IGV tool – Integrative Genomics Viewer

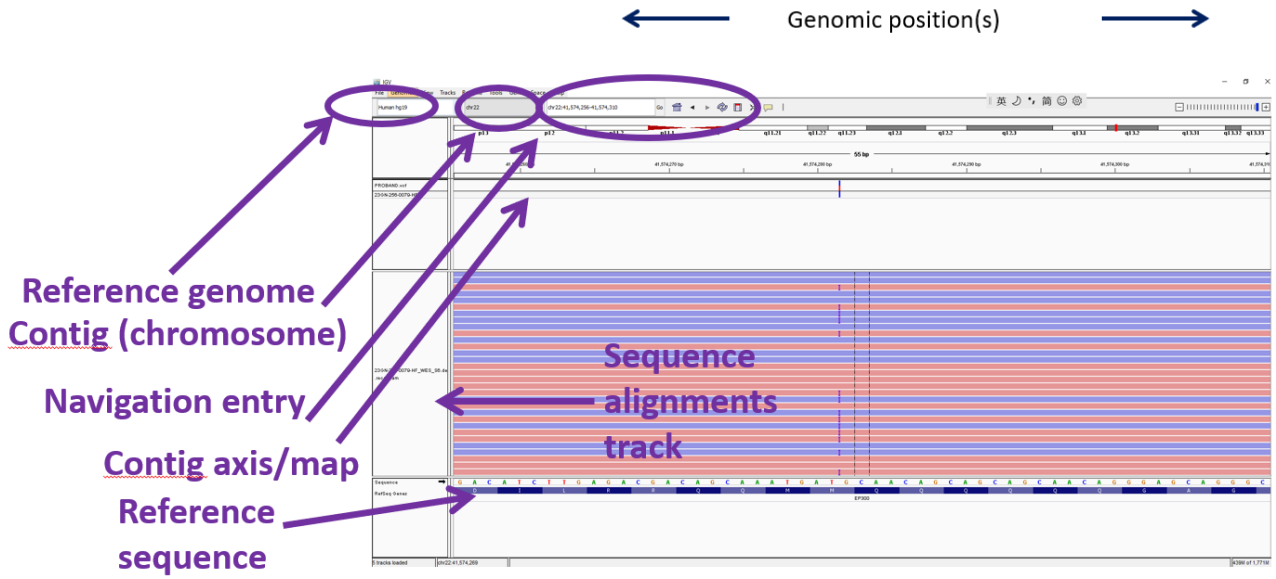
Background

- IGV is a free and widely used tool for viewing and analyzing NGS data
- Variants called by bioinformatics pipelines can be visualized in the source NGS reads, allowing real variants to be distinguished from artifacts
- The IGV Team is based at UC San Diego and the Broad Institute of MIT and Harvard

Software links

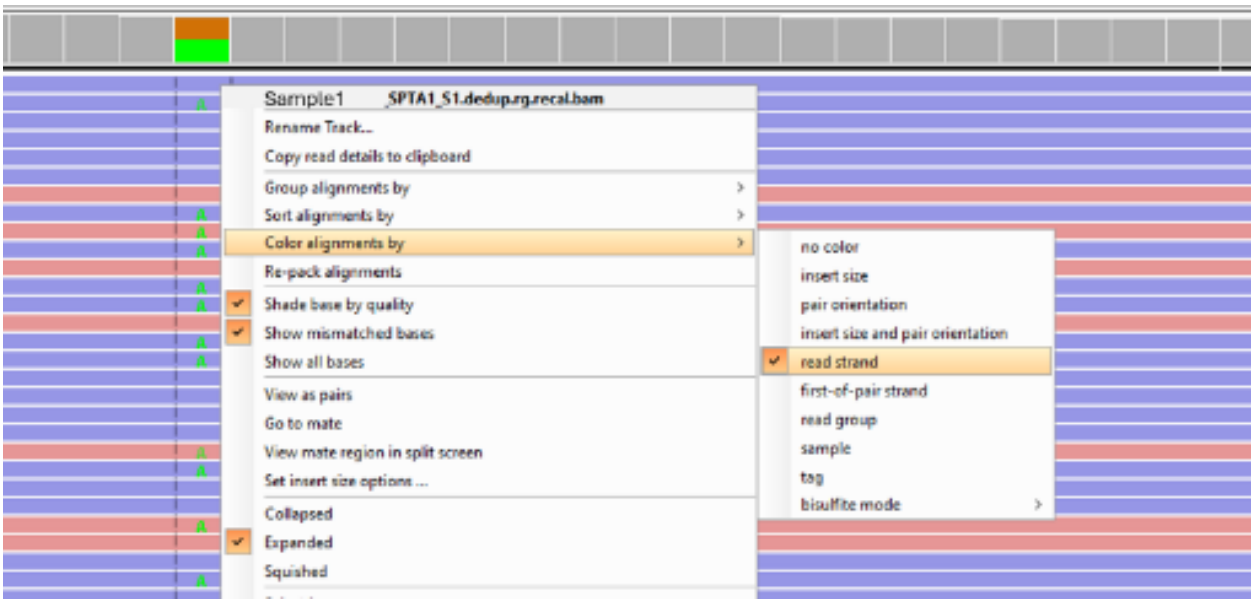
- Latest version downloads https://igv.org/doc/desktop/#DownloadPage/
- Help/feature chat group https://groups.google.com/forum/#!forum/igv-help

Basic layout:



Adding color to forward and reverse reads:

If a variant is real, it should be detected in forward and reverse NGS strands. In the default settings for IGV, forward and reverse strands might be grey, making it difficult to differentiate between them. To apply different colors to forward and reverse reads, right-click on NGS reads and choose “Color alignments by” -> “read strand” (screenshot below).



Sorting reads

Rearranging the order of reads can sometimes help evaluate the quality of a variant call. To rearrange NGS reads, right-click on them and choose “Sort alignments by” -> [choose method].

Reference sequence

- A consensus, baseline, wild-type, or comparator DNA sequence
- Used as the comparator to define ‘what has changed’ – sequence variants, structural rearrangements
- Reference alleles – REF
- Alternative alleles/variants – ALT

Reference genome

- A set of reference sequences for an organism or species under study
    - Divided into ‘contigs’ – pieces of DNA that are part of the same physical package
    - In eukaryotes, contigs include chromosomes and the mitochondrial genome
- Human contigs (major): chr1 – chr22, chrX, chrY, chrM
- Interesting point – since a reference genome is generally a consensus created by sequencing many individual members of a species, it is likely that no individual member actually has a perfect consensus sequence
    - We *all* have variants

The reference genome in IGV

- Most clinical labs are still using hg19/GRCh37
- The latest build is GRCh38
- For most clinical sequences this doesn't make much difference because clinical labs are generally reporting at the gene & protein level (c. & p. in HGVS), not the genome level
    - BRAF c.1799T>A (p.V600E) is the nomenclature of the common BRAF mutation in melanoma, papillary thyroid carcinoma, regardless of the reference genome
- FASTA file: the simplest reference genome
    - FASTA is a simple text format for representing DNA sequences
    - Can include multiple sequences (in the setting of a reference genome, multiple contigs)
- Optionally, reference genomes (GENOME files) may include:
    - gene level projections/transcript mappings
    - a cytomap/karyotype projection

The sequence alignments track in IGV

- This track contains the experimental sample's sequence.
- File types that can be loaded in the sequence alignments track include:
    - SAM – sequence alignment map file (.SAM)
        - ➢ Uncompressed human-readable format
    - BAM – binary (sequence) alignment map file (.BAM)
        - ➢ Compressed non-readable format (i.e., binary equivalent of a SAM file)
    - Both can be indexed (.SAI and .BAI)
        - ➢ Index is always binary and is a fraction of the size of the.SAM/.BAM
        - ➢ Contains information about where alignments that map to a particular position are located in the sequence alignment map

For more information, please visit:
https://igv.org/doc/desktop/#QuickStart/

References:

Hands On Exploration of NGS Data using the Integrative Genomics Viewer (IGV) (routbort.org)
https://igv.org/

Author Information and Affiliations:

Xinxiu Xu, PhD

Cincinnati Children's Hospital Medical Center

Division of Human Genetics

Date Created: 5/6/2024