Author Manuscript

HHS Public Access
Author manuscript
Nat Methods. Author manuscript; available in PMC 2019 July 11.

Published in final edited form as:
Nat Methods. 2018 December ; 15(12): 992–993. doi:10.1038/s41592-018-0225-6.

CRISPR-SURF: discovering regulatory elements by deconvolution of CRISPR tiling screen data

Jonathan Y. Hsu[1,2], **Charles P. Fulco**[3,4], **Mitchel A. Cole**[5,6,7], **Matthew C. Canver**[2,5,6,7], **Danilo Pellin**[8,9], **Falak Sher**[5,6,7], **Rick Farouni**[2,3,10], **Kendell Clement**[2,3,10], **Jimmy A. Guo**[2], **Luca Biasco**[8,9], **Stuart H. Orkin**[5,6,7,11], **Jesse M. Engreitz**[3,12], **Eric S. Lander**[3,4,13], **J. Keith Joung**[2,10], **Daniel E. Bauer**[5,6,7,*], and **Luca Pinello**[2,3,10,*]

[1]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA.

[2]Molecular Pathology Unit, Center for Cancer Research, Center for Computational and Integrative Biology, Massachusetts General Hospital, Charlestown, MA, USA.

[3]Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[4]Department of Systems Biology, Harvard Medical School, Boston, MA, USA.

[5]Division of Hematology/Oncology, Boston Childrens Hospital, Boston, MA, USA.

[6]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA, USA.

[7]Harvard Stem Cell Institute, Department of Pediatrics, Harvard Medical School, Boston, MA, USA.

[8]Gene Therapy Program, Harvard Medical School, Boston, MA, USA.

[9]Dana-Farber/Boston Childrens Cancer and Blood Disorders Center, Boston, MA, USA.

[10]Department of Pathology, Harvard Medical School, Boston, MA, USA.

[*] bauer@bloodgroup.tch.harvard.edu.

Author contributions

J.Y.H. and L.P. conceived of and developed the CRISPR- SURF framework. M.A.C., M.C.C., and F.S. performed the experiments. C.P.F., D.P., R.F., K.C., J.A.G., L.B.,S.H.0., J.M.E., and E.S.L. provided statistical and experimental expertise. J.K.J., L.P., and D.E.B. oversaw the project and offered feedback and guidance. J.Y.H., L.P.,D.E.B., and J.K.J. wrote the manuscript with input from all other authors.

Competing interests

J.K.J. has financial interests in Beam Therapeutics,Editas Medicine, Endcadia, Monitor Biotechnologies (formerly known as Beacon Genomics), Pairwise Plants, Poseida Therapeutics, and Transposagen Biopharmaceuticals. J.K.J. holds equity in EpiLogic Therapeutics. J.K.J's interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict-of-interest policies. J.K.J. is a member of the Board of Directors of the American Society of Gene and Cell Therapy, and an inventor on patents and patent applications covering CRISPR-based nucleases and gene regulatory proteins. E.S.L. serves on the Board of Directors for Codiak BioSciences and Neon Therapeutics, and on the Scientific Advisory Board of F-Prime Capital Partners and Third Rock Ventures; he is also affiliated with several nonprofit organizations, including through his service on the Board of Directors of the Innocence Project and Biden Cancer Initiative and the Board of Trustees for the Parker Institute for Cancer Immunotherapy. He has served and continues to serve on various federal advisory committees. The Broad Institute, which E.S.L. directs, holds patents and has filed patent applications on technologies related to other aspects of CRISPR.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-018-0225-6.

Data availability

The data are available at SRA under project number PRJNA494935.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[11]Howard Hughes Medical Institute, Boston, MA, USA.

[12]Harvard Society of Fellows, Harvard University, Cambridge, MA, USA.

[13]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

To the Editor –

Tiling screens that use CRISPR-Cas technologies provide a powerful approach for the mapping of regulatory elements to phenotypes of interest[1–6]. Here we present CRISPR screening uncharacterized region function (CRISPR-SURF), a deconvolution framework that can be used to identify functional regulatory regions in the genome from data generated by CRISPR-Cas nuclease, CRISPR interference (CRISPRi), or CRISPR activation (CRISPRa) tiling screens. CRISPR-SURF can be run as a stand-alone command line utility (https://github.com/pinellolab/CRISPR-SURF) or as a web application (http://crisprsurf.pinellolab.org/) (Supplementary Note 1).

The methodology underlying the CRISPR-SURF framework leverages the concept that single guide RNAs (sgRNAs) represent a functional readout for base pairs within the perturbation range. This range depends on the CRISPR screening approach used: CRISPR-Cas nucleases introduce insertion and deletion (indel) mutations of varying lengths (typically < 30 bp, although potentially varying with cell type), whereas CRISPRi and CRISPRa strategies may remodel chromatin structure across hundreds of nucleotides. Importantly, each CRISPR technology offers its own advantage: CRISPRi and CRISPRa strategies increase the likelihood of detecting regulatory elements, given their larger perturbation ranges, whereas CRISPR-Cas nucleases provide higher resolution on the boundaries of regulatory elements, given their sharper perturbation windows. Because each sgRNA perturbs variable-size regions around its target site, the sgRNA data from CRISPR tiling screens can be seen as imprecise measurements of an underlying genomic regulatory signal. To address this variable, we model these imprecise measurements by means of a convolution operation that accounts for the perturbation profiles associated with different CRISPR technologies.

CRISPR-SURF deconvolves tiling screen data to find the genomic regulatory signal that best explains the observed sgRNA scores given the perturbation profile and sgRNA spacing (Fig. 1). The CRISPR-SURF framework accounts for overlapping perturbation profiles between neighboring sgRNAs and leverages shared information to infer the underlying genomic regulatory signal even from noisy measurements. The exact sgRNA targeting coordinates are also taken into account, thus allowing for location-dependent statistical tests with a power that reflects the local density of sgRNAs in a region. This enables CRISPR-SURF to estimate perturbation-specific and position-specific statistical power for CRISPR tiling screens (Supplementary Note 2).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We evaluated the performance of CRISPR-SURF by using three published CRISPR tiling screens spanning CRISPR-Cas9[1], CRISPRi[2], and CRISPRa[3] modalities. For all three datasets, CRISPR-SURF reliably identified all of the experimentally validated regulatory elements. CRISPR-SURF further identified potentially novel regulatory regions supported by both chromatin accessibility and epigenetic marks (Supplementary Notes 3 and 4, Supplementary Figs. 1–3). We elaborate on key differences between CRISPR-SURF and the analysis methods used in these previous studies in Supplementary Notes 5 and 6.

Furthermore, we carried out two matched CRISPR tiling screens using CRISPR-Cas9 (SpCas9) and CRISPRi (dCas9-KRAB) on the *BCL11A* locus (Supplementary Note 7) and found that significant regions identified within previously validated functional enhancers[1,7,8] were narrower in the CRISPR- Cas9 screen than in the CRISPRi screen, consistent with the narrower perturbation profiles of CRISPR-Cas9 indel mutations compared with those of CRISPRi epigenetic modifications (Supplementary Fig. 4). In summary, CRISPR-SURF leverages the broad CRISPRi and CRISPRa perturbation profile for efficient enhancer discovery and the narrow CRISPR-Cas perturbation profile for high-resolution mapping of critical elements within enhancers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank R. Kurita and Y. Nakamura (RIKEN BioResource Center, Tsukuba, Japan) for sharing HUDEP-2 cells. L.P. is supported by NHGRI Career Development Award R00 HG008399 and CEGS RM1HG009490. D.E.B. is support by NIH R03 DK109232, NIH DP2 0D022716, NIH P01 HL032262, the Burroughs Wellcome Fund, and the Doris Duke Charitable Foundation. J.K.J. is supported by NIH R35 GM118158, NIH RM1 HG009490, and the Desmond and Ann Heathwood MGH Research Scholar Award. M.C.C. is supported by NIDDK Award F30-DK103359. J.M.E. is supported by NIH NHGRI 1K99HG009917-01 and the Harvard Society of Fellows.

References

1. Canver MC et al. Nature 527, 192–197 (2015). [PubMed: 26375006]
2. Fulco CP et al. Science 354, 769–773 (2016). [PubMed: 27708057]
3. Simeonov DR et al. Nature 549, 111–115 (2017). [PubMed: 28854172]
4. Korkmaz G et al. Nat. Biotechnol. 34, 192–198 (2016). [PubMed: 26751173]
5. Sanjana NE et al. Science 353, 1545–1549 (2016). [PubMed: 27708104]
6. Klann TS et al. Nat. Biotechnol. 35, 561–568 (2017). [PubMed: 28369033]
7. Bauer DE et al. Science 342, 253–257 (2013). [PubMed: 24115442]
8. Vierstra J et al. Nat. Methods 12, 927–930 (2015). [PubMed: 26322838]

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

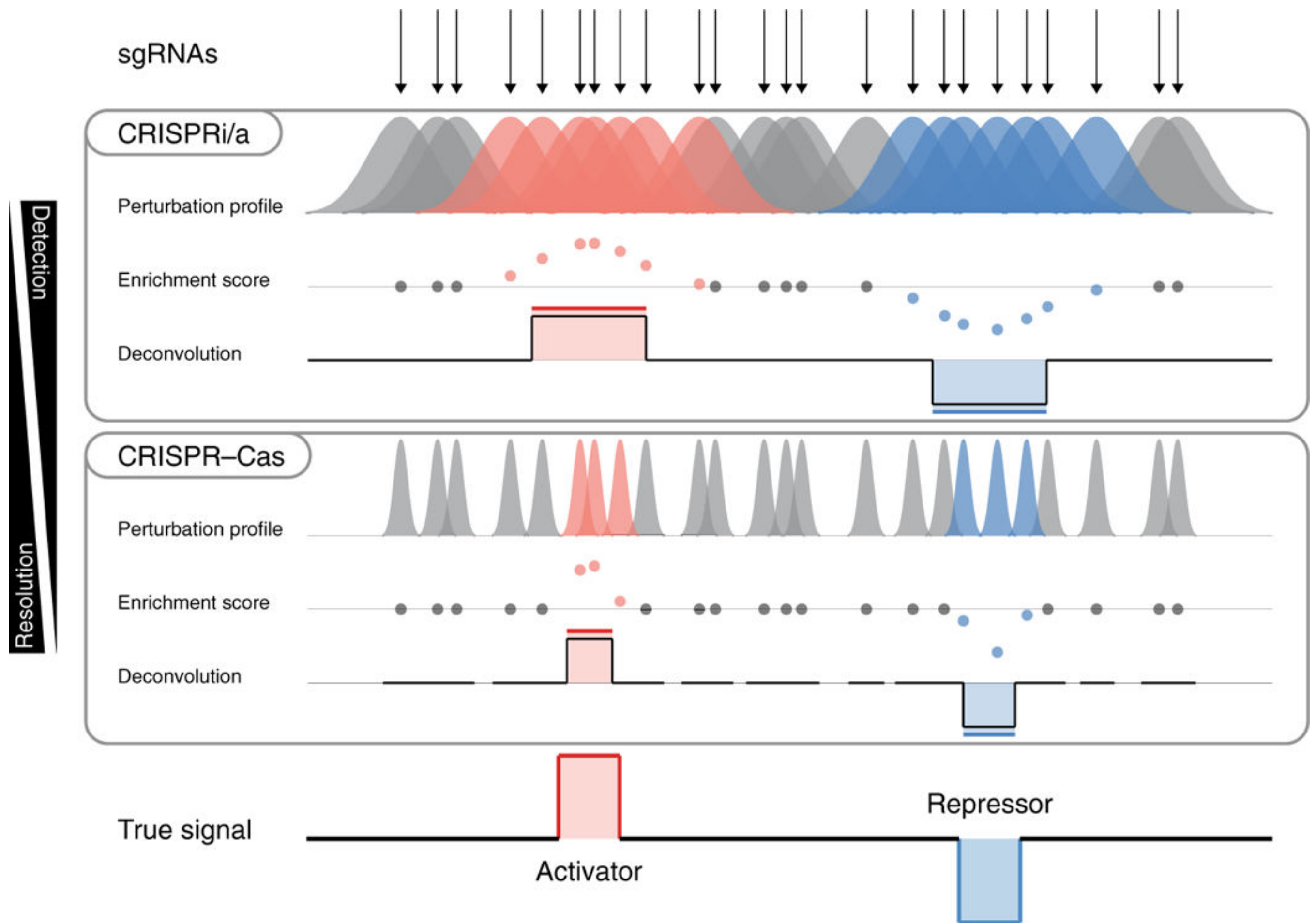

Fig. 1 |. CRISPR-SURF deconvolution framework.

An illustration of the deconvolution based on sgRNA targeting positions, different perturbation profiles (CRISPRi/a and CRISPR-Cas), and enrichment scores.