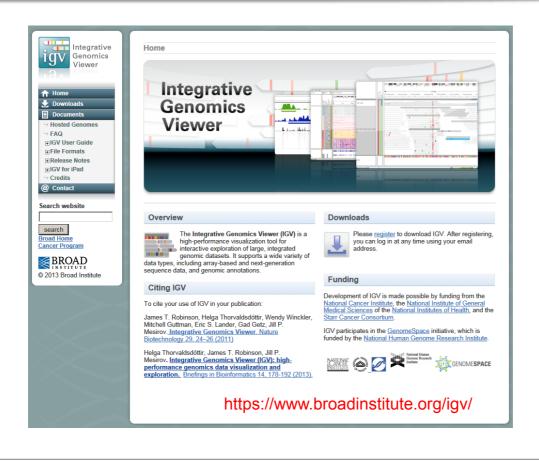
基生研ゲノムインフォマティックス・トレーニングコース 2016秋 RNA-seq入門 - NGSの基礎からde novo解析まで 実践編: RNA-seq解析パイプライン 2016.09.08-2016.09.09

NGS基本ツールIGV

基礎生物学研究所 生物機能解析センター 山口勝司

データ可視化ツール・IGVの紹介・実習



なぜIGVを取り上げるか

データ可視化ツール

- ・自分のパソコン(ローカル環境)にインストールして使うタイプ
- サーバーに構築して、ネットワークで使うタイプ

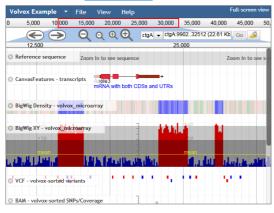
The JBrowse Genome Browser

JBrowse is a fast, embeddable genome browser built completely with JavaScript and HTML5, with optional run-once data formatting tools written in Perl.

Featured Post

Exploring structural variation using JBrowse by Richard Finkers

Latest Release - JBrowse 1.11.6



コミュニティーに広く利用、あるいは ウェブ公開を目的とするには良いが、 ネットワーク・情報セキュリティの 高度な知識も要求される。

より大容量なデータに対応できる。

管理者的な人がいて、その人がやって くれるなら、これも良いが。

もっとお手軽なものとしてIGVを紹介

可視化ツールに求められるものは何か

膨大なデータを如何に直感的に理解できるようにするか sortや絞り込みができる表データと対比双璧

- ·配列、GC ratio、遺伝子情報
- •遺伝子発現情報
- ·SNPの位置情報·頻度情報
- 様々なデータの精度情報

レファレンス配列 / gene model / gene annotationとNGSデータを並べて比較複数のデータセットを並べて比較

色々なデータ(variant, 発現, ChIP, BSseq等々)を、様々なスケールで 比較・統合的に解釈できるようにしたい

ゲノムviewerに自分のデータを乗せ、 統合的直感的に解釈できること

可視化ツールをどう選ぶか

選択の基準 genome data viewing に求められるもの 取捨選択の基準

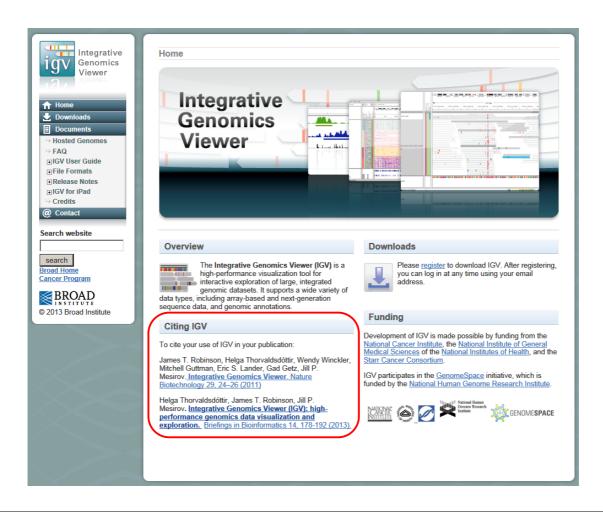
- 1. 無料 / 有料 / 基本無料
- 2. 個人的レベルの使用 / コミュニティーレベルの使用
- 3. 見るだけ/自分から色々工夫
- 4.アクセスのしやすさ・使いやすさ 導入に必要なコンピュータスペック マニュアルは分かりやすいか 情報の多さ 利用の簡便さ 使っている人が近くにいるか

Integrative Genomics Viewer(IGV)

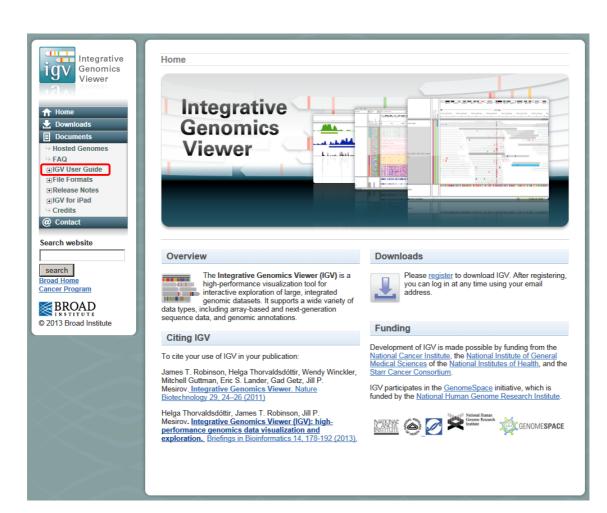
お手軽ツール

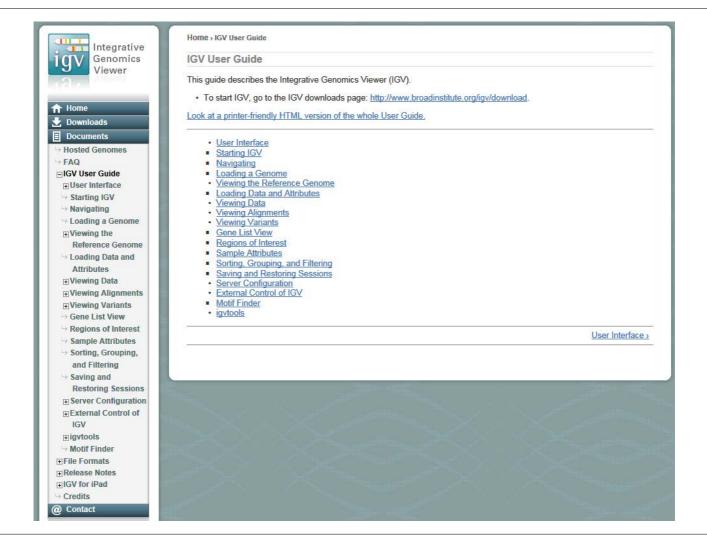
- アカデミックウェアで無料
- コミュニティーでの利用者が多いから、情報も多い
- ・javaのプログラムなので、オールプラットフォーム対応
- マニュアルは親切、サンプルデータのある
- ・WEBサーバーではなく、PCレベルでできる
- ・データ閲覧環境の共有が可能

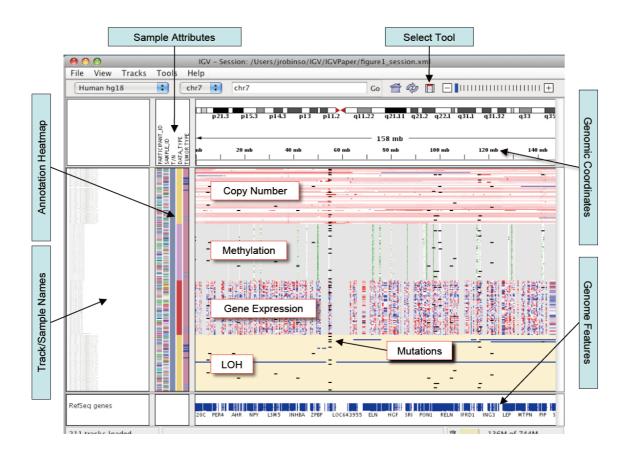
誰もが簡便に使えるものが良い。



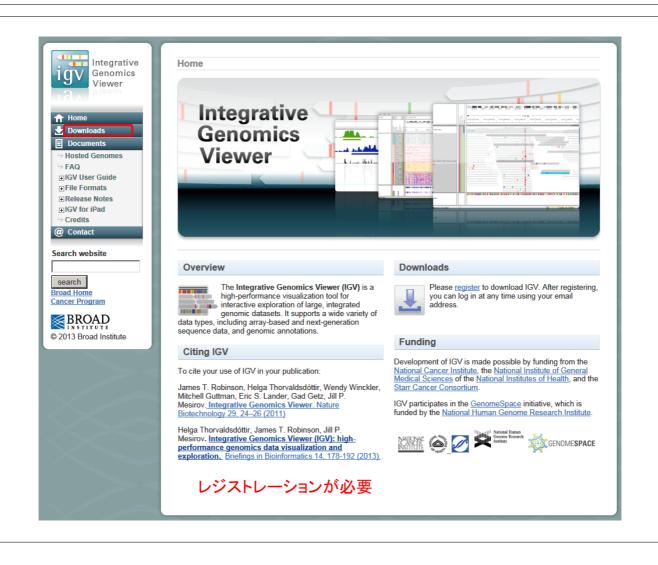


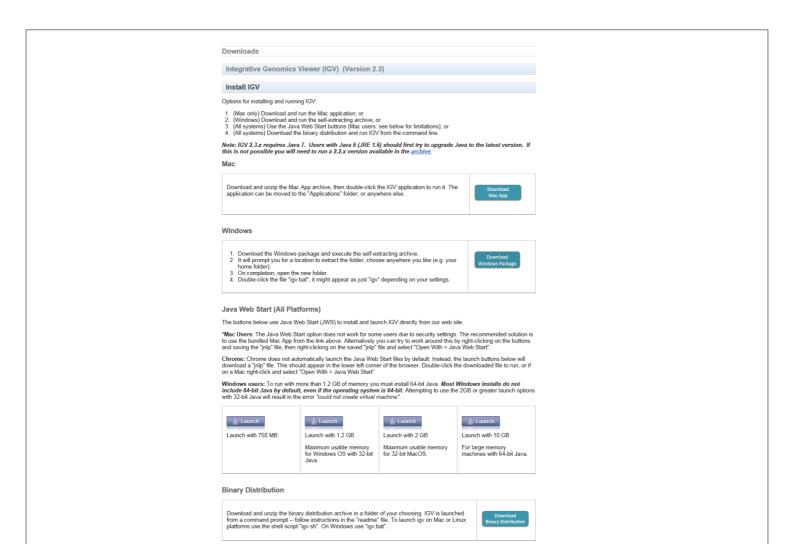


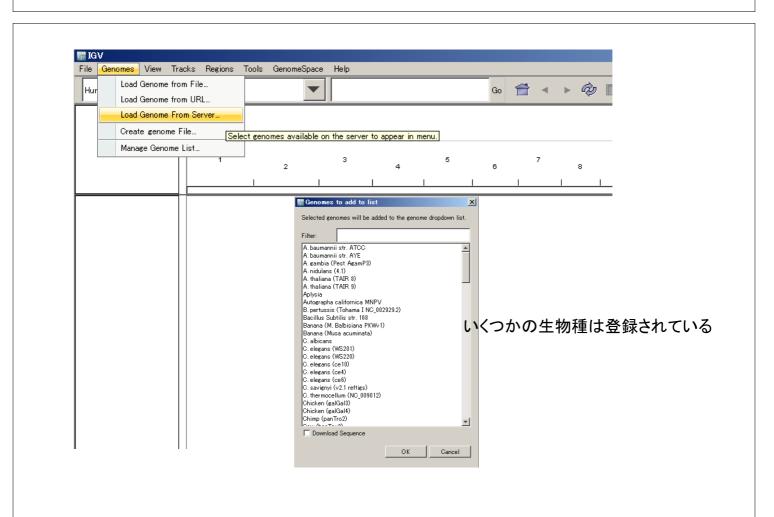


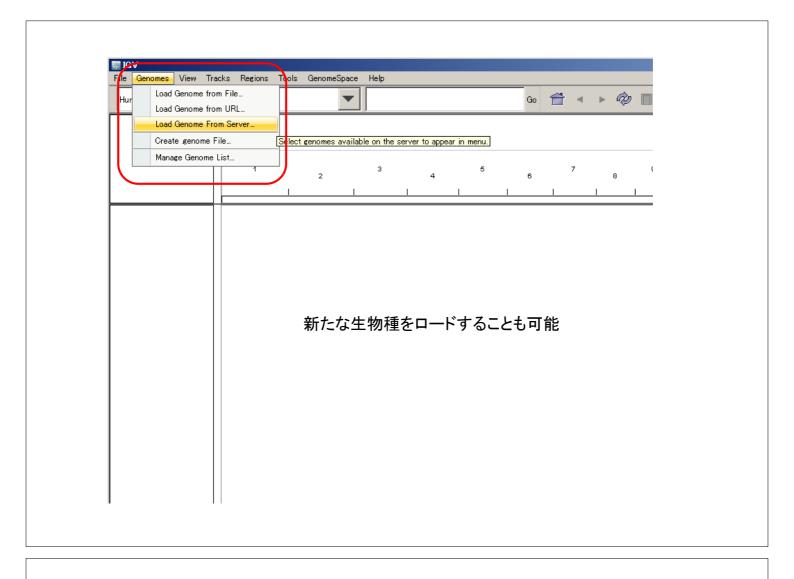


Nature Biotech. 29:24-26 (2011) Supplement figureからの抜粋



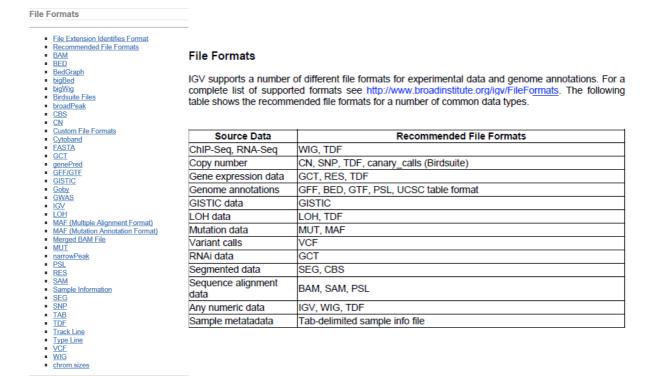




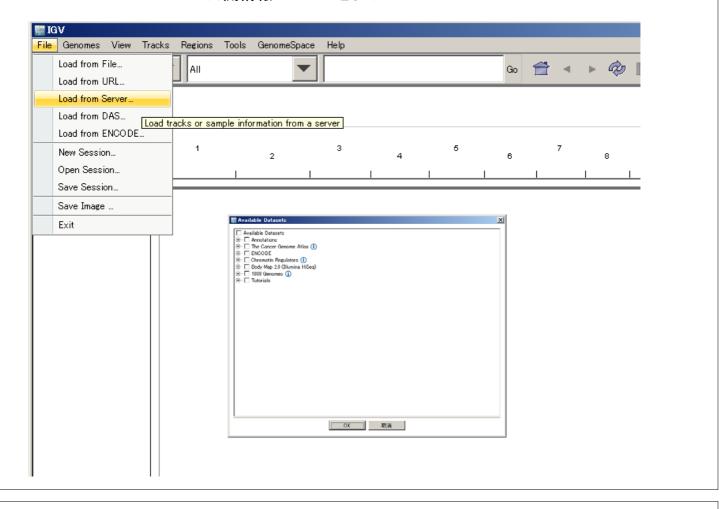


ゲノムViewerなので次世代DNAシーケンサーのデータに限定されない。 マイクロアレイの結果や、ゲノムアノテーションの情報も随時表示できる。

対応するファイル形式に応じて、表示方法が決まる。



公開情報のviewerとして



その他の便利機能

セッションの保存

表示しているデータの読み込み状況を、それごと保存。 セッションをロードすることで、意図した画面を表示できる。 データセットが揃っていること、フォルダー構造が同一である必要がある。

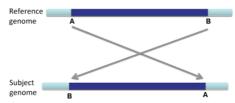
バッチ処理

重要領域の画面スナップショットを自動で取ったりできる。

new load myfile.bam snapshotDirectory mySnapshotDirectory genome hg18 goto chr1:65,289,335-65,309,335 sort position collapse snapshot goto chr1:113,144,120-113,164,120 sort base collapse snapshot



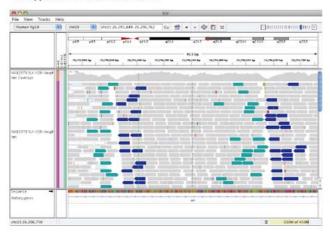
An inversion is a large section of DNA that is reversed in the subject genome compared to the reference genome.



When an inversion shows up in paired-end reads, the reads are distinctively variant from the reference genome.



This appears in IGV as shown below



Interpreting Color by Insert Size

The inferred insert size can be used to detect structural variants, such as:

- · deletions
- insertions
- · inter-chromosomal rearrangements

IGV uses color coding to flag anomalous insert sizes. When you select Color alignments>by insert size in the popup menu, the default coloring scheme is:

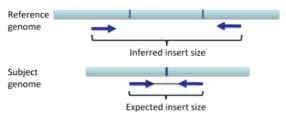
- for an insert that is larger than expected
- for an insert that is smaller than expected
- 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y for paired end reads that are coded by the chromosome on which their mates can be found

Deletions

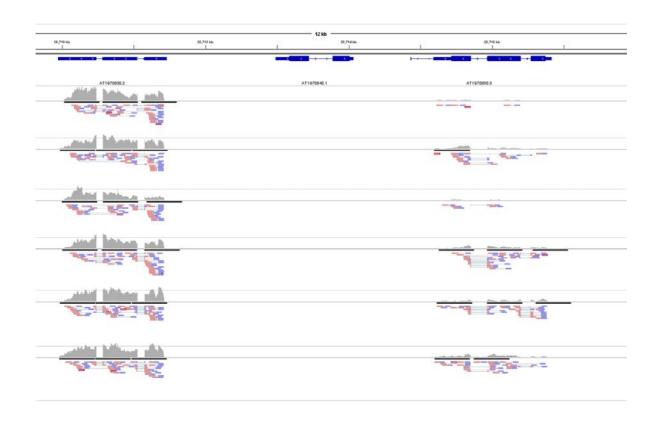
A deletion is a large section of DNA that is absent in the subject genome compared to the reference genome.



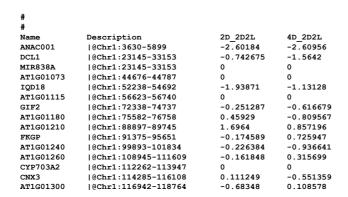
The "expected" insert size is the insert size obtained in sequencing the subject genome. The "inferred" insert size is the insert size that would result in the reference genome, assuming the same pair of reads.

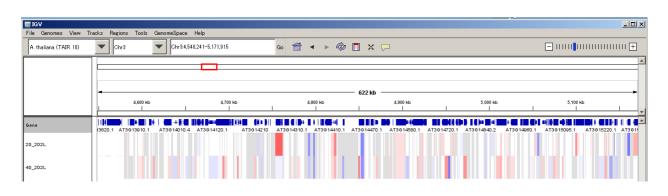


RNA-Seqのデータ表示させる

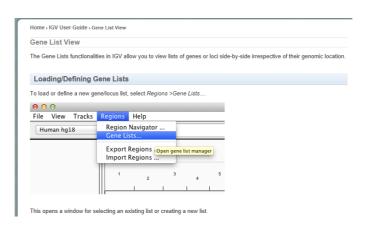


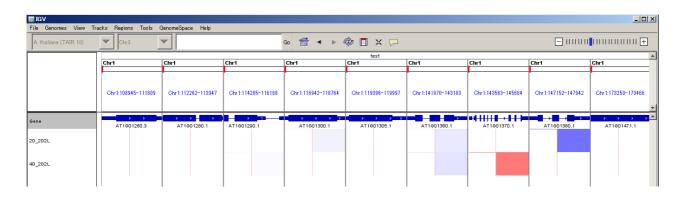
GCTファイルでgene ローカスの発現情報を図示



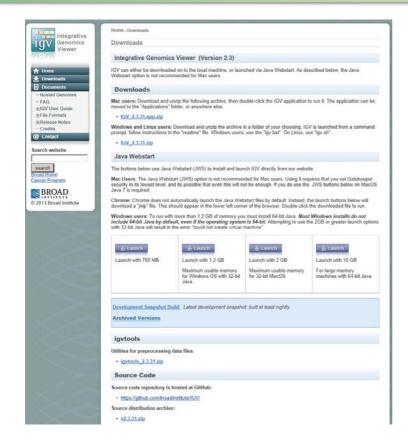


Gene listを定義して サンプルごと 条件ごと の発現・発現変動を カラーマップできる





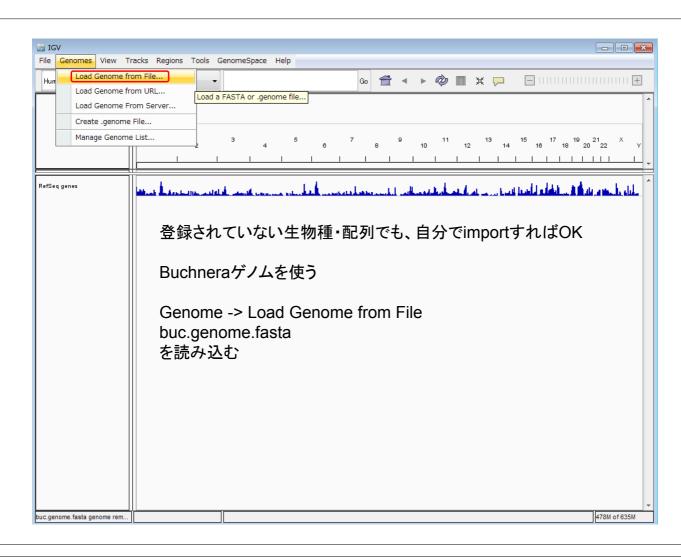
IGV実習

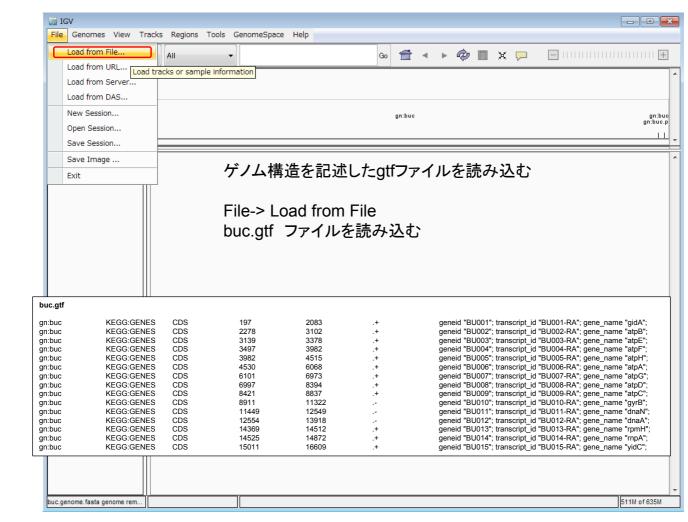


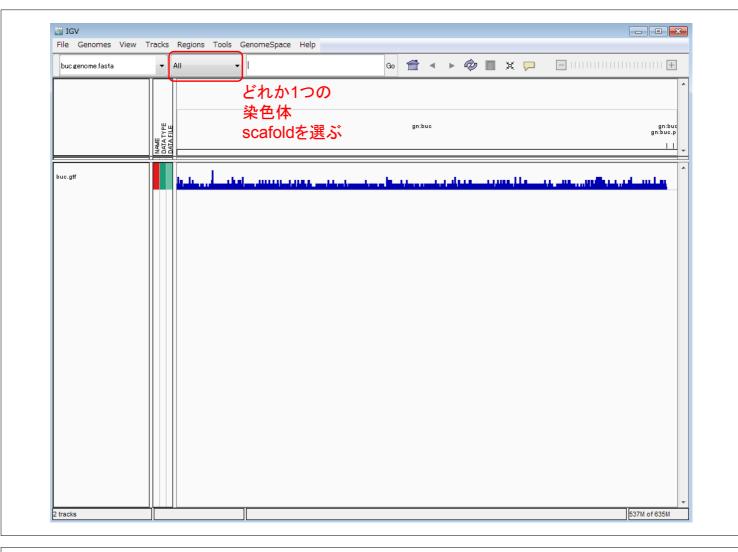
IGVの使用法を学ぶと共に 先のファイルフォーマットも 確認しよう

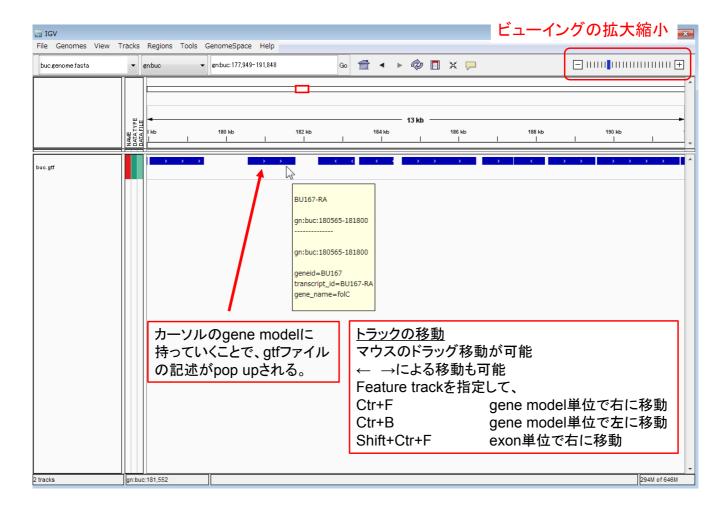
以下のファイルを確認

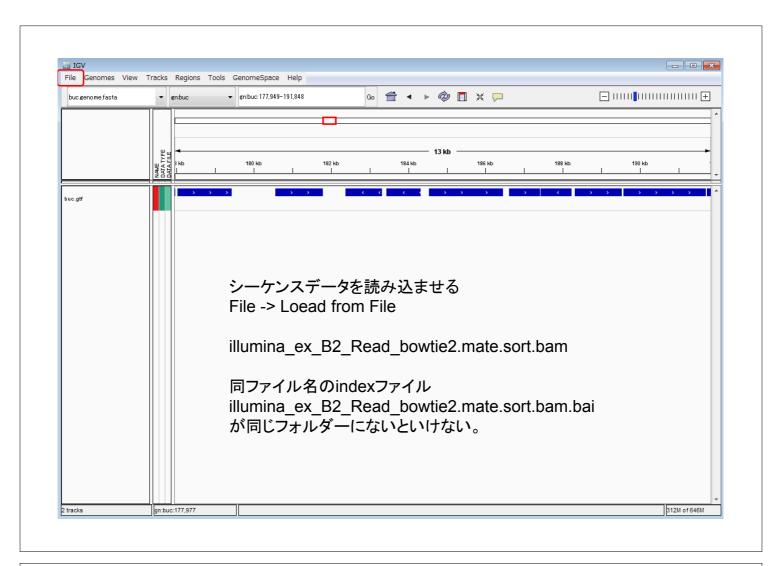
buc.genome.fasta buc.gtf buc_cg.wig illumina_ex_B2_Read_bowtie2.mate.sort.bam illumina_ex_B2_Read_bowtie2.mate.sort.bam.bai illumina_ex_B4_Read_bowtie2.mate.sort.bam illumina_ex_B4_Read_bowtie2.mate.sort.bam.bai

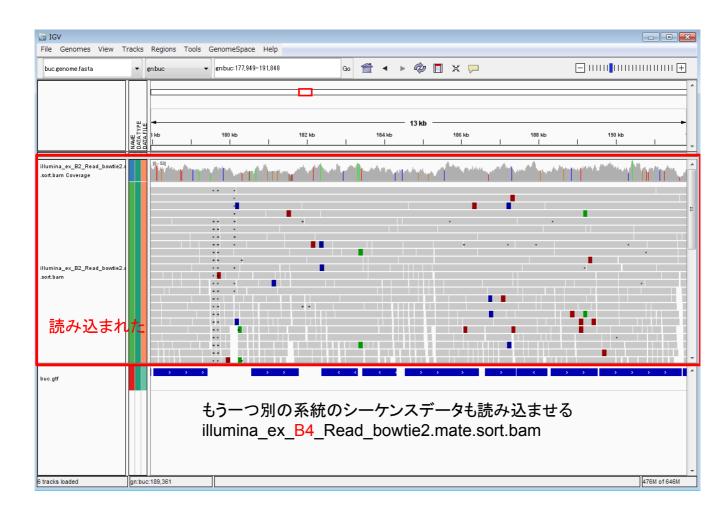


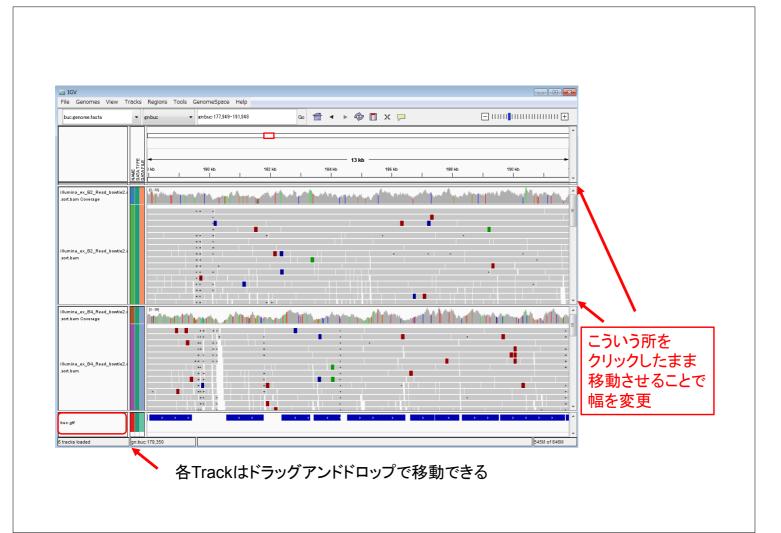


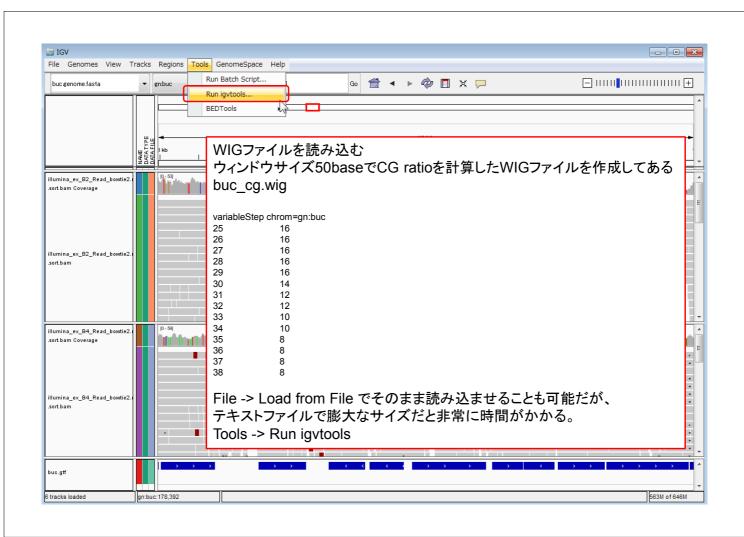


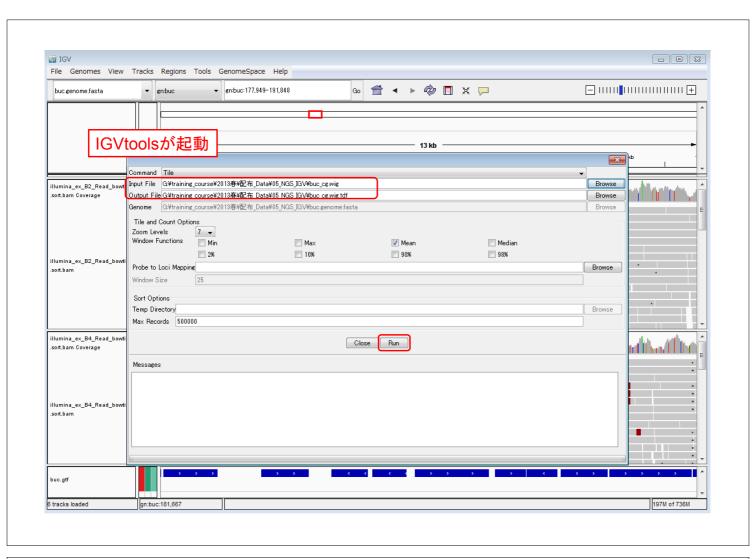


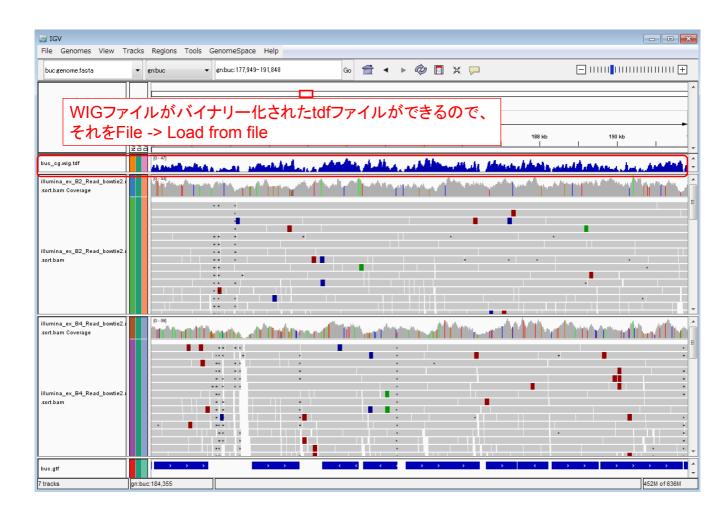


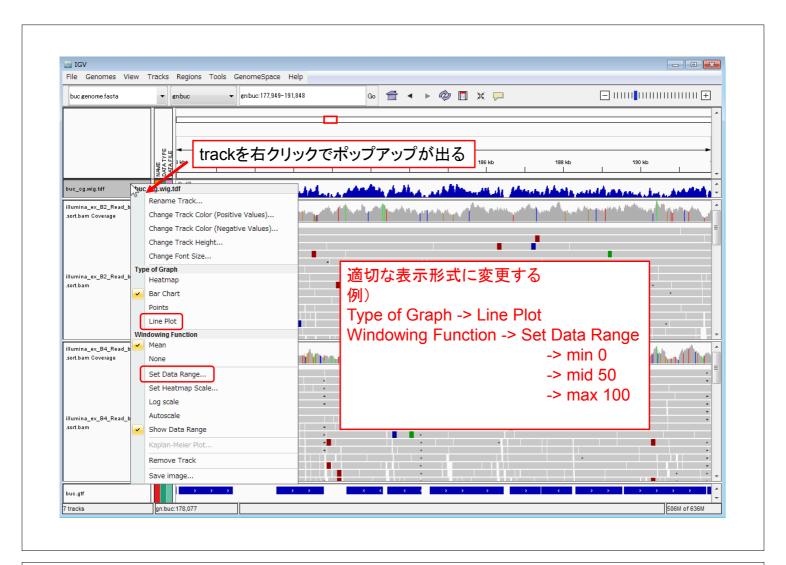


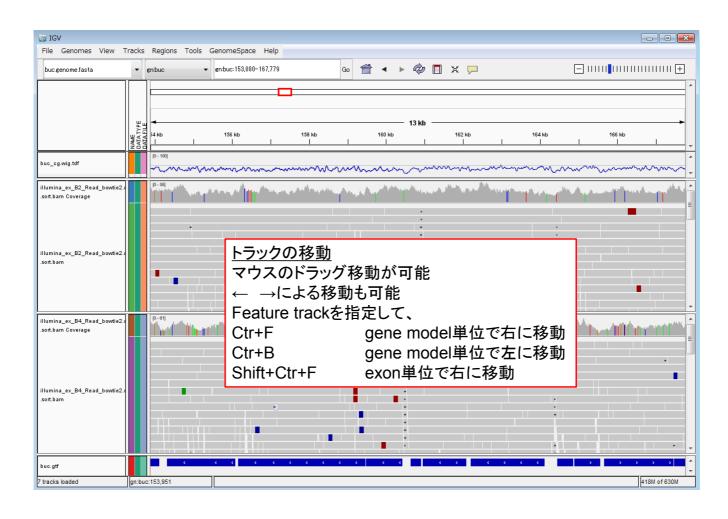


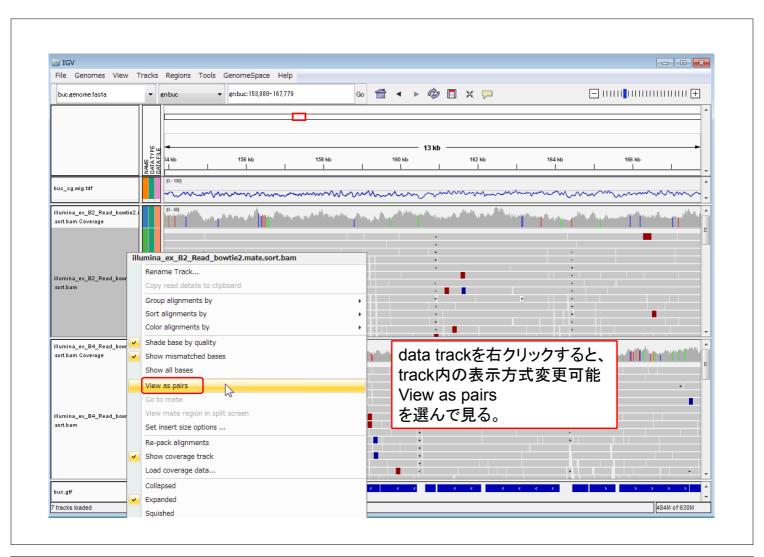


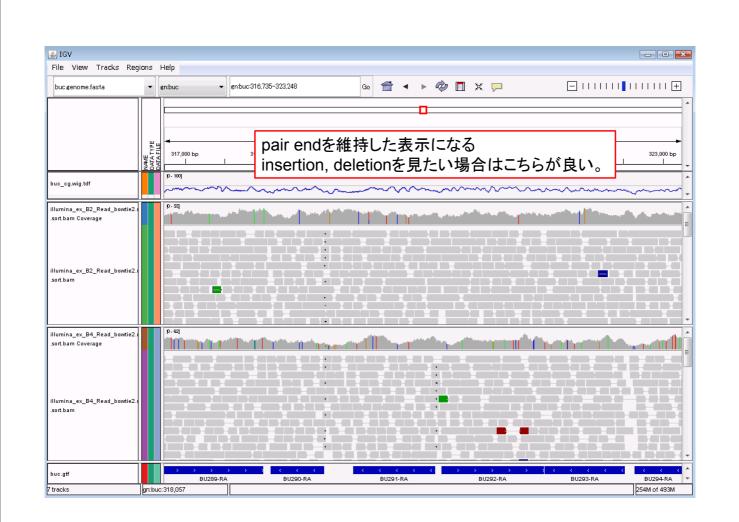


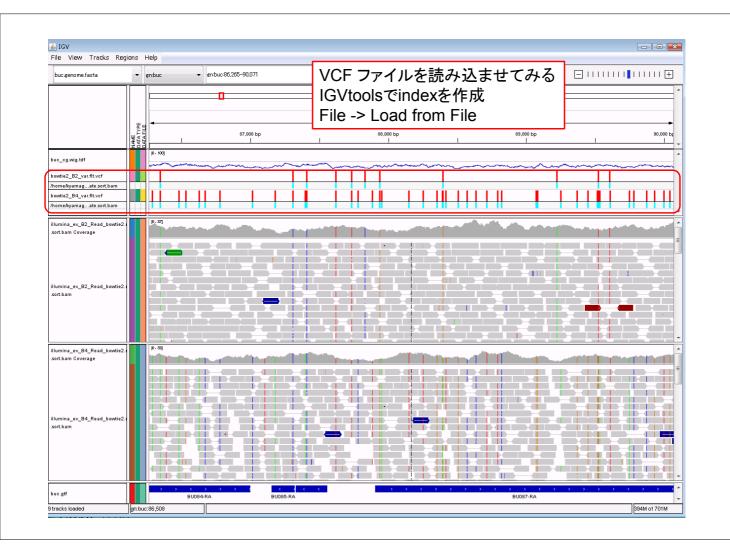


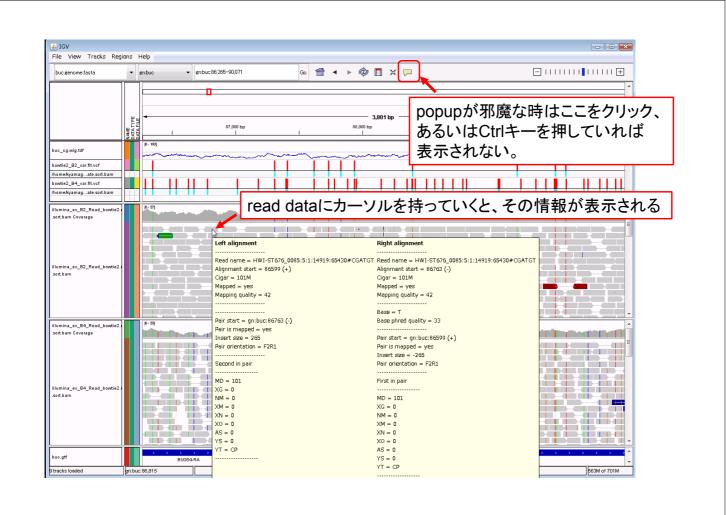












IGV紹介のまとめ

可視化ツールとして十分な機能を持つ

- -無料
- ・比較的簡単・お手軽
- ・自分で見るためにも良し、人に見せるためにも良し
- ・利用範囲は次世代DNAシーケンサーに限定しない 広くゲノミクスの解析に有用

ごく一部のみの機能を紹介しました。 ウェブサイトを見ながら復習をお勧めします。