**Project Proposal**

Team RNA-seq

We will be working on a pseudoaligner for RNA-seq reads. This will take an input set of RNA-seq reads and align them to a group of segments, which represent the transcriptome.

Basically, RNA is the expression of DNA (in biology DNA gets written to RNA which helps create proteins). The transcriptome is the set of all possible RNA-seq molecules. The letters of RNA are AUGG and they are long strings, and when a sequencing machine reads them, they chop them up into small pieces. The goal then is to compare the string of each small piece to what the original RNA is to see where each came from, and then we can count how many reads (small pieces) each RNA has, which lets us measure gene expression. Basically, we use a hashmap and do some hashing of the reads to their original location. However, what makes it hard is that the reads can have errors in them, so the algorithm has to deal with this. We take the input files and then we output what RNA each read came from.

Steffen Cornwell has done research in RNA-seq previously, so this is not too difficult of a project for the group. Existing algorithms and programs exist, but ours will be different because we will be aligning to a group of segments rather than the transcriptome. No projects exist that we know of that are in Rust.

Resources:

<http://tinyheero.github.io/2015/09/02/pseudoalignments-kallisto.html>

<http://robpatro.com/blog/?p=260>