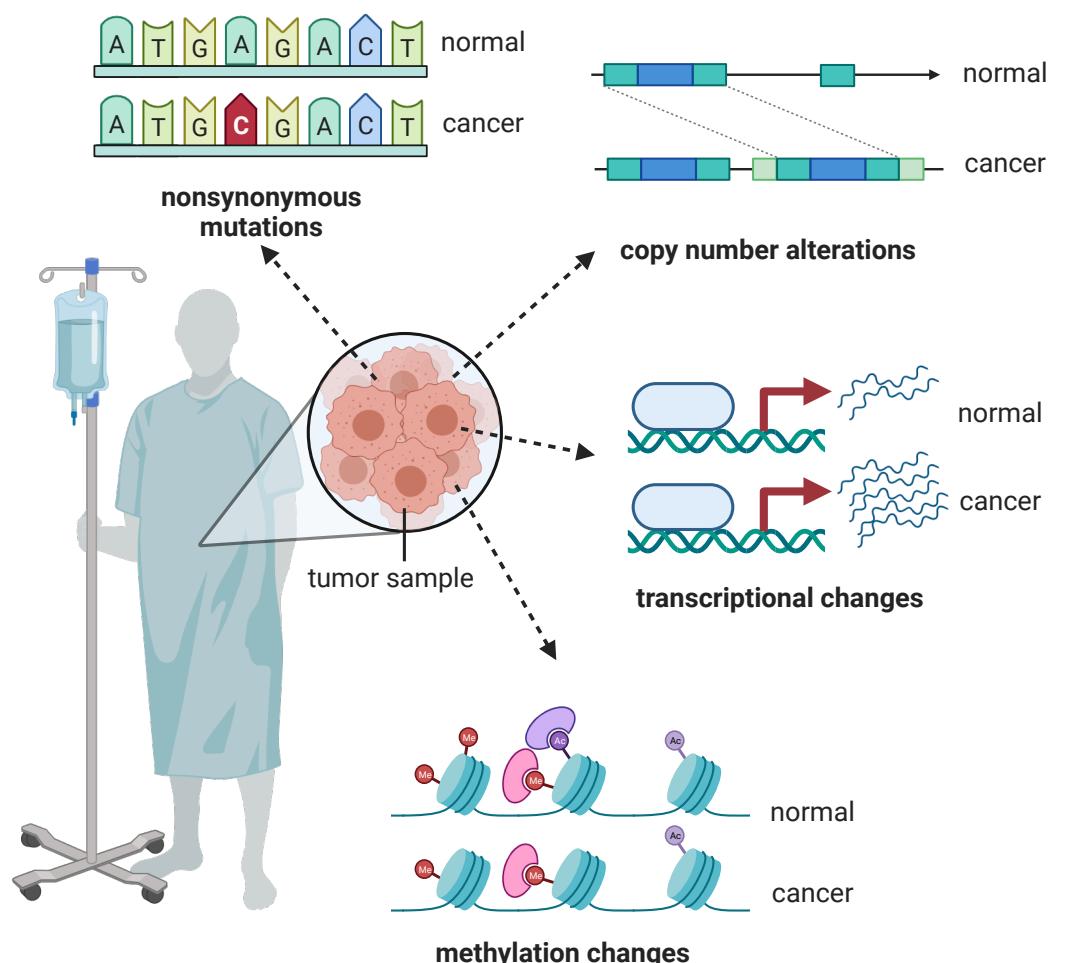


Integrative computational framework, *Dyscovr*, links mutated driver genes to metabolic dysregulation across 22 cancer types

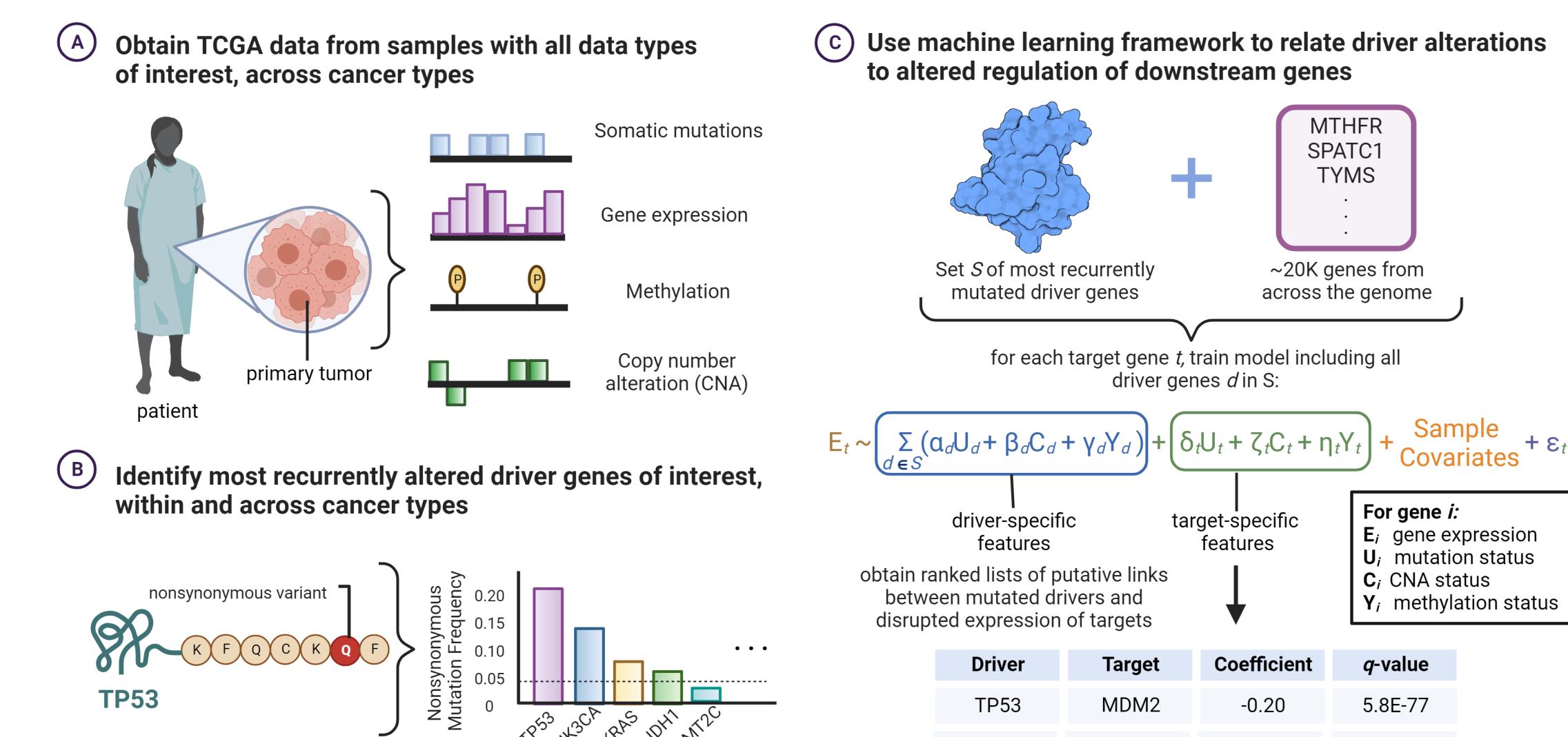
Sara Geraghty, Jacob Boyer, Mahya Fezel-Zarandi, Matthew McBride, Joshua Rabinowitz, **Mona Singh**
Lewis-Sigler Institute for Integrative Genomics, Princeton University

INTRODUCTION

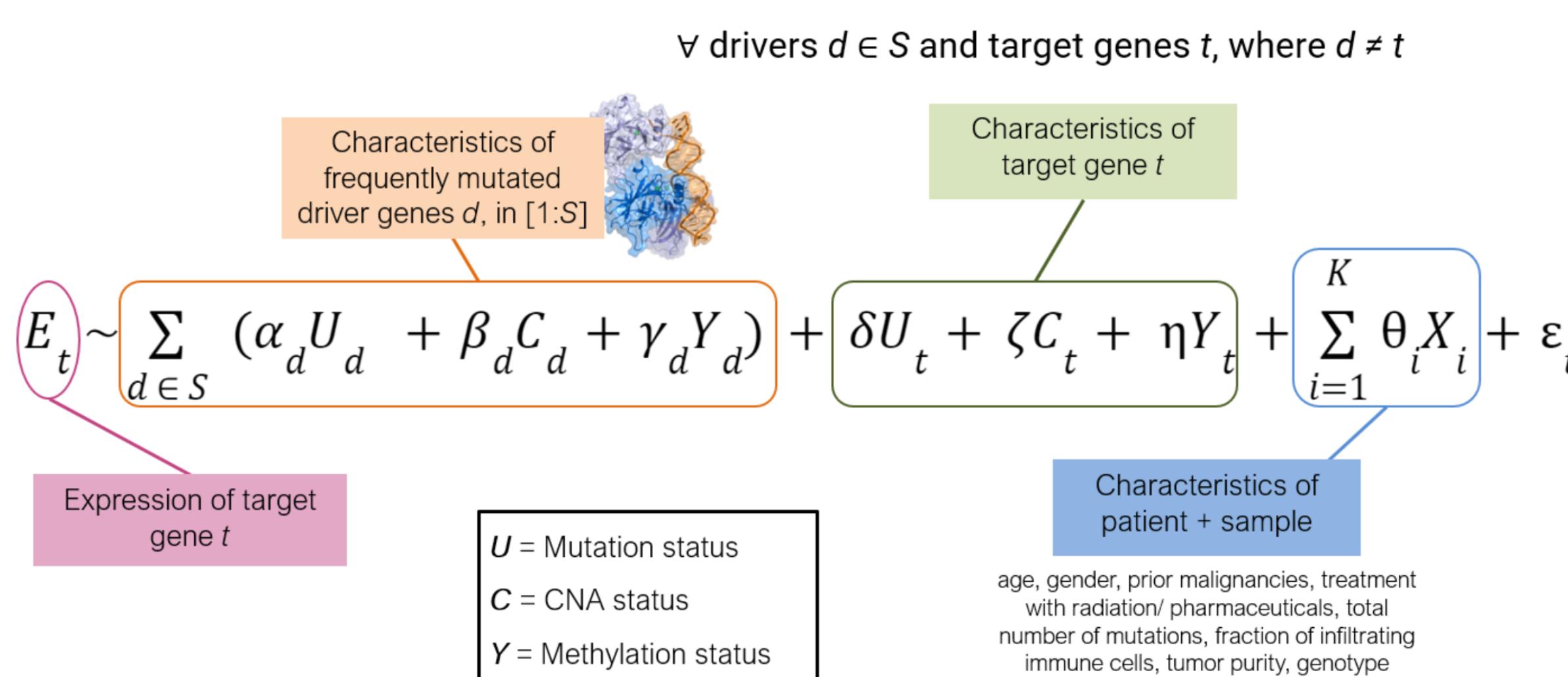


Consortia like the Cancer Genome Atlas (TCGA) are rich sources of multiomic data. With this data, we've made substantial strides in identifying genes that drive cancer ("cancer drivers"). However, there remain a variety of ways in which these drivers can impact downstream molecular phenotypes, complicating the development of effective therapeutics that target them. Here, I detail a machine learning (ML) model that, within and across 22 cancer types, uses this data to deconvolve the cascade of molecular changes that occur following a driver mutation.

METHODOLOGICAL OVERVIEW

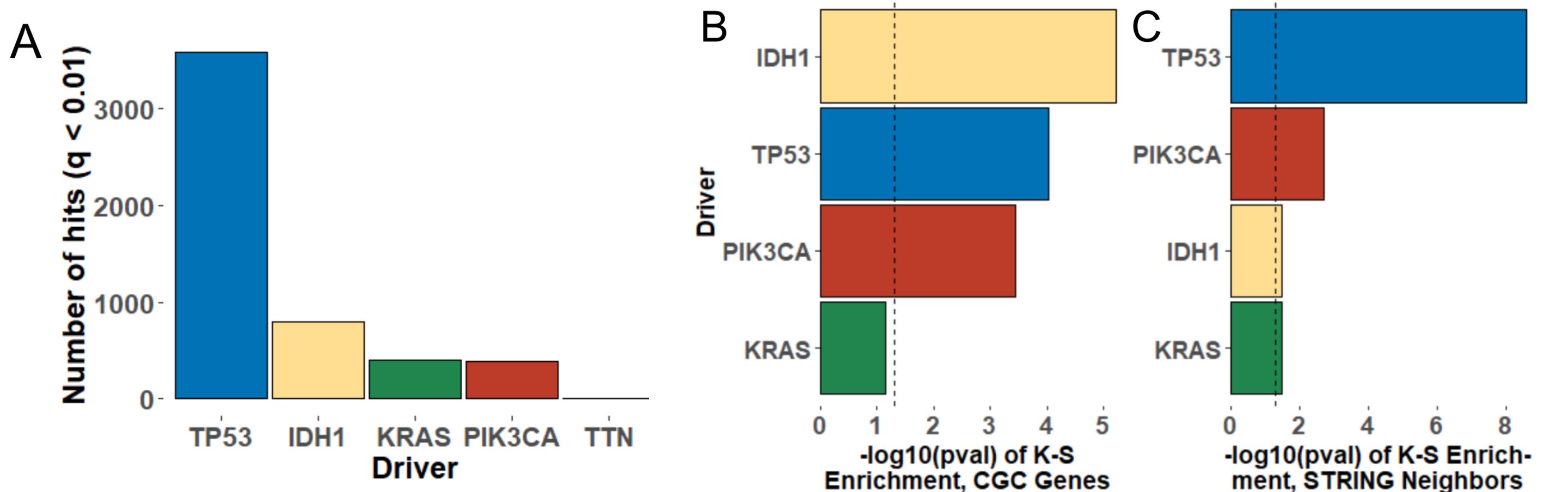


Our framework, *Dyscovr*, relates recurrently altered cancer drivers, d , to transcriptional changes in candidate target genes t across the genome while taking other molecular phenotypes (e.g. copy number and methylation) and clinical features (e.g. age, gender, TMB, tumor purity, genotypic variation, etc.) into account. *Dyscovr* outputs a ranked list of correlations between the mutation status of a driver gene and the expression of a given target gene pan-cancer or within a cancer type.

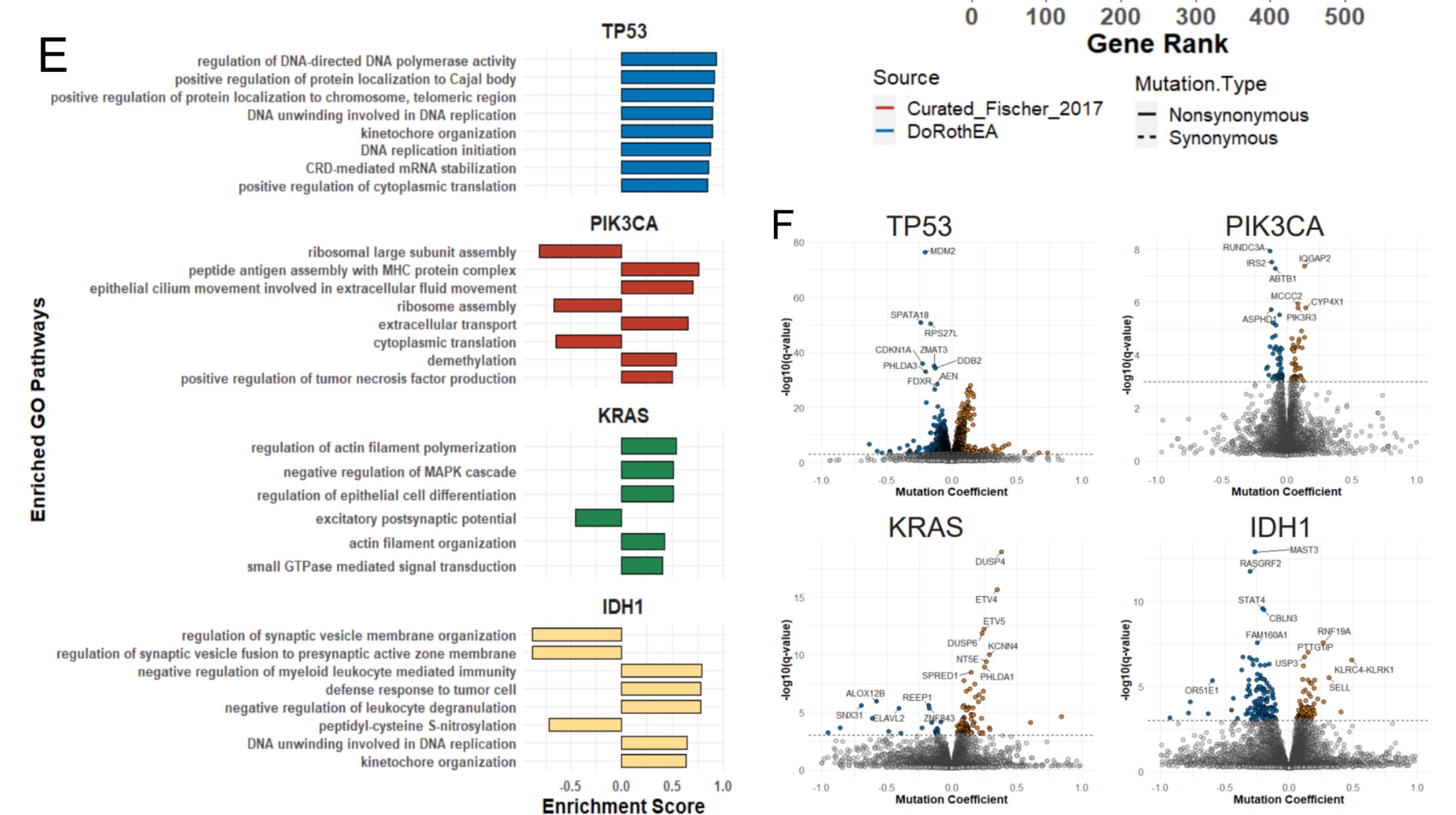


DYSCOVR IDENTIFIES THOUSANDS OF CANCER-RELEVANT CORRELATIONS PAN-CANCER

Pan-cancer, *Dyscovr* identifies hundreds to thousands of significant correlations for each driver gene mutated in at least 5% of all cancers, *TP53*, *IDH1*, *KRAS*, and *PIK3CA* ($q < 0.01$), but not for highly mutated passenger gene *TTN* (A). These correlations are enriched in cancer genes from the Cancer Gene Census (CGC) (B) and each drivers' set of neighbors from the functional interaction network STRING (C), suggesting genuine functional relevance.

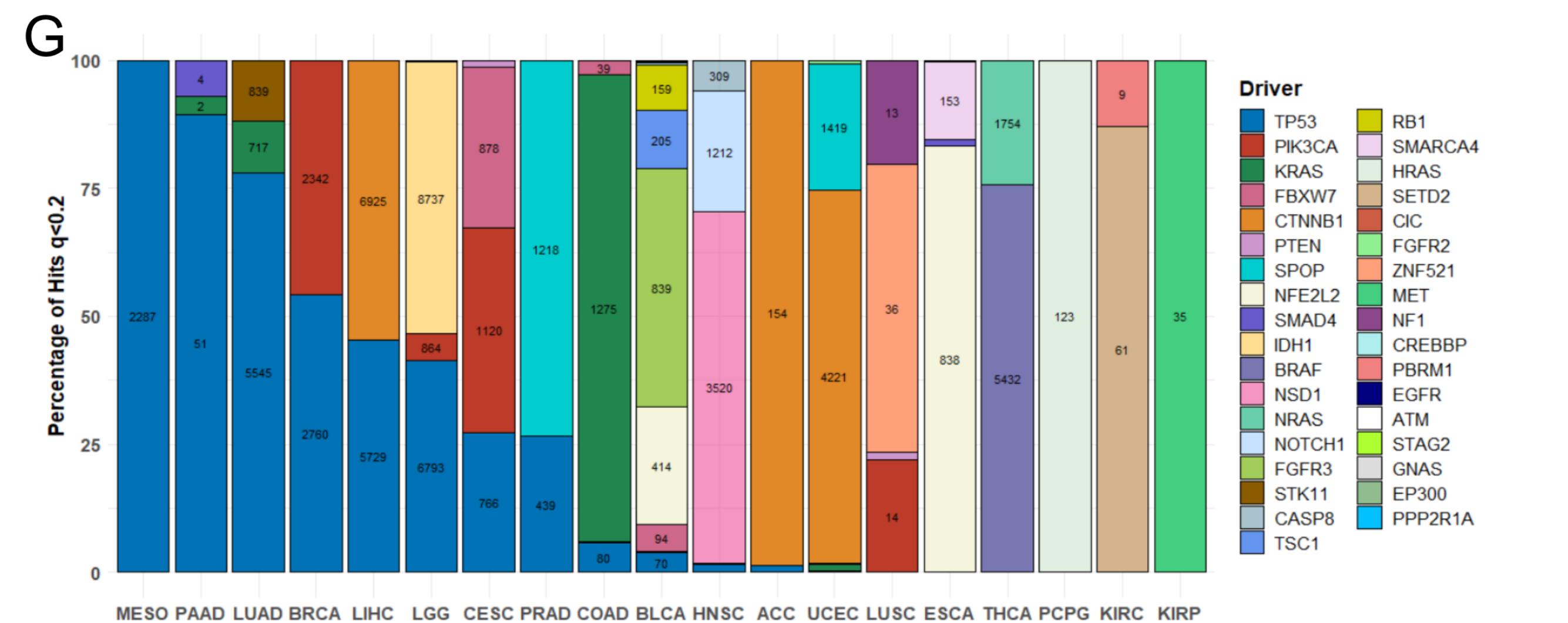


In the case of *TP53*, which has gold-standard sets of known targets from curation (e.g. Fischer, 2017, 10.1038/ Onc.2016.502), and databases like DoRothEA (10.1101/gr.240663.118), we see strong enrichment of known targets when considering nonsynonymous mutations, but not silent mutations (D). For *TP53*, *PIK3CA*, *KRAS*, and *IDH1*, pan-cancer correlations from *Dyscovr* are enriched in cancer-relevant GO pathways such as DNA replication initiation, actin filament polymerization, demethylation, and mRNA stabilization (E).

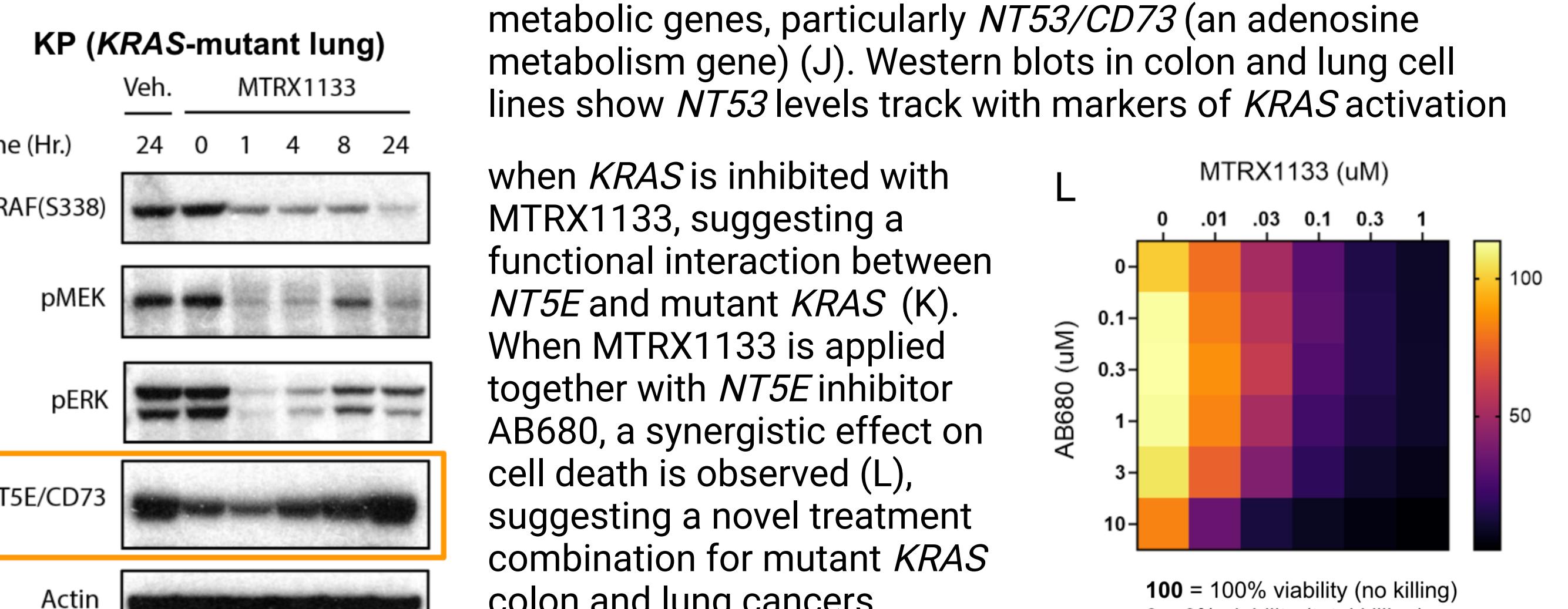


Dyscovr's pan-cancer correlations include known driver mutation-target expression relationships, such as *TP53* mutation and *MDM2* downregulation, *KRAS* mutation and *ETV4* upregulation, and *PIK3CA* mutation and *PIK3R3* upregulation (F). Many of these correlations, however, are novel.

DYSCOVR UNCOVERS EXPERIMENTALLY VERIFIABLE CORRELATIONS IN 22 CANCER TYPES



We ran *Dyscovr* individually across 22 TCGA cancer types, identifying thousands of cancer type-specific correlations ($q < 0.2$). Each cancer type has its own landscape of recurrently mutated drivers, and significant correlations for these drivers (G): a rich resource of therapeutic prospects. These correlations are consistent when run on external breast cancer samples from METABRIC (10.1038/nature10983); *Dyscovr*'s results for *TP53* and *PIK3CA* (the most frequent BRCA drivers) in METABRIC overlap significantly (H) and are strongly correlated (I) with TCGA-BRCA, suggesting our results are reliable across patient cohorts.



ACKNOWLEDGEMENTS

A heartfelt thanks to the Singh lab for their advice and support. This work is funded by the National Science Foundation through a Graduate Research Fellowship (GRFP) and Ludwig Cancer Research.