CNVpytor: a tool for CNV/CNA detection and analysis from read depth and allele imbalance in whole genome sequencing

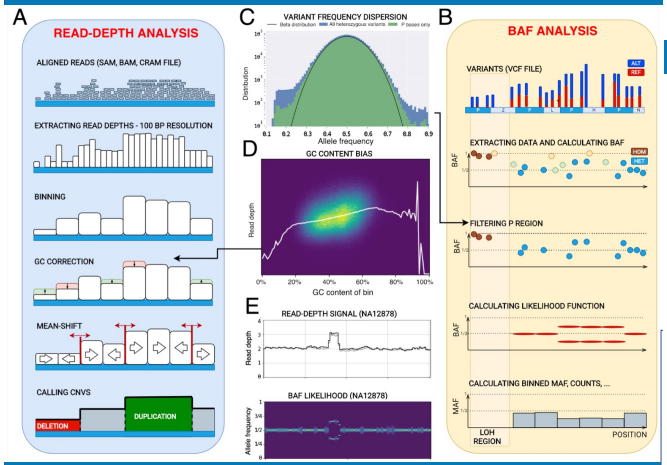
Milovan Suvakov[1] , Arijit Panda[1] , Colin Diesh[2] , Ian Holmes[2], Alexej Abyzov [1,*]

[1] Department of Quantitative Health Sciences , Mayo Clinic, Rochester, MN, USA
[2] Department of Bioengineering, University of California, Berkeley, USA
* Correspond author. Email: abyzov.alexej@mayo.edu

Abstract

Detecting copy number variations (CNVs) and copy number alterations (CNAs) based on whole genome sequencing data is important for personalized genomics and treatment. CNVnator is one of the most popular tools for CNV/CNA discovery and analysis based on read depth (RD). Herein, we present an extension of CNVnator developed in Python -- CNVpytor. CNVpytor inherits the reimplemented core engine of its predecessor and extends visualization, modularization, performance, and functionality. Additionally, CNVpytor uses B-allele frequency (BAF) likelihood information from single nucleotide polymorphism and small indels data as additional evidence for CNVs/CNAs and as primary information for copy number neutral losses of heterozygosity. CNVpytor is significantly faster than CNVnator—particularly for parsing alignment files (2 to 20 times faster)—and has (20-50 times) smaller intermediate files. CNV calls can be filtered using several criteria and annotated. Modular architecture allows it to be used in shared and cloud environments such as Google Colab and Jupyter notebook. Data can be exported into JBrowse, while a lightweight plugin version of CNVpytor for JBrowse enables nearly instant and GUI-assisted analysis of CNVs by any user. CNVpytor release and the source code are available on GitHub at https://github.com/abyzovlab/CNVpytor under the MIT license.

Working Method and Scheme

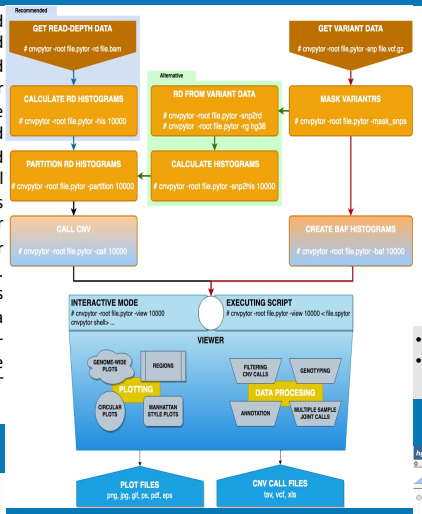


Tutorial

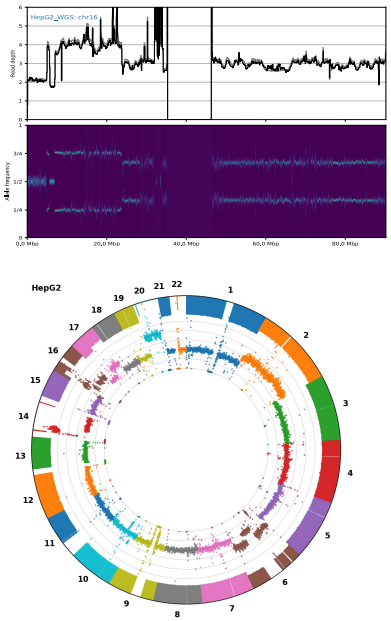
SCAN ME

- Google Colab: Using CEPH trio example dataset
  - https://github.com/abyzovlab/CNVpytor/blob/master/examples/Colab.ipynb
- CNVpytor as python Module:
  - https://github.com/abyzovlab/CNVpytor/blob/master/examples/CNVpytor.ipynb

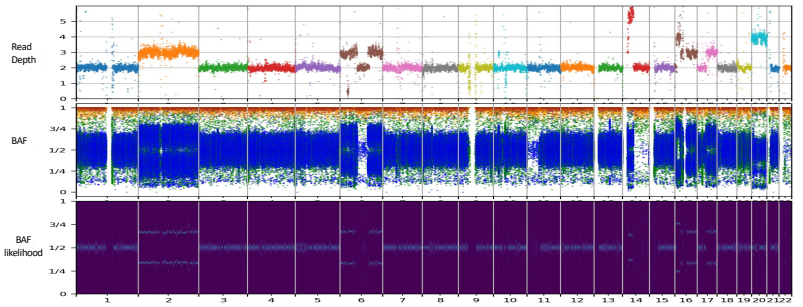
Pipeline



Region and circular plot



Plotting: HepG2 cell line



- Manhattan plot for HepG2 cell line. For read depth and BAF likelihood, 1000000 bin size is used.
- Amplification of whole chromosome amplification of chr2 and chr20 is visible where chr2 has three copies and chr20 has four copies.

Integration with JBrowse

- Jbrowse[3] view (chromosome 6) of K562 cell line where amplification is visible on both side of the chromosome.
- CNVpytor[2] data, stored in HDF5 files, can be exported to visualize in Jbrowse.
- An Jbrowse plugin[4] is developed for on-fly analysis and visualization of CNV from VCF file.

Installation

```
pip install git+https://github.com/abyzovlab/CNVpytor.git
```

References

1.Suvakov,M. *et al.* (2021) CNVpytor: a tool for CNV/CNA detection and analysis from read depth and allele imbalance in whole genome sequencing. *Biorxiv*, 2021.01.27.428472.
2. Abyzov,A. *et al.* (2011) CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res*, **21**, 974–984.
3. Buels,R. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol*, **17**, 66.
4. https://github.com/abyzovlab/CNVpytorVCF