Determining Whether a Gene Is Expressed

Since ABC doesn't accept RNA-seq data, the way it determines whether a gene is expressed is by looking at the activity (RPM) in the gene. We quantile the gene activity and consider the top X% (70% at the time of writing) as being expressed. This is an arbitrary number that we came up with in order to capture ~14.5K genes as being expressed.

There have been discussions about whether we can use a better threshold instead of using quantiles. Similar to how TPM thresholds are used for determining expression in RNA-seq data. A potential plan here is to do correlation analysis between TPM and activity RPM (or gene quantile) to find a better threshold. We have to consider that ATAC RPM and DHS RPM may look different, which may require different thresholds.

Processing HiC Data

For processing .hic files, this can get quite complicated. The steps in the code are outlined here: https://docs.google.com/document/d/1t3M7oxDMomFatwy2bO_H5pBdbVubGBdQygbRCc9C5eE/edit#heading=h.ox6gen3lgb7v