liftr & sbgr kickstart

Nan Xiao @road2stat

BioC 2015

Two R/Bioconductor Packages

- liftr: Dockerize R Markdown Documents
- sbgr: R Client for SBG and CGC Platform API

liftr

liftr.me

Idea

A Framework and Tool for
Dockerizing R Markdown Documents,
with Rabix Support.

Alternative Approach

R Markdown Documents
as Rabix Input

Rendered HTML/
PDF/Docx Reports

Knitr as a report generation tool embedded in Rabix.

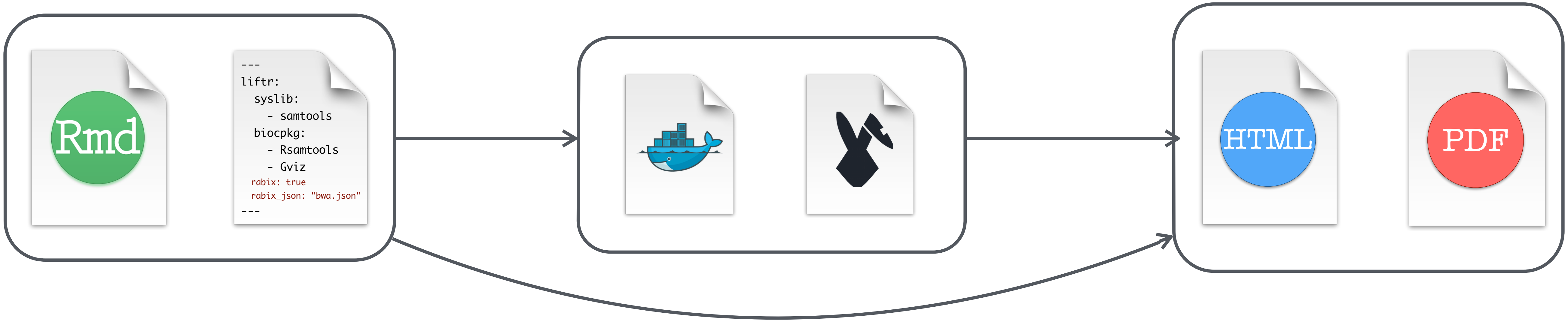
```
rabix --input foo.vcf foo.Rmd --output foo.html
```

Alternative Approach

- Pros
  - Default report binding with specific tools
- Cons
  - Hard to modify the report template
  - Hard to write report for particular workflows

The Road We Took

`lift("foo.Rmd")` → `drender("foo.Rmd")` → Read and share!



```
---
liftr:
  syslib:
    - samtools
  biocpkg:
    - Rsamtools
    - Gviz
  rabix: true
  rabix_json: "bwa.json"
---
```

Rmd Documents
with `liftr`
options in metadata

Generated Dockerfile
(Rabixfile)

Rendered HTML/
PDF/Docx Reports

Separate containers for Rabix and dockerized Rmd.

The Road We Took

- Pros
  - Run with/without Rabix
  - Easy to share and modify reports
- Cons?
  - Separate containers needed for rendering reports and running Rabix tools

The Road We Took

- Make bioinformatics tools/workflows with Rabix.
- Make statistical data analysis reports with R Markdown.
- Make the best use of each tool.

Demo

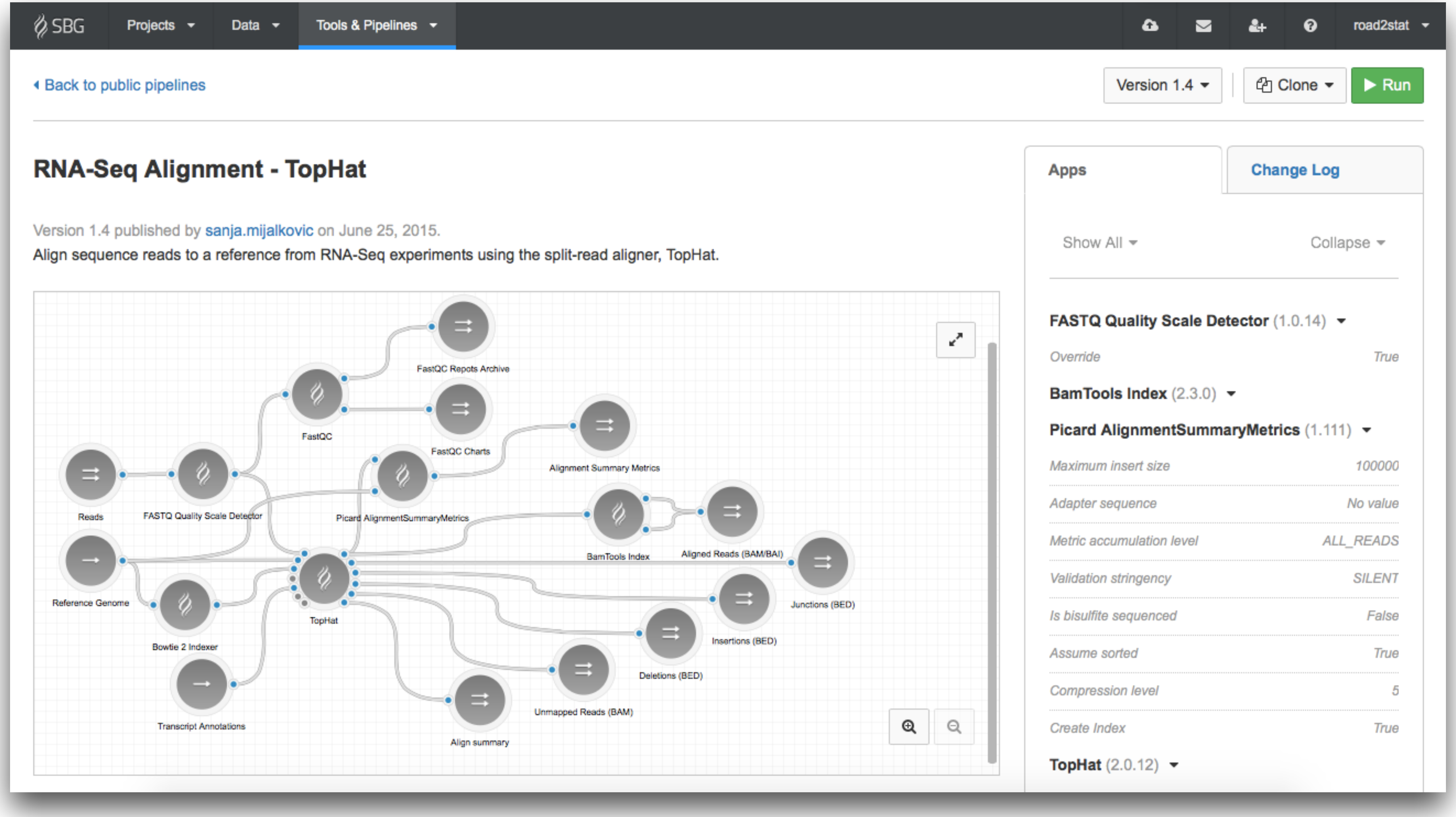
```
vignette("liftr-intro")
```

sbgr

github.com/road2stat/sbgr

R API Client

- SBG IGOR platform
- NCI Cancer Genomics Cloud (CGC) platform



RNA-Seq Alignment - TopHat

Version 1.4 published by sanja.mijalkovic on June 25, 2015.
Align sequence reads to a reference from RNA-Seq experiments using the split-read aligner, TopHat.

FASTQ Quality Scale Detector (1.0.14)

Override	*True*

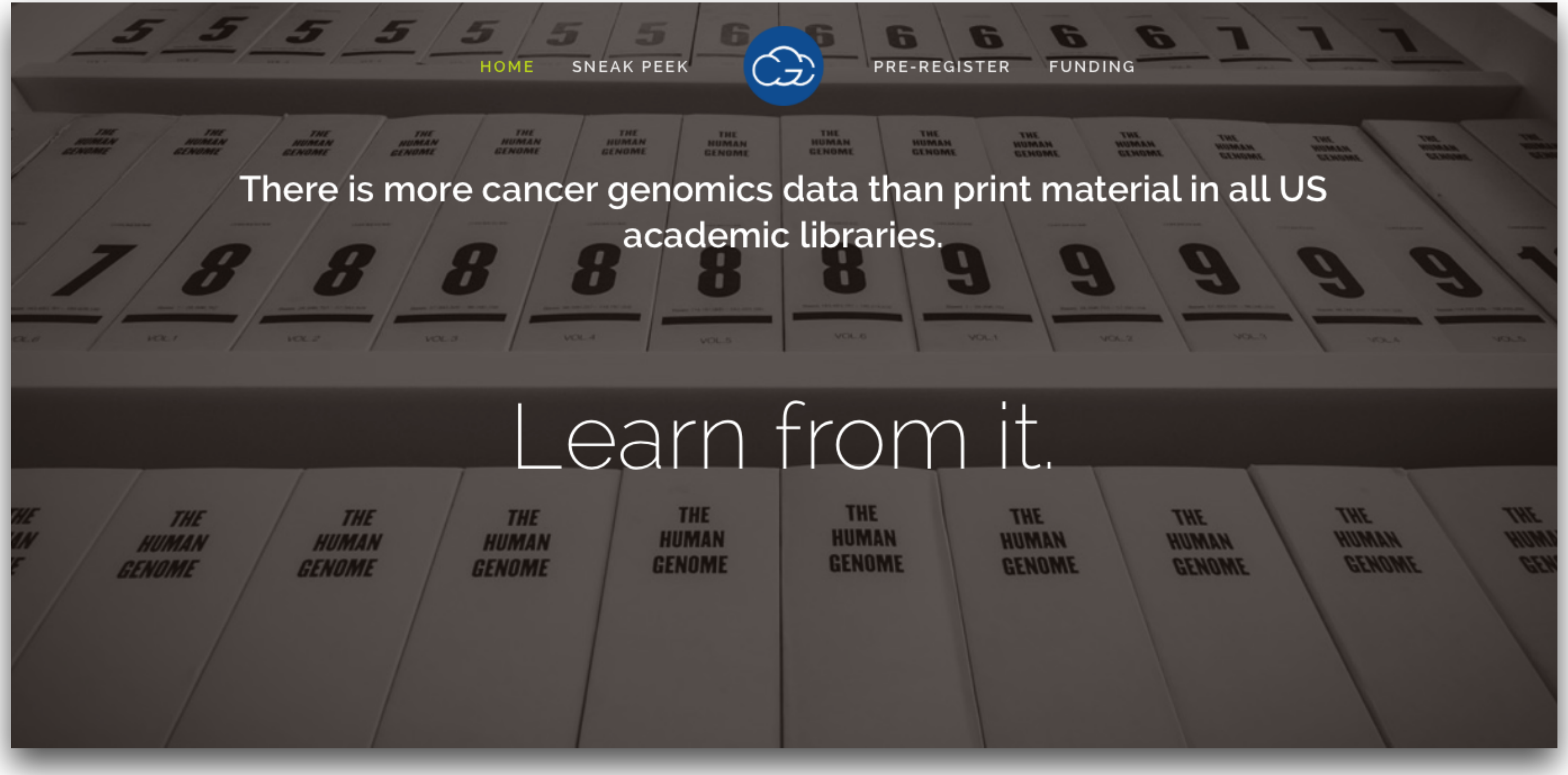
BamTools Index (2.3.0)

Picard AlignmentSummaryMetrics (1.111)

Maximum insert size	*100000*
Adapter sequence	*No value*
Metric accumulation level	*ALL_READS*
Validation stringency	*SILENT*
Is bisulfite sequenced	*False*
Assume sorted	*True*
Compression level	*5*
Create Index	*True*

TopHat (2.0.12)

SBG IGOR Platform
igor.sbgenomics.com



There is more cancer genomics data than print material in all US academic libraries.

Learn from it.

NCI Cancer Genomics Cloud Platform
genomicscloud.org

Demo

```
vignette("sbgr")
```

Our Goal

- Making bioinformatics / data analysis pipelines much more reproducible with Docker, Rabix, and knitr.
- Sharing portable bioinformatics pipelines, and reproducible data analysis reports, as a community.

Welcome to the poster session @ 5:15

liftr: http://liftr.me

sbgr: https://github.com/road2stat/sbgr