Supplementary Materials

Molecular Biology of the Cell

Caicedo *et al*.

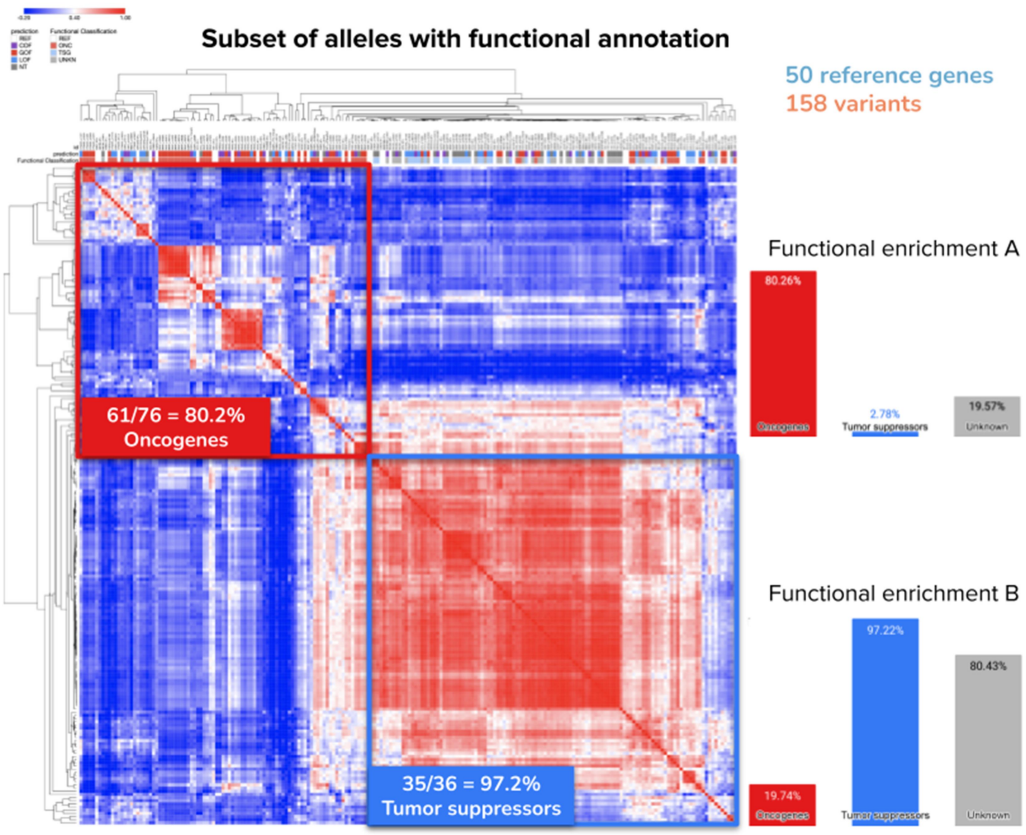
Cell Painting predicts impact of lung cancer variants

Juan C. Caicedo, John Arevalo, Federica Piccioni, Mark-Anthony Bray, Cathy L. Hartland, Xiaoyun Wu, Angela N. Brooks, Alice H. Berger, Jesse S. Boehm, Anne E. Carpenter*, Shantanu Singh*

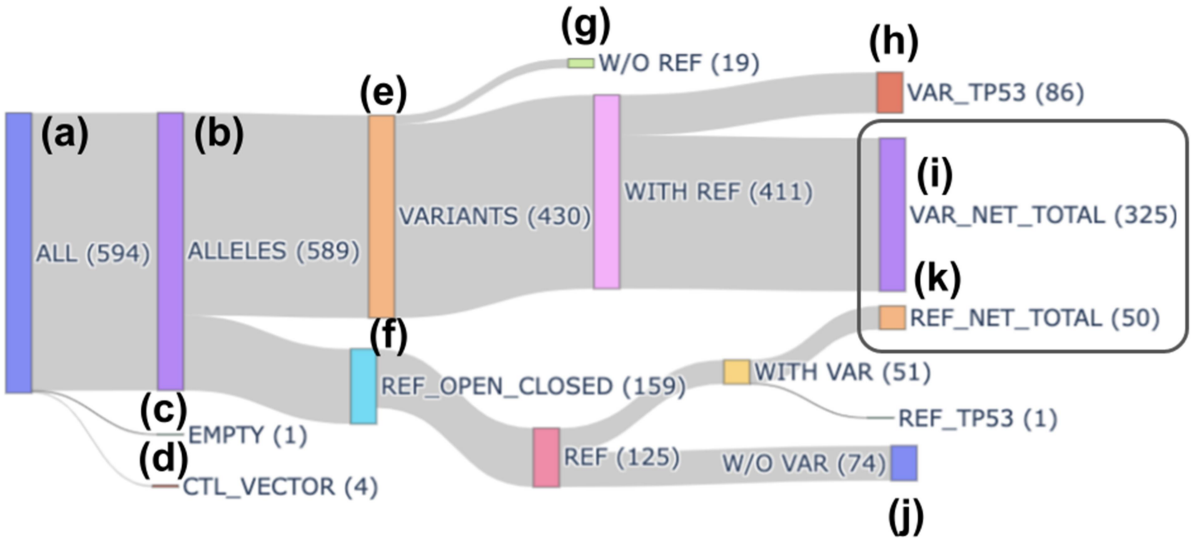
* corresponding authors: shantanu@broadinstitute.org, anne@broadinstitute.org

Supplementary material

All figures and data presented in this manuscript are available in an interactive format at http://broad.io/cmvip.



Supplementary Figure 1. Correlation matrix of Cell Painting profiles for a subset of alleles with functional annotation. After ordering rows and columns according to the hierarchical clustering, the matrix can be divided in two parts: one enriched with oncogenes and the second enriched with tumor suppressor genes. Enrichment here is defined as the proportion of alleles in one functional category that are present in the group with respect to all alleles of that category.



Supplementary Figure 2. Disaggregation of the metadata in the raw dataset. The *x_mutation_status* column tags the content of each assay. It contains 594 unique values (a), 589 (b) corresponds to alleles, EMPTY (c) corresponds to the control tag, and ctl_vectors (d) are not considered in this analysis. Out of the 589, 430 (e) are VARIANTS and 159 (f) are REFERENCE alleles. We discarded 19 variants (g) without reference, and 86 TP53 Variants (h) resulting in the 325 alleles(i). We also discarded 74 references that do not match with any variant (j), ending up with 50 (k) REFERENCE alleles. The enclosed subset represents the data used in this study.

Classification of 20 benchmark alleles

Allele	Reference	L1000	CellP	Prior annotation
ARAF_p.S214F	Imielinski et al., 2014	COF	LOF	GOF
ARAF_p.S214C	Imielinski et al., 2014	COF	COF	GOF
ARAF_p.D429A	Imielinski et al., 2014	GOF	COF	LOF
CTNNB1_p.S37C	Palacios et al., 1998	GOF	GOF	GOF
EGFR_p.ELREA746del	Greulich et al., 2005	COF	COF	GOF
EGFR_p.L858R	Greulich et al., 2005	NT	LOF	GOF
KEAP1_p.G333C	Hast et al., 2014	LOF	LOF	LOF
KRAS_p.G12Y	Prior et al., 2012	COF	GOF	GOF
KRAS_p.G12A	Prior et al., 2012	GOF	GOF	GOF
KRAS_p.G12C	Prior et al., 2012	GOF	GOF	GOF

Allele	Reference	L1000	CellP	Prior annotation
KRAS_p.G13V	Prior et al., 2012	GOF	COF	GOF
KRAS_p.G12R	Prior et al., 2012	GOF	GOF	GOF
KRAS_p.G12S	Prior et al., 2012	GOF	GOF	GOF
KRAS_p.G12D	Prior et al., 2012	GOF	COF	GOF
KRAS_p.G12F	Prior et al., 2012	GOF	GOF	GOF
NRAS_p.Q61L	Prior et al., 2012	COF	GOF	GOF
RIT1_p.F82L	Berger et al., 2014	COF	COF	GOF
RIT1_p.R122L	Berger et al., 2014	GOF	LOF	GOF
STK11_p.G242W	Olschwang et al., 2001	LOF	LOF	LOF
STK11_p.D194Y	Engel et al., 2014	LOF	LOF	LOF

Supplementary Table 1. Benchmark classification.

False Positives Rate on mock REF vs VAR pairs

Control Allele	FPR
EGFR_WT.c	7.70%
EGFR_p.L858R	7.80%
EGFR_p.T790M, p.L858R.o	8.00%
KEAP1_WT.c	6.60%
KRAS_p.G12V	7.10%
MDM2_WT.c	5.40%
NFE2L2_WT.c	7.60%
STK11_WT.c	7.90%
EMPTY	2.70%
Average	**6.75%**

Supplementary Table 2. Results of the false-positive analysis with mock alleles.