

Piano: RNA-Seq Fusion Calling Pipeline

Introduction

The Piano pipeline integrates two key components for analysis. The primary component is the Forte pipeline (<https://github.com/mskcc/forte>), specifically version 0b3d9275. Following the Forte analysis, additional post-processing steps are performed by the Piano pipeline (<https://github.com/soccin/piano>).

Output Structure

The pipeline generates two main output directories: `forte` and `post`. The structure is organized as follows:

```
{outdir}
├── forte/      # Primary analysis results
└── post/      # Post-processing results and reports
```

Forte Directory

The `forte` directory contains all standard Forte pipeline outputs, excluding BAM files. For a comprehensive description of these files, please refer to the official Forte documentation at <https://github.com/mskcc/forte/blob/main/docs/output.md>

Post Directory

The post-processing directory, `post`, contains the consolidated fusion analysis report in an Excel workbook named `{ProjectNumber}_FusionTableV4.xlsx`. This workbook is organized into two sheets that present the fusion analysis results at different confidence levels.

The first sheet, HC.Events (High Confidence Events), contains fusion events detected by at least two of the three Forte callers. These represent the most reliable fusion predictions and are recommended for primary analysis. The second sheet, ALLEvents, provides a comprehensive list of all detected fusion events, including predictions from any single caller. This complete list is particularly useful for exploratory analysis or when investigating specific regions of interest.