

Appendix II: Software Links

Google Cloud SDK:

- Site: <https://cloud.google.com/sdk/>
- Documentation: <https://cloud.google.com/docs/>

QC:

Quake:

- Paper: <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2010-11-11-r116>
- Code Source: <http://www.cbcb.umd.edu/software/quake/>

Reference Genome Alignment:

CloudBurst:

- Paper: <http://bioinformatics.oxfordjournals.org/content/25/11/1363>
- Code Source: <https://sourceforge.net/projects/cloudburst-bio/>

Brisera:

- Code Source: <https://github.com/bbengfort/brisera>

BlastReduce:

- Paper: <http://www.cs.umd.edu/grad/scholarlypapers/papers/MichaelSchatz.pdf>
- Code Source: <http://www.cbcb.umd.edu/software/blastreduce/> [no longer posted]

Seal:

- Code Source: <https://sourceforge.net/projects/biodoop-seal/?source=recommended>
- Program Site: <http://biodoop-seal.sourceforge.net/>

De Novo Genome Assembly:

Contrail

- Description:
<http://www.homolog.us/blogs/blog/2011/09/08/contrail-a-de-bruijn-genome-assembler-that-uses-hadoop/>
- Kmer script description:
<http://www.homolog.us/blogs/blog/2011/08/31/using-hadoop-for-transcriptomics-an-example-to-get-started/>
- Code Source: <http://contrail-bio.svn.sourceforge.net/>

CloudBrush:

- Code Source: <https://github.com/ice91/CloudBrush>
- Commentary:
<http://stackoverflow.com/questions/31375801/error-java-lang-illegalargumentexception-comparison-method-violates-its-genera>)

Additional Analysis:

MrMC-MinH:

- Code Source: <http://cs.gmu.edu/~mlbio/MrMC-MinH/>

SeqPig:

- Program Description and Code: <http://seqpig.sourceforge.net>

ADAM:

- Project Site: <http://bdgenomics.org/>
- Tech Report: <http://www.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-207.html>
- Code Source and Manual: <https://github.com/bigdatagenomics/adam>

Avocado:

- Project Site: <http://bdgenomics.org/projects/avocado/>
- Code Source: <https://github.com/bigdatagenomics/avocado>

Data links:

- Staph reference file obtained from NCBI
(http://www.ncbi.nlm.nih.gov/assembly/GCA_000284535.1)
- FastQ files for methicillin-resistant *Staphylococcus aureus* (MRSA) reads obtained from Sanger Institute
(<http://www.sanger.ac.uk/resources/downloads/bacteria/staphylococcus-aureus.html>).
- Human Microbiome Data: <http://hmpdacc.org/HMASM/>.

Big Data platforms:

Hadoop/MapReduce:

<http://hadoop.apache.org/>

<https://www-01.ibm.com/software/data/infosphere/hadoop/mapreduce/>

Pig Latin:

<https://pig.apache.org/docs/r0.9.1/basic.html>

Apache Spark: <http://spark.apache.org/>