

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) alterations. More specifically: the set of gene-level and the downstream pathway dysregulations detected 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. 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.

Sniffles is known to be reliable, fast and least memory consuming (<https://doi.org/10.1186/s13059-024-03324-5>; improved over sniffles (<https://doi.org/10.3390/genes15070925>)

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. (studies that used the same tool: <https://doi.org/10.1007/s00439-024-02670-5>, <https://doi.org/10.1038/s41586-020-2371-0>, <https://doi.org/10.1016/j.jbi.2019.103286>, 10.1109/ICTer51097.2020.9325461)

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).

Results

While Reactome terms only showed three significant entries, KEGG and GO showed the disruptions in pathways involved in overlapping processes. Key processes included:

- **Mitochondrial respiration / ATP synthesis**
Disruption in oxidative phosphorylation is expected, due to the The Warburg effect (<https://doi.org/10.1186/s12943-024-02119-3>)
- **PI3K/Akt (phosphatidylinositol) signaling**
Key pathway for cell proliferation and growth (<https://doi.org/10.1038/nrc839>)
- **Nervous system, Neurodegenerative diseases**
Known to be affected in cancers (<https://doi.org/10.2174/0115672050389561250429112631>)
- **Oxidative-stress**
Known to play role in cancer development

(<https://doi.org/10.2174/1381612825666190215121712>, <https://doi.org/10.1007/s10552-019-01242-7>)

Among GO terms, purine related mutations were enriched, possibly suggesting chromatin restructuring (<https://doi.org/10.3390/ijms252011050>)