Unbiased linkage of SNV to expression in single cells for regulatory variant discovery influencing cancer drug resistance

I. Salas-González[1], T. Morozova[1], D. Arvapalli[1], S. Velivela[1], G. Harton[1], J. Zawistowski[1], J. Blackinton[1], V. Weigman[1], J. A. A. West[1]

[1]BioSkryb Genomics, Inc., Durham, NC

BioSkryb GENOMICS

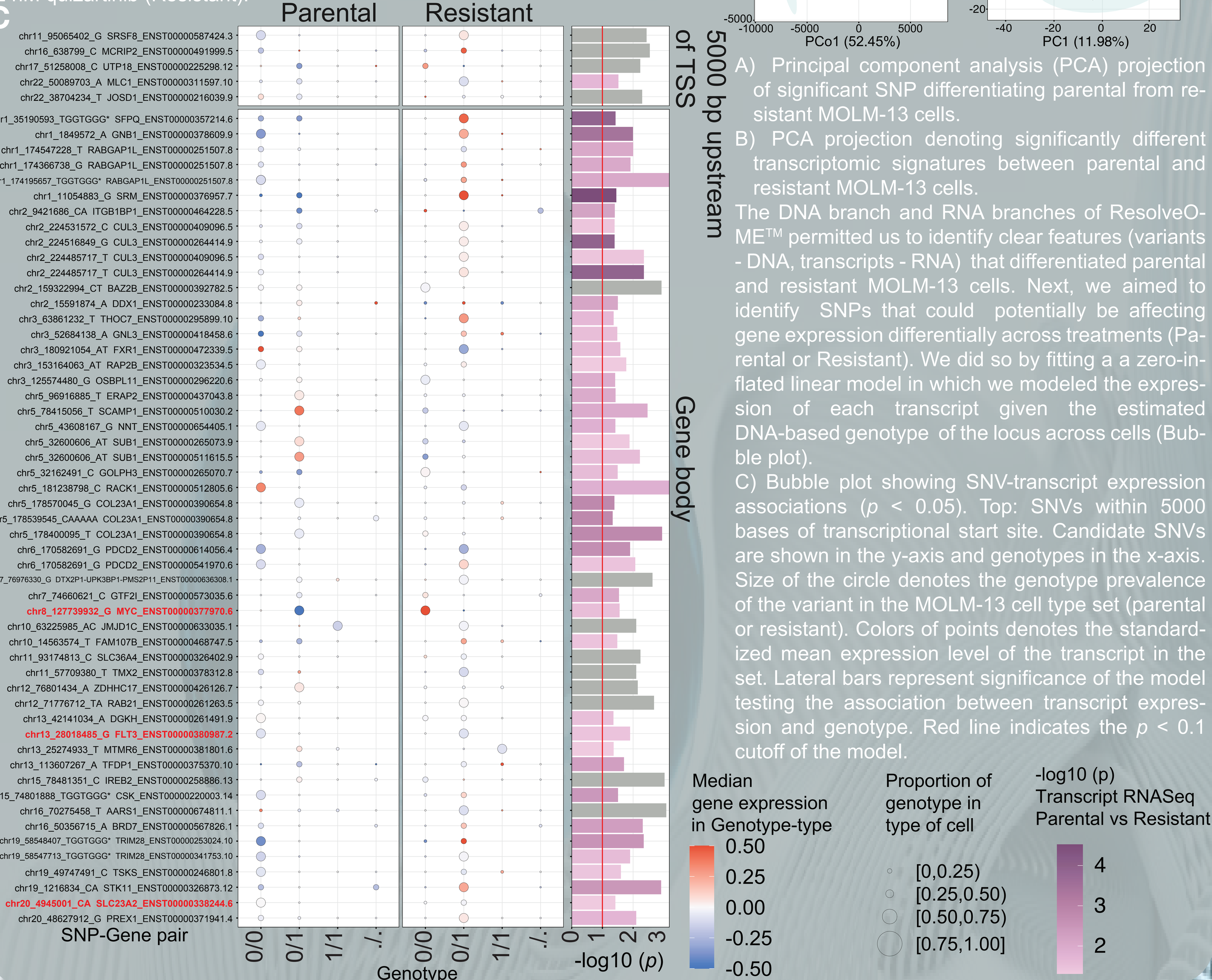
ASHG American Society of Human Genetics

Abstract

Single-cell RNAseq methodologies have revolutionized the ability to define cell type identity and transcriptional phenotypic diversity within a tumor sample, yet in most studies DNA-level variation contributing to driving this diversity of gene expression remains uncharacterized. We employed ResolveOME chemistry to bridge this gap within an individual cell, by taking inventory of genome-wide single nucleotide variation (SNV) and cross referencing it with full-transcript RNAseq data in an acute myeloid leukemia (AML) cell line model of drug resistance. The catalog of genome-wide SNV in conjunction with transcriptomic data in this model powered an association screening to identify regulatory SNVs with biased prevalence in single cells resistant to the FLT3 inhibitor quizartinib vs treatment-naïve single cells, which then were correlated with differential expression of genes proximal to the variant in the same cell. Initially, we focused on nucleotide variation within core promoter regions and within gene bodies. Fitting a zero-inflated linear model to a matrix of expression and genotype across all single cells revealed an intronic heterozygous variant in MYC in parental single cells, while, in contrast, the alternate allele was absent in quizartinib-resistant cells. Upregulation of MYC transcript was correlated with the absence of this variant in resistant single cells, suggesting that the single nucleotide change may have intronic enhancer activity as opposed to intronic regulation of splicing. Additionally, the screen uncovered a heterozygous single nucleotide change within 5kb 5' of the transcriptional start site of the mRNA binding factor PABPC4 as a candidate promoter variant not present in parental single cells yet harbored by 50% of the quizartinib resistant single cells displaying PABPC4 expression relative to parental cells. Differential expression analysis uncovered the enhancer factor CEBPA as upregulated in the resistant single cells, and the genomic component of ResolveOME identified two SNVs 20 kb upstream of the locus representing putative CEBPA enhancer variants influencing the differential gene expression. The co-identification of CEPBA and PABPC4 through this multi-omic approach suggest that quizartinib resistance in this model is mediated in part by global gene regulation through enhancer modulation and through mRNA stability/translational regulation, respectively. We are furthering enhancer variant detection in intergenic space by overlaying ChIP-seq, chromatin accessibility, and transcription factor binding site data to prioritize candidates contributing to resistance in this AML drug resistance model.

Drug Resistance: linking variants' impact on expression

To demonstrate the utility of concurrent genomic and transcriptomic information in single cells in the context of drug resistance, we created a model by exploiting the presence of an internal tandem duplication (ITD) mutation in MOLM-13 cells. Since the ITD mutation, found in ~20% of AML patients, hyperactivates FLT3 signaling and results in poor prognosis and relapse, we treated non-resistant (Parental), drug-sensitive cells with a continual dose of 2 nM quizartinib (Resistant).



A) Principal component analysis (PCA) projection of significant SNP differentiating parental from resistant MOLM-13 cells.

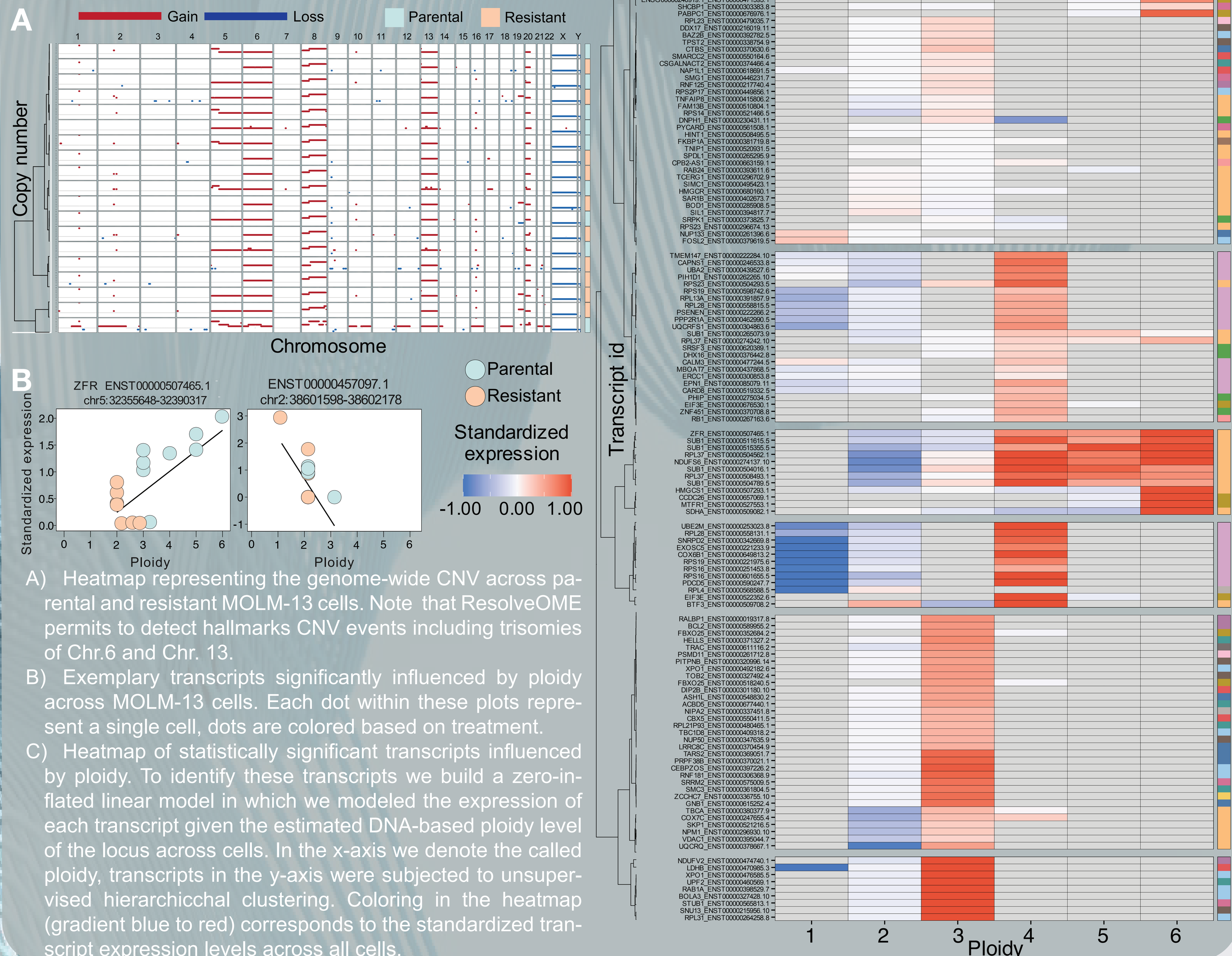
B) PCA projection denoting significantly different transcriptomic signatures between parental and resistant MOLM-13 cells.

The DNA branch and RNA branches of ResolveOME™ permitted us to identify clear feature (variants - DNA, transcripts - RNA) that differentiated parental and resistant MOLM-13 cells. Next, we aimed to identify SNPs that could potentially be affecting gene expression differentially across treatments (Parental or Resistant). We did so by fitting a zero-inflated linear model in which we modeled the expression of each transcript given the estimated DNA-based genotype of the locus across cells (Bubble plot).

C) Bubble plot showing SNV-transcript expression associations ($p < 0.05$). Top: SNVs within 5000 bases of transcriptional start site. Candidate SNVs are shown in the y-axis and genotypes in the x-axis. Size of the circle denotes the genotype prevalence of the variant in the MOLM-13 cell type set (parental or resistant). Colors of points denotes the standardized mean expression level of the transcript in the set. Lateral bars represent significance of the model testing the association between transcript expression and genotype. Red line indicates the $p < 0.1$ cutoff of the model.

Drug Resistance: expression impacted by gene copy

Given the concurrent characterization of DNA and RNA from the same cell, we had the power to associate genome-wide the influence of copy number (CNV) events with respect to gene expression. To do so, we identified clusters of genes that showed clear positive and negative association with ploidy across the dataset. Further, we also identified genes with more cryptic patterns of influence suggesting alternative mechanisms influencing gene expression such as epigenetics.



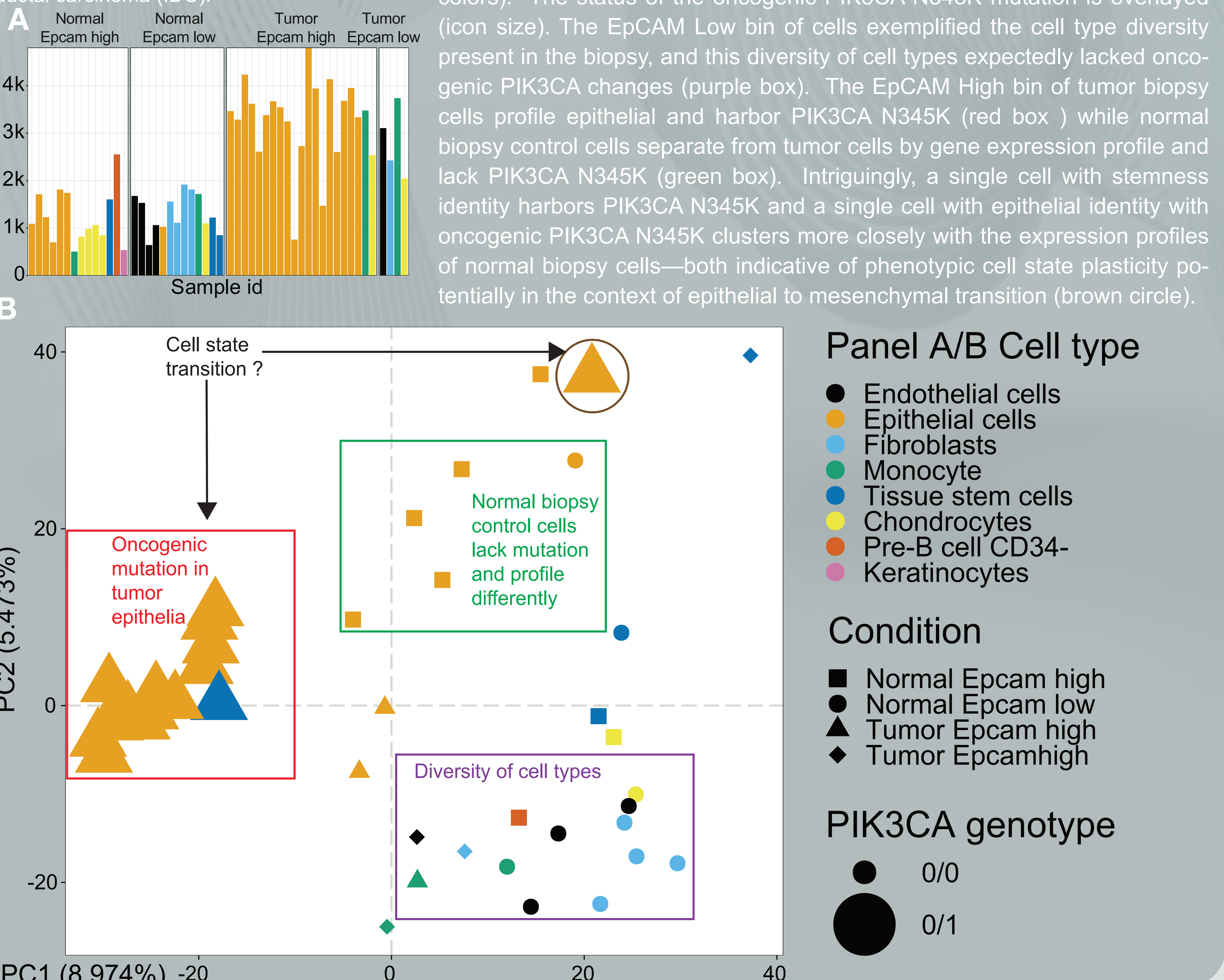
A) Heatmap representing the genome-wide CNV across parental and resistant MOLM-13 cells. Note that ResolveOME permits to detect hallmarks CNV events including trisomies of Chr.6 and Chr. 13.

B) Exemplary transcripts significantly influenced by ploidy across MOLM-13 cells. Each dot within these plots represent a single cell, dots are colored based on treatment.

C) Heatmap of statistically significant transcripts influenced by ploidy. To identify these transcripts we build a zero-inflated linear model in which we modeled the expression of each transcript given the estimated DNA-based ploidy level of the locus across cells. In the x-axis we denote the called ploidy, transcripts in the y-axis were subjected to unsupervised hierarchicchal clustering. Coloring in the heatmap (gradient blue to red) corresponds to the standardized transcript expression levels across all cells.

Tumor Heterogeneity: characterizing cell states

We demonstrate analogous multi-omic utility in elucidating single-cell oncogenic mechanisms in primary human cancer. To do so, we characterized genomic and transcriptomic contributions to the transition of premalignant ductal carcinoma in situ (DCIS) to invasive ductal carcinoma (IDC).

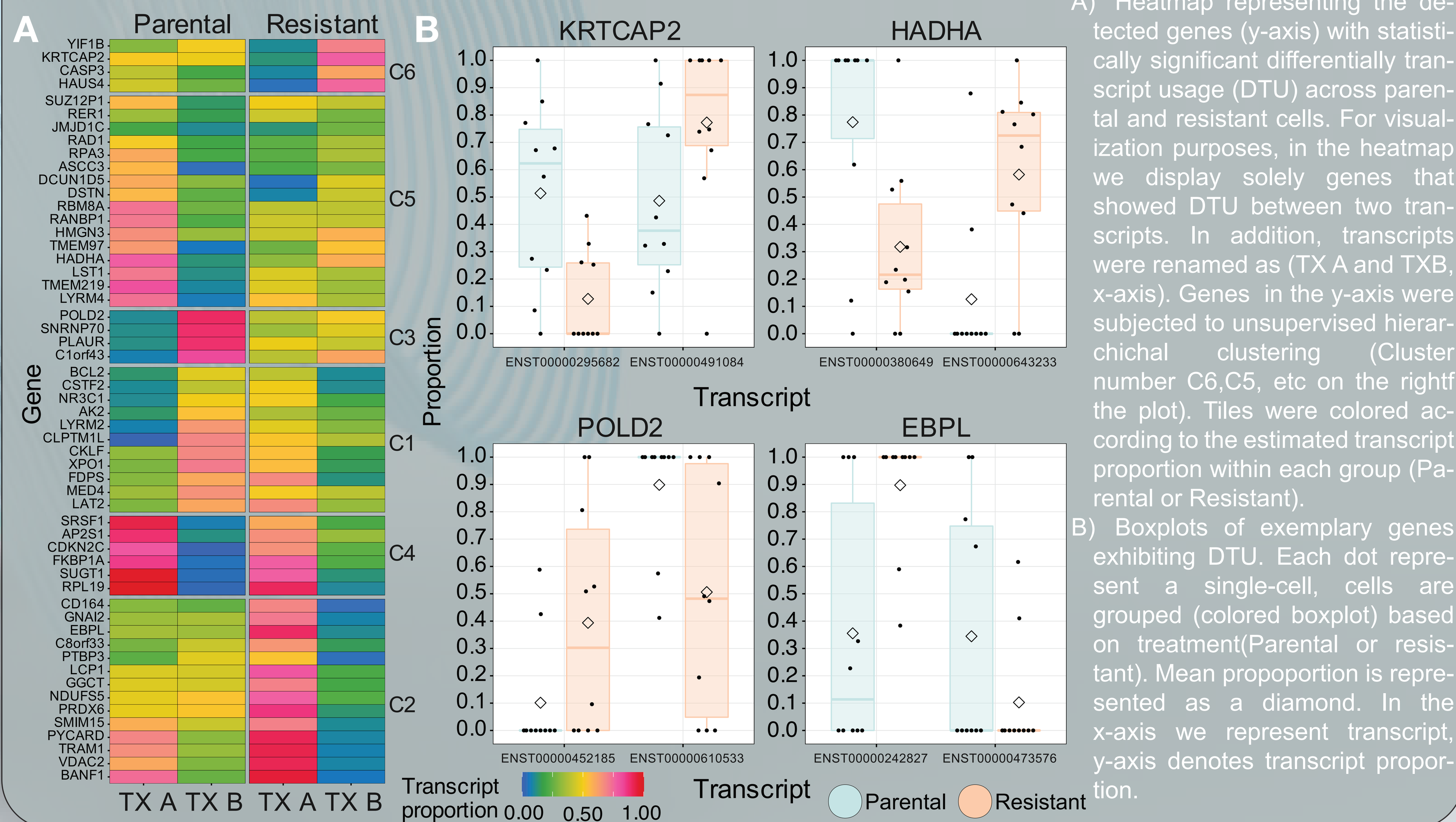


A) Bargraphs denoting the number of detected genes by cell. Colors within bar represent cell type identity as determined by gene expression profile.

B) Principal component analysis of DCIS/IDC single cell gene expression overlayed with PIK3CA mutation status. Each point represents a single cell. The Human Cell Atlas database was utilized for cell type identification (icon colors). The status of the oncogenic PIK3CA N345K mutation is overlayed (icon size). The EpCAM Low bin of cells exemplified the cell type diversity present in the biopsy, and this diversity of cell types expectedly lacked oncogenic PIK3CA changes (purple box). The EpCAM High bin of tumor biopsy cells profile epithelial and harbor PIK3CA N345K (red box) while normal biopsy control cells separate from tumor cells by gene expression profile and lack PIK3CA N345K (green box). Intriguingly, a single cell with stemness identity harbors PIK3CA N345K and a single cell with epithelial identity with oncogenic PIK3CA N345K clusters more closely with the expression profiles of normal biopsy cells—both indicative of phenotypic cell state plasticity potentially in the context of epithelial to mesenchymal transition (brown circle).

Drug Resistance: Differential isoform presentation

The ability of ResolveOME™ to characterize the full-length transcriptome of a single cell opens the possibility to identify differential transcript usage (DTU) events that could be driving physiological phenotypes. To do so, we performed DTU characterization in the MOLM-13 model described previously. Briefly, a DTU event across two treatment groups (e.g. Parental and Resistant MOLM-13 cells) tests for proportional differences in the expressed transcript composition of a gene, thus comparing how much each transcript contributes to the gene's total expression between conditions.

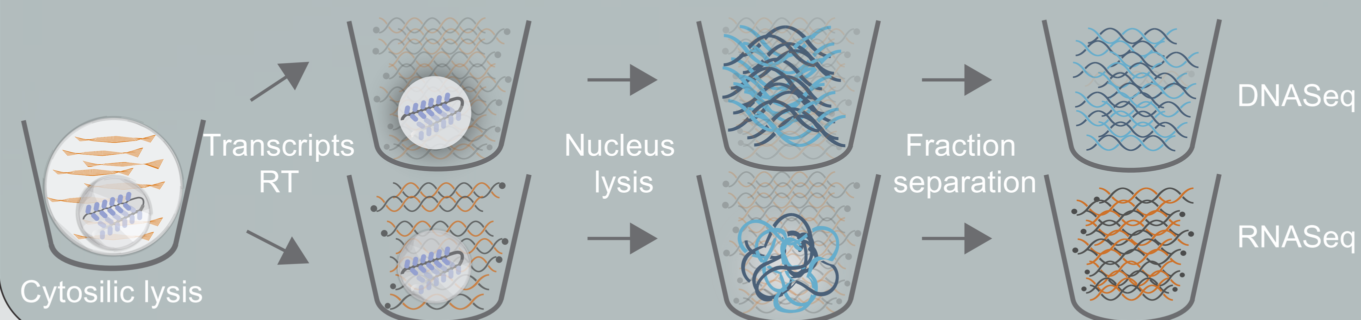


A) Heatmap representing the detected genes (y-axis) with statistically significant differentially transcript usage (DTU) across parental and resistant cells. For visualization purposes, in the heatmap we display solely genes that showed DTU between two transcripts. In addition, transcripts were renamed as (TX A and TXB, x-axis). Genes in the y-axis were subjected to unsupervised hierarchichal clustering (Cluster number C6,C5, etc on the rightf the plot). Tiles were colored according to the estimated transcript proportion within each group (Parental or Resistant).

B) Boxplots of exemplary genes exhibiting DTU. Each dot represent a single-cell, cells are grouped (colored boxplot) based on treatment(Parental or resistant). Mean propoportion is represented as a diamond. In the x-axis we represent transcript, y-axis denotes transcript proportion.

ResolveOME workflow

ResolveOME permits to simultaneously characterize the genomic and transcriptomic data from the same individual cell vastly increases the complexity of putative mechanisms of drug resistance and oncogenesis



ResolveOME™