

Screw: tools and hardware for building reproducible single-cell epigenomics workflows

Kieran O'Neill, Chelsey Fang, Benjamin Decato, Alexander
Goncarencu, Azhar Khandekar, Ben Busby, Aly Karsan

Reproducible Research

Reproducible research requires having the exact **data**, **code** and **software** used to perform an analysis.

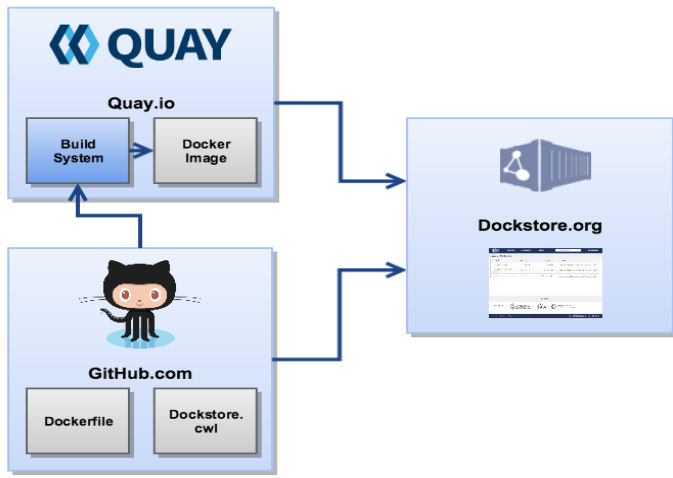
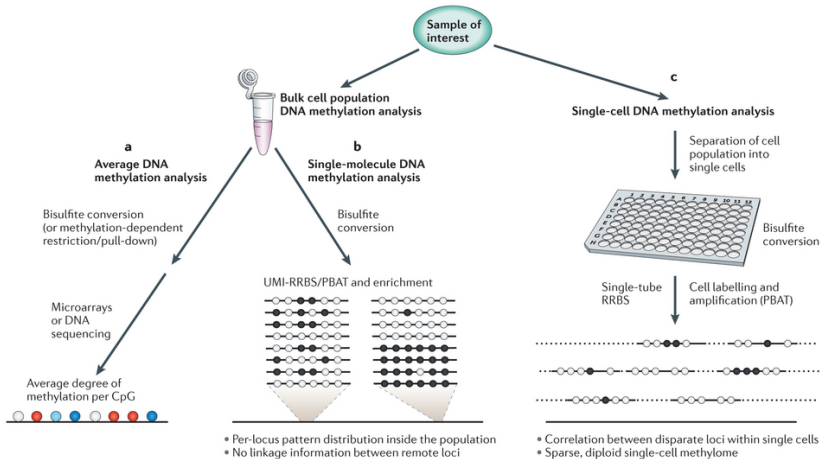


Image from dockstore.org, used under Apache-2.0

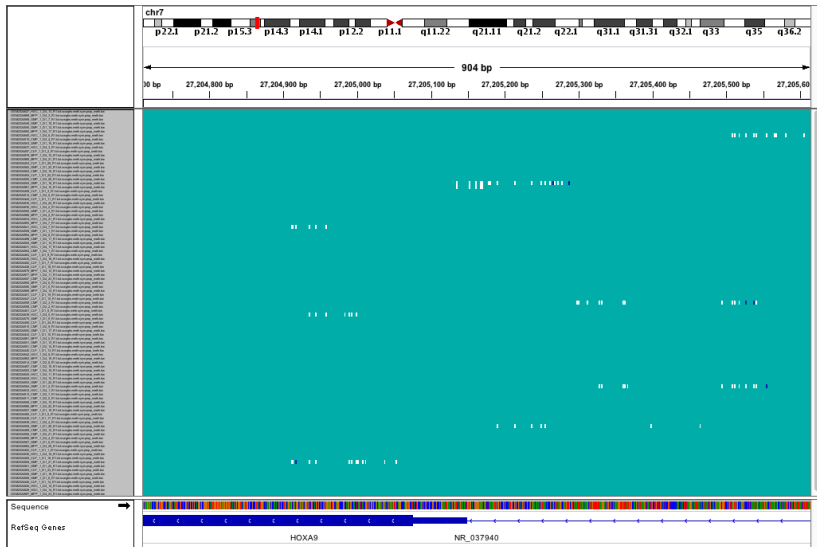
Single-cell Whole Genome Bisulfite Seq



Nature Reviews | Genetics

Schwartzman et al (2015) Nature Reviews in Genetics (used by permission)

Single-cell Whole Genome Bisulfite Seq Data



Preprocessing: Example problem – lack of standards!

Farlik 2013

chr19	89480	-	1	0	CG	CGT
chr19	93048	+	0	1	CG	CGT
chr19	109711	-	1	0	CG	CGG
chr19	109828	-	1	0	CG	CGT

Farlik 2016

chr10	48037	1	1
chr10	48132	0	1
chr10	48140	0	1
chr10	48143	0	1

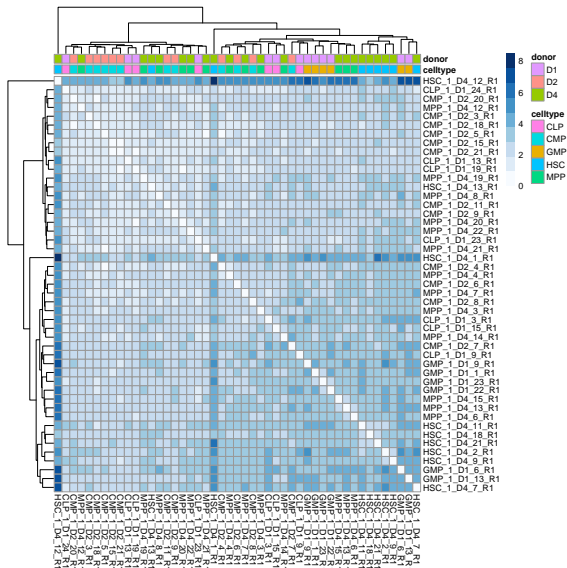
NovoMethyl

chr1	3113716	3113718	100	1	1	1
chr1	3327525	3327527	0	0	1	1
chr1	3327549	3327551	100	1	1	1
chr1	3642497	3642499	100	1	1	1

Screw

chr6	149393	*	CpG	0.5	2
chr6	149422	*	CpG	0.5	2
chr6	190375	*	CpG	0.5	2
chr6	190632	*	CpG	0.5	2

Output: Distance matrix and clustering



Status and Plans

- ▶ So far we just have one example data set, with some basic preprocessing and heatmap generation.
- ▶ But about half a dozen SC-WGBS (or SC-RRBS) public data sets now exist.
- ▶ We intend to use Screw to perform a meta-analysis of these.
- ▶ More vis, more tools, differential methylation calling.

Acknowledgements

NCBI Genomics Hackathon: Benjamin Decato, Alexander Goncarencu, Azhar Khandekhar, Ben Busby (hackathon organiser)

Summer intern: Chelsey Fang



Code, examples, slides: <https://github.com/Epigenomics-Screw>
Poster: A-153 (Sunday 6pm)