An Augmented, Extended Cancer Gene Panel Integrating DNA and RNA Sequencing Comprehensively Assesses Neoplastic Genetic Lesions

Michael James Clark, Sean Michael Boyle, Elena Helman, Deanna Church, Gabor Bartha, Jason Harris, Stephen Chervitz, Anil Patwardhan, Sarah Garcia, Scott Kirk, Richard Chen, John West

Personalis, Inc. | 1350 Willow Road, Suite 202 | Menlo Park, CA 94025

Contact: michael.clark@personalis.com

The Complexity of Cancer Mutations

Through next-generation sequencing, we have come to understand cancer as a genomic disorder affecting key genes in cell growth and survival pathways. Cancer is a disease of driver mutations, often in multiple genes, which come in a plethora of forms. Cancer mutations includes small variations, copy number alterations (CNAs), gene fusions, alternative splicing, changes in gene expression, and more. We must assess the different types of cancer mutations to accurately determine prognosis and make therapeutic decisions. Tumor purity and heterogeneity reduce the variant allele frequency (VAF) of mutations in cancer samples to boot, so we must also be robust against low representation alleles.

## Assay Dictates Sensitivity to Different Variant Types

Assay	Typical Sequencing Amount	Mean Depth	Variant Types Detectable								
			SNVs	Indels	CNAs	Fusion	Gene Expression	Splice Isoforms	Allelic Expression	Low Detection Limit	Cross-Validation
Exome	12Gb	100x	●	●	●						
Transcriptome	14Gb	--	●	●		●	●	●	●		
Whole Genome	>100Gb	30-60x	●	●	●						
Cancer Panel (DNA+RNA)	15Gb	500x	●	●	●	●	●	●	●	●	●

Different assays are capable of detecting different variant types. In cancer there are a wide array of different variant types that can act as driver mutations. In order to detect all of the major variant types, one must assess both DNA and RNA. Moreover, given the mixed purity and heterogeneity of neoplastic tissues, a limit of detection down into the single digit percentile range is vital.

A comprehensive, high depth cancer panel applied to both DNA and RNA from the same sample is well equipped to assess all of these features, and has the unique advantage of internal cross-validation.

Designing an Optimized Cancer Panel

Use an Optimized Enrichment Strategy

BRAF, NF1, SDHD, TP53, MAP2K1
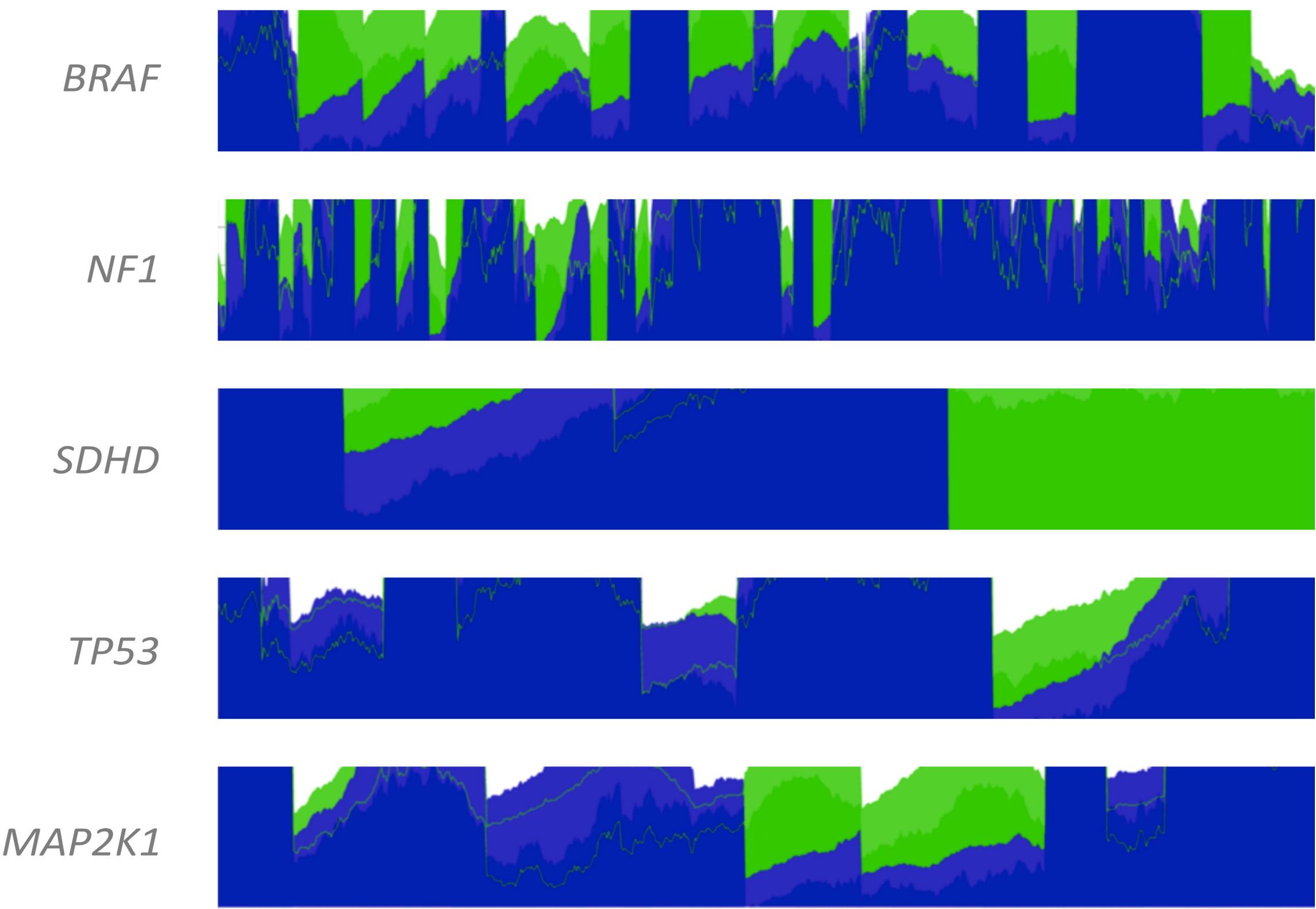
Standard enrichment strategies typically suffer in regions with a molecular structure that does not match the majority of the genome's sequence, such as regions of high GC content.

We developed a novel technology called ACE (Accuracy and Content Enhanced) that fills in the gaps that would typically be seen with a standard enrichment assay. ACE uses a custom-designed probe set and includes multiple enrichment steps.

In the figure to left, the exons of five major cancer genes are plotted along the horizontal axis, and the mean read depth across 16 samples is plotted along the vertical axis. In blue is standard exome coverage depth at 12Gb of sequencing. In green is the augmented coverage depth added by ACE.

Using ACE enrichment strategies allows us to more completely capture all targeted regions.

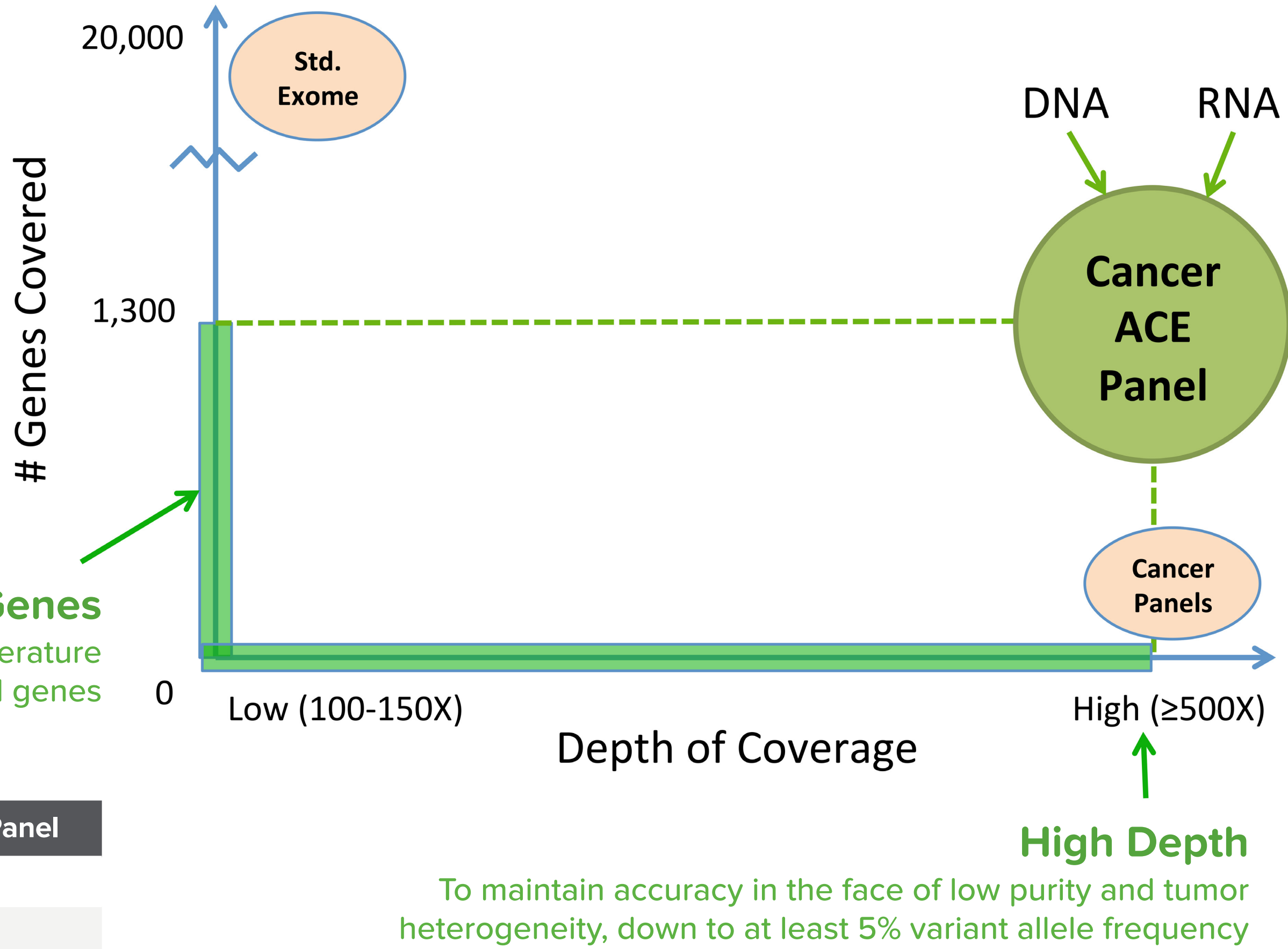### Choose Genes to Include on the Panel

We took a data-driven approach to generating a list of all cancer-relevant genes by referring to numerous databases and the literature, as well as key cancer pathways. We also scrutinized the list for any key cancer genes that may have been missed manually. We arrived at a panel design that includes over 1,300 cancer-related genes.

Panels can be sequenced to very high depth. While a panel of over 1,300 genes is considered quite large in general, it is still of a size that allows for sequencing to mean depths of over 500x with high sensitivity for alleles with VAF<5%.

One final advantage to an enrichment-based panel design is the ability to target both RNA and DNA, facilitating integrated analysis.
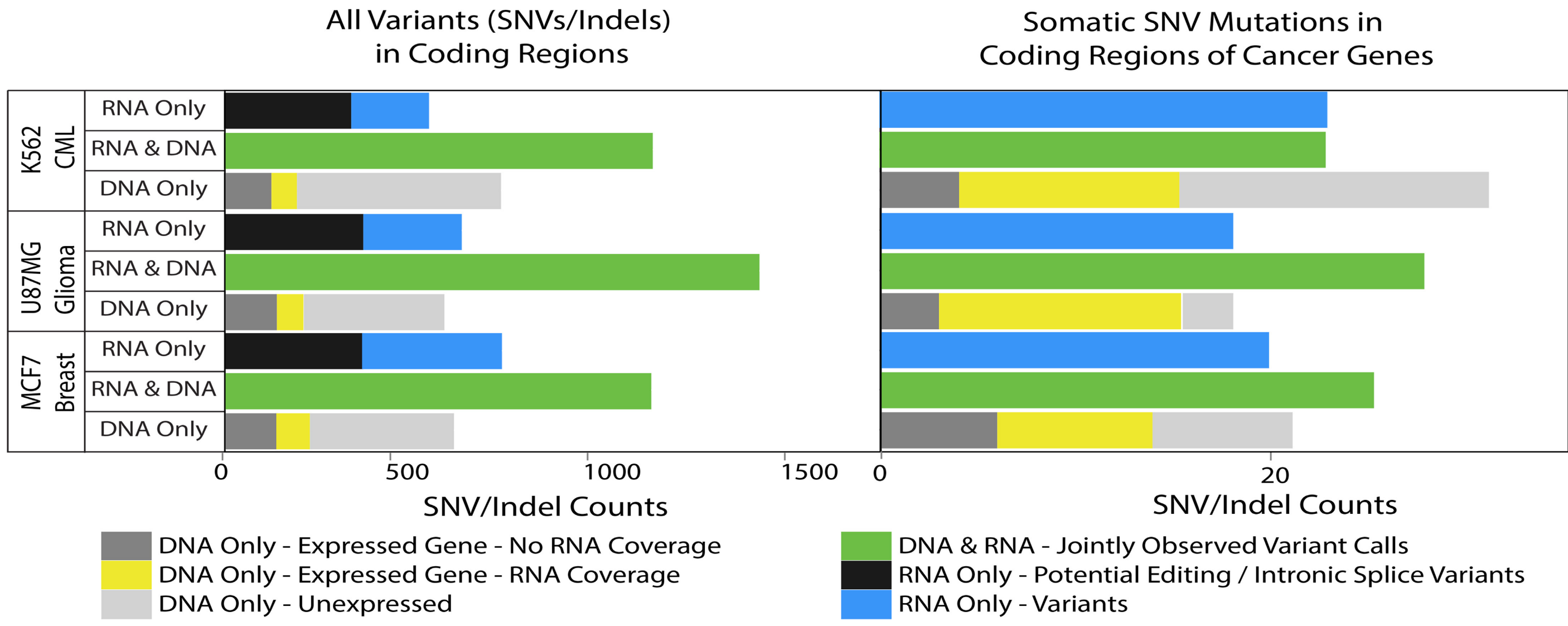

>1,300 Cancer-Related Genes
Clinically actionable genes Consensus driver genes Current literature Public databases Key cancer pathways Clinical trial genes

High Depth
To maintain accuracy in the face of low purity and tumor heterogeneity, down to at least 5% variant allele frequency
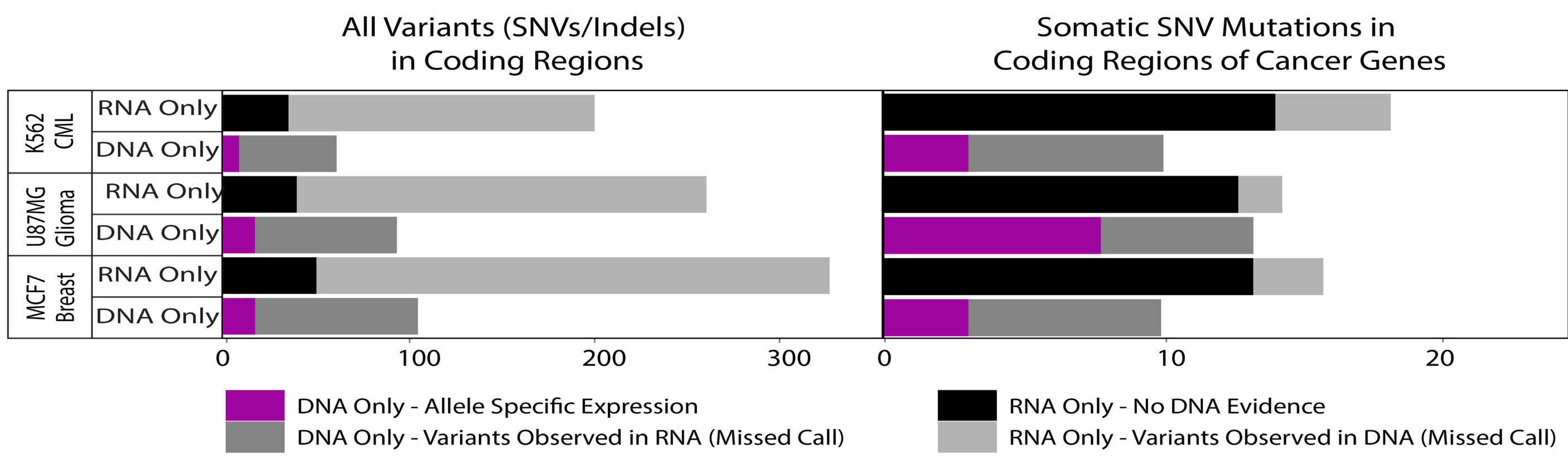
Cancer pathways are well established. Including genes on the cancer gene panel in an objective, data-driven manner results in the inclusion of a substantially larger number of cancer pathway genes than is seen on more limited cancer gene panels (anonymized as "A" and "B" in the adjacent table).

Pathway	Panel A	Panel B	ACE Cancer Panel
Apoptosis	17	18	28
Cell Cycle	28	31	44
JAK-STAT	26	33	45
MAPK	37	46	70
miRNA	51	53	175
Notch	5	9	15
PI3K-AKT	67	67	106
P53	18	19	25
Transcriptional	30	35	91

Integrating Small Variant Calls from RNA and DNA

All Variants (SNVs/Indels) in Coding Regions / **Somatic SNV Mutations in Coding Regions of Cancer Genes**

DNA Only - Expressed Gene - No RNA Coverage
DNA Only - Expressed Gene - RNA Coverage
DNA Only - Unexpressed
DNA & RNA - Jointly Observed Variant Calls
RNA Only - Potential Editing / Intronic Splice Variants
RNA Only - Variants

By performing small variant calling in both DNA and RNA, we can cross-validate calls in one using calls in the other. The majority of all variants in a cancer cell line set are called in both RNA and DNA, shown to the right.

All Variants (SNVs/Indels) in Coding Regions / **Somatic SNV Mutations in Coding Regions of Cancer Genes**

DNA Only - Allele Specific Expression
DNA Only - Variants Observed in RNA (Missed Call)
RNA Only - No DNA Evidence
RNA Only - Variants Observed in DNA (Missed Call)

Through a closer look at the variants called only in RNA or DNA alone, we are able to recover missed calls in one molecule or the other. Integrating the calling of variants between the RNA and DNA allows us to recover these variants that otherwise would have been missed if assessing only one of the two molecules. In addition, variants displaying skewed allelic expression may alter diagnostic conclusions.

Integrating Copy Number and Gene Expression
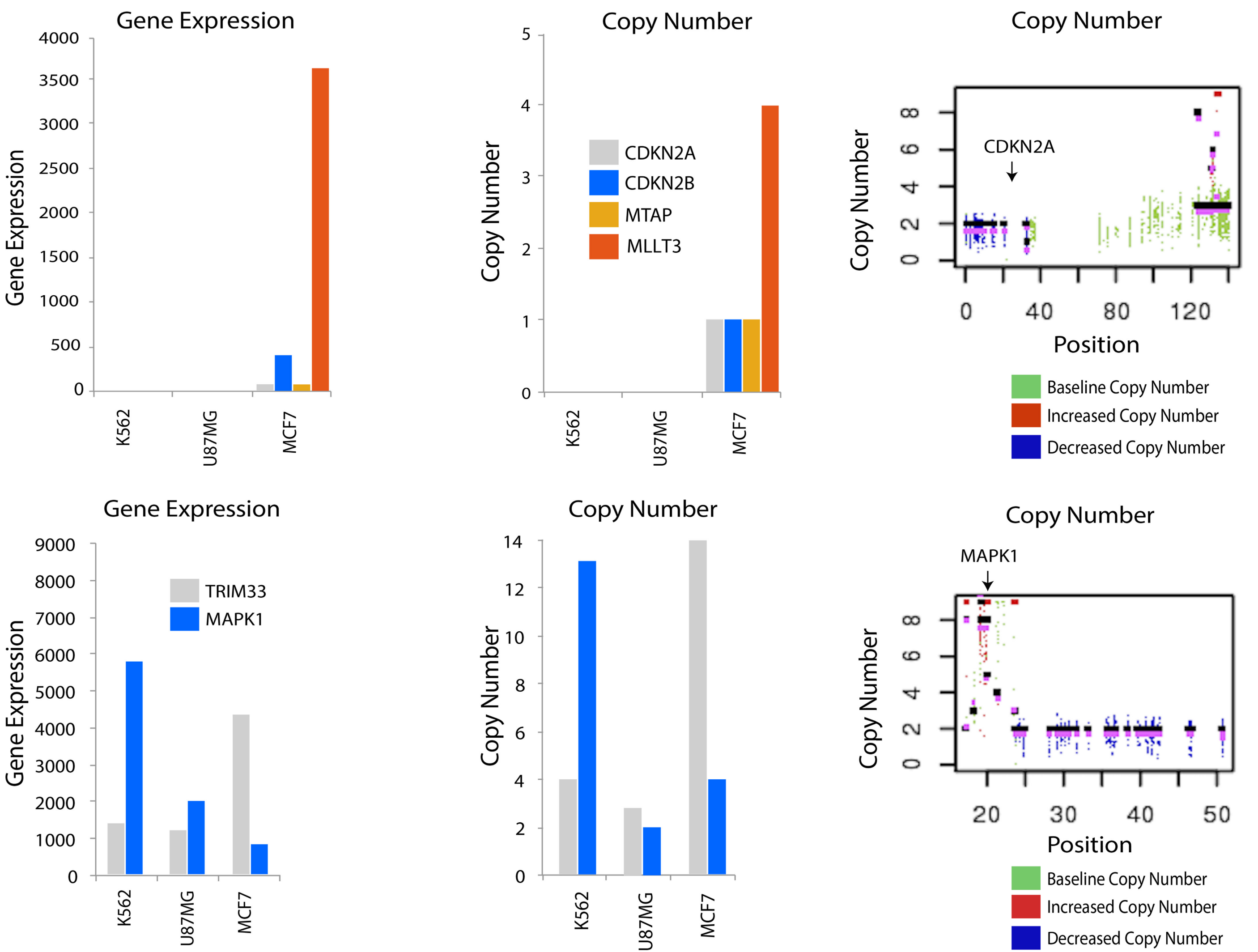
Copy number state in DNA can correlate with expression in the RNA.

As expected, complete gene deletions lead to zero expression of the gene.

Genes present at one copy in a cell line are expressed at a lower level than genes present at four copies.

Similarly, hyper-amplified genes demonstrate substantially higher expression than non-amplified genes in the same cell line.

These findings are not unexpected. It is known that copy number and gene expression can correlate with one another, and hyper-amplification and complete deletion are well described mutational mechanisms in cancer samples.

However, without looking at the RNA in conjunction with the DNA, there is no way to be certain that any given CNA will affect RNA expression.

Gene Fusions Correlate with Copy Number Alteration Boundaries

K562 - CML / **U87MG - Glioma** / **MCF7 - Breast**

K562 - Chr22 / **K562 - Chr9** / **MCF7 - Chr17**
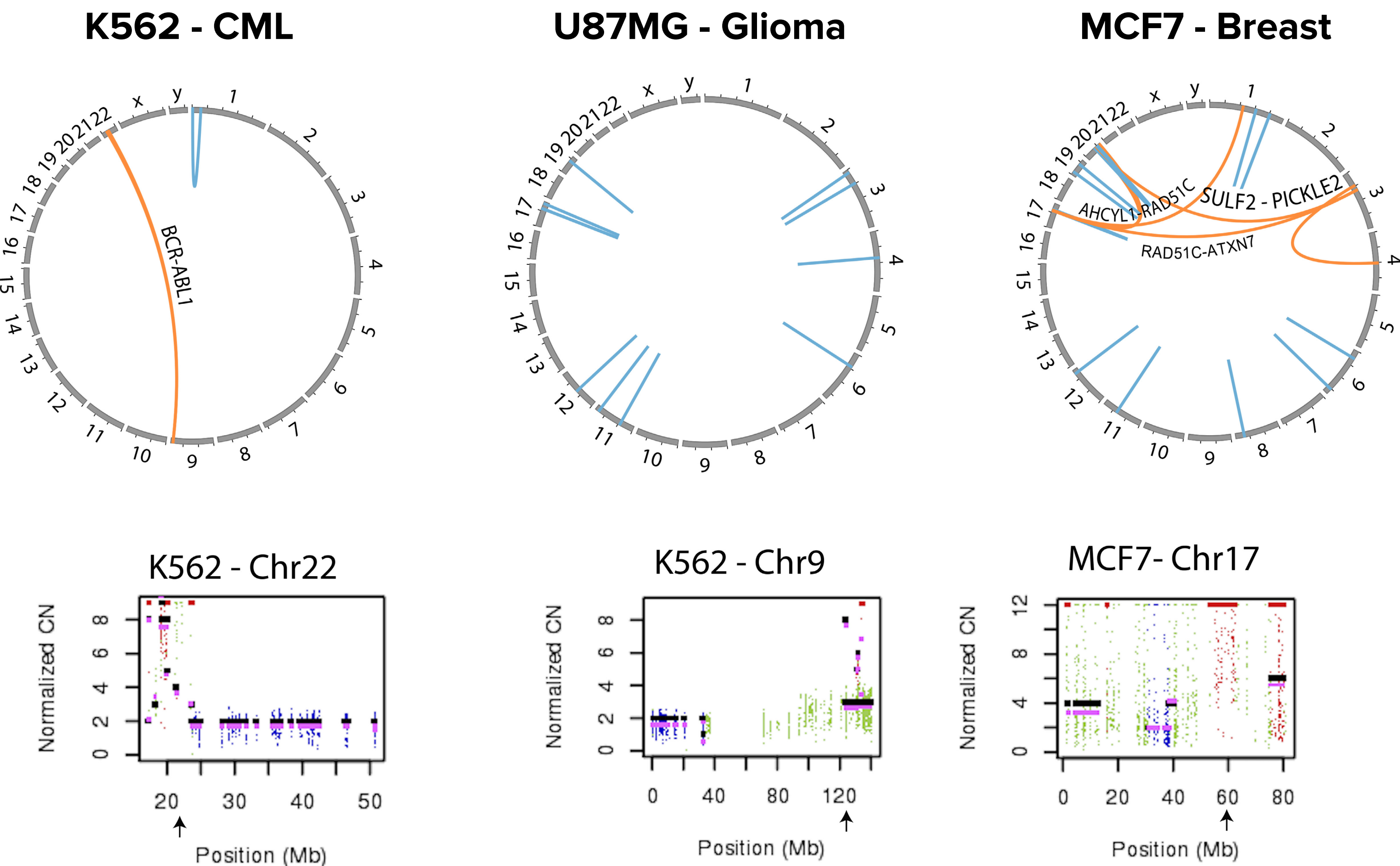
Another set of variant types that benefit from integrated analysis are gene fusions from the RNA with CNA calls from the DNA.

Gene fusions are mediated by chromosomal translocations. In the circle plots to the left, fusion genes are represented as orange lines connecting two genes together.

Cross-referencing our findings from the fusion analysis in RNA with the CNA calling from our gene panel, we are able to see clearly that the fusion events are happening right at positions where we see a shift in copy number, demonstrating that the fusions are not the only consequence of the translocation events in the DNA. Rather, these translocations also flank and are potentially facilitating copy number alterations.