Linking Structural Variation to Tumor Phenotype in KPC Mouse Pancreatic Cancer via long read Nanopore Sequencing

Background

Structural variants (SVs), including deletions, duplications, inversions, and translocations are genomic alterations that can drive tumor behavior by disrupting gene coding regions, enhancers, or chromatin structure. SVs are especially relevant in cancers like pancreatic ductal adenocarcinoma (PDAC), where they can silence tumor suppressors (*Cdkn2a*, *Smad4*), activate oncogenes (*Myc*), or create fusion genes with novel functions (doi:10.1371/journal.pgen.1010514, doi:10.1038/s41587-023-02024-y).
Long-read sequencing from Oxford Nanopore enables sensitive detection of SVs that may be missed by short-read platforms, providing a clearer link between genomics and tumor phenotype (doi:10.1038/s41587-023-02024-y).

The KPC mouse model (LSL-Kras^G12D/+; LSL-Trp53^R172H/+; Pdx-1-Cre) recapitulates many aspects of human PDAC, including invasive growth, metastasis, and stromal remodeling (doi:10.1002/cpph.2). Investigating tumor-specific genomic alterations in this model offers a controlled system to understand how specific SVs may drive cancer phenotypes.

In this study, **"phenotype"** refers to the observable traits and behaviors of the pancreatic tumors that arise as a consequence of the underlying structural-variant (SV) landscape. More specifically:vthe set of gene-level and the downstream pathway dysregulations you detect by VEP-based annotation and enrichment analysis

Research Question

Which structural variants in KPC pancreatic tumors disrupt genes and regulatory regions and how do these alterations affect known pathways?

Proof of concept study

Xu, L., Wang, X., Lu, X., Liang, F., Liu, Z., Zhang, H., Li, X., Tian, S., Wang, L. and Wang, Z. (2023) 'Long-read sequencing identifies novel structural variations in colorectal cancer', *PLOS Genetics*, 19(2), e1010514. doi: 10.1371/journal.pgen.1010514.
The authors performed Oxford Nanopore long-read sequencing on 21 colorectal tumor–normal sample pairs to profile somatic structural variants (SVs).

Methods

Data Acquisition

- **Raw reads**: Oxford Nanopore FASTQ files (sample accessions ERR4351540 and ERR4351539) were downloaded via SRA Toolkit (prefetch → fasterq-dump).
- **Reference genome**: GRCm39 FASTA was obtained from NCBI (GCF_000001635.27).

Quality Control

Per-sample long-read QC was performed using NanoPlot and NanoStat. Read length distributions, quality score histograms and basic summary metrics were generated and inspected to ensure library integrity.

Read Alignment

Reads were aligned to the GRCm39 reference using minimap2 (map-ont preset). The resulting SAM

streams were converted to BAM, filtered to retain mapped reads, coordinate-sorted, and indexed via samtools.

Structural Variant Calling

- **Per-sample "mosaic" SVs**: Sniffles2 was run in mosaic mode to detect low-frequency (~10–20 % VAF) insertions, deletions, inversions and duplications. Outputs were individual per-sample VCFs.
- **Cohort SVs**: All per-sample .snf intermediate files were merged by Sniffles2 to produce a multi-sample VCF for population-level SV discovery.

SV Visualization
Sniffles2's plotting module was used to generate per-sample and cohort SV summary plots, facilitating rapid inspection of SV size distributions and genomic locations.

Variant Annotation
Both per-sample and cohort VCFs were annotated with Ensembl VEP (v114) in offline, cache-based mode.

Post-annotation Filtering & Impact Scoring
VEP outputs were imported into R and consolidated across samples. Each SV consequence was assigned an IMPACT category (HIGH, MODERATE, LOW, MODIFIER) based on a standard VEP hierarchy of SO terms. Distinct SV–gene pairs were then summarized, and a high/moderate-impact shortlist was extracted for downstream interpretation.

Pathway Enrichment Analysis
High/moderate-impact gene lists were converted from MGI symbols to Entrez IDs and subjected to functional enrichment:

- GO Biological Process (adjusted $q < 0.05$),
- KEGG pathways (p adjusted < 0.05),
- Reactome ($p < 0.05$, human-readable IDs).

Top (10) enriched terms tabulated for interpretation.

Results

Filtering to SO terms of HIGH or MODERATE impact (e.g. stop_gained, frameshift, missense, inframe indels) yielded 8 690 unique variant-transcript pairs affecting 2 412 distinct Ensembl genes.

Enrichment analysis on the filtered gene set revealed:

- GO Biological Processes: striking over-representation of purine and ribonucleotide biosynthesis/metabolism (e.g. "purine nucleoside triphosphate biosynthetic process", "ATP biosynthetic process"), indicating disruption of nucleotide supply pathways critical for rapid cell division.
- KEGG pathways: significant enrichment in oxidative phosphorylation (mmu00190), phosphatidylinositol signaling (mmu04070), and chemical carcinogenesis – reactive oxygen species (mmu05208), alongside multiple neurodegenerative disease pathways (mmu05012, mmu05016, mmu05022).
- Reactome: core hits in aerobic respiration and respiratory electron transport (R-MMU-1428517) and Complex I biogenesis (R-MMU-6799198).

Our structural variant (SV) calls in the KPC model preferentially disrupt genes governing energy metabolism and nucleotide biosynthesis, two pillars of cancer cell proliferation. Disruption of phosphatidylinositol signaling genes aligns with activation of the PI3K–AKT–mTOR axis, a known driver of PDAC growth and chemoresistance. Enrichment in oxidative stress/ROS pathways underscores how genomic instability and metabolic reprogramming contribute to redox imbalance in tumors. Although neurodegenerative disease pathways appeared, these likely reflect shared stress-response networks rather than a direct link to pancreatic cancer. Overall, the SV-driven perturbations map onto core cancer-hallmark processes,uncontrolled proliferation, metabolic flexibility, and oxidative stress adaptation,and thus constitute a coherent molecular phenotype in KPC tumors.