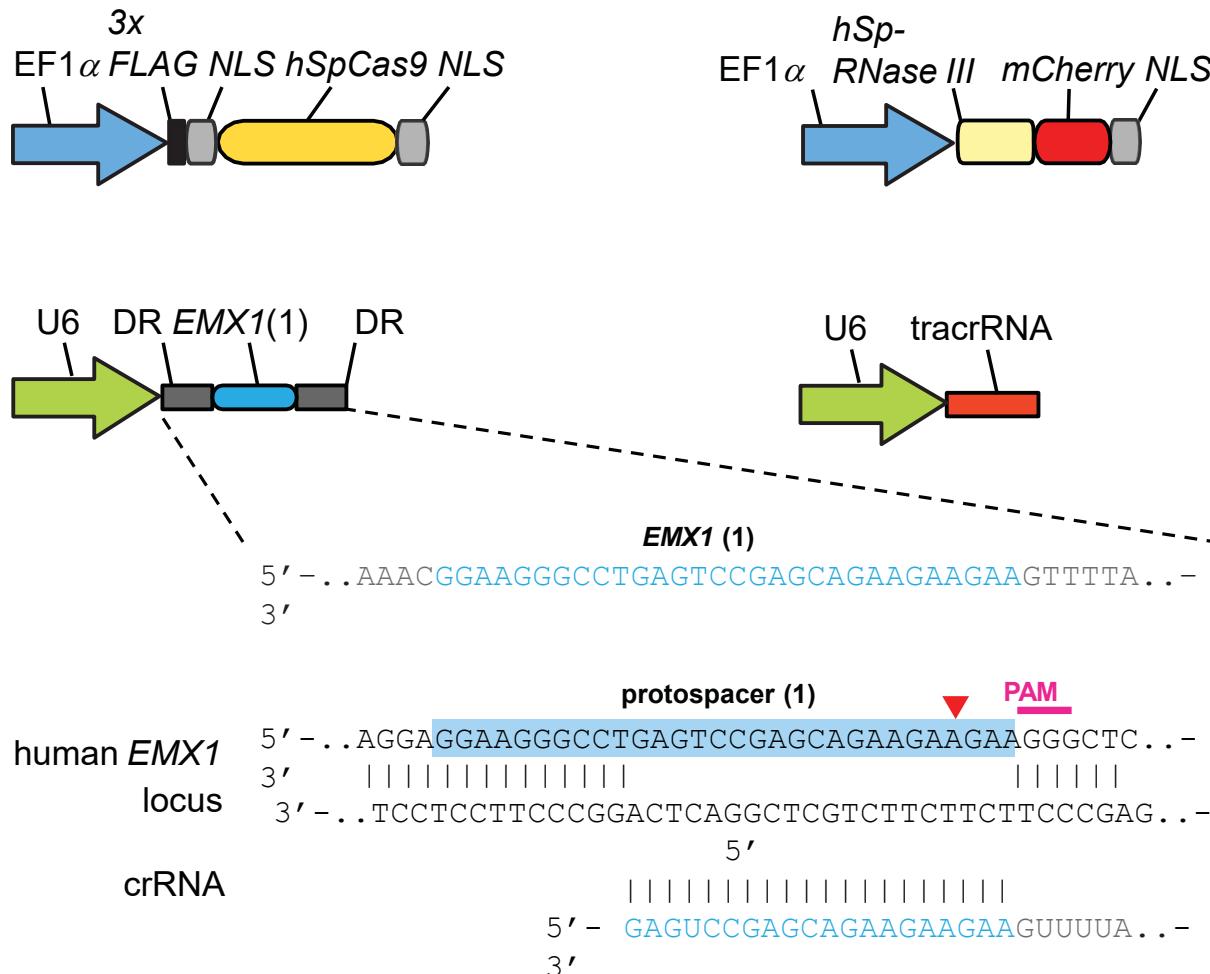
3. Endogenous genome editing: cutting and template-based repair

- Endonucleases: ZNFs, TALENs. Repair pathways. Diversity of outcomes.
- CRISPR-Cas9: origins, discovery, optimization, base-editing
- Next-generation: Perturb-seq, multiplexing, delivery, applications

Bringing CRISPR into mammalian cells

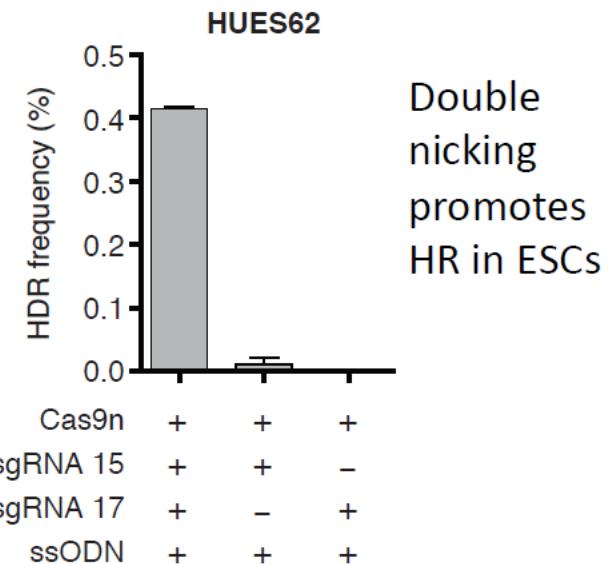
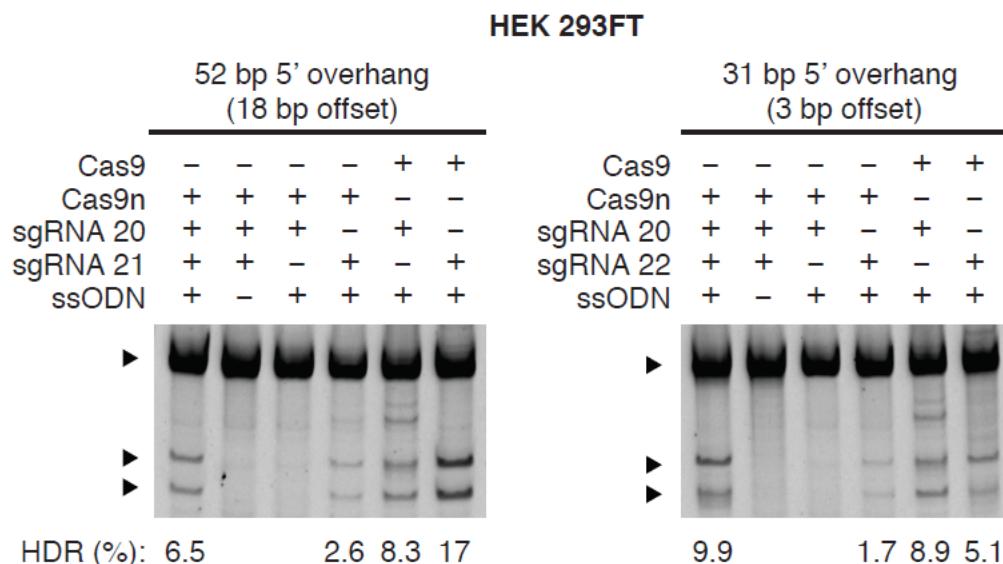
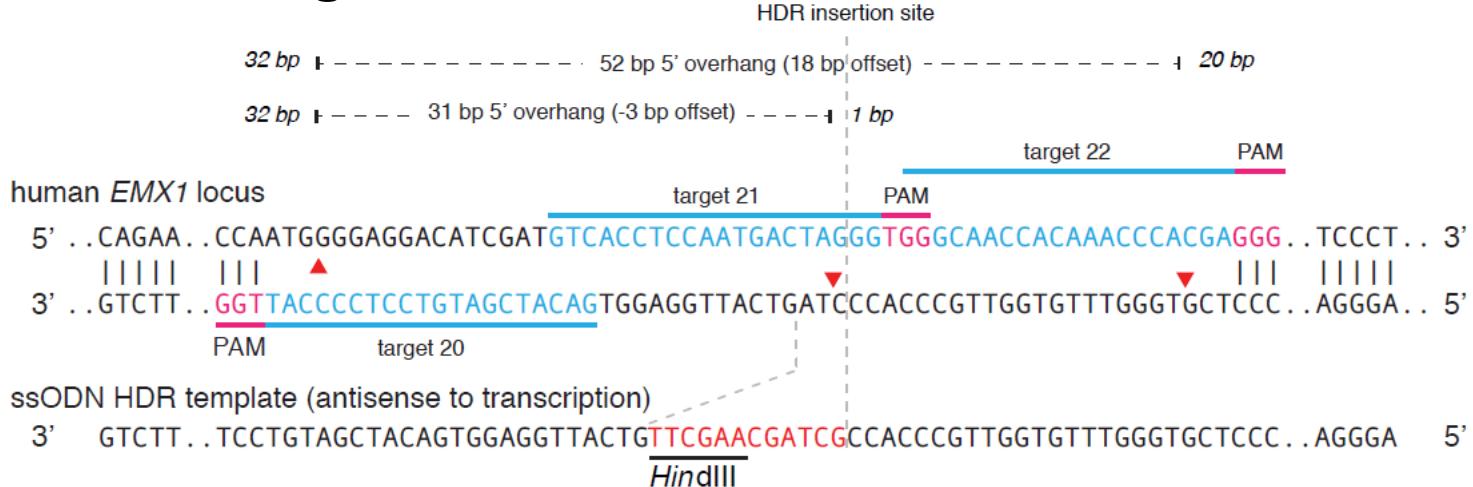


Expressing a minimal CRISPR system in mammalian cells



Cong et al., Science 2013

Double nicking stimulates efficient HR



Ran*, Hsu* et al., Cell
2013

Computational Tool for Identification of Guide Sequences

Optimized CRISPR Design x
crispr.mit.edu/job/8296402127

CRISPR Design Alpha version .02 About Contact

The CRISPR design tool is an alpha release undergoing active development.... [show more](#)

Optimized CRISPR Design

For query sequence "test" [... show info](#)

Downloadable results: [summary.pdf](#) [offtargets.csv](#) [... show warning](#)

Interactive results: mouse over a guide or explore below for details

HUMAN unmapped

all guides

scored by inverse likelihood of offtarget binding
mouse over for details [... show legend](#)

score	sequence
95	AGTCCAGCAGCTCGCTCCGC CGG
93	GCTGGCTCGTCGTAGTTTTT TGG
84	CTCGTCGGAGCTGCAGGGAC CGG
78	CGAGGGGACCATGTGGCTC AGG
84	GAGTGCTGGACTGTGGTG AGG
79	CTTAGCCACCCCTGAGGCCACA TGG

guide #3 quality score: 84

guide sequence: CTCGTCGGAGCTGCAGGGAC CGG
on-target locus: chr???:??
number of offtarget sites: 45 (9 are in genes)

top 20 genome-wide off-target sites

sequence	score	mismatches	UCSC gene	locus
CTGAACGGAGCTGCAGGGACAAG	2.3	3MMs [2:3:4]	uc001wlj.2	chr14:-24527872
CTAGGTGGAGCTGCAGGGACAAG	1.4	3MMs [2:4:5]		chr9:+13461052
CGCGGGCGGAGCTGCAGGGGACTGG	0.8	3MMs [1:4:14]		chr19:-8578876
CCCCGGCGGAGCTGCAGGGACGGG	0.8	3MMs [1:4:14]		chr12:+66135639
CITGACCGGAGCTGCAGGGACAGG	0.6	3MMs [2:4:17]		chr7:+55656361
ATACTCCCACTGCAGGGACAAG	0.6	5MMs [0:2:3:6:7]		chr1:+221356693
GIGTCCCTGAGCTGCAGGGACAAG	0.6	5MMs [0:2:3:4:6]		chr8:-68542438
CGGGACAAAGCTGCAGGGACAAG	0.6	5MMs [1:2:4:6:7]		chr8:+145989888
GATGGCGAGACTGCAGGGACAAG	0.6	5MMs [0:1:2:4:6]		chr7:-38780575
ATCCACCGAAGCTGCAGGGACAAG	0.6	5MMs [0:3:4:6:9]		chr7:-94260264
GCAGTCAGAACCTGCAGGGACAAG	0.5	5MMs [0:1:2:6:9]		chr2:+73311138
GGCTTCGGAGCTGCAGGGCTGG	0.5	4MMs [0:1:3:18]		chr11:-63891720

-Integrating up-to-date off-target analysis data

-154GB RAM server

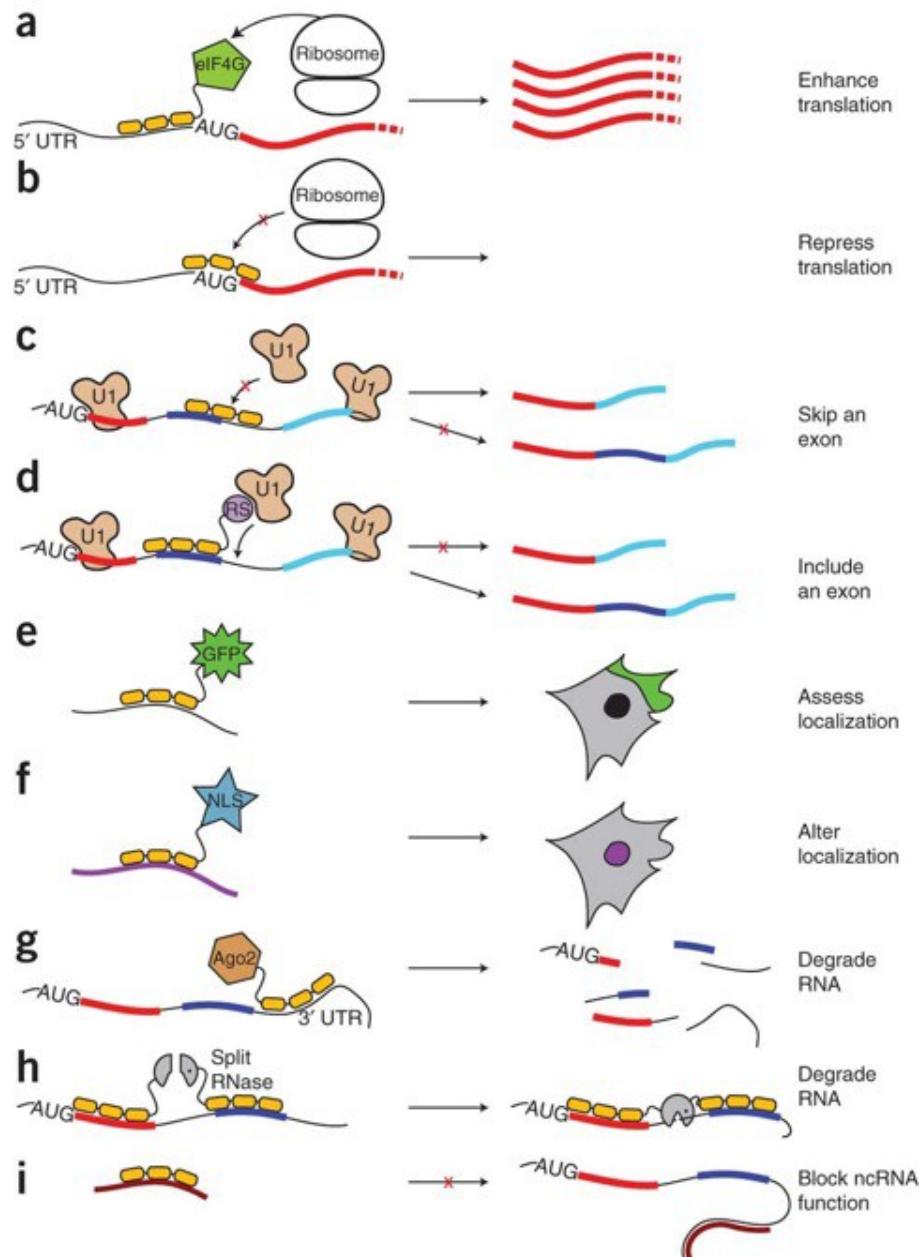
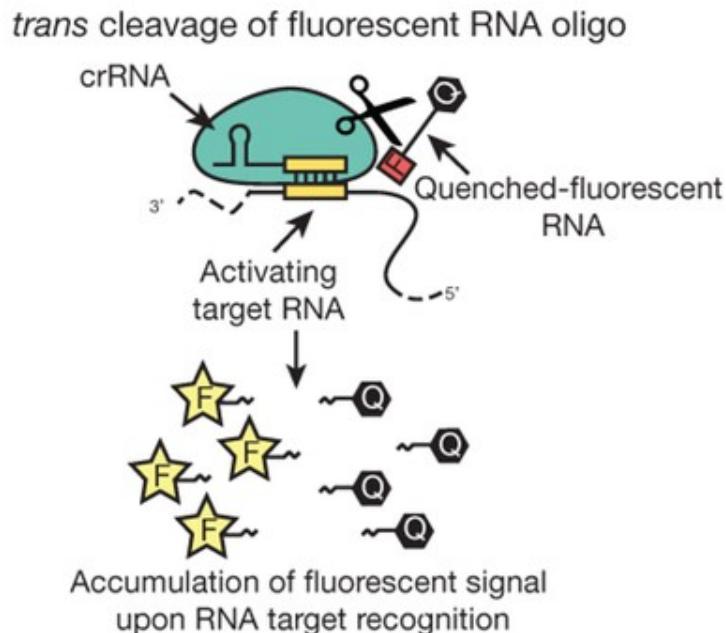
-real-time genome-wide searches

-design reagents for off-target analysis

<http://tools.genome-engineering.org>

Hsu et al., Nature Biotechnology 2013

Applications for CRISPR associated RNA effectors



Mackay et al., *Nature Structural & Molecular Biology* 2011

Multiple CRISPR nucleases can cleave mammalian DNA

	PAM Location	PAM sequence	Cleave own CRISPR array?	Cut Type	Notes
<i>S. pyogenes</i> Cas9	3' of target	NGG	No	Blunt	Most consistent
<i>S. aureus</i> Cas9	3' of target	GGRRT	No	Blunt	Fits into AAV
<i>L. bacterium</i> Cpf1	5' of target	TTTV	Yes	Staggered	Much shorter sgRNA

High-throughput expts: MPRA, SHARPR, HiDRA, Perturb-seq

1. High-throughput synthesis: Massively Parallel Reporter Assays (MPRA)

- MPRA technology: in vitro synthesis, reporter design, transfection
- 1 site X deep: single-base dissection, combinatorial changes, high-res map
- 2k sites x shallow: motif-guided perturbations, activators/repressors

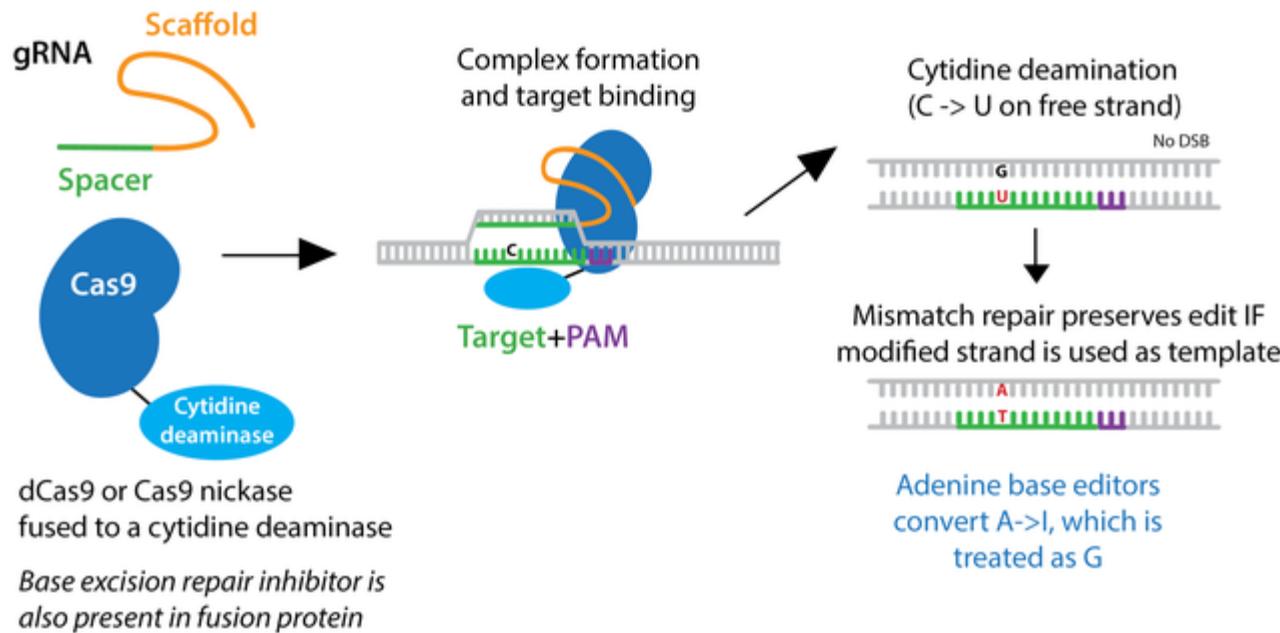
2. Next-generation applications of MPRA + STARR-Seq, SHARPR, HiDRA

- 10k x deep: high-resolution dissection by tiling + deconvolution (Sharpr)
- 10M x deep: HiDRA, no synthesis (STARR), targeted capture (ATAC)

3. Endogenous genome editing: cutting and template-based repair

- Endonucleases: ZNFs, TALENs. Repair pathways. Diversity of outcomes.
- CRISPR-Cas9: origins, discovery, optimization, **base-editing**
- Next-generation: Perturb-seq, multiplexing, delivery, applications

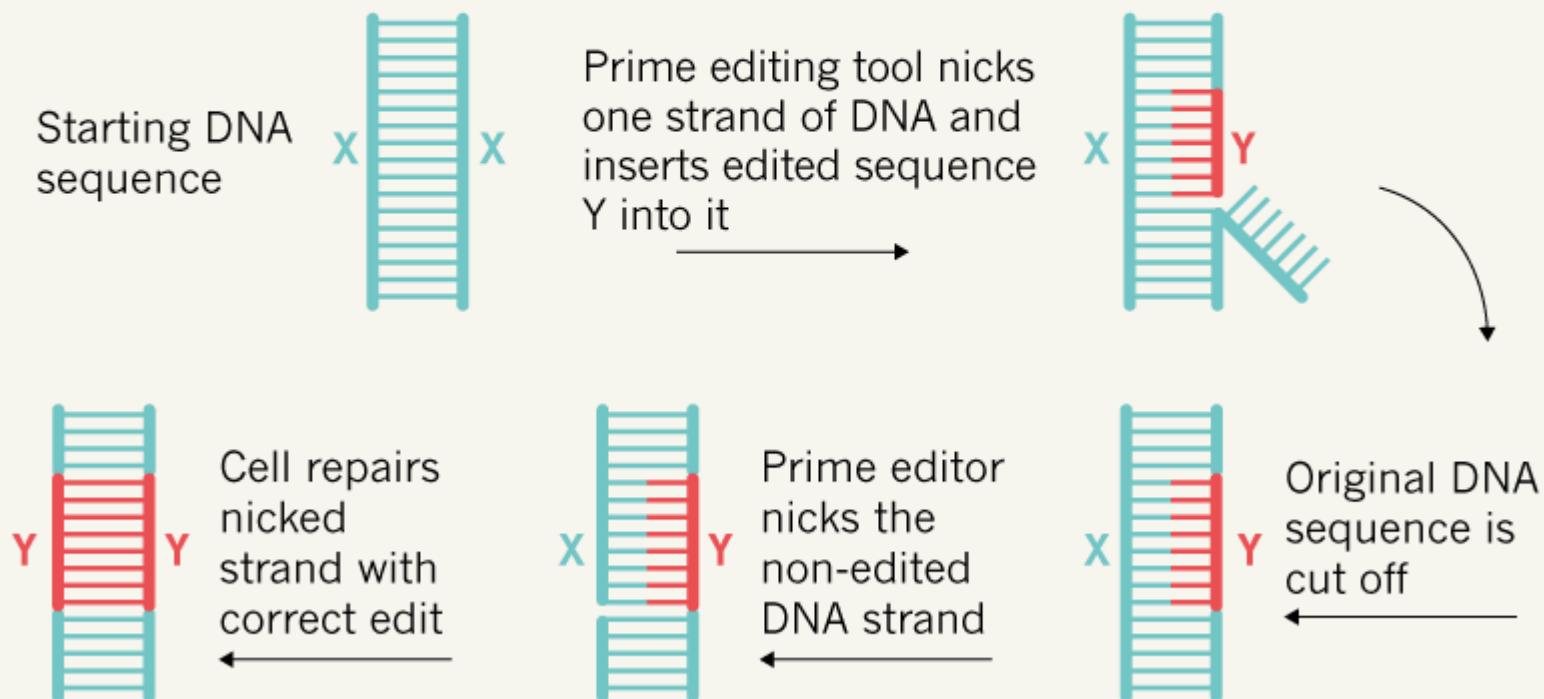
From “genome vandalism” to single-base direct editing



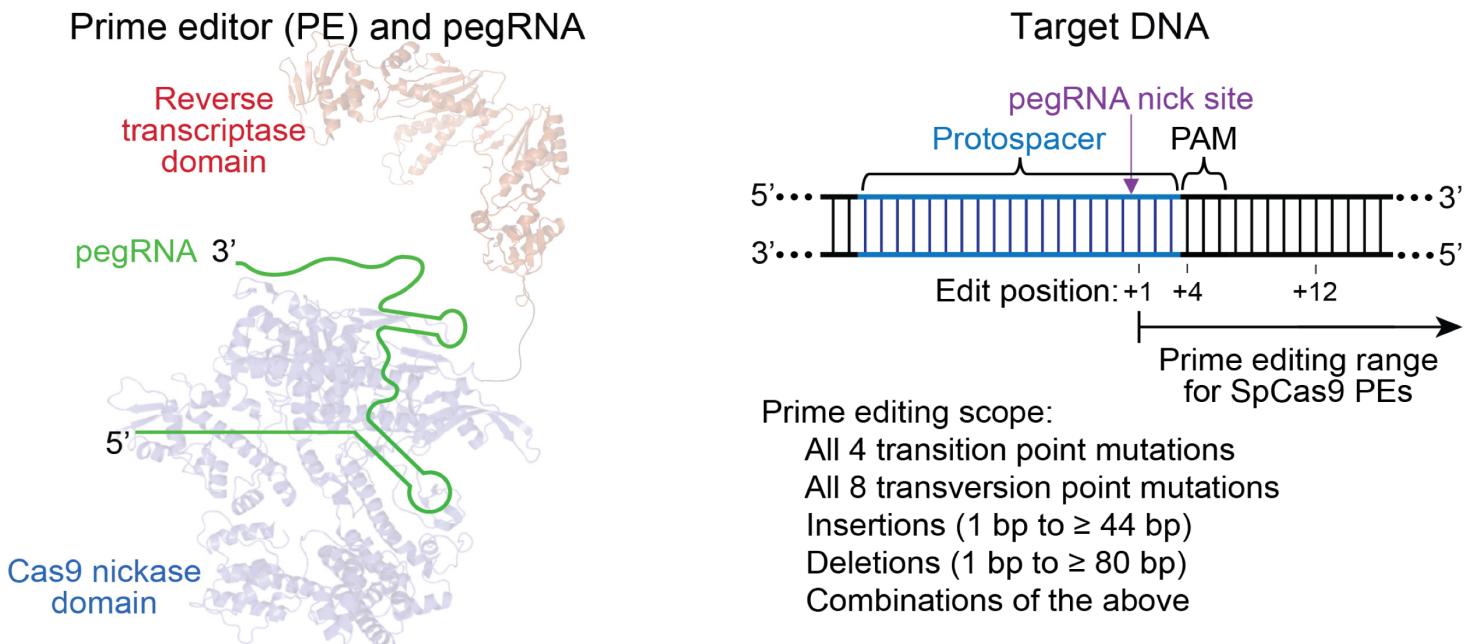
Prime editing

PRECISION EDITOR

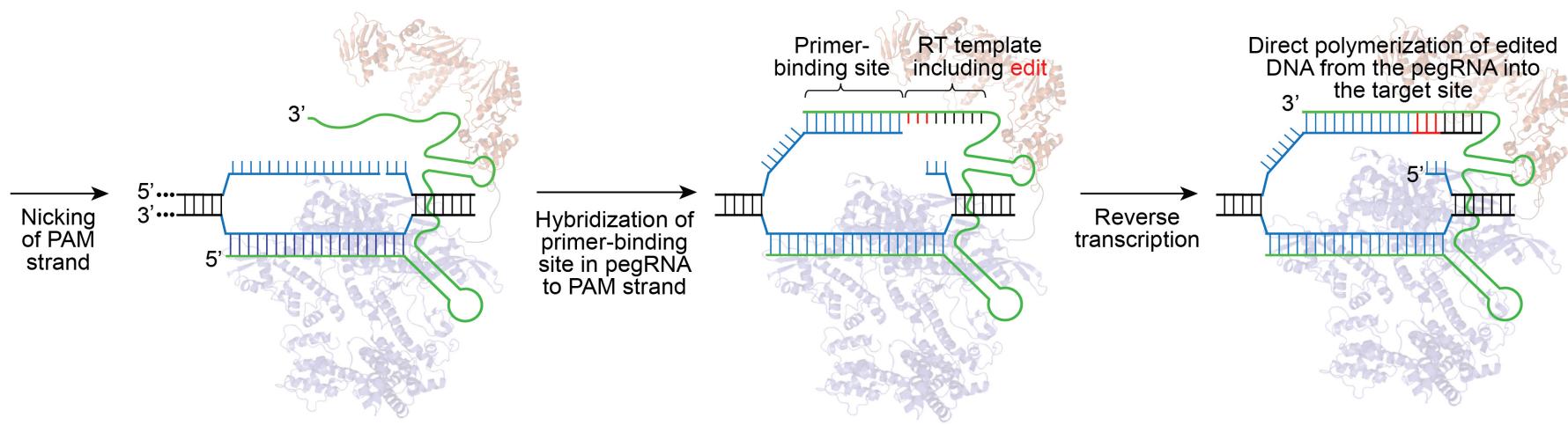
Prime editing reduces the number of unintended changes to a genome by inserting the edits researchers want to make into the DNA itself. This contrasts with CRISPR-Cas9, which relies on the cell's repair system to make the changes.



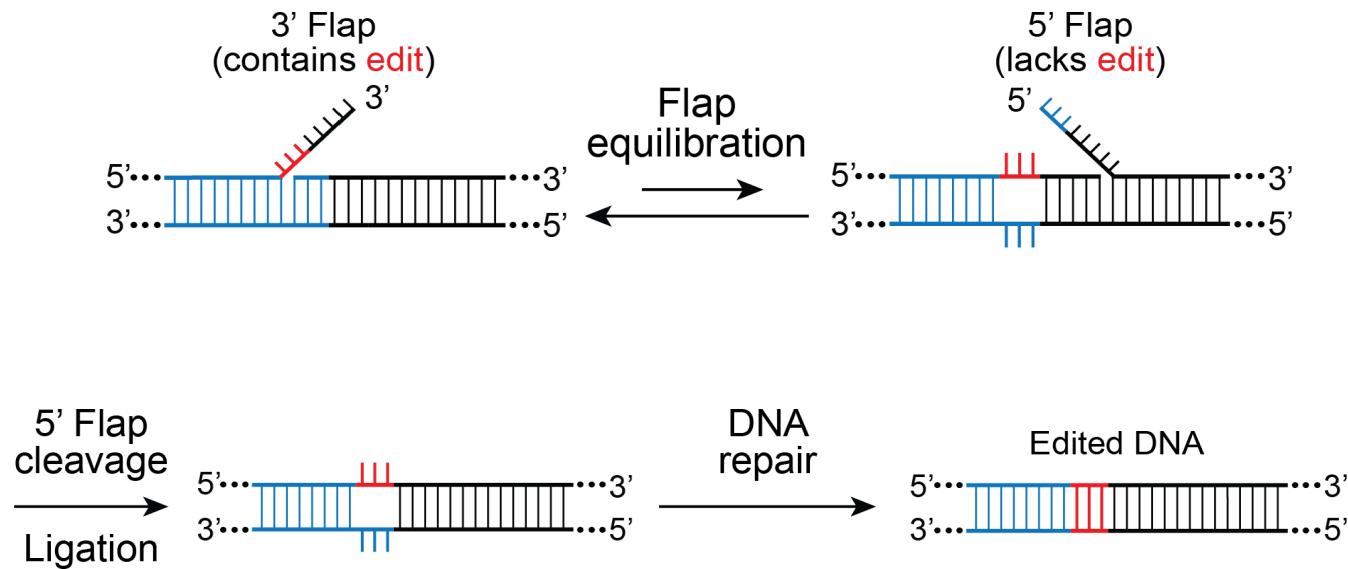
The prime editing system



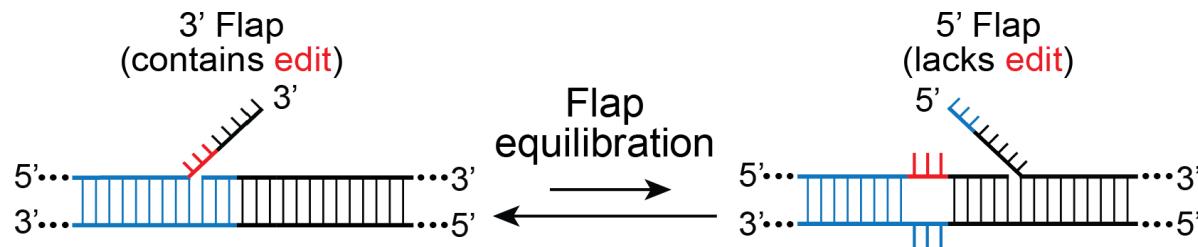
The prime editing system



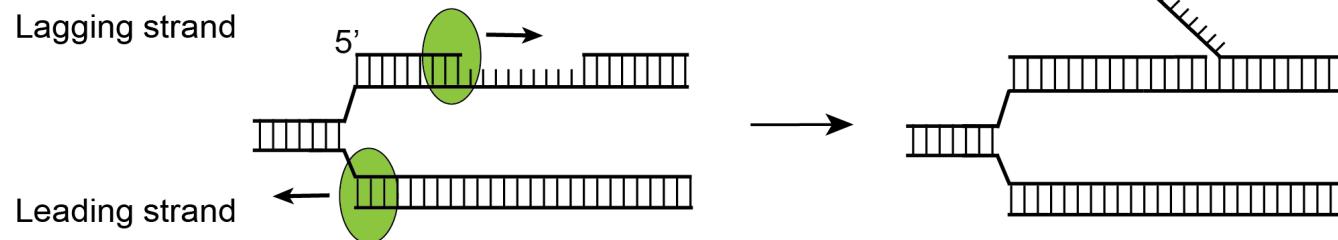
Resolution of DNA flaps generated by target-primed reverse transcription



Resolution of DNA flaps generated by target-primed reverse transcription

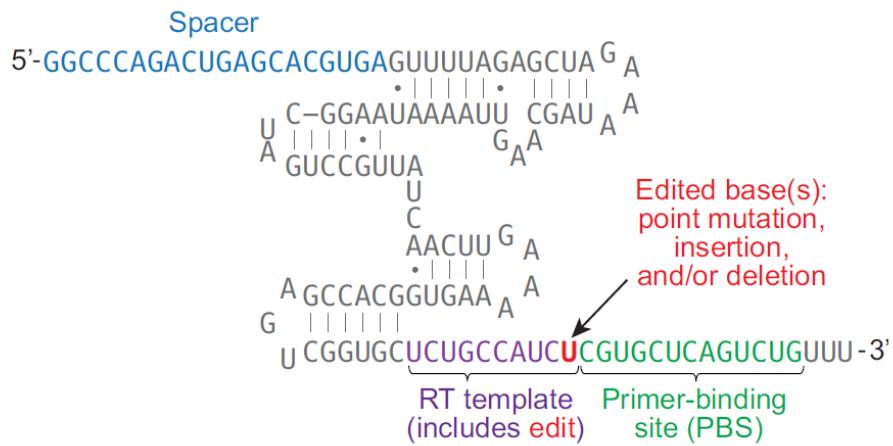


Resembles 5' flaps generated during lagging strand DNA synthesis and long-patch base excision repair.

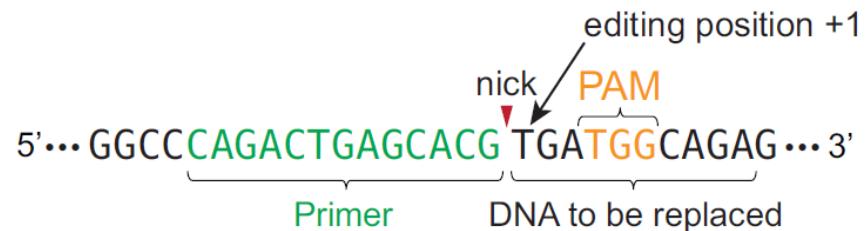


PE1: Cas9(H840A)–M-MLV RT(wt) fusion

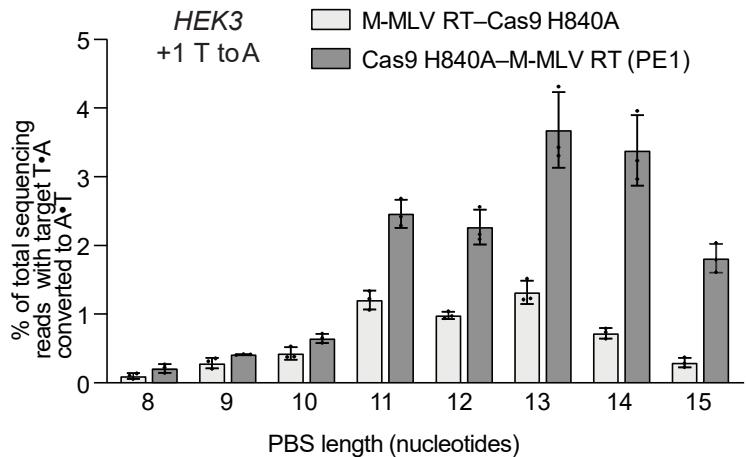
pegRNA



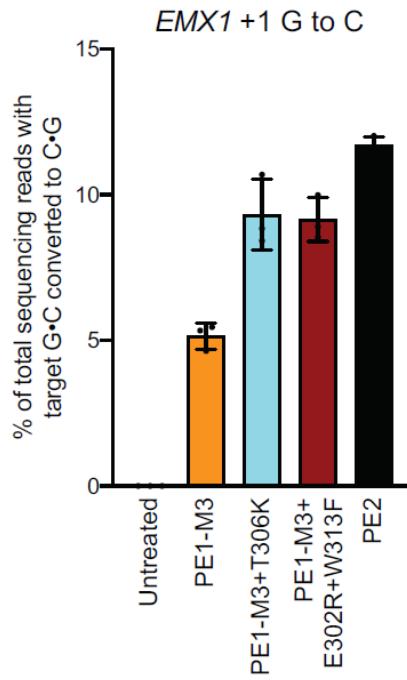
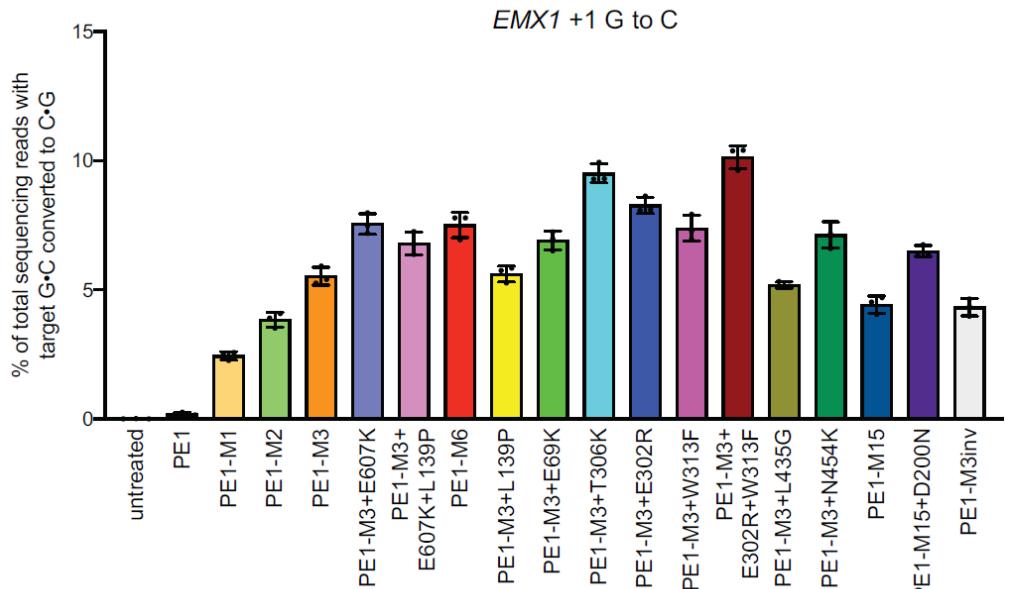
- Cas9–M-MLV(RT) fusions and pegRNAs support genome editing in human cells.



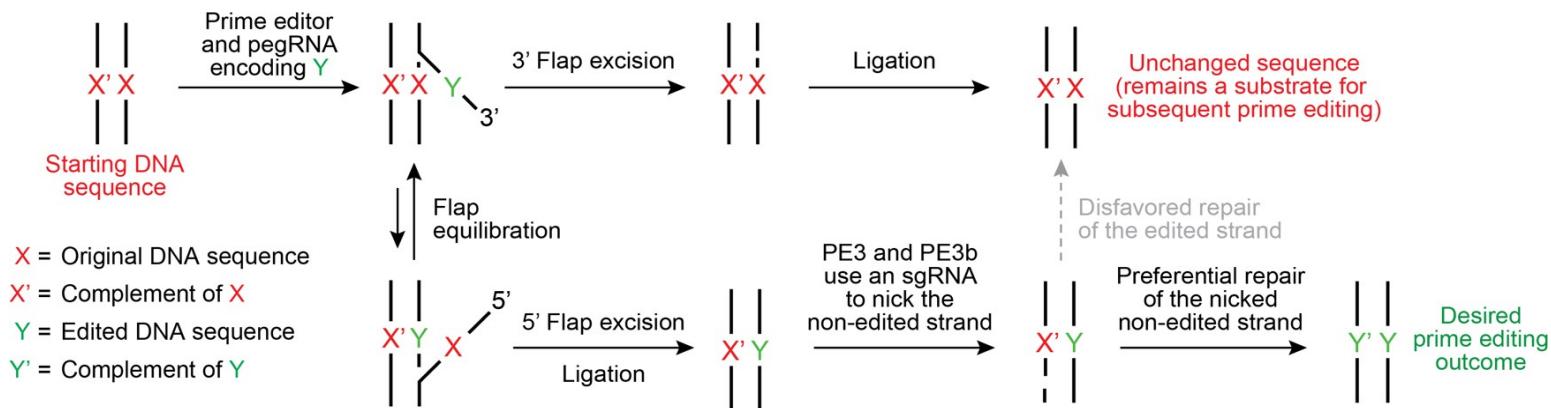
HEK293T cells



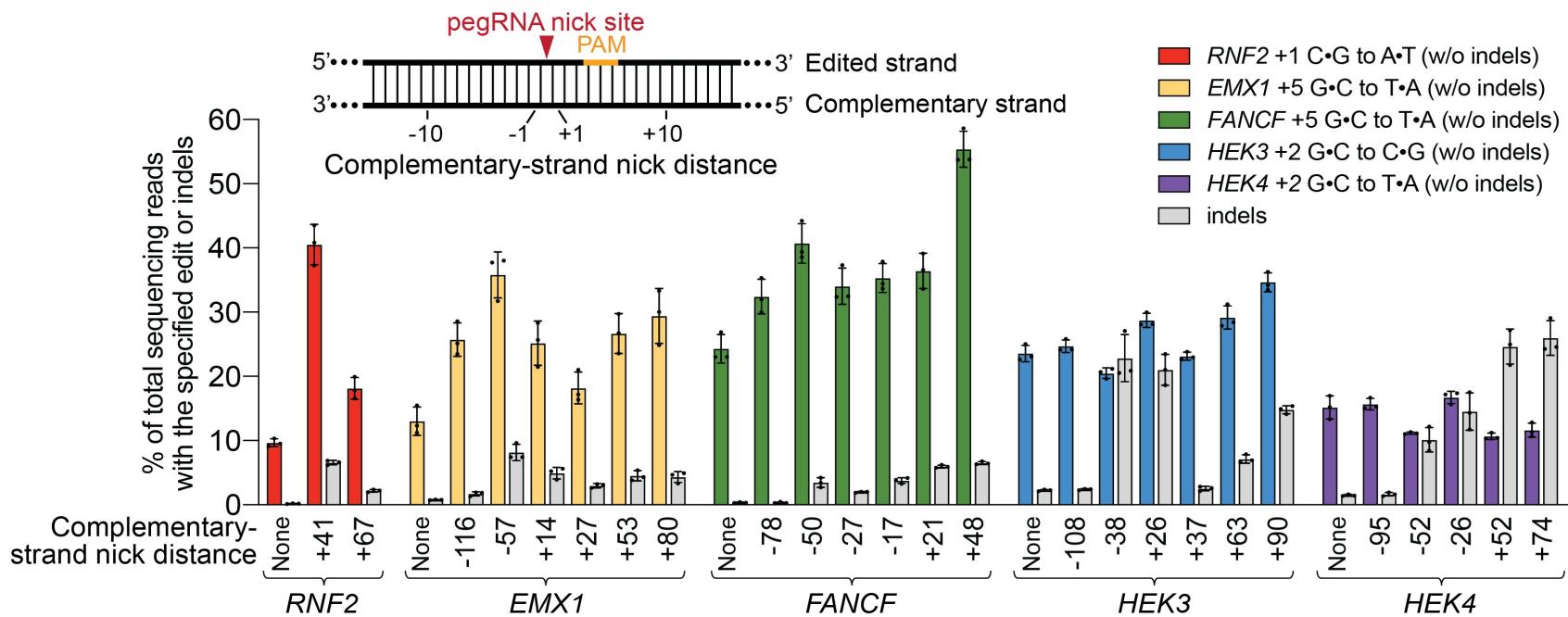
PE2: higher editing with mutant RT



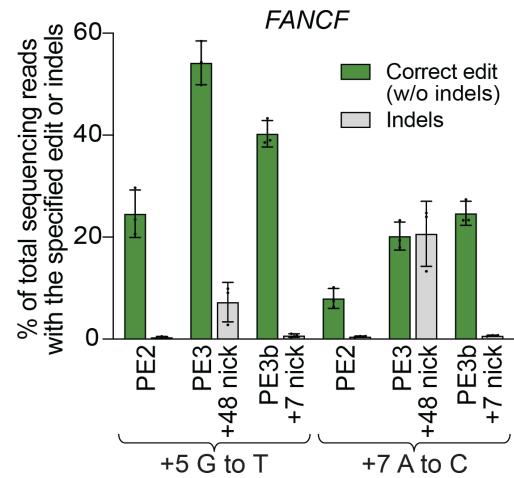
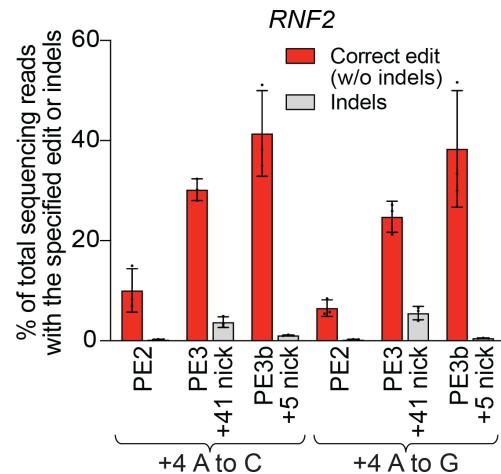
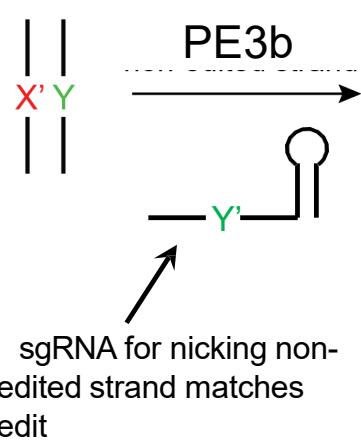
Biasing DNA repair to incorporate the edit



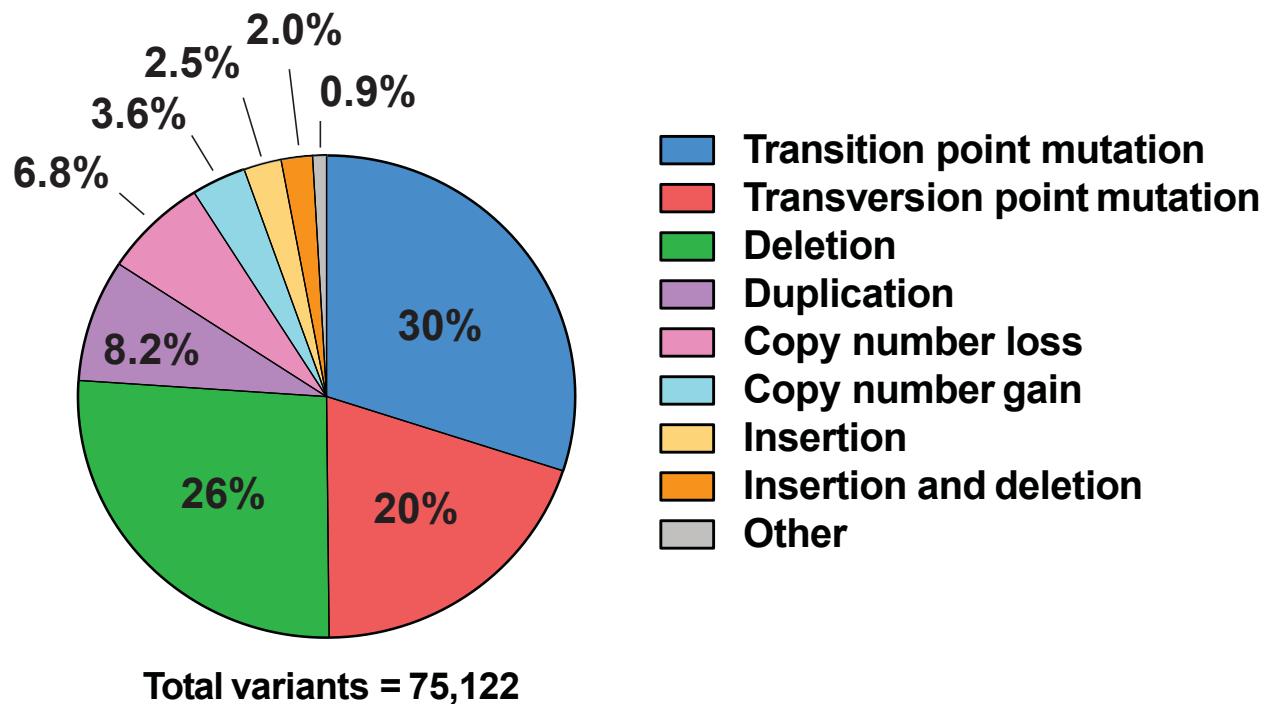
PE3: nicking the unedited strand



PE3b: edit-specific nicking lowers indels



Human pathogenic variants from ClinVar



- ~82% of human disease variants are point mutations (SNPs), ≤ 10-bp deletions / duplications / insertions / indels.

ClinVar database: Landrum et al., *Nucleic Acids Res.* 44, D862 (2016), accessed July, 2019.

High-throughput expts: MPRA, SHARPR, HiDRA, Perturb-seq

1. High-throughput synthesis: Massively Parallel Reporter Assays (MPRA)

- MPRA technology: in vitro synthesis, reporter design, transfection
- 1 site X deep: single-base dissection, combinatorial changes, high-res map
- 2k sites x shallow: motif-guided perturbations, activators/repressors

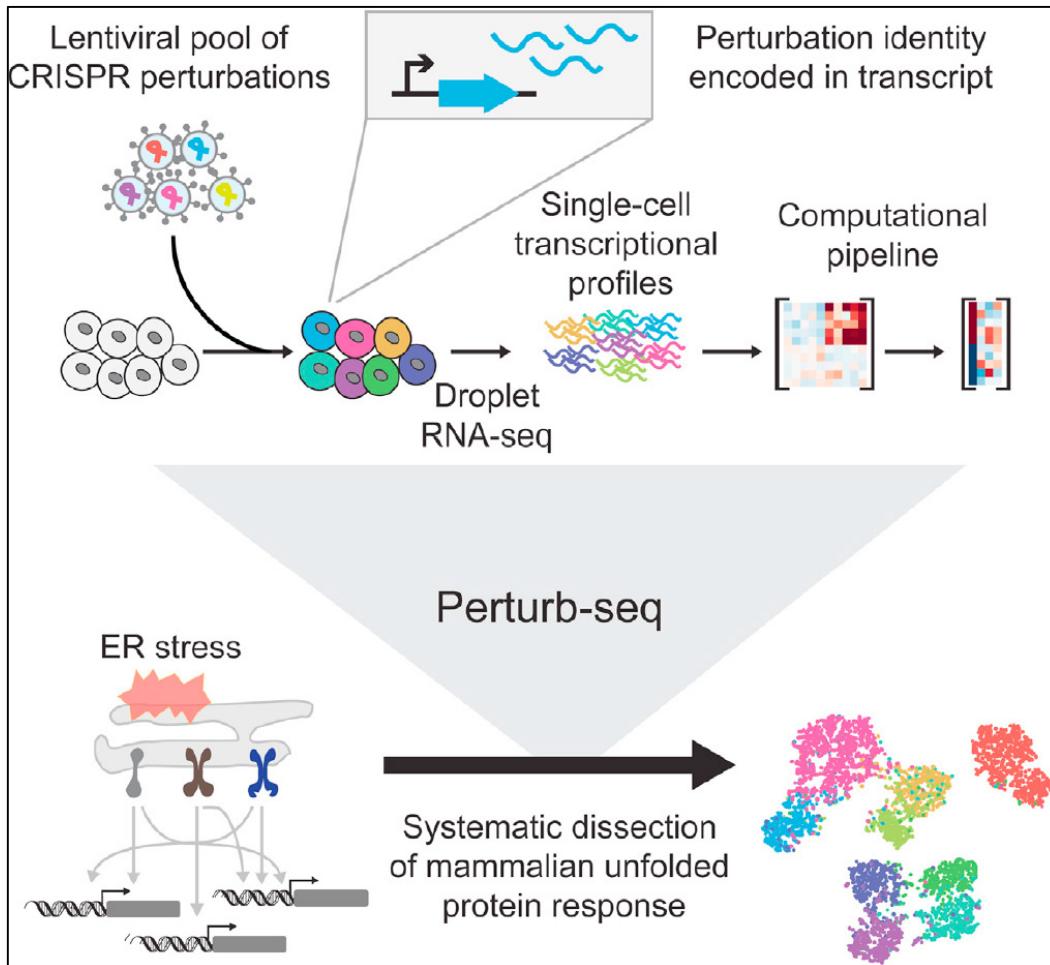
2. Next-generation applications of MPRA + STARR-Seq, SHARPR, HiDRA

- 10k x deep: high-resolution dissection by tiling + deconvolution (Sharpr)
- 10M x deep: HiDRA, no synthesis (STARR), targeted capture (ATAC)

3. Endogenous genome editing: cutting and template-based repair

- Endonucleases: ZNFs, TALENs. Repair pathways. Diversity of outcomes.
- CRISPR-Cas9: origins, discovery, optimization, base-editing, applications
- Therapeutic applications, multiplexing, delivery, next-generation tools

Perturb-Seq: Genome-wide systematic perturbations

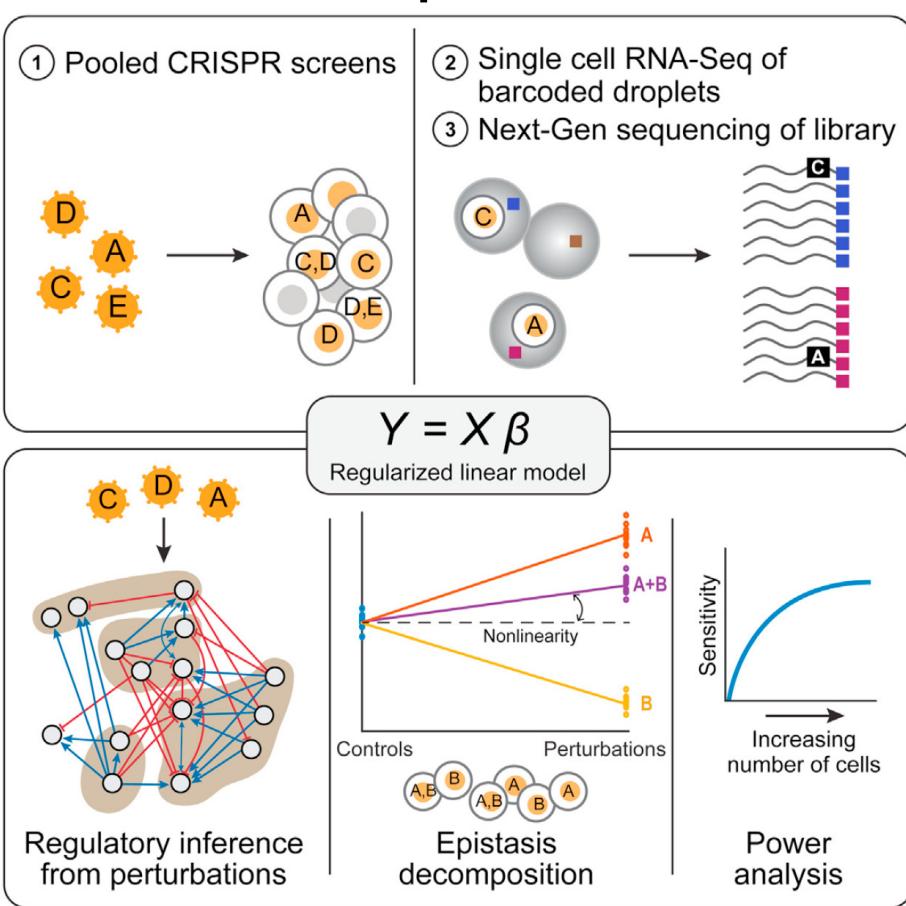


A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response

Britt Adamson,^{1,2,3,4,12} Thomas M. Norman,^{1,2,3,4,12} Marco Jost,^{1,2,3,4,5} Min Y. Cho,^{1,2,3,4} James K. Nuñez,^{1,2,3,4} Yuwen Chen,^{1,2,3,4} Jacqueline E. Villalta,^{1,2,3,4} Luke A. Gilbert,^{1,2,3,4} Max A. Horlbeck,^{1,2,3,4} Marco Y. Hein,^{1,2,3,4} Ryan A. Pak,^{1,8} Andrew N. Gray,^{5,14} Carol A. Gross,^{5,6,7} Atray Dixit,^{9,10} Oren Parnas,^{10,13} Aviv Regev,^{10,11} and Jonathan S. Weissman^{1,2,3,4,15,*}

- Perturb-seq allows parallel screening with rich phenotypic output from single cells
- Simultaneous delivery and identification of up to three CRISPR perturbations
- Genome-scale screens dissect the mammalian unfolded protein response
- Analytical methods separate perturbation responses from confounding effects

Perturb-Seq: Genome-wide systematic perturbations

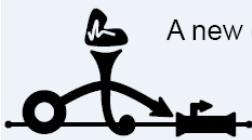


Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens

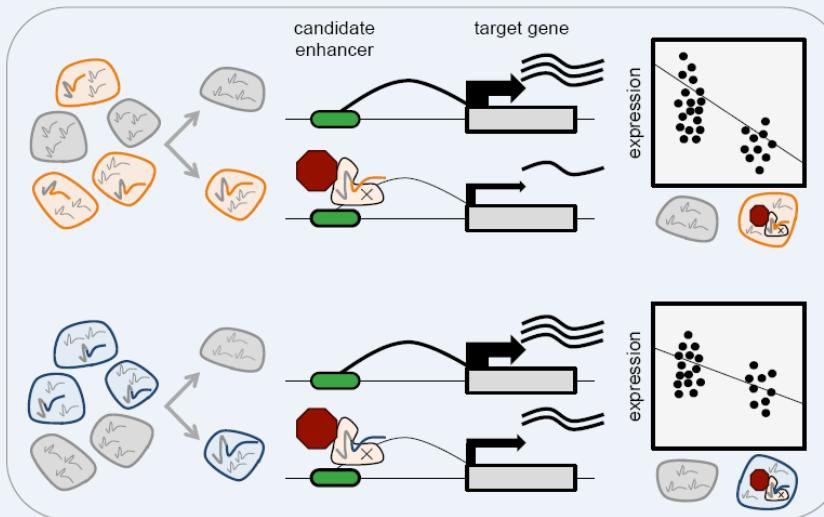
Atray Dixit,^{1,2,9} Oren Parnas,^{1,9,10} Biyu Li,¹ Jenny Chen,^{1,2} Charles P. Fulco,^{1,4} Livnat Jerby-Arnon,¹ Nemanja D. Marjanovic,^{1,3} Danielle Dionne,¹ Tyler Burks,¹ Raktima Raychowdhury,¹ Britt Adamson,⁵ Thomas M. Norman,⁵ Eric S. Lander,^{1,4,6} Jonathan S. Weissman,^{5,7} Nir Friedman,^{1,8}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

- Pooled CRISPR screen with scRNA-seq readout
- Integrated model of perturbations, single cell phenotypes, and epistatic interactions
- Effect of TFs on genes, programs, and states in LPS response in immune cells
- Downsampling assessment of feasibility of genome-wide or combinatorial screens

Genome-wide systematic perturbations

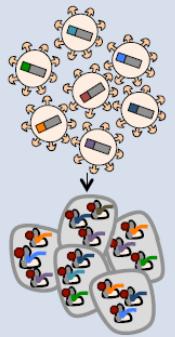


A new eQTL-inspired association framework for cellular genetic screens

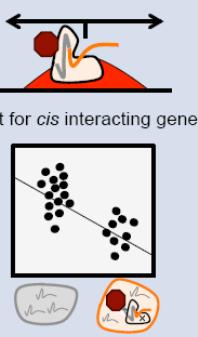


Identified 664 enhancer-gene pairs to gain insight into rules of human gene regulation

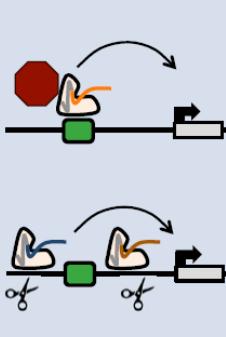
Multiplexed ~28 guide-RNAs per cell



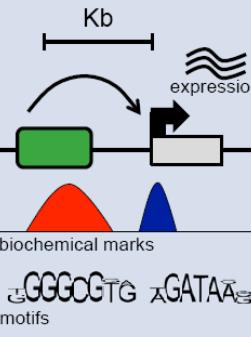
Screened 5,920 candidate enhancers



Validation of individual enhancer-gene pairs



Which marks and motifs predict pairing?



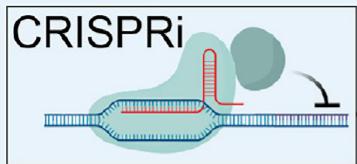
A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens

Molly Gasperini,^{1,*} Andrew J. Hill,¹ José L. McFaline-Figueroa,¹ Beth Martin,¹ Seungssoo Kim,¹ Melissa D. Zhang,¹ Dana Jackson,¹ Anh Leith,¹ Jacob Schreiber,² William S. Noble,^{1,2} Cole Trapnell,^{1,3} Nadav Ahituv,⁴ and Jay Shendure^{1,3,5,6,*}

- Perturbed 5,920 human candidate enhancers for impact on gene expression
- Multiplexed ~28 CRISPRi perturbations per single-cell transcriptome
- Adapted the eQTL analytical framework to identify 664 cis human enhancer-gene pairs
- Characterized genomic features associated with these enhancer-gene pairs

Genome-wide screens: phenotype=survival

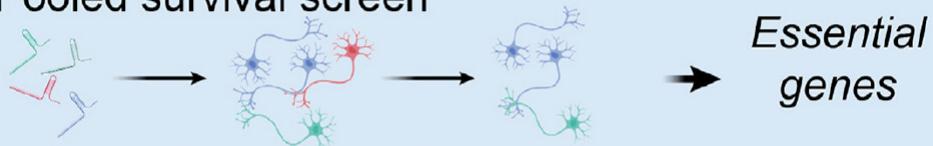
Systematic elucidation of gene function
in human neurons



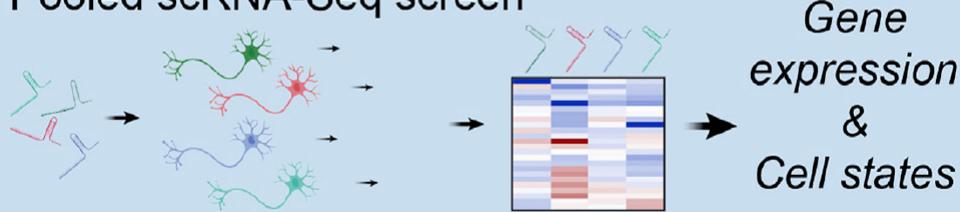
iPSC

Neuron

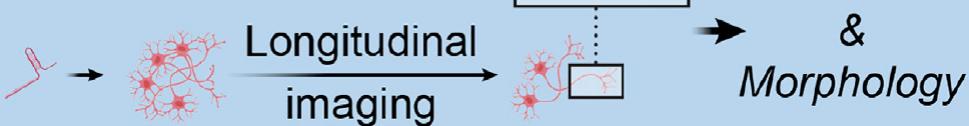
1) Pooled survival screen



2) Pooled scRNA-Seq screen



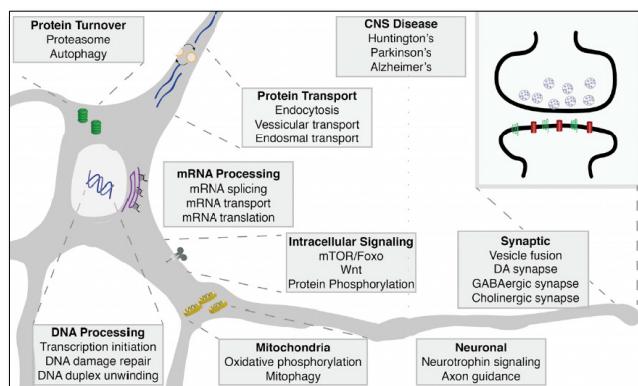
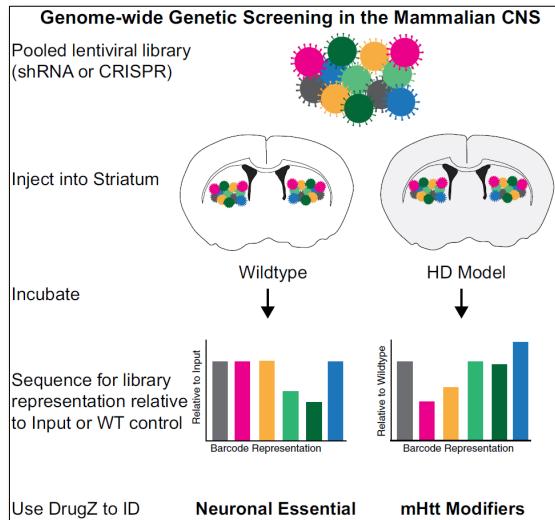
3) Arrayed high-content screen



A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens

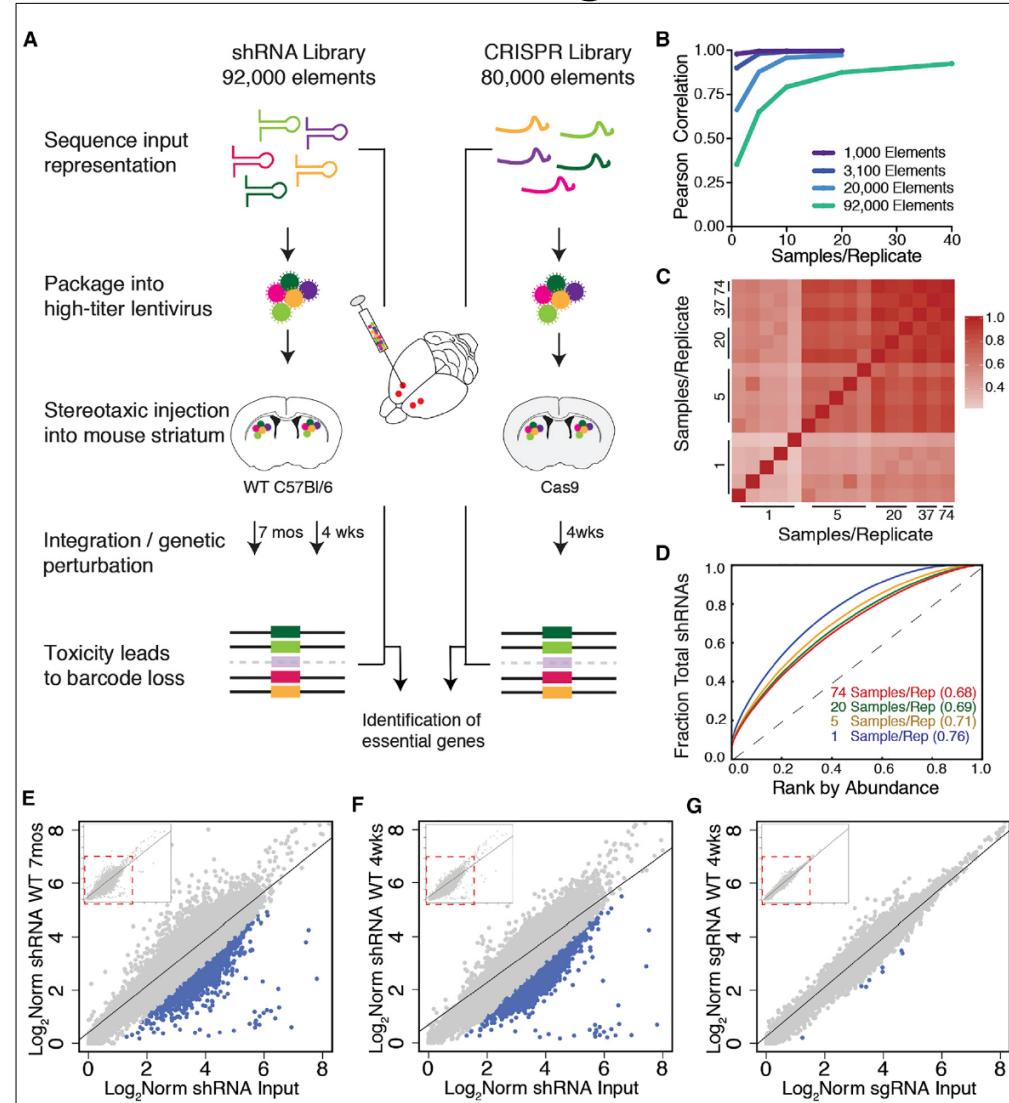
Molly Gasperini,^{1,*} Andrew J. Hill,¹ José L. McFaline-Figueroa,¹ Beth Martin,¹ Seungsoo Kim,¹ Melissa D. Zhang,¹ Dana Jackson,¹ Anh Leith,¹ Jacob Schreiber,² William S. Noble,^{1,2} Cole Trapnell,^{1,3} Nadav Ahituv,⁴ and Jay Shendure^{1,3,5,6,*}

Genome-wide CRISPR/siRNA screening *in vivo*



Genome-wide *In Vivo* CNS Screening Identifies Genes that Modify CNS Neuronal Survival and mHTT Toxicity

Mary H. Wertz,^{2,3} Mollie R. Mitchem,^{2,3} S. Sebastian Pineda,^{3,7,8} Lea J. Hachigian,^{1,2,3} Hyeseung Lee,^{2,3} Vanessa Lau,^{2,3} Alex Powers,^{2,3} Ruth Kulicke,^{2,3} Gurrein K. Madan,⁴ Medina Colic,⁴ Martine Therrien,^{2,3} Amanda Vernon,^{1,2,3} Victoria F. Beja-Glasner,^{1,3,5} Mudra Hegde,³ Fan Gao,^{2,6} Manolis Kellis,^{3,7} Traver Hart,⁴ John G. Doench,³ and Myriam Heiman^{1,2,3,8,*}

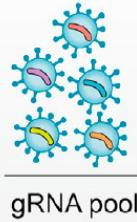


CRISP-Seq: Genome-wide systematic perturbations

Cell

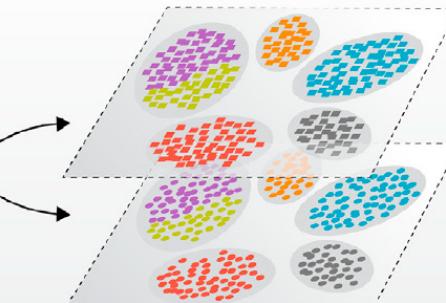
CRISP-seq - a new method for profiling the perturbation and transcriptome in the same cell

Cas9-GFP



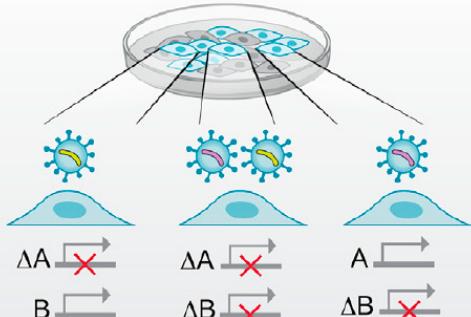
Genotype: gRNA sequencing

■ ΔA ■ ΔB ■ ΔC ■ ΔD ■ ΔE ■ w.t.

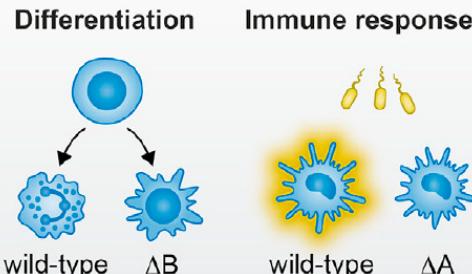


Phenotype: single-cell RNA-seq

Uncovering genetic interactions



In vivo screening of immune pathways



Resource

Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq

Diego Adhemar Jaitin,^{1,5} Assaf Weiner,^{1,2,5} Ido Yofe,^{1,5} David Lara-Astiaso,¹ Hadas Keren-Shaul,¹ Eyal David,¹ Tomer Meir Salame,³ Amos Tanay,⁴ Alexander van Oudenaarden,² and Ido Amit^{1,6,*}

¹Department of Immunology, Weizmann Institute, Rehovot 76100, Israel

²Hubrecht Institute-KNAW, Royal Netherlands Academy of Arts and Sciences, 3584 CT Utrecht, the Netherlands

³Flow Cytometry Unit, Department of Biological Services, Weizmann Institute of Science, Rehovot 76100, Israel

⁴Department of Computer Science and Applied Mathematics, Department of Biological Regulation, Weizmann Institute, Rehovot 76100, Israel

⁵Co-first author

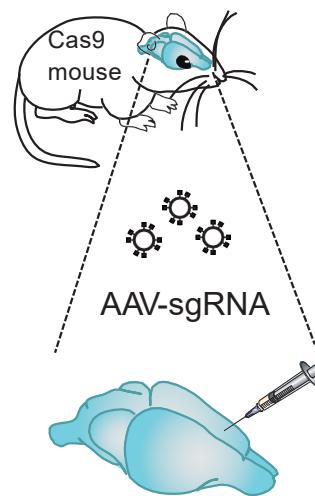
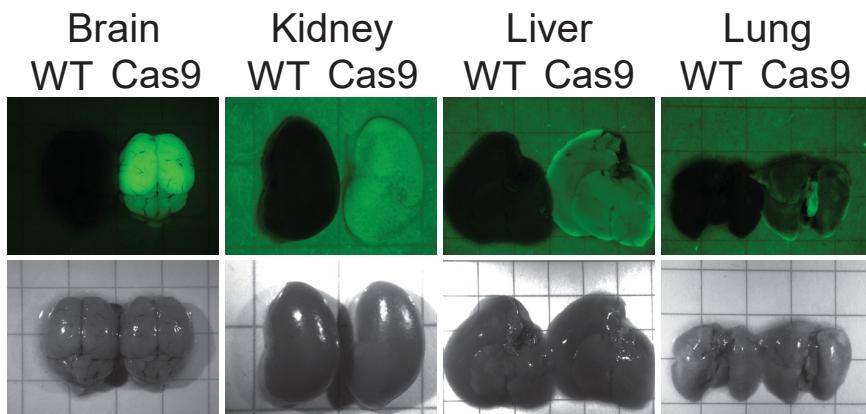
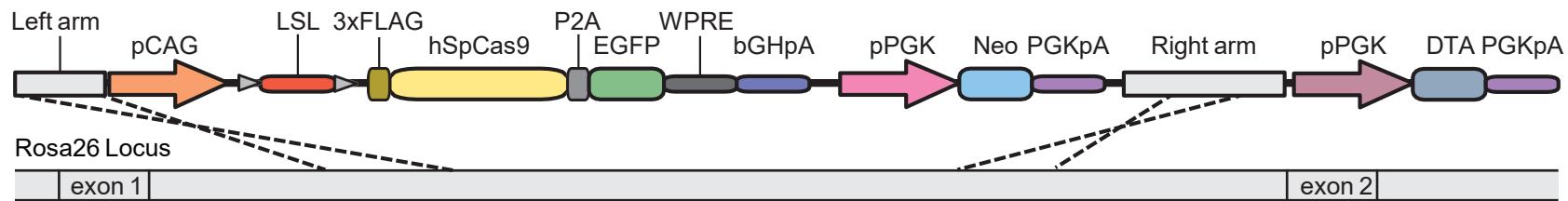
⁶Lead Contact

*Correspondence: ido.amit@weizmann.ac.il

<http://dx.doi.org/10.1016/j.cell.2016.11.039>

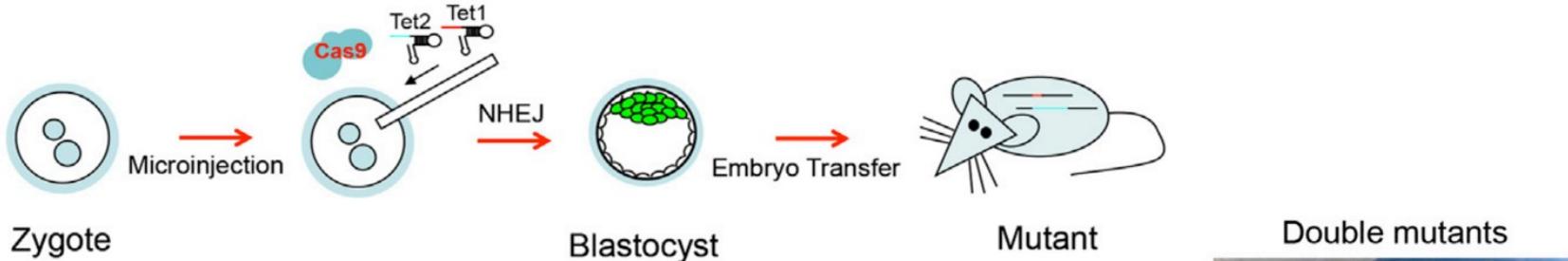
- A new method for profiling the perturbation and transcriptome in the same cell
- An integrated approach for CRISPR-pooled screens with single-cell transcriptomics
- Immune development and signaling-dependent circuits revealed by CRISP-seq
- CRISP-seq uncovers gene modules regulated by Stat1 and Rela in myeloid cells

Cas9 mouse

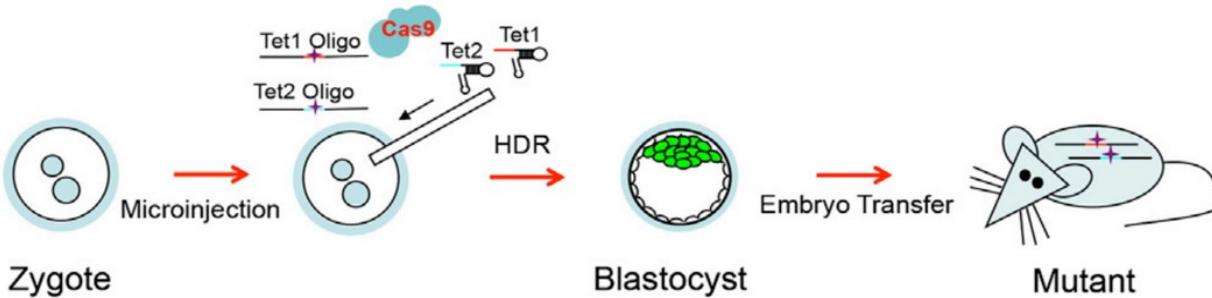


Rapid generation of genetically engineered mice

Targeted Mutations (Deletion / Insertion)



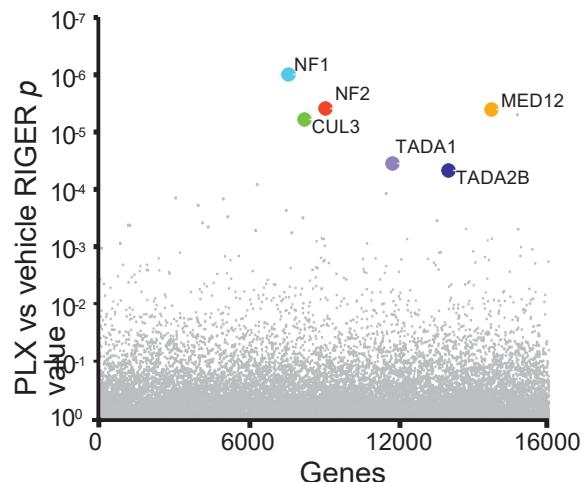
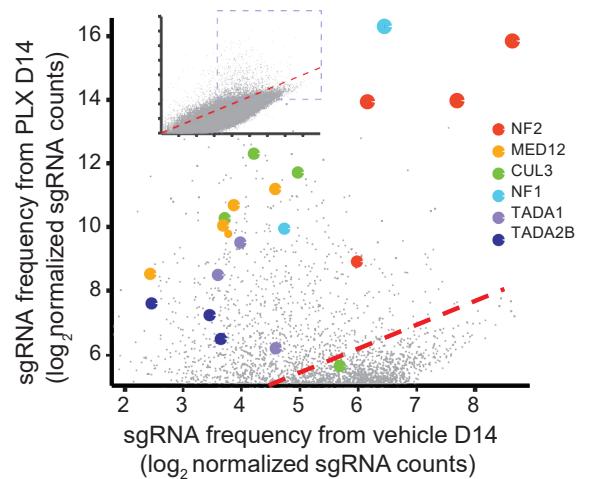
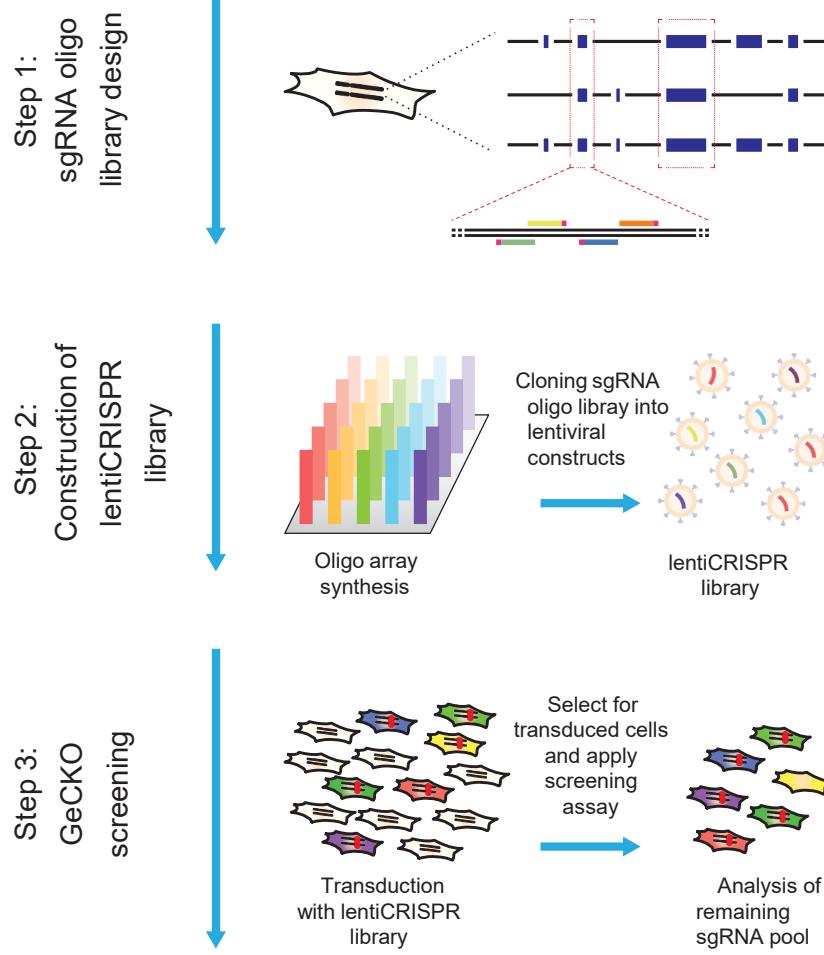
Predefined Precise Mutations



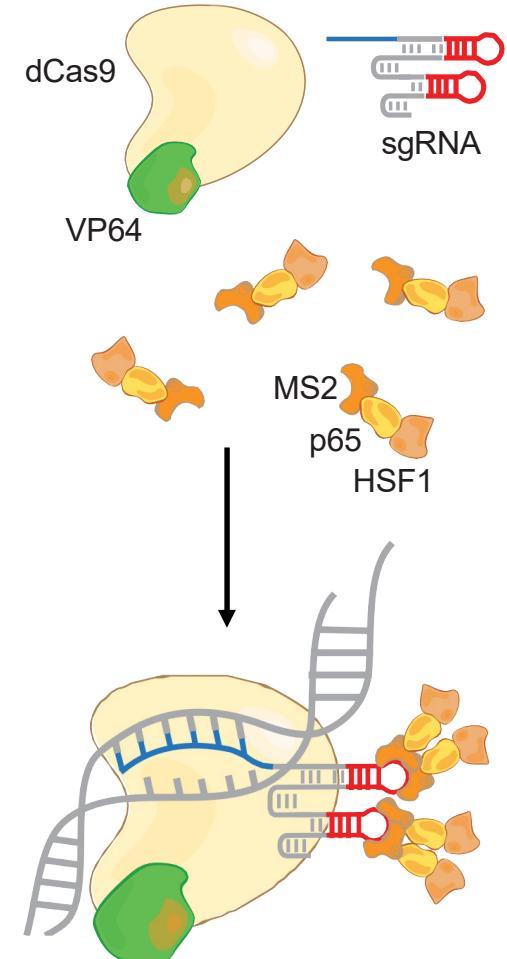
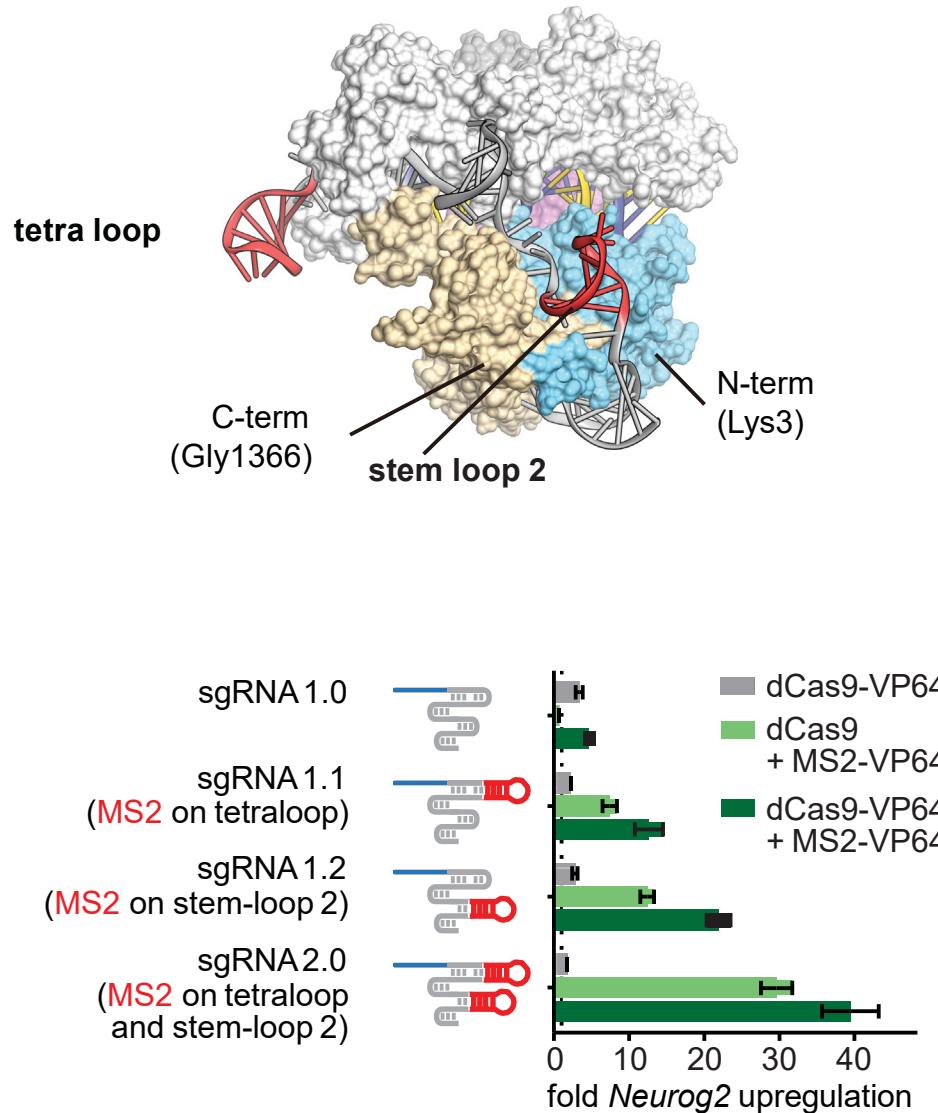
Rapid production of transgenic animals: 3 weeks vs. 2 years of breeding

Wang et al., Cell 2013

Cas9 knockout screen to identify genes involved in vemurafenib resistance in melanoma

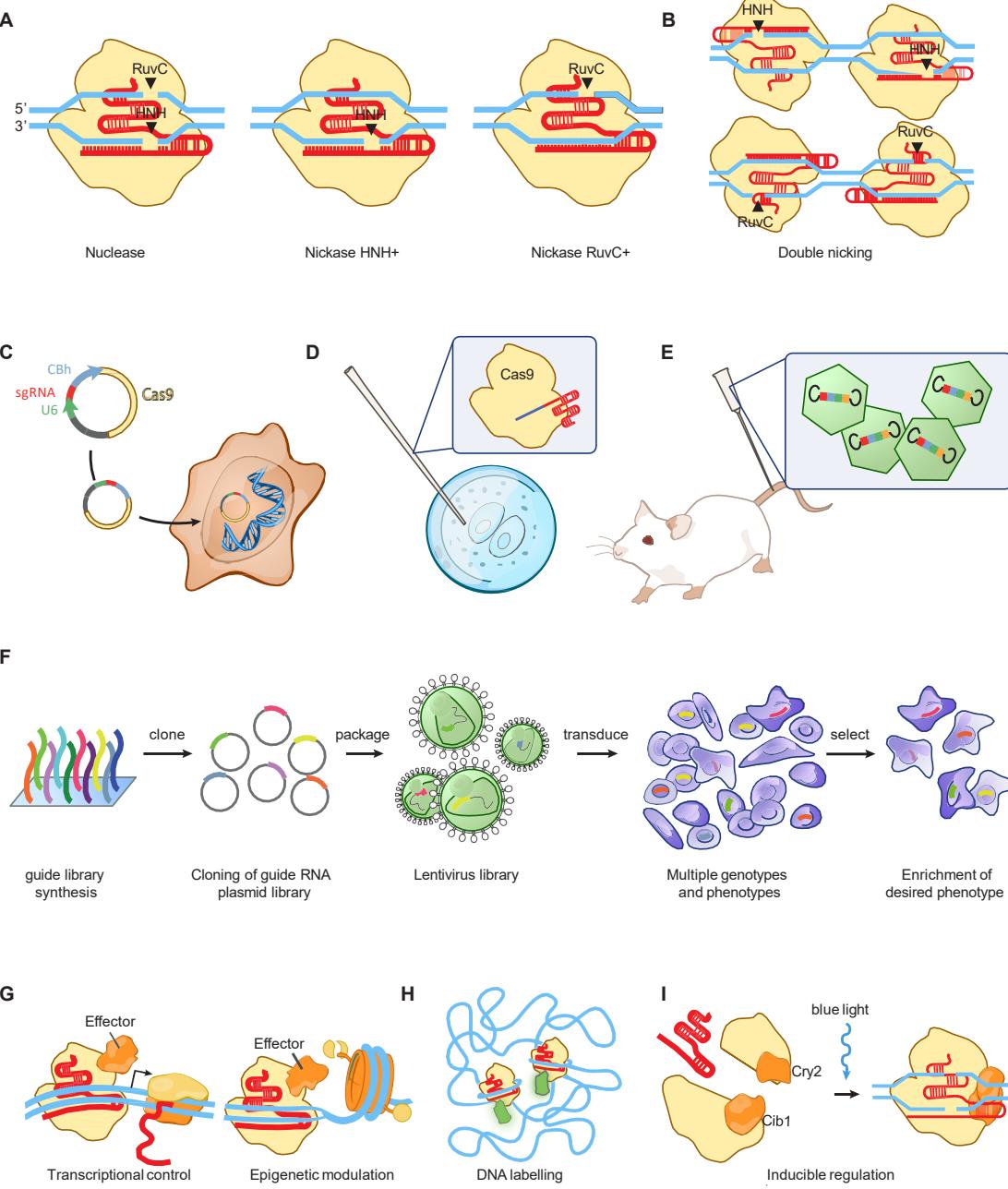


dCas9 coupled with transcriptional activators



Summary of Applications

- Cas9 as a nuclease for genome editing
- Cas9 as a screening approach
- Cas9 for transcriptional control



High-throughput expts: MPRA, SHARPR, HiDRA, Perturb-seq

1. High-throughput synthesis: Massively Parallel Reporter Assays (MPRA)

- MPRA technology: in vitro synthesis, reporter design, transfection
- 1 site X deep: single-base dissection, combinatorial changes, high-res map
- 2k sites x shallow: motif-guided perturbations, activators/repressors

2. Next-generation applications of MPRA + STARR-Seq, SHARPR, HiDRA

- 10k x deep: high-resolution dissection by tiling + deconvolution (Sharpr)
- 10M x deep: HiDRA, no synthesis (STARR), targeted capture (ATAC)

3. Endogenous genome editing: cutting and template-based repair

- Endonucleases: ZNFs, TALENs. Repair pathways. Diversity of outcomes.
- CRISPR-Cas9: origins, discovery, optimization, base-editing, applications
- Therapeutic applications, multiplexing, delivery, next-generation tools