RiboPipe: A Ribosome Profiling and RNA Sequencing Pipeline Optimized for Easy Assembly and Data Analysis

Jordan A. Berg, Jonathan R. Belyeu, Jared P. Rutter

ABSTRACT

With the advent of high-throughput sequencing, ribosome profiling is coming of age. Previously, assembly and analysis of this data was hindered by the need to manually assemble raw data for several samples with options from a variety of assemblers, followed by data structure formatting and quality control, and general and targeted analysis. As ribosome profiling is still a maturing technique, a standard set of protocols both at the bench and in silico are yet to be widely adopted. In an effort to standardize the in silico portion of ribosome profiling and improve accessibility of the computational aspects of this sequencing technology for the general public, we developed “RiboPipe,” a flexible, automated pipeline optimized for rapid assembly of raw single-end, short (<100nt) sequence data and quality analysis of libraries.

BACKGROUND

Ribosome profiling is an emerging and developing technology that takes advantage of the recent developments in high-throughput sequencing to measure protein translation dynamics in cells in different genetic backgrounds, during time courses, etc. These metrics are obtained by sequencing both the ribosome footprint, or the RNA protected by the ribosome and accepted as an output of what the cell was translating at the time of harvest, and total mRNA. Briefly, to perform ribosome profiling, cultures are grown in the appropriate conditions, harvested, and flash frozen, then lysed for downstream RNA isolation. The next step is to isolate the ribosome footprints by performing an RNase-mediated digestion of RNA and preserving the mRNA footprint protected by the ribosome as the ribosome complex is somewhat impervious to RNase digestion. Footprint RNAs are then separated from ribosomal RNAs by a variety of methods (see references for examples), rRNA depleted via targeted probes or other rRNA-depletion kits (RiboZero, etc) and integrated into an Illumina sequencing library. With a total RNA aliquot from each sample, RNA is similarly rRNA-depleted by a kit-based method, fragmented to produce a similarly-sized library to the ribosome footprints, and integrated into an Illumina sequencing library.