Connecting computational steps for NGS, and beyond.

The analysis of Next-Generation Sequencing (NGS) data consists in applying 3rd party tools in sequences of inter-dependent steps.

While this is often referred to as a "pipeline", we instead see the process as much less linear than a pipe. We propose here a framework to implicitly build and manipulate a directed graph of steps in an interactive and incremental way that we believe to be natural and intuitive for bioinformaticians and computational biologists.

Our "railroadtracks" framework is designed to be a modular ensemble of loosely coupled components, able to integrate arbitrary sets of 3rd party tools such as aligners, read counters, and differential expression methods in the case of RNA-Seq into models. The persistence of the graph connecting steps is also enabling reproducibility and the computation of variants of a process while keeping their computational cost relatively low.