BIO392
Bioinformatics of Genome Variations

Genome Variation | Function | Data Formats | Resources | Privacy

Michael Baudis **UZH** **SIB**
Computational Oncogenomics

University of Zurich UZH

Task: Exploration of different file formats

- Which genomic file formats exist & what are their use cases?
    - SAM
    - BAM
    - CRAM
    - VCF
    - FASTA
    - MPEG-G

Task: Estimate Storage Requirements for 1000 Genomes

- How much computer storage is required for 1000 Genomes

  - WES & WGS

  - Different file formats

    - SAM
    - BAM
    - CRAM
    - VCF
    - FASTA
  - Associated costs
    - Cost factors
    - Raw Storage costs

Please provide 1-page size estimates and reasoning for the use of the different file types (i.e. which would you use for storing called variants, which for full archival purposes, browser visualisation), for 3-5 formats.
Submit your files (.md) per pull request to your Github directory.

BIO392 HS 2020
Github Activity

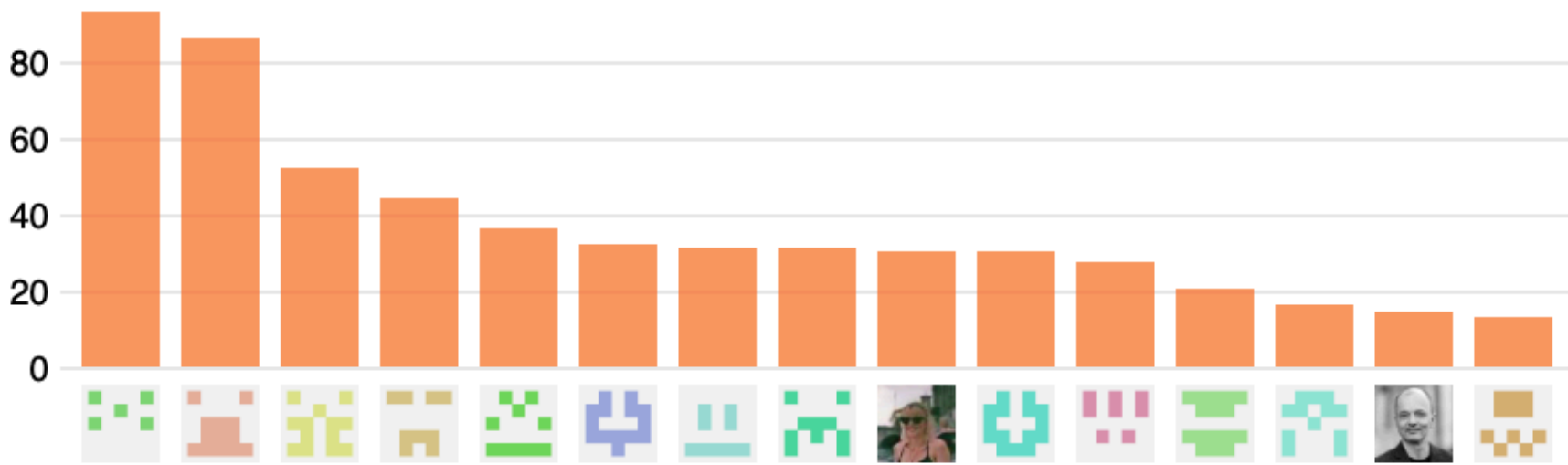September 15, 2020 – September 22, 2020

Period: 1 week ▾

Overview

13 Active Pull Requests

0 Active Issues

⑂ **13**
Merged Pull Requests

⇅ **0**
Open Pull Requests

⊘ **0**
Closed Issues

⊙ **0**
New Issues

Excluding merges, **22 authors** have pushed **604 commits** to master and **608 commits** to all branches. On master, **98 files** have changed and there have been **4,025 additions** and **26 deletions**.

⑂ **13** Pull requests merged by **8** people

samtools

🔗 http://samtools.github.io

🔍 Find a repository...   Type: All ▾   Language: All ▾

samtools

Tools (written in C using htslib) for manipulating next-generation sequencing data

⚫ C   ⑂ 449   ⭐ 979   ⊙ 167   ⑃ 21   Updated 4 hours ago

bcftools

This is the official development repository for BCFtools. To compile, the develop branch of htslib is needed: git clone --branch=develop git://github.com/samtools/htslib.git htslib

⚫ C   ⑂ 165   ⭐ 329   ⊙ 168   ⑃ 3   Updated 4 hours ago

htslib

C library for high-throughput sequencing data formats

⚫ C   ⑂ 348   ⭐ 491   ⊙ 113   ⑃ 20   Updated 4 days ago

hts-specs

Specifications of SAM/BAM and related high-throughput sequencing file formats

🟢 TeX   ⑂ 140   ⭐ 393   ⊙ 114   ⑃ 38   Updated 15 days ago

Top languages

⚫ C   🟠 Java   🔵 Perl   🟣 CSS   🟢 TeX

People   14 >

Report abuse

Task: Reading up on Genome Technologies

- General NGS technologies

- count based vs. intensity based as principle

- bonus: dig deeper for some molecular-cytogenetic techniques:
  - SNP, aCGH arrays
  - SKY, M-FISH
  - chromosomal CGH