

Galaxy-qld users meetup

December 7, 2016

QBP/IMB, UQ

Igor Makunin, UQ RCC

Public data on Galaxy-qld

- Why use public data?
- What is available?
- What do you need?

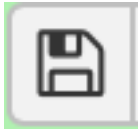
Slides are available through the GVL-Qld blog
<https://genomicsvirtuallab.wordpress.com>

This Galaxy service is maintained by the Research Computing Centre, UQ

Site news and announcements are available on the GVL_QLD Twitter and GVL_QLD blog

Why use public data?

Galaxy-qld storage:



49 Tb Volume storage (user data)

1 Tb Volume storage for indices

No external backup

User allocation:

UQ: 1 Tb

Qld and Australian users: 600 Gb

Datasets imported from Data Libraries are shared:

- no data coping
- not counted towards user quota
- tested datasets, e.g. gene annotations

dbSNP consumes 0% of user quota

Both datasets were imported from a Data Library

Using 0%

History

search datasets

Unnamed history
2 shown, 2 deleted

47.88 GB

4: dbSNP_147.hg19.vcf

~170,000,000 lines
format: vcf, database: hg19

dbSNP_147, hg19

display at UCSC [main](#)
display with IGV [web](#) [current](#) [local](#)
display at RViewer [main](#)

1. Chrom	2. Pos	3. ID	4. Ref
##fileformat=VCFv4.0			

3: dbsnp_144.hg38.vcf

Public data on Galaxy-ql

Build-in data

Genome indices for BWA, bowtie2/tophat2, Lastz

- 12 species, 17 genome assemblies

Databases for SnpEff: Variant analysis

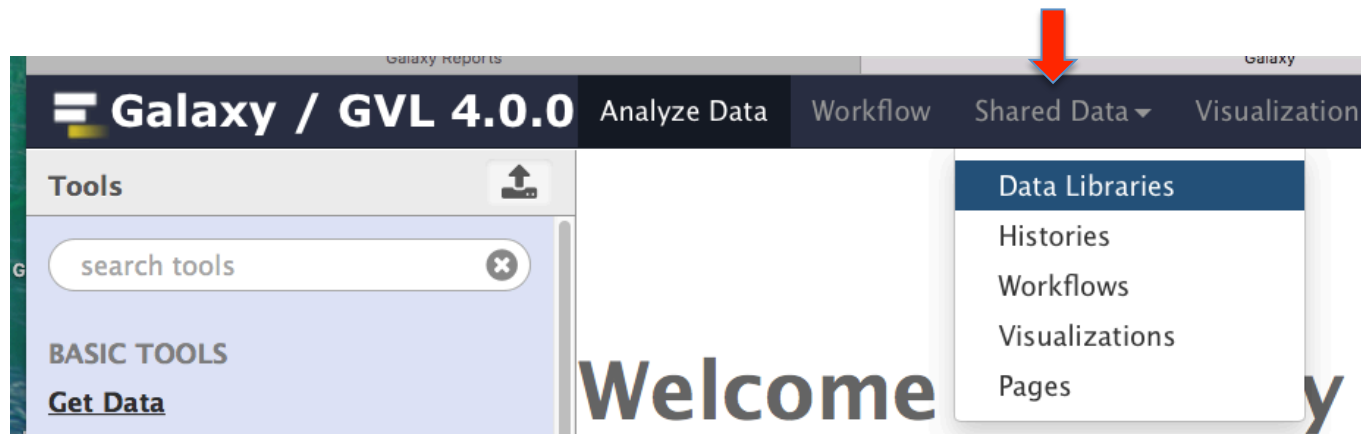
Data Libraries

Human variation data

Gene annotations (four species)

Tutorial datasets

Data Libraries



Human gene annotations	GenCode gene annotation for the human genome	Downloaded from http://www.gencodegenes.org
Human variation data	dbSNP data	dbSNP
Metagenomics 16S rRNA Data	Data for the GVL Metagenomics 16S rRNA tutorial	Data for the GVL Metagenomics tutorial
Mouse gene annotations	GenCode gene annotation for the mm10 / GRCm38 genome assembly	Ensemble 78 gene models for the mouse genome
Solanum tuberosum gene annotations	Potato gene annotations from Potato Genome Sequencing Consortium	Potato gene annotations from Potato Genome Sequencing Consortium

Data Libraries

DATA LIBRARIES

« 0 1 2 »

showing 2 of 2 items

include deleted



to History

Download ▾

Delete

Details

Help

[Libraries](#) / Human variation data

[name](#) ↓^A_Z

description

data type size time updated (UTC)



[dbsnp_144.hg38.vcf](#)

Coordinates: hg38/GRCh38. Downloaded from GATK bundle. chr was added to contig names, e.g. 1 was changed to chr1.

vcf

22.3
GB

2016-03-11 05:59
AM



[dbSNP_147.hg19.vcf](#)

Coordinates: hg19/GRCh37. Downloaded from ftp.ncbi.nih.gov/snp/organisms/human_9606_b147_GRCh37p13/VCF/GATK/All_20160601.vcf

vcf

25.6
GB

2016-10-08 12:07
AM



Plans

We are keen to add new public data to Galaxy-qld.
Preference for big and popular datasets

Potential options:

- public NCBI Blast+ dataset
- genome indices for other species
- gene annotations
- ???

Share data through GenomeSpace Australia

<https://genomespace.genome.edu.au>

Proposed topics for the next meeting

Date: February 1, 2017

Time: 10:00 – 11:00 AM

Location: QBP/IMB Seminar Room Small 3.146

Topic:

De novo transcriptome assembly in Galaxy:

- data preparation, QC, read trimming
- Trinity read normalization
- assembly by Trinity
- read count (RSEM, eXpress)

Email Igor at i.makunin@uq.edu.au