

Interoperable, collaborative multi-platform variant calling with bcbio

Brad Chapman
Bioinformatics Core, Harvard Chan School

<https://bcb.io>

<http://j.mp/bcbiolinks>

22 July 2017

Mike Lin Retweeted



DNAnexus, Inc. @dnanexus · 13 Jun 2013

#BigData Parking: "There's no reason **to move data** outside the **#cloud**. You can do **analysis** right there." ow.ly/m14Ke **#genomics**



Stuart Watt @morungos · 4 Mar 2014

Big upcoming change in **genomics**: **data** sets are now too large **to download** for **analysis**. **Move code to the data**, not vice versa **#ibcretreat2014**



Rob Schaefer @CSciBio · Jul 17

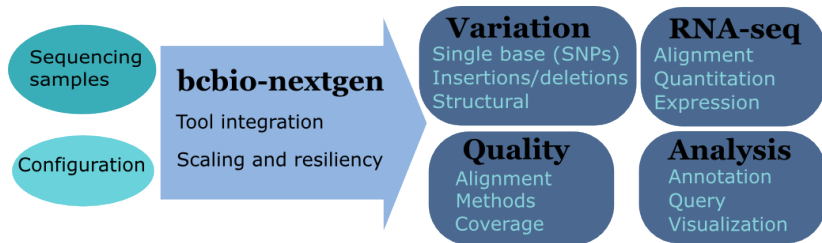
huge problem: moving **analysis** to the data, not the other way around.
[@ewanbirney](#) **#ISAG2017** **#BigData**



Why do we transfer data around?

- Lots of work to setup and configure an analysis
- Hard to port scalable analysis to new environment

bcbio: community built analyses



<https://github.com/chapmanb/bcbio-nextgen>

Support the work you do

☐ Pipelines

☐ Germline variant calling

Basic germline calling

Population calling

Cancer variant calling

Structural variant calling

RNA-seq

single-cell RNA-seq

smallRNA-seq

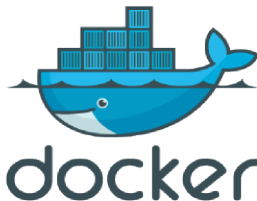
ChIP-seq

<https://bcbio-nextgen.readthedocs.org/en/latest/contents/pipelines.html>

Better abstractions = more interoperability



COMMON
WORKFLOW
LANGUAGE



<https://bcbio-nextgen.readthedocs.io/en/latest/contents/cwl.html>

Supported now



Common Workflow Language reference implementation

Toil

Toil is an open-source pure-Python workflow engine



Arvados™



CANCER GENOMICS CLOUD
SEVEN BRIDGES

In progress

DNAneXus

 broadinstitute / cromwell

 Code

 Issues 286

 Pull req

Workflow Execution Engine using WDL



nextflow

Bringing analysis to your data

- Community built analysis – work together
- Interoperable workflows – the right abstractions
- Platform support – real practical progress

<https://bcb.io>

<http://j.mp/bcbiolinks>