A tiling-deletion-based genetic screen for *cis*-regulatory element identification in mammalian cells

Yarui Diao[1,10], Rongxin Fang[1,2,10], Bin Li[1,10], Zhipeng Meng[3,4], Juntao Yu[1,5], Yunjiang Qiu[1,2], Kimberly C Lin[3,4], Hui Huang[1,6], Tristin Liu[1], Ryan J Marina[6], Inkyung Jung[7], Yin Shen[8], Kun-Liang Guan[3,4] & Bing Ren[1,4,9]

NATURE METHODS | VOL.14 NO.6 | JUNE 2017 | 629

"The Two Faces of Promoter"

JClub 09.13.2017
Donghoon Lee

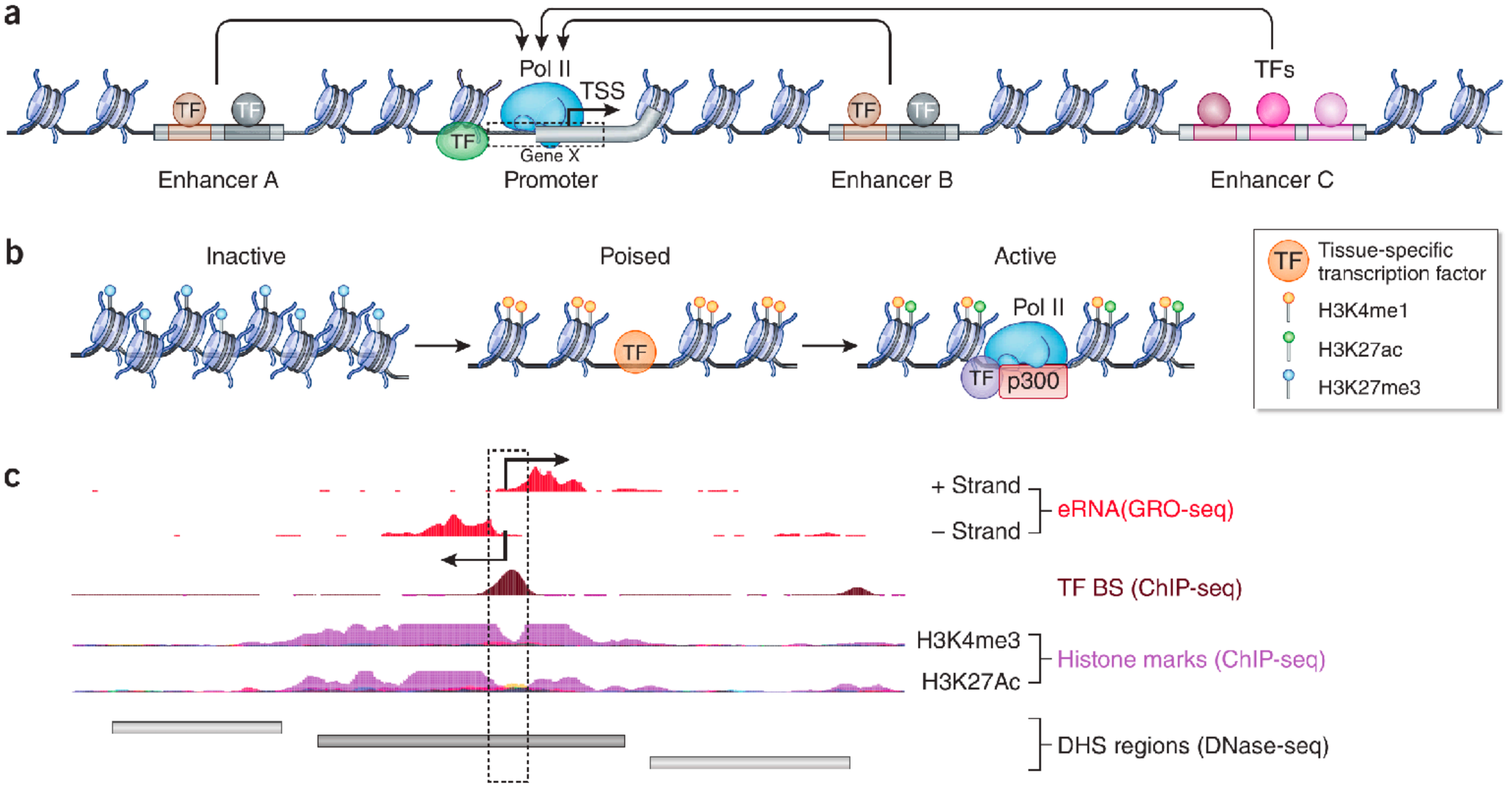
one page summary

- CRISPR-based noncoding genomic screen on ***POU5F1*** gene (Oct4), one of Yamanaka factors, in human embryonic stem cells (hESCs)
- Discovered 45 regulatory sequences of ***POU5F1***
- Surprisingly, 17 of them were promoters of functionally unrelated genes
- Survey of these enhancer-like promoters and found that they form extensive spatial contacts with the ***POU5F1*** promoter

previously..

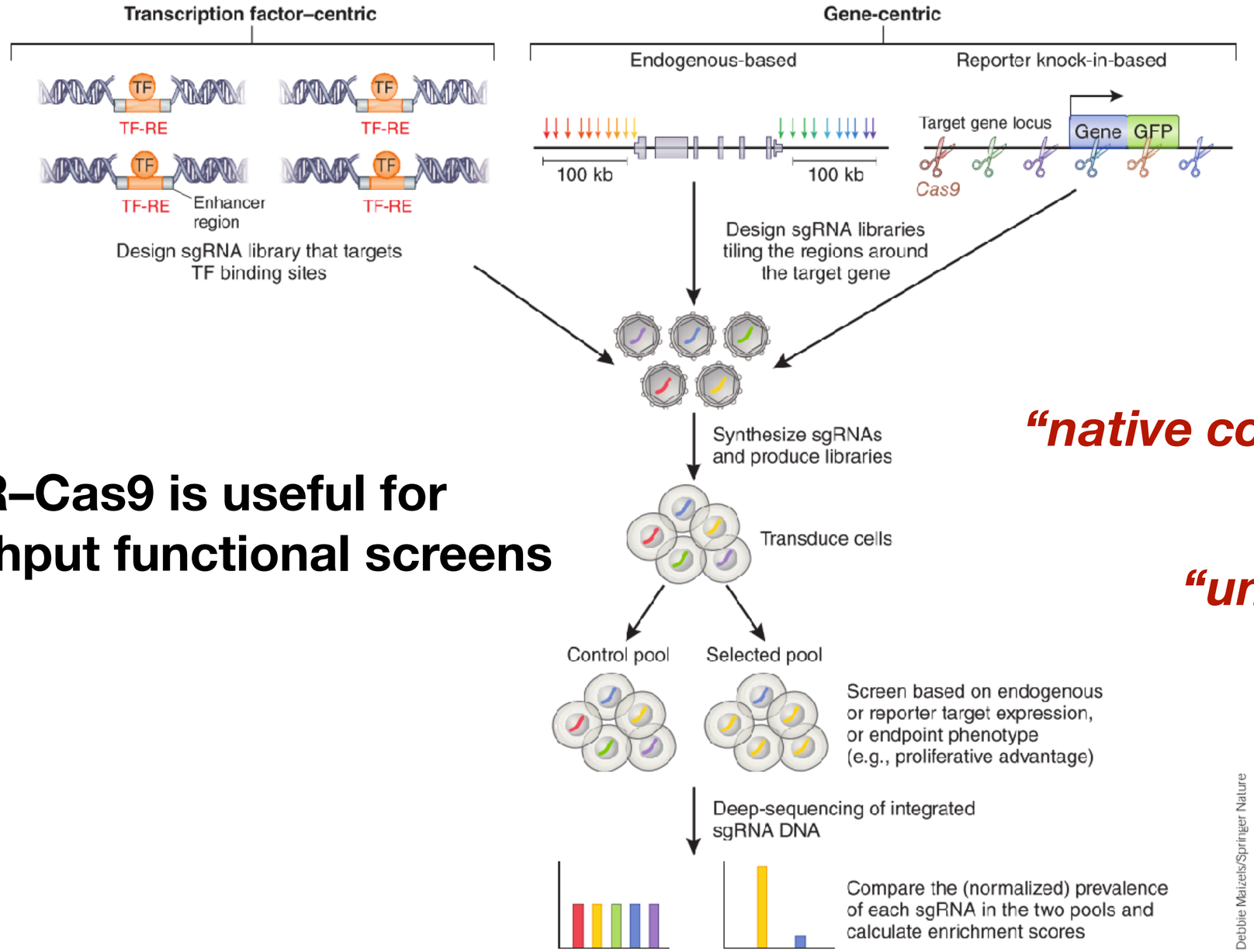
- “promoters can function as enhancers in ectopic enhancer activity assays” (Zabidi et al 2015, Arnold et al 2013)
- “promoters form long-range contact with other promoters” (Li et al 2012)
- “mutation in promoter affect distal gene expression” (Rajagopal et al 2016)

➡However, it was unclear whether promoters can function as distal-acting enhancers *in vivo*



Elkon R, Agami R. Characterization of noncoding regulatory DNA in the human genome. Nat. Biotechnol. 2017 35:732–746. Available from: http://www.nature.com/doifinder/10.1038/nbt.3863

CRISPR–Cas9 is useful for high-throughput functional screens

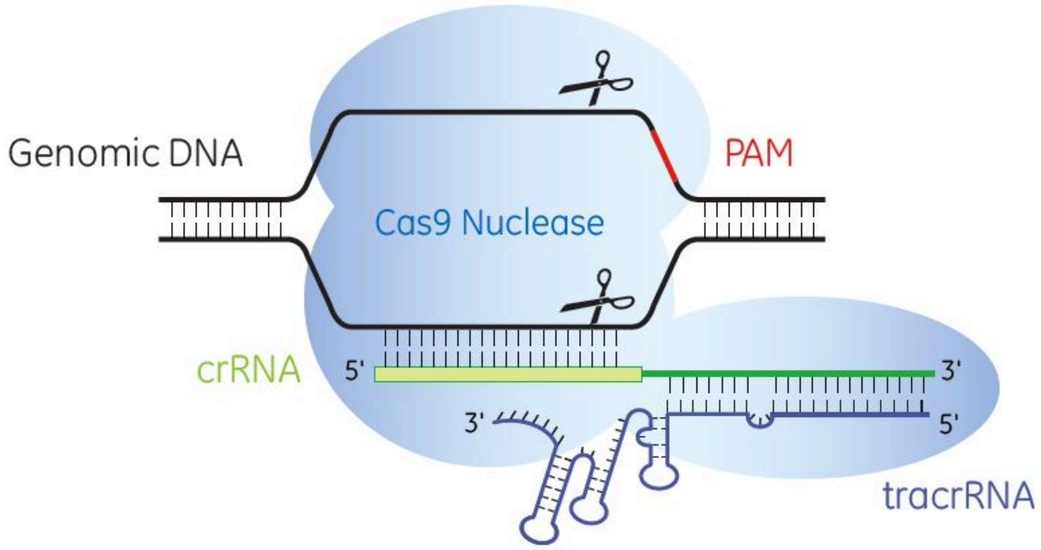


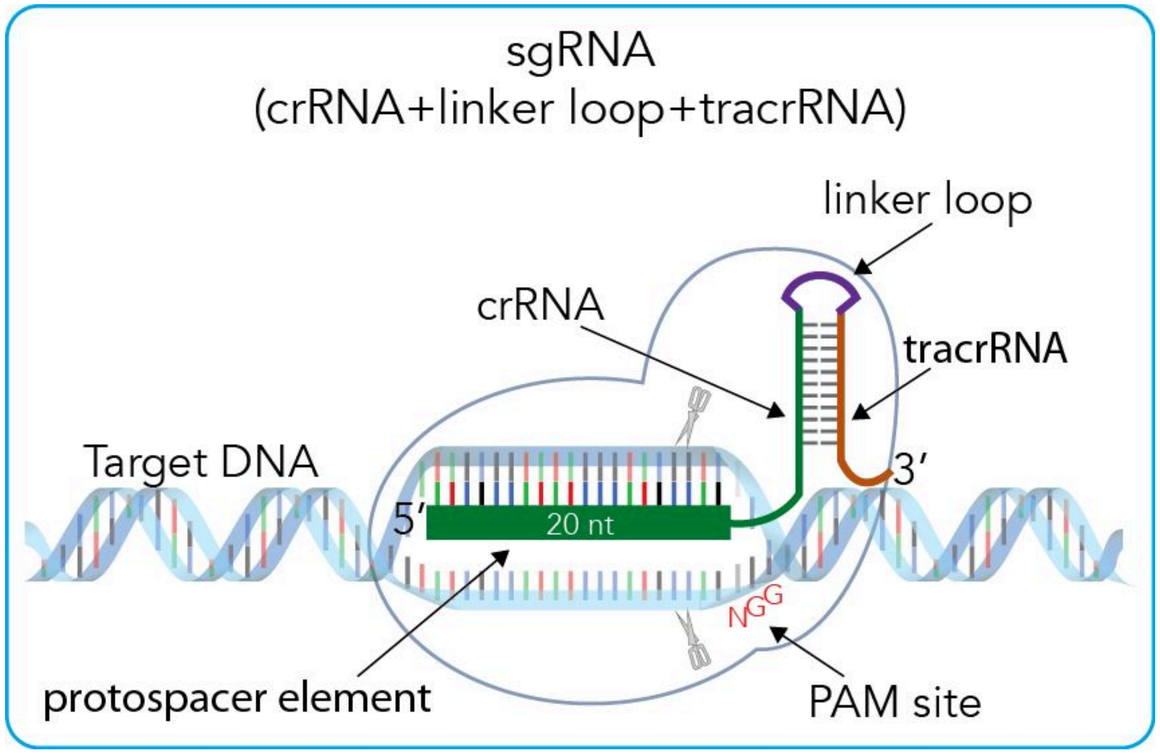
"native context (in vivo)"

"unbiased"

Elkon R, Agami R. Characterization of noncoding regulatory DNA in the human genome. Nat. Biotechnol. 2017 35:732–746. Available from: http://www.nature.com/doifinder/10.1038/nbt.3863

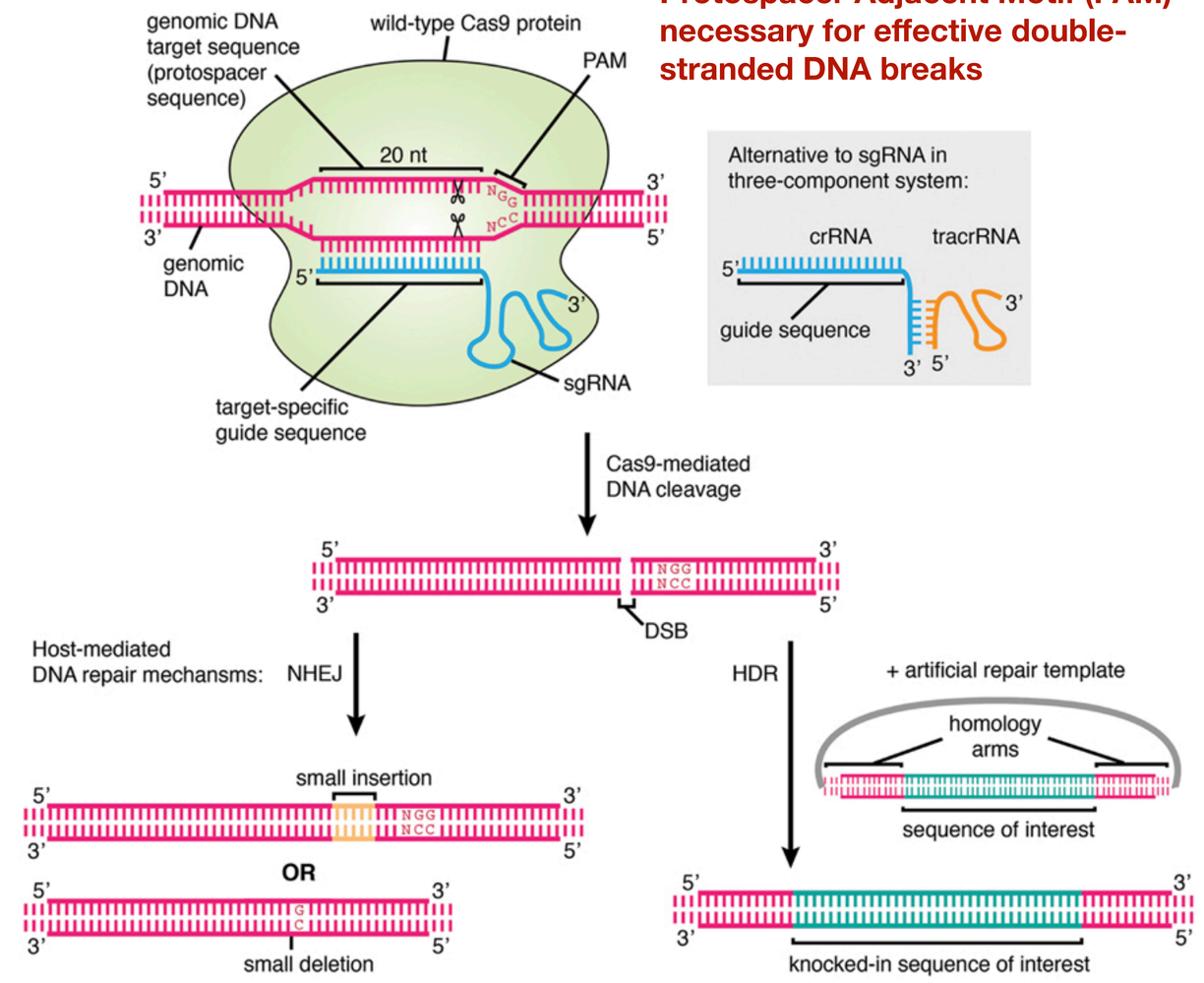
CRISPR-Cas9 mediated genome editing





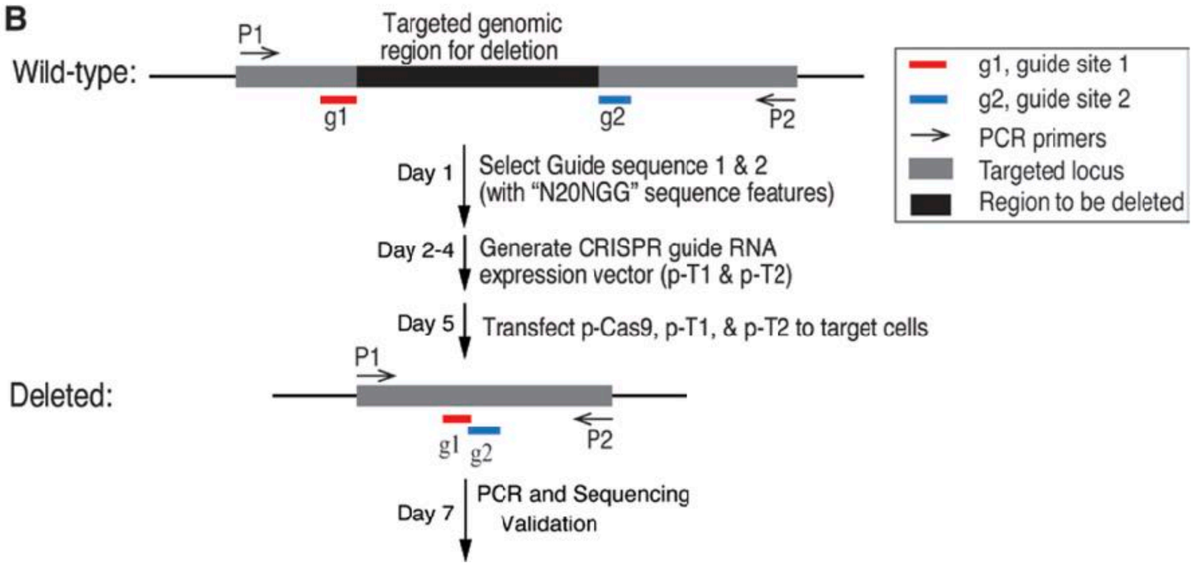
- 20bp guide RNA (gRNA)
- 3bp Protospacer adjacent motif (PAM) (NGG for cas9)

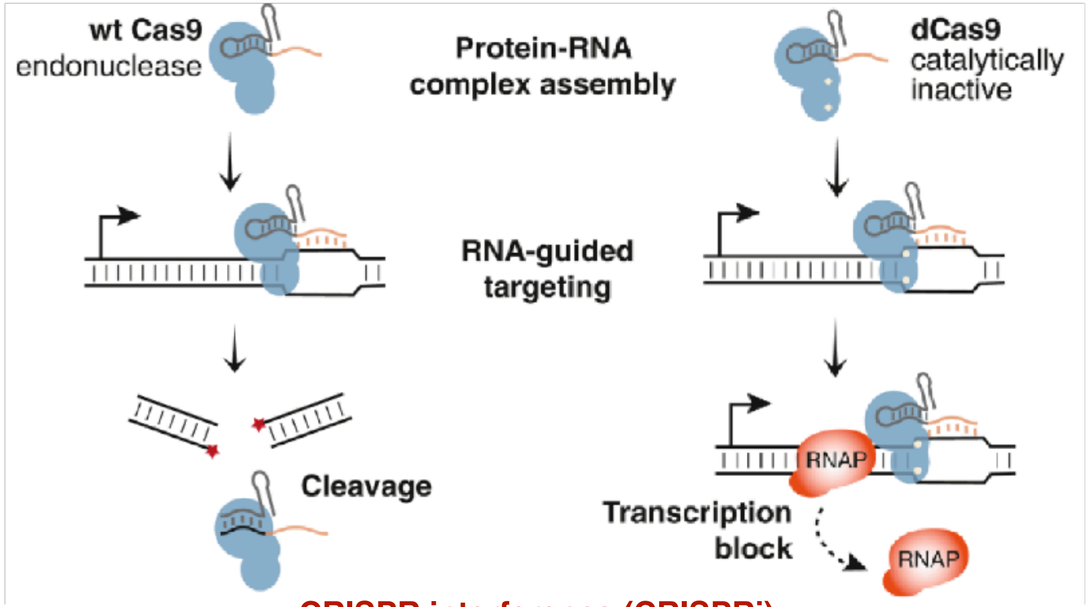
Protospacer Adjacent Motif (PAM) necessary for effective double-stranded DNA breaks



Non-homologous end joining (NHEJ)

Homology directed repair (HDR)





CRISPR interference (CRISPRi)

Table 1 | Comparison of CREST-seq data to published functional screens of noncoding regulatory sequences

Reference	Target region	Total oligos	Oligo density (per kb)	Coverage	Able to distinguish *trans* or *cis*?	
Canver *et al.*[10]	4.2 kb, 3 DHSs and 1 exon	582	137	~1×	No	
Korkmaz *et al.*[11]	685 p53 ChIP-seq peaks **CRISPR-based mutagenesis**	1,116	N.A.	1.3–1.6 oligos per ChIP-seq peak	No	**Jan 2016**
	73 ChIP-seq peaks for ER-α expressing enhancer RNA	97	N.A.			
	2-kb *CDKN1A* locus	197	98.5	<93.6%		
Rajagopal *et al.*[12]	40-kb *Tdgf1* locus	3,908	98	<93.1%	No	
	Rpp25, *Nanog* and *Zfp42* loci	3,908	N.A.	N.A.		
Diao *et al.*[13]	37.6 kb, 174 putative enhancers in 1-Mb *POU5F1* locus	1,964	52	<49.4%	No	
Sanjana *et al.*[14]	200-kb *NF1* locus **CRISPR-based mutagenesis**	6,682	33.4	<31.7%	No	**Sep 2016**
	200-kb *NF2* locus	6,934	34.6	<32.9%		
	200-kb *CUL3* locus	4,699	23.5	<22.3%		
Fulco *et al.*[15]	1.29-Mb *GATA1* and *MYC* loci **CRISPRi**	98,000	76	~64×	No	**Nov 2016**
CREST-seq	2-Mb *POU5F1* locus **CRISPR-based deletion**	11,600	5.7	20×	Yes	**Jun 2017**

Here CREST-seq is compared with published screens of noncoding regulatory elements. The following aspects are compared: the size of the screen region, the total number of oligos required to construct the library, the average number of oligos per kilobase in each screen, and the estimated coverage of the target region. To estimate the coverage of the target region, we assumed that the PAMs were equally distributed across the genome and that each gRNA created a mean insertion/deletion size of 9.5 ± 13.7 bp. To compute the coverage of the CRISPRi screen using dCas9–KRAB, we assumed that the average size of H3K9me3 peaks introduced by dCas9–KRAB was about 850 bp. N.A., not available.

CRISPR-based mutagenesis is limiting because..

1. Not all sequences suitable for CRISPR-Cas9 mediated genome editing, owing to the PAM requirement of gRNAs (required for targeting/cutting)

2. They make point mutations or short indels; you need unrealistically large number of sgRNAs to interrogate human genome

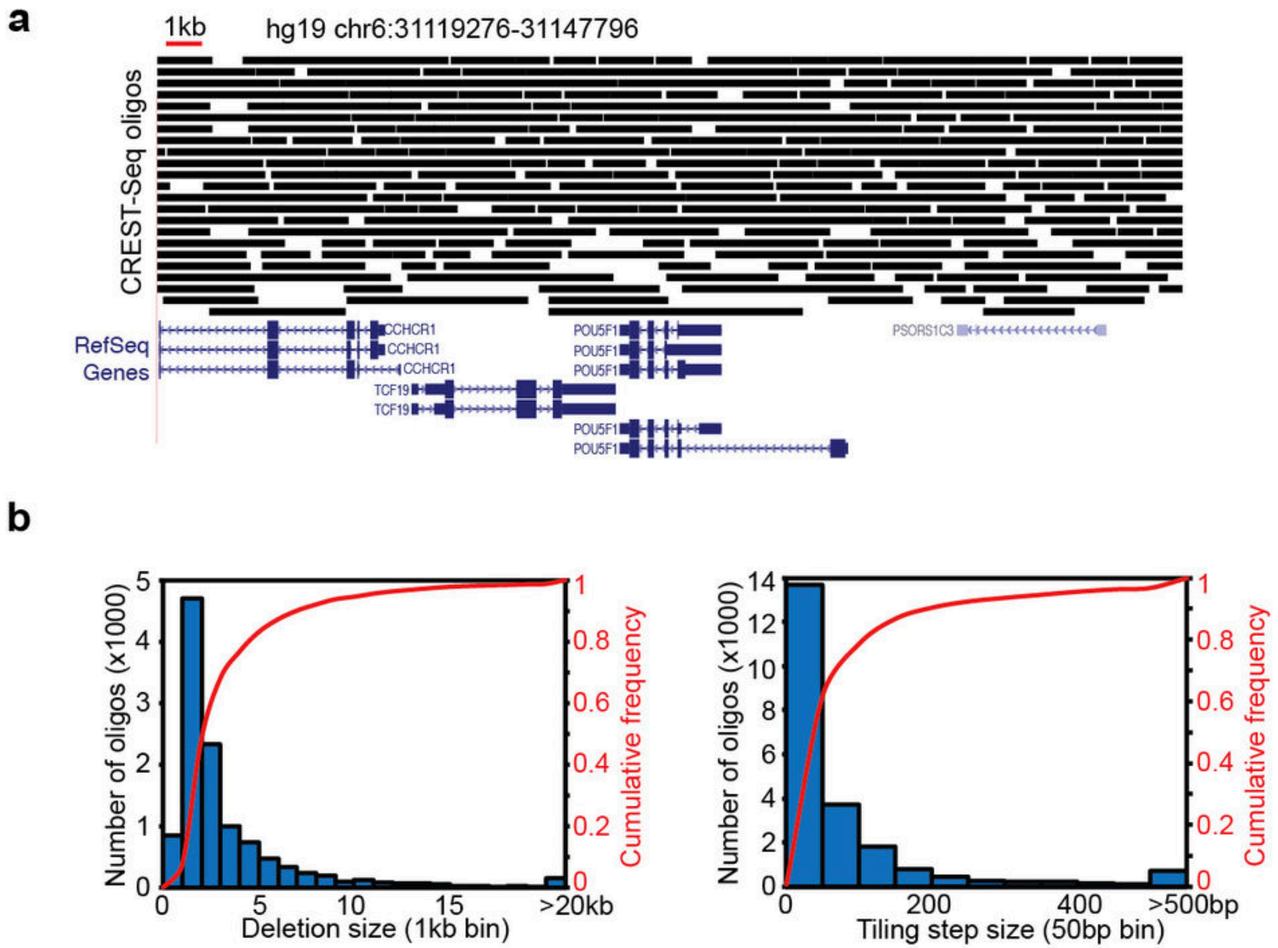
3. Challenging to distinguish cis- and trans- regulatory elements

CREST-seq

***C**is-**R**egulatory **E**lement Scan by **T**iling-deletion and sequencing*

“Yet another CRISPR-based high-throughput functional screens in native context”

2-Mb POU5F1 locus in an hESC line



Oct-4

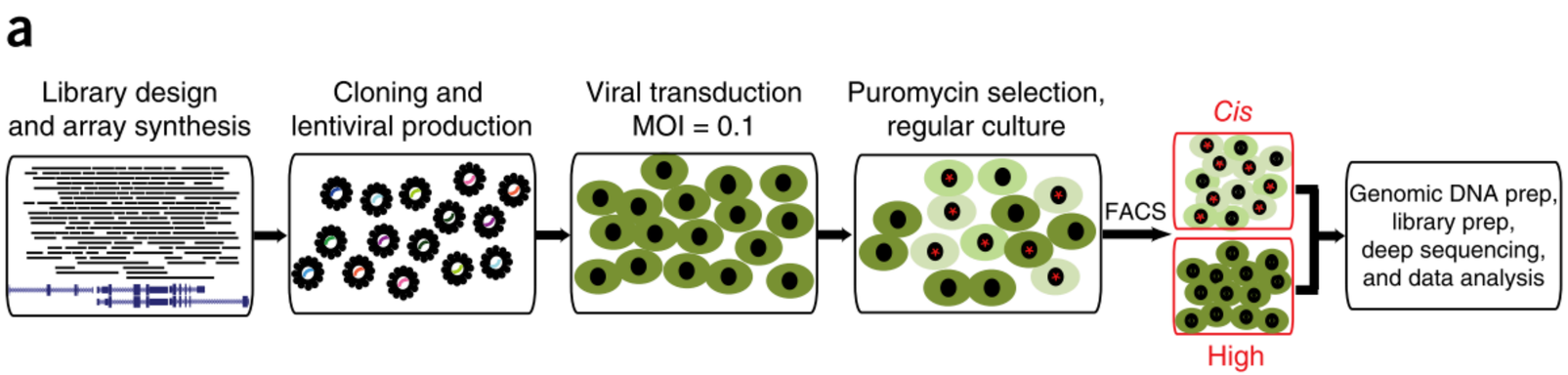
one of Yamanaka factors (Oct4, Sox2, cMyc, and Klf4)

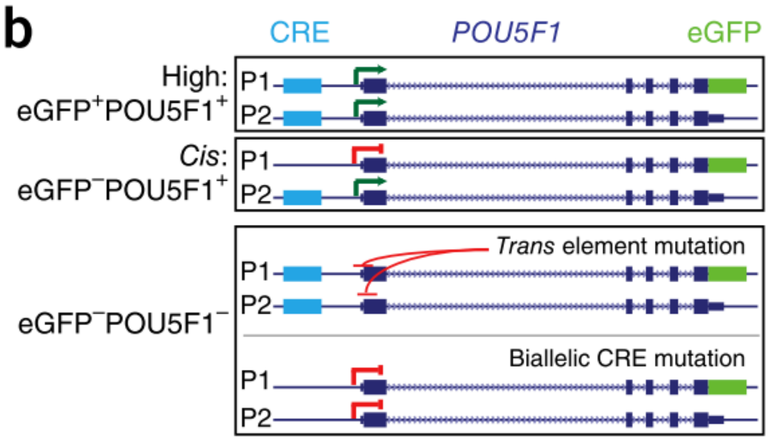
critically involved in the self-renewal of undifferentiated

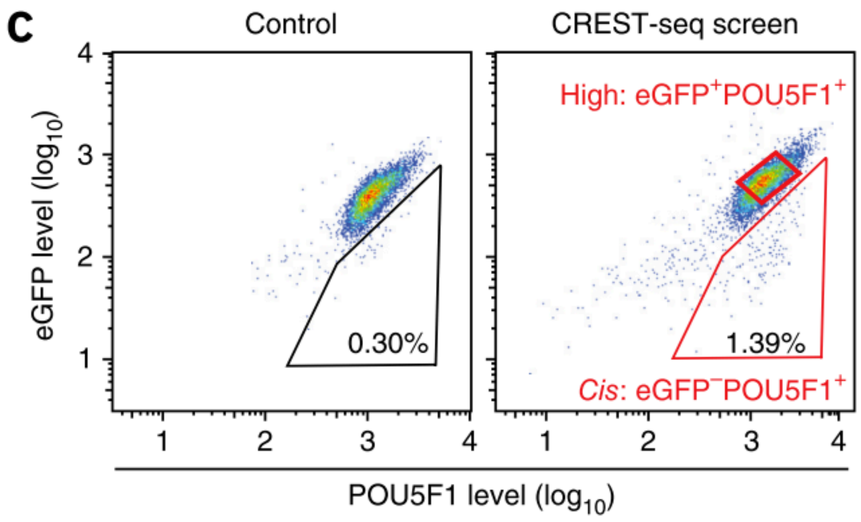
Oct-4 expression must be closely regulated; too much or too little will cause differentiation of the cells

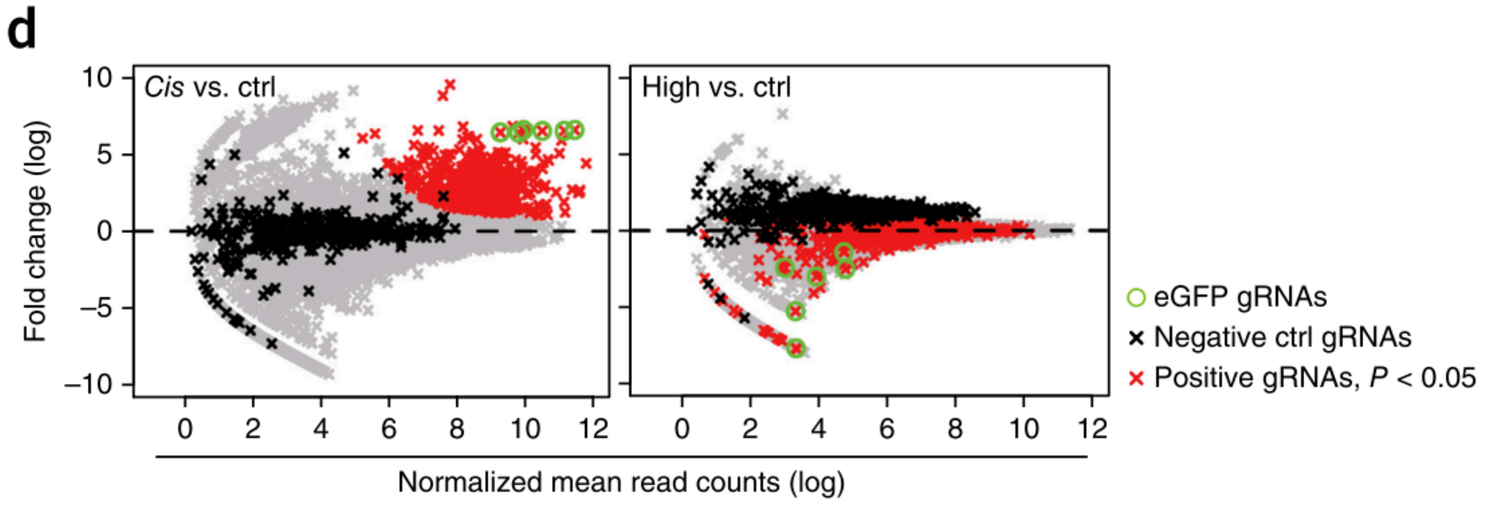
- average size of each deletion was ~2 kb, with an overlap of 1.9 kb between two adjacent deletions
- each nucleotide in the locus was covered by ~20 distinct genomic deletions on average

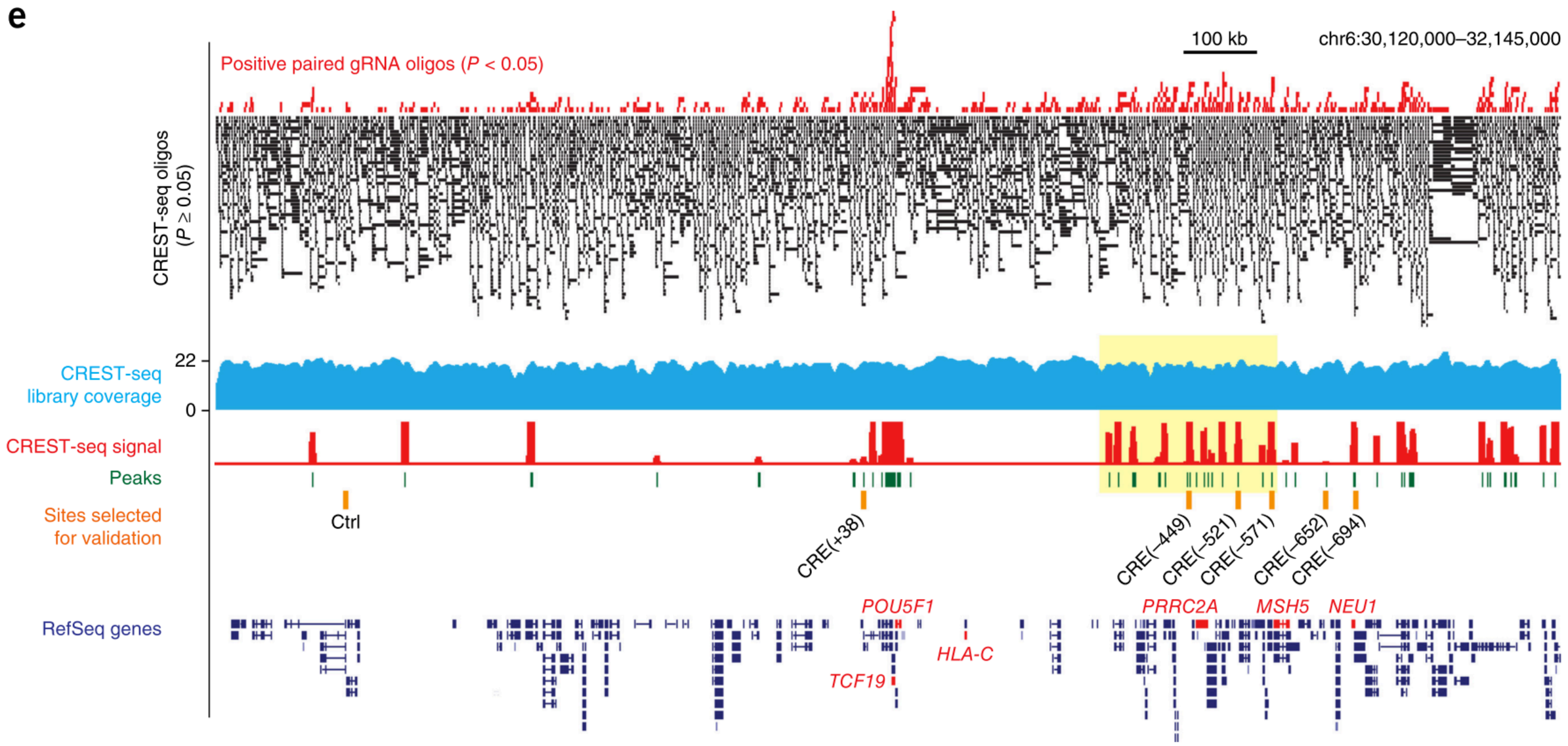
Experiment Strategy



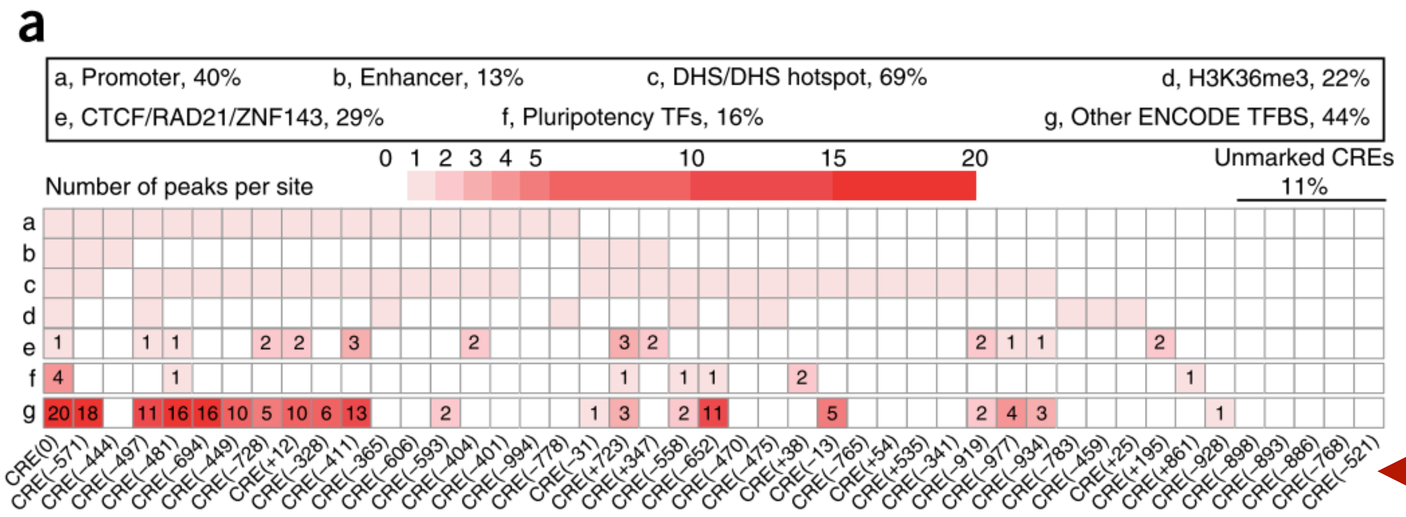




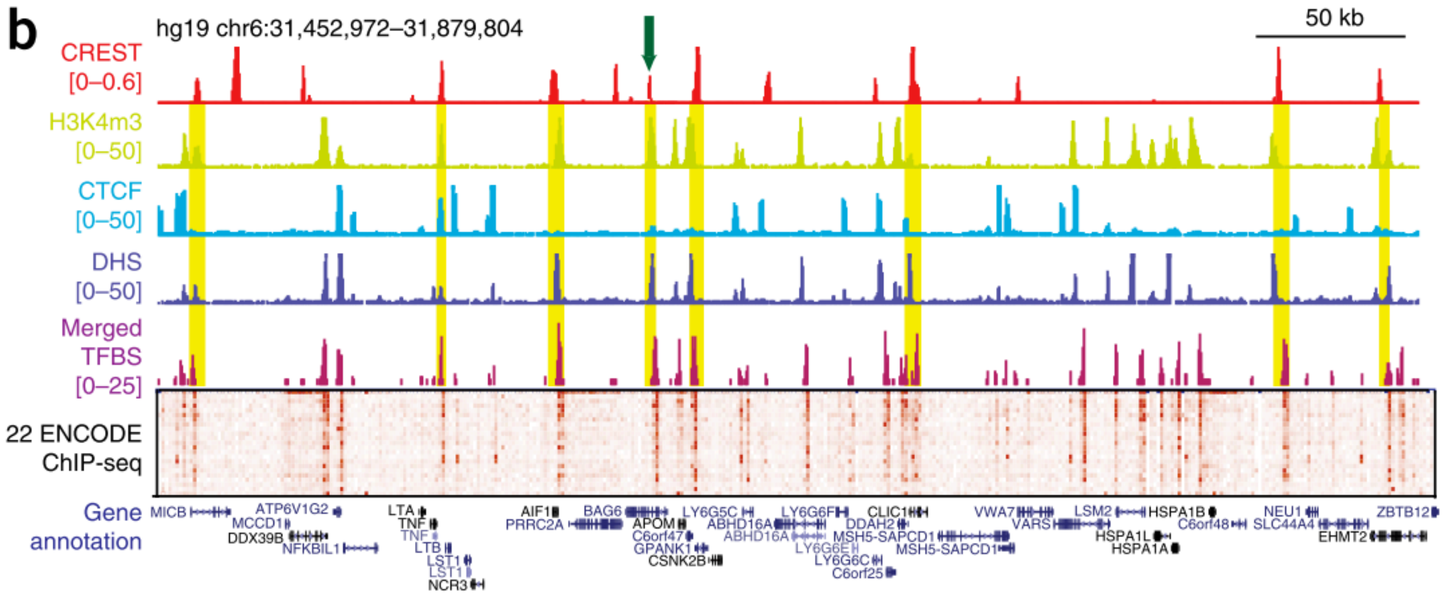




- 45 genomic regions with a significant score, 41 novel



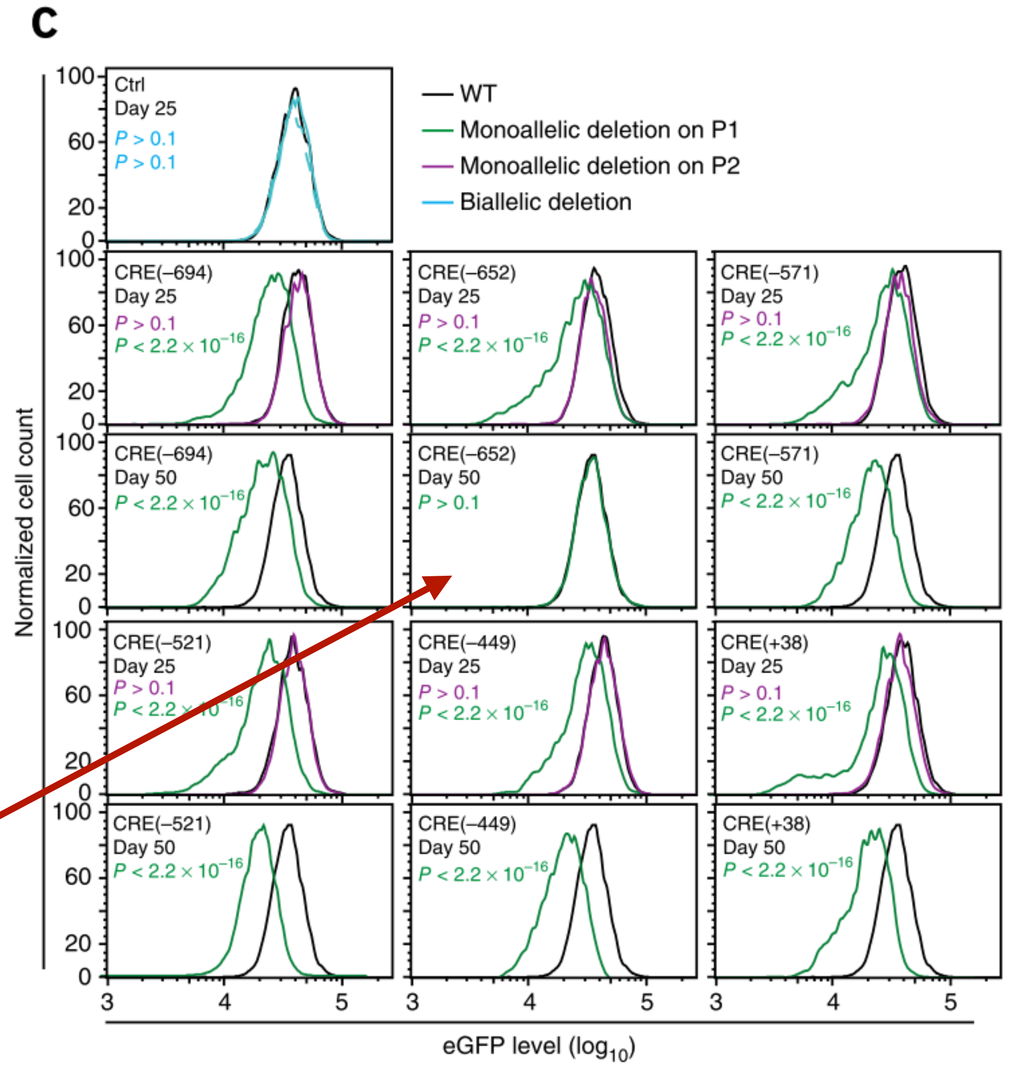
five CREs lacked any canonical chromatin signatures



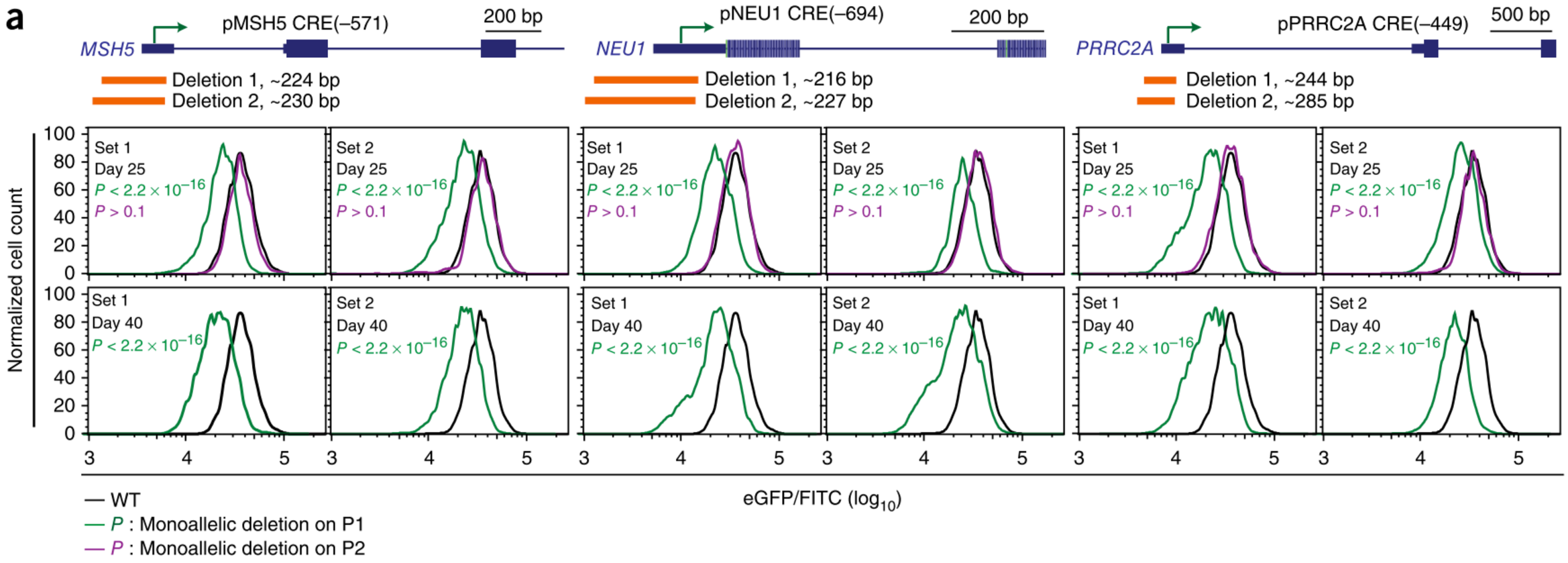
In-depth validation of 6 candidate

- cell clones with monoallelic deletion on the P1 (eGFP-tagged) allele showed a significant reduction in eGFP gene expression (P < 2.2 × 10–16)
- cell clones with monoallelic deletions on the P2 (wild-type) allele showed normal eGFP gene expression

temporarily phenotypic enhancer?

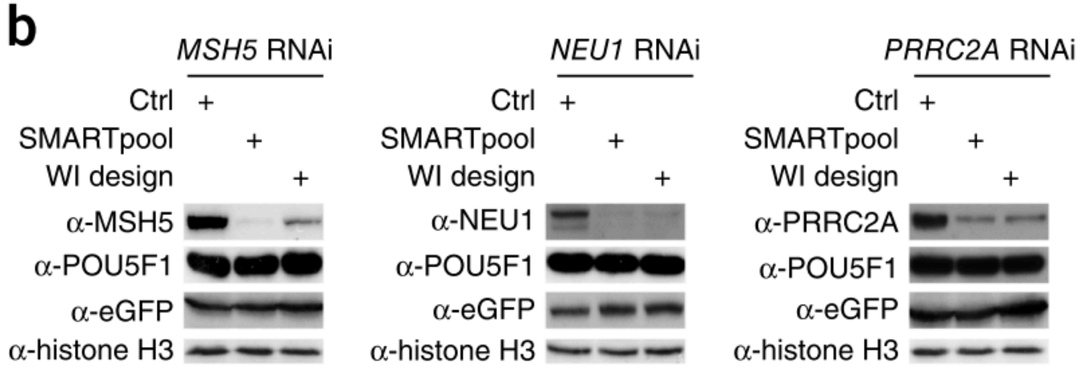


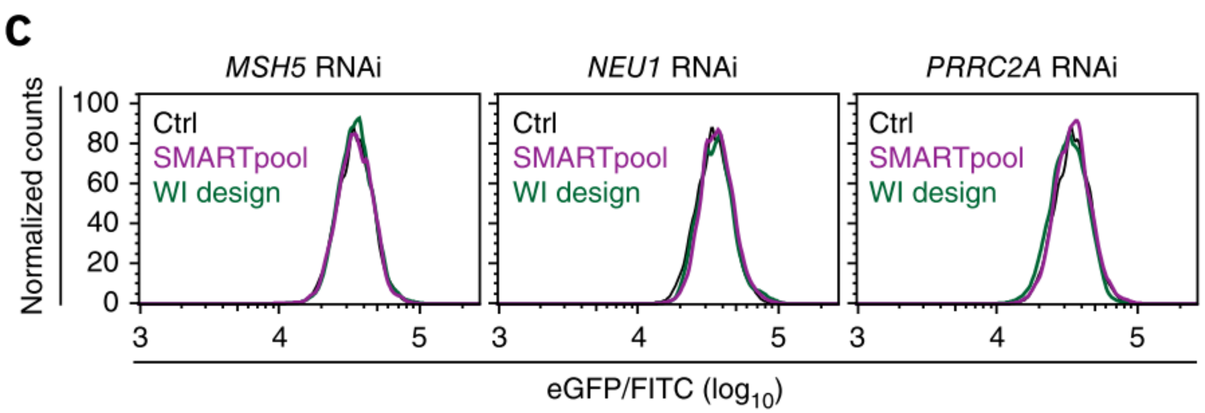
Potential distal regulatory effect of promoter-proximal elements

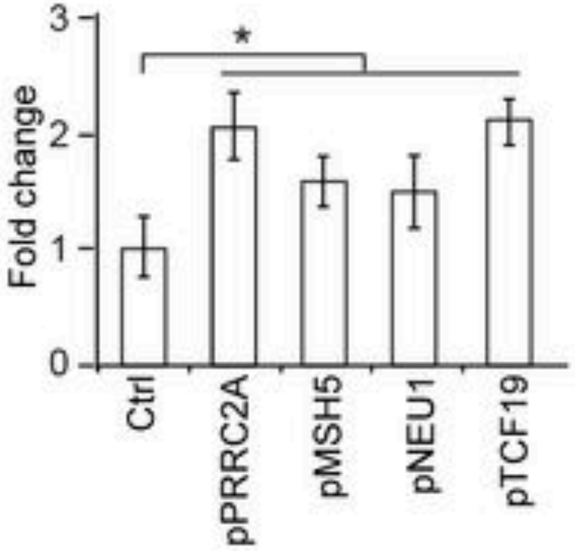


enhancer activity of promoters were recapitulated using short (2-300bp) TSS deletions

Potential *trans* effect of the RNA or protein products of the proximal genes

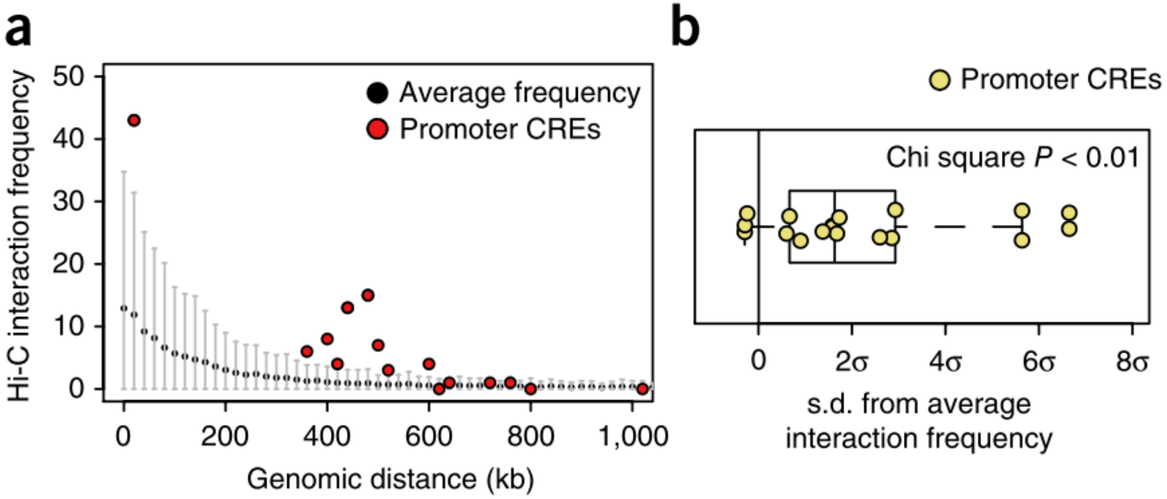


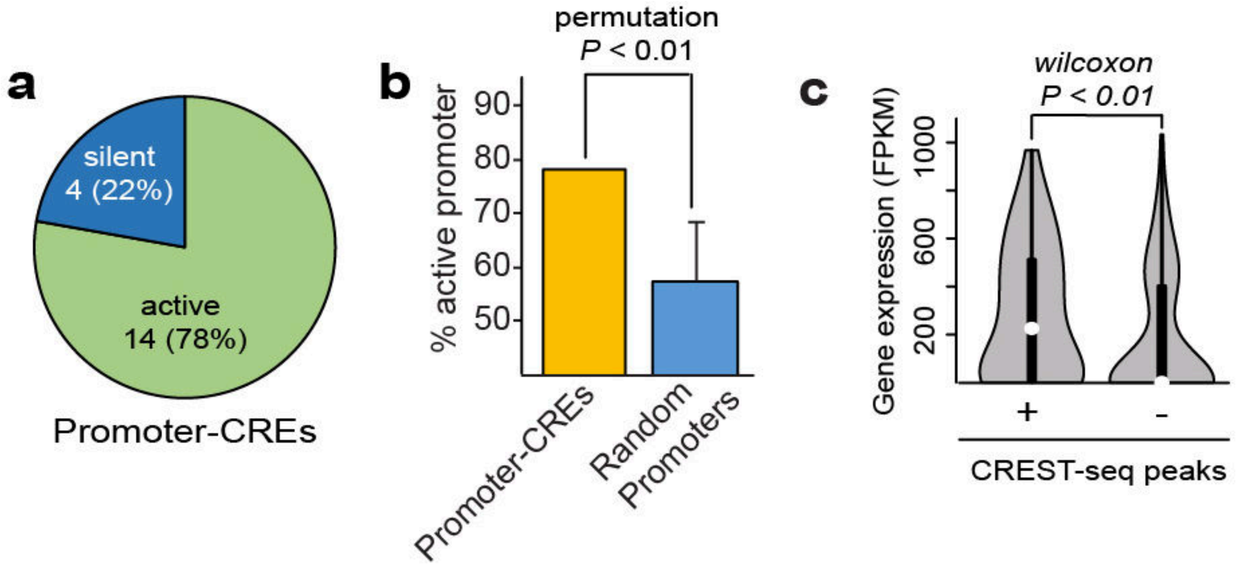




- KD (both siRNA and shRNA-based) of MSH5, NEU1, PRRC2A did not have effect on eGFP gene expression

enhancer-like promoters show extensive Hi-C contacts





14 out of 17 were significant (P < 0.01)

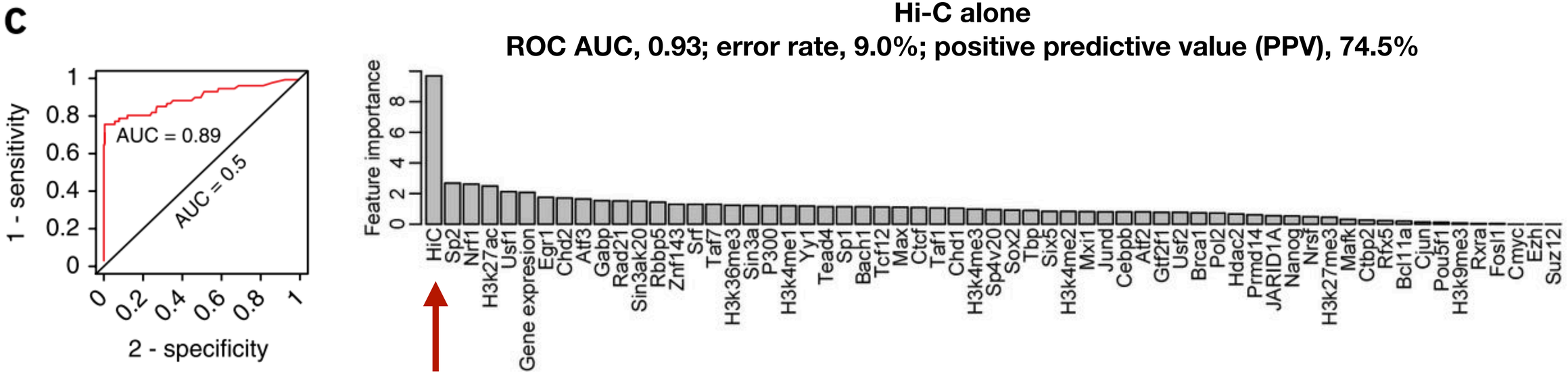
active promoter signatures (Pol2/H3K4me3/H3K27ac)

Hi-C was the single most important feature

random forest classifier
[Hi-C, TFBS, histone, expression]

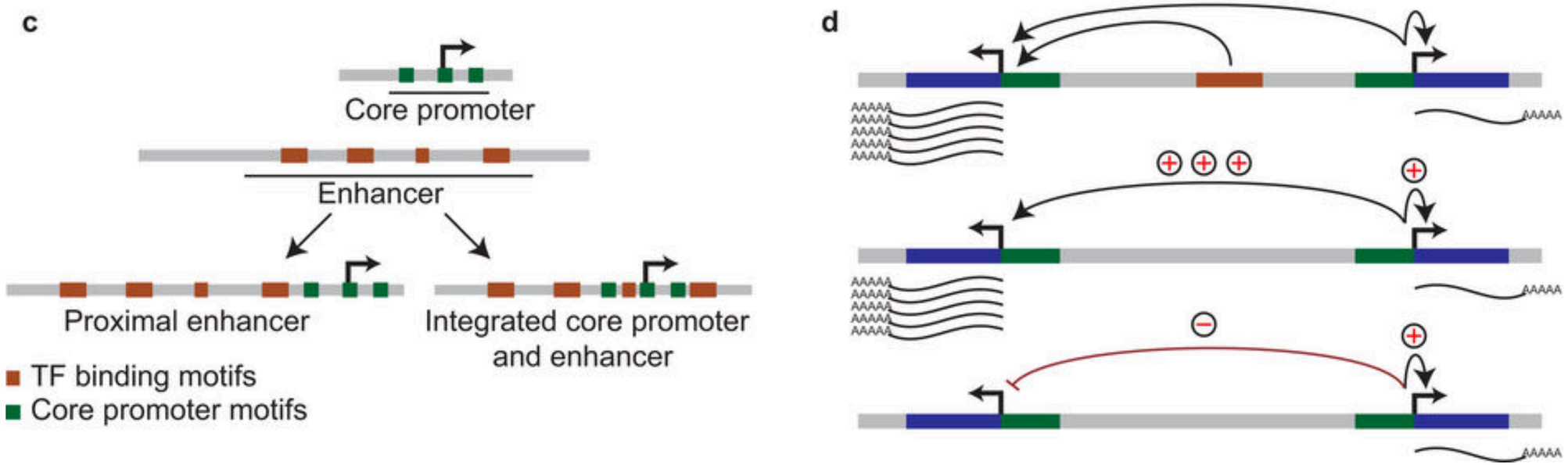
ALL FEATURES
ROC AUC, 0.89; error rate, 6.3%; positive predictive value (PPV), 97.2%

Hi-C alone
ROC AUC, 0.93; error rate, 9.0%; positive predictive value (PPV), 74.5%



PPV = nTP / (nTP + nFP)

We should be cautious about promoters having single proximal target genes



- ~40% CREs of POU5F1 correspond to promoters of other genes
- promoters could be '*independently*' work as distal enhancers
- extensive promoter–promoter interactions in mammalian cells should be noted