Integrated Analysis of Transcriptomes and Exomes in Cancer Samples Improves Interpretation and Reveals Additional Therapeutic Insights

Sean Michael Boyle, Michael James Clark, Elena Helman, Deanna Church, Shujun Luo, Scott Kirk, Parvin Sripakdeevong, Mirian Karbleashvili, Mark Pratt, Michael Snyder, Russ Altman, Rich Chen, & John West

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

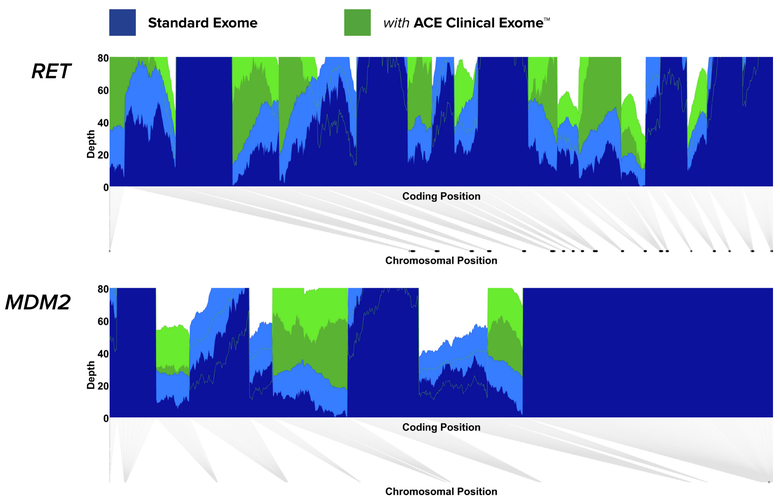
Contact:

sean.boyle@personalis.com

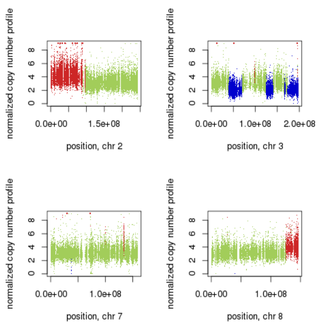
Introduction

Next-generation sequencing is being increasingly applied as a method for cancer analysis in both research and clinical settings. Many basic research studies utilize gene panel, exome, genome, or transcriptome sequencing to assess the genetic basis of tumor progression. These analyses not only help to determine cancer progression, but can also guide therapeutic decision-making. While gene panel, exome, and whole genome sequencing of tumors are widely used for clinical guidance and translational research, transcriptome sequencing has yet to be widely adopted in the clinical environment. However, analyzing the transcriptome allows for observation of expression changes as well as direct detection of functional gene fusions, SNVs, and Indels. To assess these unique features of RNA, we performed whole transcriptome sequencing along with ACE exome sequencing of cancer samples. Our sample set included commonly used cell lines which contain known variants and gene fusions.

Analyzing Cancer Samples Using ACE Exomes

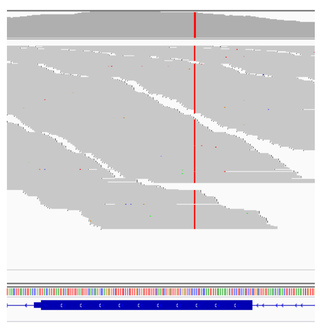


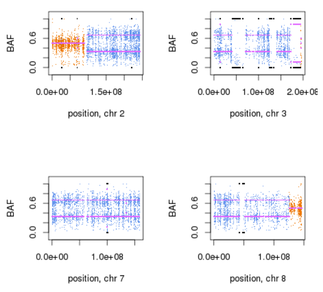
Applying Personalis ACE Exome to Finish Genes - The ACE Exome was specifically designed to increase both gene coverage and the ability to detect variants in medically important genes. Here the ACE Exome, which is displayed in green, fills in important regions of cancer genes which are missed by traditional enrichment techniques. We applied this enrichment for each of our DNA analyses.



Copy Number Analysis - We apply cutting edge techniques to identify copy number change between tumor and normal tissue samples.

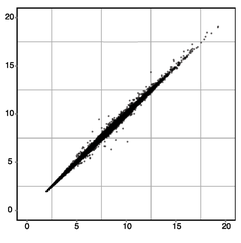
SNV / Indel Variant Calling - A carefully selected set of core tools call both germline and somatic variants in normal and/or tumor tissue samples. Identified variants are annotated from a very large set of genomic and pharmacogenomic databases.





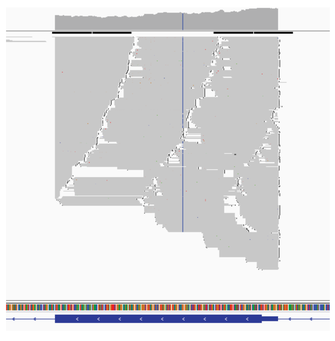
Loss of Heterozygosity - LOH is a very important genomic feature in cancer and, when interpreted with Copy Number Analysis, provides a snapshot of structural genomic shifts within a tumor.

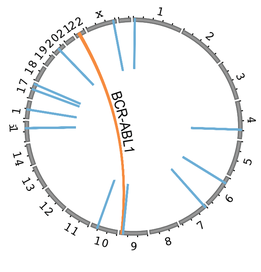
Analyzing Cancer Samples Using Transcriptomes



Gene Expression Analysis - We apply well known and thoroughly vetted approaches to calculate gene expression between either tumor/normal samples or groups of tumor samples.

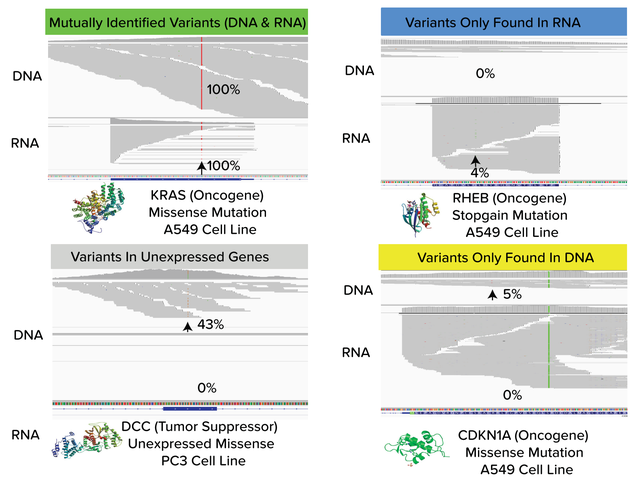
SNV / Indel Variant Calling - A carefully selected set of core tools call both germline and somatic variants in normal and/or tumor tissue samples. Identified variants are annotated from a very large set of genomic pharmacogenomic databases.



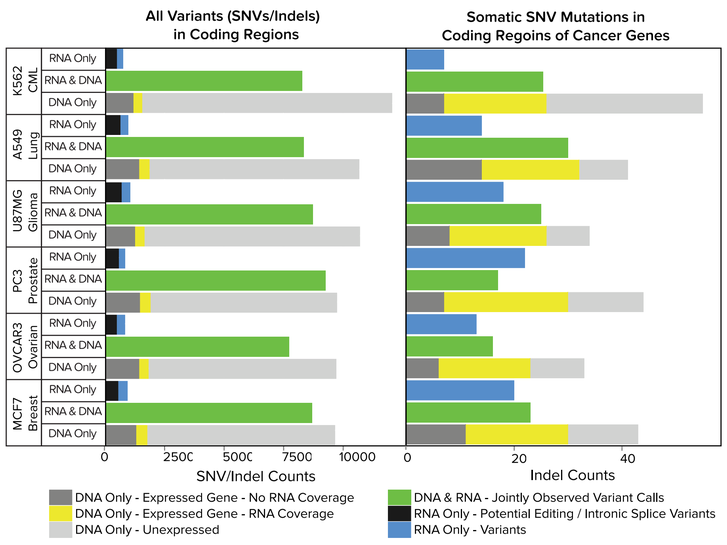


Gene Fusion Detection - Structural rearrangements can lead to functionally important gene fusions. We identify these fusions at base resolution and annotate them 2015105020151050 for improved downstream interpretation.

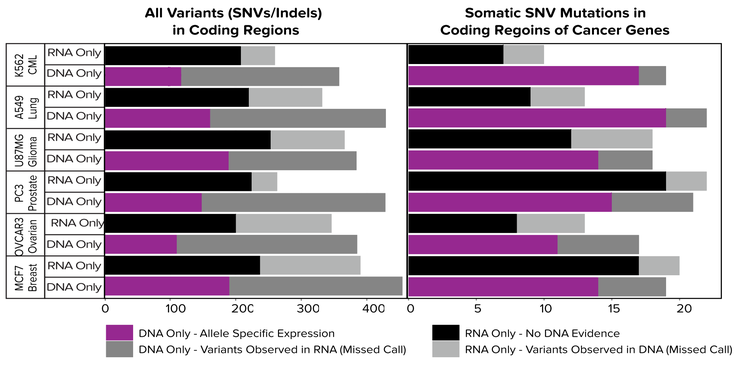
Integrating RNA and DNA Variant Calling



Integrated Analysis Allows for Sub-classification of Variants - By identifying variants in both RNA and DNA it becomes possible to determine which variants have the strongest evidence (DNA & RNA), are in genes which are unexpressed (DNA only), or were not found to be expressed (RNA only).

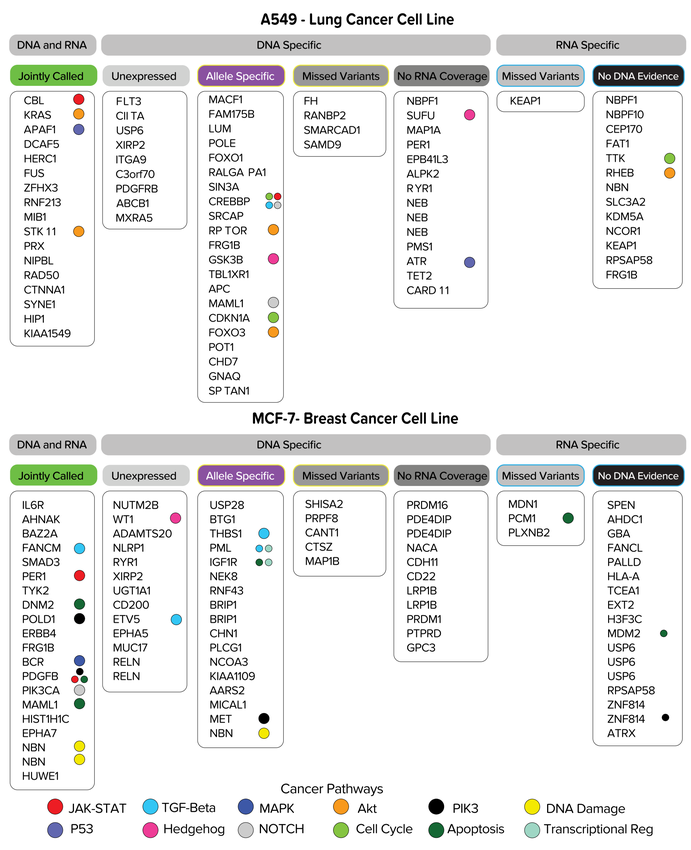


Sub-classification of Variants Improves Insight - We call and annotate SNVs and Indels for several ATCC cell lines, analyzing those found in coding regions of any gene and, separately, only genes known to be associated with cancer. Many variants found only in DNA are unexpressed and the majority of expressed variants are called in both RNA and DNA.



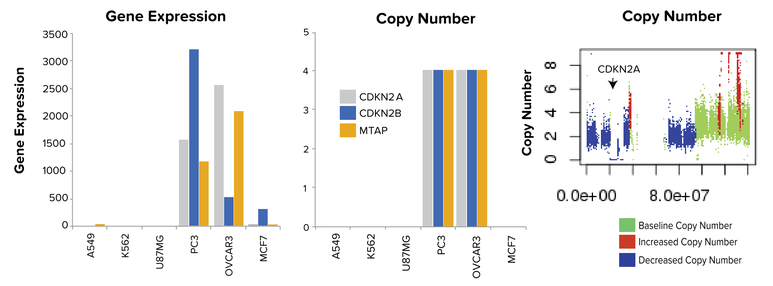
Insights from RNA and DNA Specific Variant Calls - We next analyze variants that are exclusive to RNA or DNA in order to better understand why they are detected in only one class. Interestingly, many are found to be allele specific.

Leveraging Integration for Improved Insight

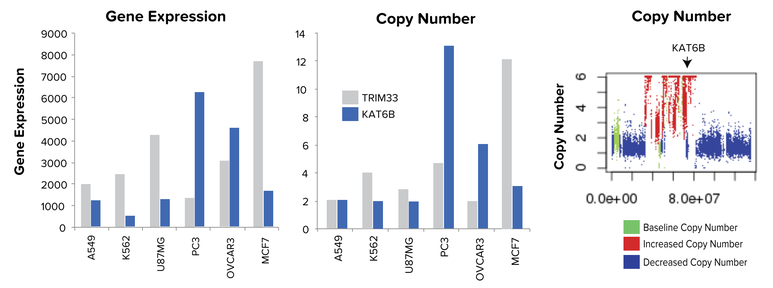


Inspecting RNA and DNA Specific Variants at the Pathway Level - Pathway analysis is a critical aspect of variant interpretation in cancer samples. Overlaying well-known cancer pathways onto each variant classification, which is only possible through joint RNA-DNA analysis, empowers researchers, allowing them to focus on variants which are most likely driving cancer progression.

Integrating Gene Expression and CNVs



Combined Focal Deletion and Lack of Expression Increase Confidence - The region containing all three genes is predicted to be fully deleted in A549, K652, and MCF7, and only single copy in U87MG. Through comparison with transcriptome data we observe a lack of gene expression in A549 and K562 and only a slight expression of CDKN2B in MCF7. Interestingly, all three genes in U87MG are unexpressed regardless of a single gene copy prediction. Thus, only through an integrated analysis can the entire picture be accurately assessed.



Combined Copy Amplification and Increased Gene Expression - The gene KAT6B is hyper-amplified in PC3 (13 predicted copies) and moderately amplified in OVCAR3 (6 predicted copies). Additionally, the gene TRIM33 is hyper-amplified in MCF7 (12 predicted copies) and moderately amplified in PC3 and K652 (5 and 4 predicted copies). By integrating gene expression analysis into copy number analysis, we observe that the samples containing hyper-amplified copy numbers also have the highest gene expression levels. However, while moderate gene copy number amplification (OVCAR3, U87MG, and K562) does relate with increased expression, the trends are slightly more complex. Interpreting the intricacies of such a cancer sample requires the integration of both transcriptome expression and DNA copy number analysis.

Identifying Gene Fusions from Transcriptome Data

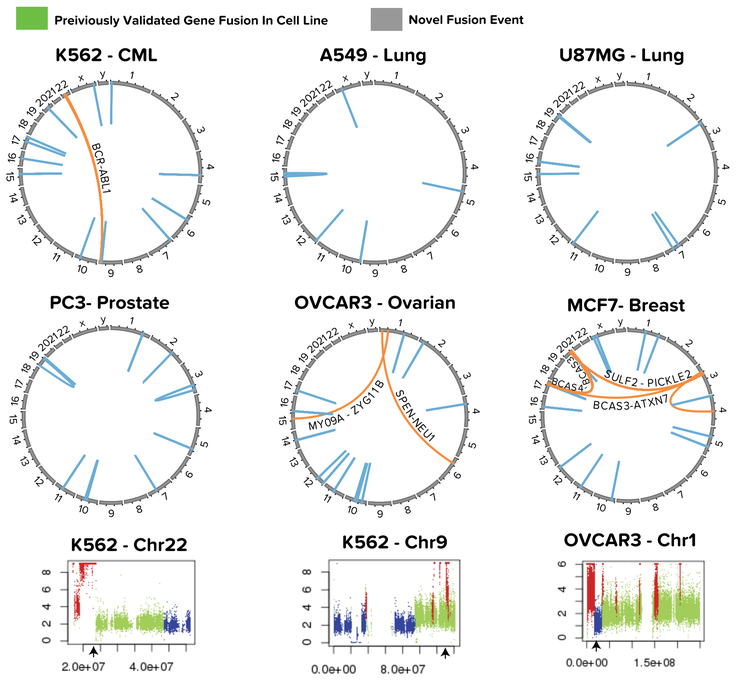
K562 - Chronic Myelogenous Leukemia

Gene 1	Gene 2	Fusion Annotation	Spanning Read Pairs	Spanning Reads	5' Parter Loc	3' Partner Loc
BCR	ABL1	known fusion, Cdb2, cgp, ticdb	29	14	22:23632600:+	9:133729451:+
BAG6	SLC44A4	known fusion	21	9	6:31619433:-	6:31833561:-
NUP214	XKR3	known fusion	8	12	9:134074402:+	22:17288973:-

MCF7- Breast Cancer

Gene 1	Gene 2	Fusion Annotation	Spanning Read Pairs	Spanning Reads	5' Parter Loc	3' Partner Loc
BCAS4	BCAS3	known fusion, Cdb2	218	24	20:494 11710:+	17:59445688:+
ARFGEF2	SULF2	known fusion	162	26	20:47538547:+	20:46365686:-
RPS6KB1	VMP1	100kbp, known fusion, Cdb2	87	18	17:57992064:+	17:57917129:+
GCN1L1	MSI1	short distance, known fusion	49	16	12:120628101:-	12:120785317:-
RSBN1	AP4B1-AS1	no protein product, antisense	47	17	1: 114354332:-	1:114442494:+
TANC2	CA4		39	13	17:6 1151375:+	17:58232675:+
C16orf45	ABCC1		26	9	16:15528616:+	16:16170183:+
SULF2	PRICKLE2	known fusion	24	9	20:46414087:-	3:64085601:-

Preiviously Validated Gene Fusion In Cell Line
Novel Fusion Event



Identifying Well-Documented Gene Fusions In Cancer Cell Lines - We apply our carefully selected and vetted gene fusion identification approach to identify gene fusions in several widely studied cancer cell lines. Gene fusions were annotated and filtered based upon previous observations in cancer samples, lack of occurrence in healthy samples, physical distance, and existence in protein-coding genes.

Summary

We assessed important genetic alterations and limits of detection in these samples through analysis of RNA-derived variant calls, gene expression levels, and gene fusion events. These calls were then filtered and annotated with our cancer gene database.

We found that transcriptome analysis improves upon assessment of the impact of CNVs. We performed gene expression analyses to accurately quantify differential regulation, something that can only be assumed when analyzing copy number changes (CNVs) from DNA alone. By combining expression and DNA-based CNV analyses we were able to generate more accurate CNV interpretations.

Additionally, we called variants directly from the RNA, and identified variants that are truly expressed in the tumor. We observed variants in important cancer driver genes which were selectively identified in either genomic or transcriptomic samples through allele specific expression, poorly covered regions, or lack of expression.

There are a growing number of therapeutically relevant gene fusions, and assessing these fusions is critically important for accurate tumor mutational analysis. We identified important gene fusion events at the expression level and cross-referenced these with exome findings, where present. In many cases, these fusion events would have been missed by exome or gene panel analyses alone.

Taken together, these joint analyses demonstrate how incorporation of both transcriptome and exome sequencing substantially bolsters accuracy in cancer analysis --important features would have been missed through individual analysis. By assessing transcriptome and exome together, we greatly increase our interpretive capabilities, leading to changes in both research results and clinical decisions.

Personalis®
Pioneering Genome-Guided Medicine