

# Module 2: Lab Practical

Michael Brudno

Estimated completion time: 1-2 hours

# Objectives

- Get a practical feel for using letter-space and color-space data
- Do some small assembly jobs
- See how various genomes are very different

# Task 1: Get the Tools

- Download Velvet  
<http://www.ebi.ac.uk/~zerbino/velvet/>
- Download SHRiMP:  
<http://compbio.cs.toronto.edu/shrimp>
- Learn the wget command in Linux:  
wget <http://something>
- Don't be afraid of the source code:  
compiling is as simple as “make”

# Playing with Velvet

- Read the README file (or at least scan it)
- The real Solexa data from a human BAC is in `~brudno/human_bac/`
- Run velvet:  

```
../velvet_0.6.04/velveth . 19 split_reads/reads.fa  
../velvet_0.6.04/velvetg .
```

Try varying the k-mer parameter. what's the biggest contig? What is the N50 score? What is it in your case? Which chromosome is this BAC from?

# SHRiMP

- Read the README file (or at least scan it)
- Get the corresponding chromosome (from `~brudno/chr.fa`)
- Use SHRiMP to map the reads to the reference bactig. How long did it take? Try again the reference chromosome (use `-B` flag)
- Kill the job (Ctrl-C)
- Try only with a few reads (use the `head` command)

# SHRiMP

- Use the probcalc program to get out only reads with top hits (with normalized odds>0.5)
- What does this command do:

```
sort -nk 4 tophits | awk '{if ($4 > s) print s,$4; s = $5}'
```

How is this number related to the number of contigs from velvet?

- Look at the alignment file. Find the alignment and try to understand it. Run the line through prettyprint to get human-readable alignments

# Colospace

- Going colospace with SHRiMP is just replacing `-ls` with `-cs` :)
- Go to the color-space directory, map the reads to the reference genome
  - Wait! Don't decompress the genome
  - Compress the reads, if you want
- Run `probcalc`. What are the rates produced? Why do we have separate error rates and SNP rates? What happened in letter space?
- Go into the README file and use the SHRiMP parameters from there. Do the rates change?