

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

The screenshot shows the IGV website homepage. On the left is a sidebar with navigation links: Home, Downloads, Documents, Hosted Genomes, FAQ, IGV User Guide, File Formats, Release Notes, Credits, and Contact. Below these is a search bar and copyright information (© 2013-2018 Broad Institute, and the Regents of the University of California). The main content area has a large banner image of the IGV interface. Below the banner are sections for Overview, Downloads, Citing IGV, and Funding. The Overview section describes IGV as a high-performance visualization tool for large, integrated genomic datasets. The Downloads section provides a link to download the desktop application and igvtools. The Citing IGV section lists publications where IGV is cited. The Funding section mentions support from the National Cancer Institute (NCI) and the National Institutes of Health.

Home

Integrative Genomics Viewer

Overview

The Integrative Genomics Viewer (IGV) is a high-performance visualization tool for interactive exploration of large, integrated genomic datasets. It supports a wide variety of data types, including array-based and next-generation sequence data, and genomic annotations.

Downloads

Download the IGV desktop application and igvtools.

Citing IGV

To cite your use of IGV in your publication:

James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration*. *Briefings in Bioinformatics* 14, 178-192 (2013).

Funding

Development of IGV has been supported by funding from the National Cancer Institute (NCI) of the National Institutes of Health, the Informatics Technology for Cancer Research (ITCR) of the NCI, and the Starr Cancer Consortium.

IGV participates in the GenomeSpace initiative, which is funded by the National Human Genome Research Institute.

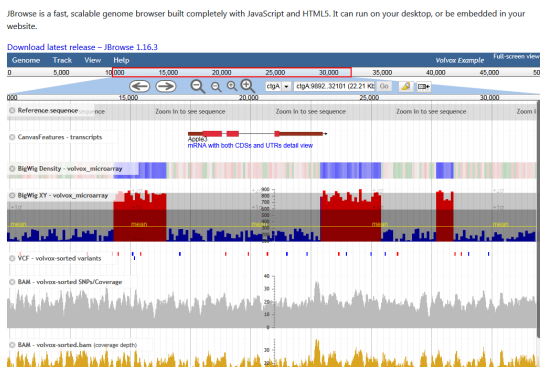
<https://www.broadinstitute.org/igv/>

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、ネットワーク上で使うタイプ

The JBrowse Genome Browser



後者はコミュニティで広く利用、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壁

多様なデジタル情報

- ・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
- ・ChIP-seq, RAD-seq, BS-seq・・・

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

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、
統合的直感的に比較・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

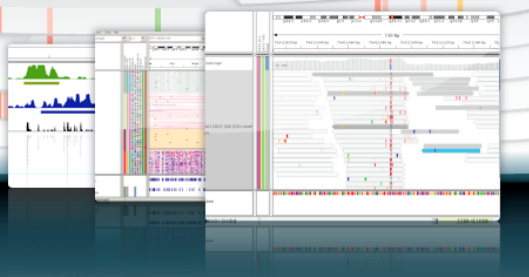
Integrative Genomics Viewer(IGV)

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。

Integrative Genomics Viewer



Overview

The Integrative Genomics Viewer (IGV) is a high-performance visualization tool for interactive exploration of large, integrated genomic datasets. It supports a wide variety of data types, including array-based and next-generation sequence data, and genomic annotations.

Funding

Development of IGV has been supported by funding from the [National Cancer Institute \(NCI\)](#) of the [National Institutes of Health](#), the [Informatics Technology for Cancer Research \(ITCR\)](#) of the NCI, and the [Starr Cancer Consortium](#).

IGV participates in the [GenomeSpace](#) initiative, which is funded by the [National Human Genome Research Institute](#).

Downloads



Download the IGV desktop application and igvtools.

Citing IGV

To cite your use of IGV in your publication:

James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. [Integrative Genomics Viewer](#). *Nature Biotechnology* 29, 24–26 (2011)

Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov. [Integrative Genomics Viewer \(IGV\): high-performance genomics data visualization and exploration](#). *Briefings in Bioinformatics* 14, 178–192 (2013).

Integrative genomics viewer

James T Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S Lander, Gad Getz & Jill P Mesirov

Affiliations | Corresponding authors

Nature Biotechnology 29, 24–26 (2011) | doi:10.1038/nbt.1754
Published online 10 January 2011

To the Editor:

Rapid improvements in sequencing and array-based platforms are resulting in a flood of diverse genome-wide data, including data from exome and whole-genome sequencing, epigenetic surveys, expression profiling of coding and noncoding RNAs, single nucleotide polymorphism (SNP) and copy number profiling, and functional assays. Analysis of these large, diverse data sets holds the promise of a more comprehensive understanding of the genome and its relation to human disease. Experienced and knowledgeable human review is an essential component of this process, complementing computational approaches. This calls for efficient and intuitive visualization tools able to scale to very large data sets and to flexibly integrate multiple data types, including clinical data. However, the sheer volume and scope of data pose a significant challenge to the development of such tools.

Journal home

Current issue

For authors

Subscribe

E-alert sign up

RSS feed



Subscribe today, save 50% and receive 51 weekly issues of **Nature** in print, online and mobile.

Citations to this article
Crossref (10) Scopus (12) Web of Science (0)

Science jobs from [naturejobs](#)

Faculty Position

Harvard Medical School

Ramalingaswami Re-Entry Fellowship

Ministry of Science & Technology, Government of India

Integrative Genomics Viewer



Overview

The Integrative Genomics Viewer (IGV) is a high-performance visualization tool for interactive exploration of large, integrated genomic datasets. It supports a wide variety of data types, including array-based and next-generation sequence data, and genomic annotations.

Downloads



Download the IGV desktop application and igvtools.

Citing IGV

To cite your use of IGV in your publication:

James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. [Integrative Genomics Viewer](#). *Nature Biotechnology* 29, 24–26 (2011)

Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov. [Integrative Genomics Viewer \(IGV\): high-performance genomics data visualization and exploration](#). *Briefings in Bioinformatics* 14, 178–192 (2013).

Funding

Development of IGV has been supported by funding from the [National Cancer Institute \(NCI\)](#) of the [National Institutes of Health](#), the [Informatics Technology for Cancer Research \(ITCR\)](#) of the NCI, and the [Starr Cancer Consortium](#).

IGV participates in the [GenomeSpace](#) initiative, which is funded by the [National Human Genome Research Institute](#).

IGV User Guide

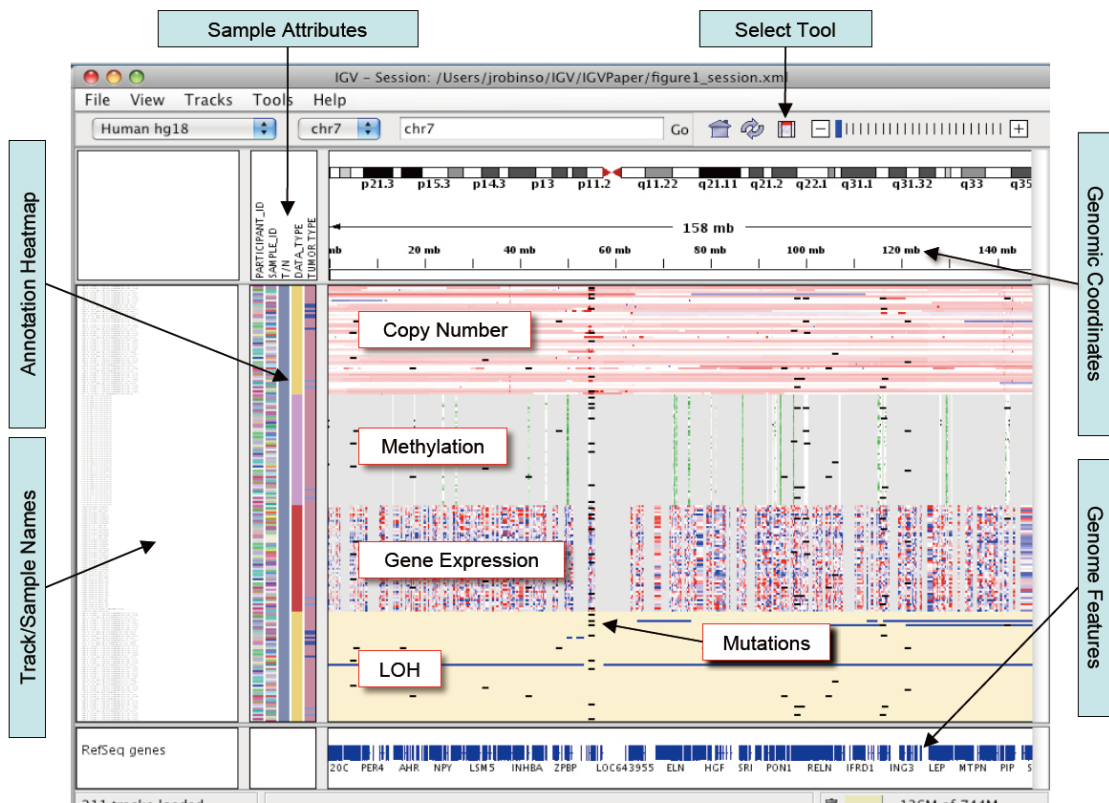
This guide describes the Integrative Genomics Viewer (IGV).

- To start IGV, go to the IGV downloads page: <http://www.broadinstitute.org/igv/download>

[Look at a printer-friendly HTML version of the whole User Guide.](#)

- User Interface
- Starting IGV
- Navigating
- Loading a Genome
- External Control of IGV
- Viewing the Reference Genome
- Loading Data and Attributes
- Viewing Data
- Viewing Alignments
- Viewing Variants
- Gene List View
- Regions of Interest
- Sample Attributes
- Sorting, Grouping, and Filtering
- Saving and Restoring Sessions
- Server Configuration
- igvtools
- Motif Finder
- BLAT search

[User Interface »](#)



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

igv Integrative Genomics Viewer

Home
Downloads
Documents

Hosted Genomes
FAQ
IGV User Guide
File Formats
Release Notes
Credits
Contact

Search website

search

© 2013-2018
Broad Institute, and
the Regents of the
University of California

Home

Integrative Genomics Viewer

Overview

The Integrative Genomics Viewer (IGV) is a high-performance visualization tool for interactive exploration of large, integrated genomic datasets. It supports a wide variety of data types, including array-based and next-generation sequence data, and genomic annotations.

Funding

Development of IGV has been supported by funding from the [National Cancer Institute \(NCI\)](#) of the [National Institutes of Health](#), the [Informatics Technology for Cancer Research \(ITCR\)](#) of the NCI, and the [Starr Cancer Consortium](#).

IGV participates in the [GenomeSpace](#) initiative, which is funded by the [National Human Genome Research Institute](#).

Downloads

Download the IGV desktop application and igvtools.

Citing IGV

To cite your use of IGV in your publication:

James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. [Integrative Genomics Viewer \(IGV\): high-performance genomics data visualization and exploration](#). *Nature Biotechnology* 29, 24–26 (2011)

Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov. [Integrative Genomics Viewer \(IGV\): high-performance genomics data visualization and exploration](#). *Briefings in Bioinformatics* 14, 178–192 (2013).

Downloads

Integrative Genomics Viewer - IGV 2.4

NOTE: IGV 2.4.x releases require [Java 8](#). For Java 11 see the [development snapshot build](#).

Install IGV

Download IGV Mac App

Download and unzip the Mac App Archive, then double-click the IGV application to run it.
The application can be moved to the *Applications* folder, or anywhere else



Download IGV on Windows

Download and unzip the Archive, then double-click the *igv.bat* file to run IGV.
See *readme.txt* to run IGV from the command line



For high DPI screens: Use the [development snapshot build of IGV](#).

Download IGV to run on Linux / MacOS command line

Download and unzip the Archive. See the downloaded *readme.txt* for further instructions.



Launch IGV using Java Web Start



1.2 GB

Max memory for
Windows with 32-bit Java



2 GB

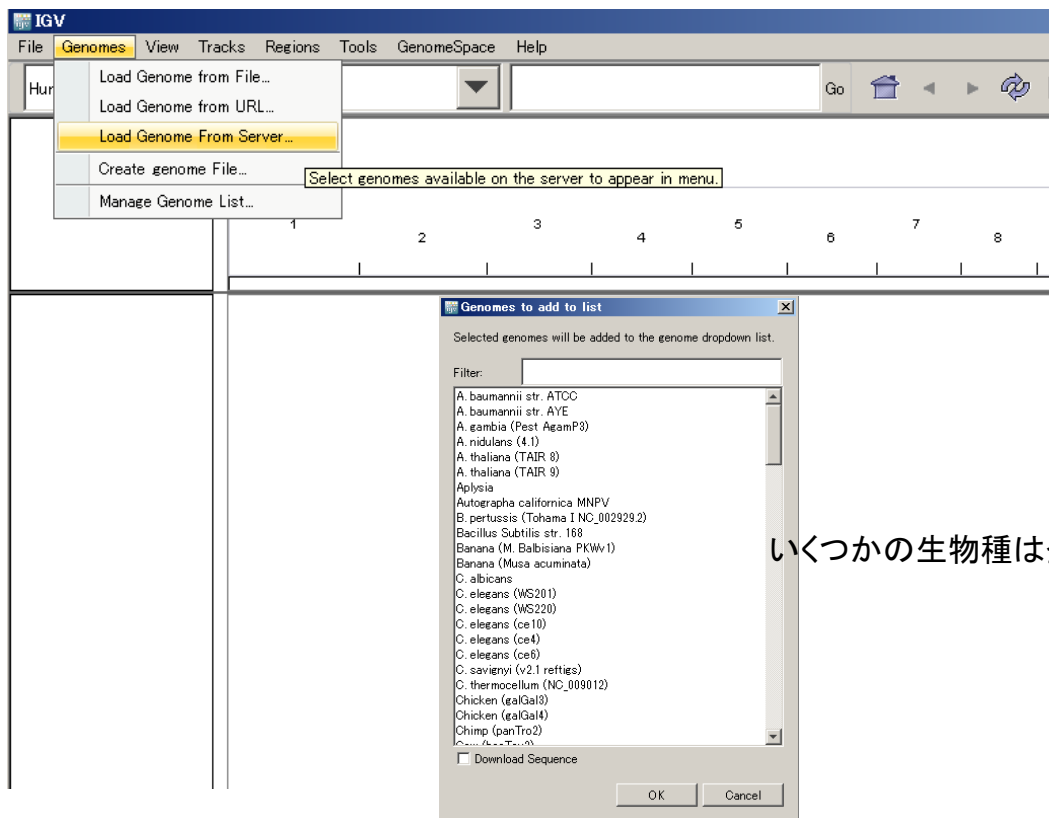


10 GB

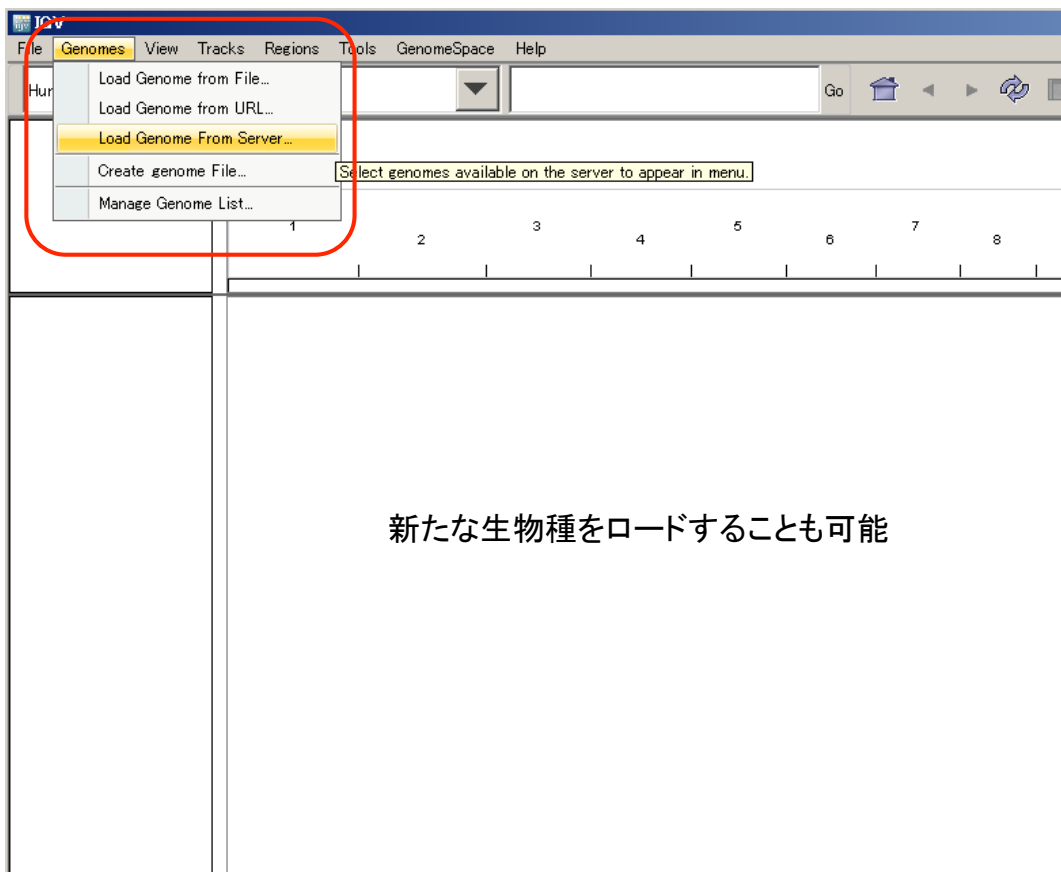
Only for large memory
machines with 64-bit Java

Mac users note: If you are notified of security errors that prevent launching IGV, try right-clicking on the downloaded .jnlp file and select *Open With > Java Web Start*.

Windows users note: Most Windows installs do not include 64-bit Java by default, even if the operating system is 64-bit. Attempting to use the 2GB or greater launch options with 32-bit Java will result in the error "could not create virtual machine".



いくつかの生物種は登録されている



新たな生物種をロードすることも可能

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。

File Formats

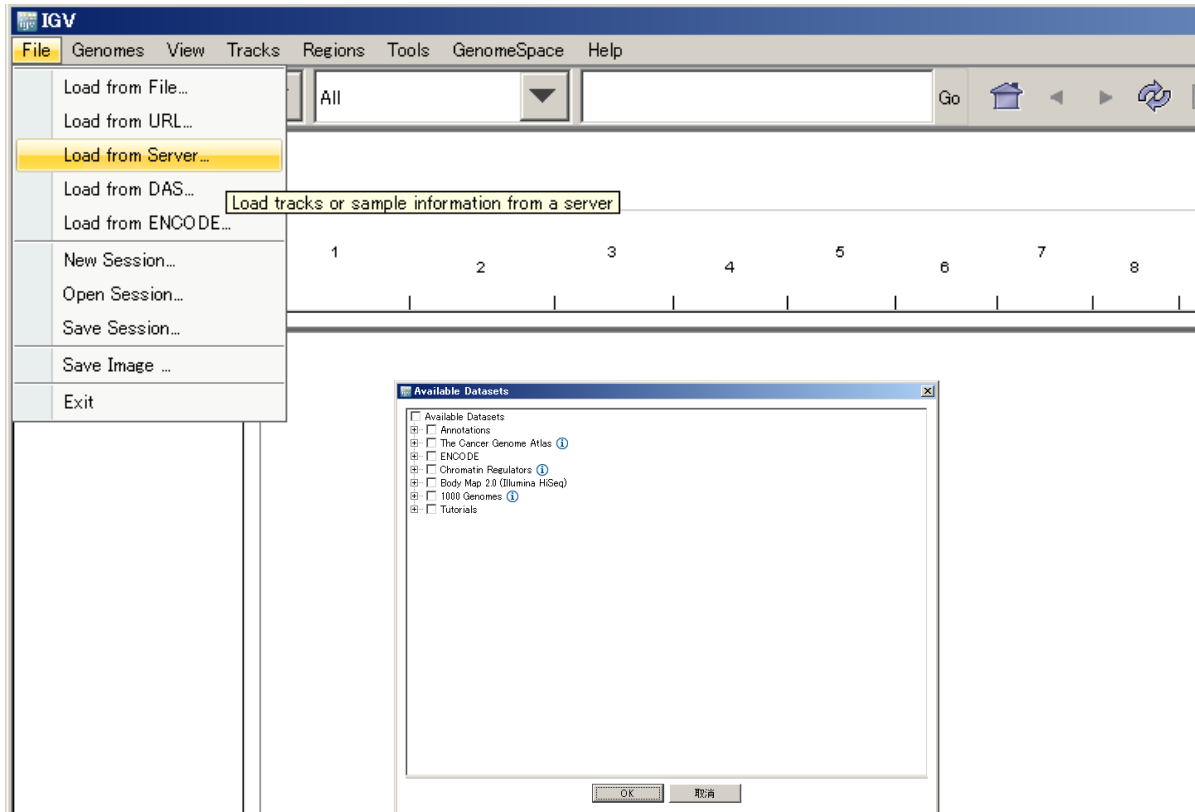
- [File Extension Identifies Format](#)
- [Recommended File Formats](#)
- BAM
- BED
- BedGraph
- bigBed
- bigWig
- Birdsuite Files
- broadPeak
- CBS
- Chemical Reactivity Probing Profiles
- chrom.sizes
- CN
- Custom File Formats
- Cytoband
- FASTA
- GCT
- CRAM
- genePred
- GFF/GTF
- GISTIC
- Goby
- GWAS
- IGV
- LOH
- MAF (Multiple Alignment Format)
- MAF (Mutation Annotation Format)
- Merged BAM File
- MUT
- narrowPeak
- PSL
- RES
- RNA Secondary Structure Formats
- SAM
- Sample Info (Attributes) file
- SEG
- SNP
- TAB
- TDF
- Track Line
- Type Line
- VCF
- WIG

File Formats

IGV supports a number of different file formats for experimental data and genome annotations. For a complete list of supported formats see <http://www.broadinstitute.org/igv/FileFormats>. The following table shows the recommended file formats for a number of common data types.

Source Data	Recommended File Formats
ChIP-Seq, RNA-Seq	WIG, TDF
Copy number	CN, SNP, TDF, canary_calls (Birdsuite)
Gene expression data	GCT, RES, TDF
Genome annotations	GFF, BED, GTF, PSL, UCSC table format
GISTIC data	GISTIC
LOH data	LOH, TDF
Mutation data	MUT, MAF
Variant calls	VCF
RNAi data	GCT
Segmented data	SEG, CBS
Sequence alignment data	BAM, SAM, PSL
Any numeric data	IGV, WIG, TDF
Sample metadata	Tab-delimited sample info file

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。

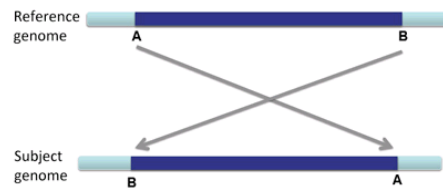
バッチ処理

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。

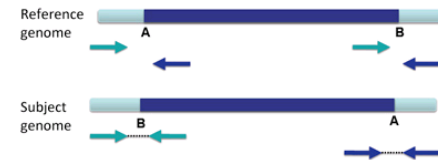
```
new
load myfile.bam
snapshotDirectory mySnapshotDirectory
genome hg18
goto chr1:65,289,335-65,309,335
sort position
collapse
snapshot
goto chr1:113,144,120-113,164,120
sort base
collapse
snapshot
```

Inversions

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:

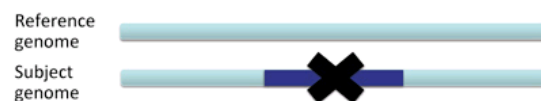
- deletions
- insertions
- inter-chromosomal rearrangements

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:

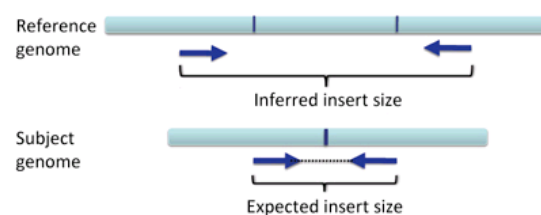
-  for an insert that is larger than expected
-  for an insert that is smaller than expected
-  for paired end reads that are coded by the chromosome on which their mates can be found

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.

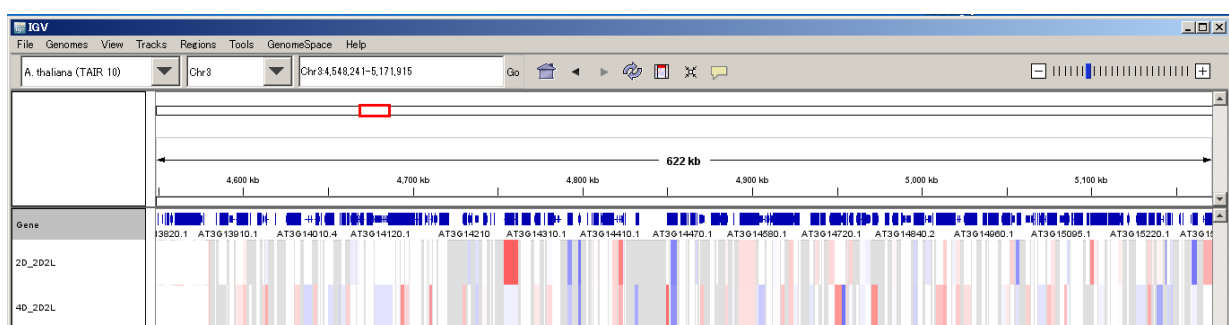


RNA-Seqのデータ表示させる

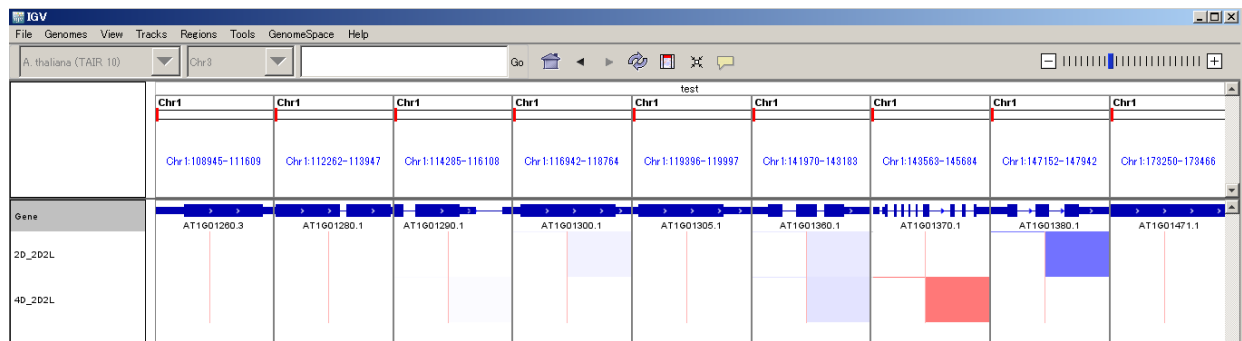
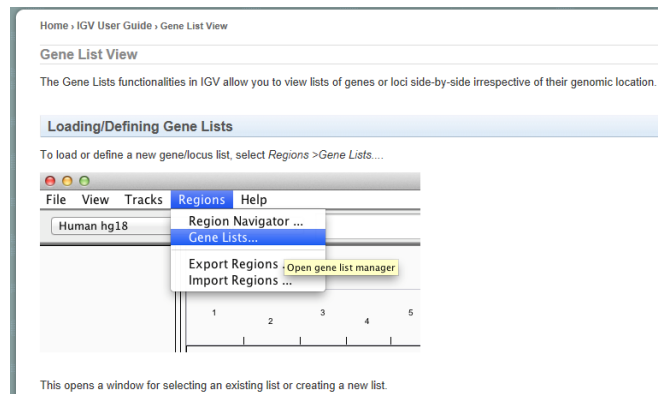


GCTファイルでgene ロークスの発現情報を図示

#				
#	Name	Description	2D_2D2L	4D_2D2L
	ANAC001	@Chr1:3630-5899	-2.60184	-2.60956
	DCL1	@Chr1:23145-33153	-0.742675	-1.5642
	MIR838A	@Chr1:23145-33153	0	0
	AT1G01073	@Chr1:44676-44787	0	0
	IQD18	@Chr1:52238-54692	-1.93871	-1.13128
	AT1G01115	@Chr1:56623-56740	0	0
	GIF2	@Chr1:72338-74737	-0.251287	-0.616679
	AT1G01180	@Chr1:75582-76758	0.45929	-0.809567
	AT1G01210	@Chr1:88897-89745	1.6964	0.857196
	FKGP	@Chr1:91375-95651	-0.174589	0.725947
	AT1G01240	@Chr1:99893-101834	-0.226384	-0.936641
	AT1G01260	@Chr1:108945-111609	-0.161848	0.315699
	CYP703A2	@Chr1:112262-113947	0	0
	CNX3	@Chr1:114285-116108	0.111249	-0.551359
	AT1G01300	@Chr1:116942-118764	-0.68348	0.108578



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

Downloads

Integrative Genomics Viewer - IGV 2.4

NOTE: IGV 2.4.x releases require Java 8. For Java 11 see the development snapshot build.

Install IGV

Download IGV Mac App

Download and unzip the Mac App Archive, then double-click the IGV application to run it. The application can be moved to the Applications folder or anywhere else.



Download IGV on Windows

Download and unzip the Archive, then double-click the igv.bat file to run IGV. See readme.txt to run IGV from the command line.



For high DPI screens: Use the development snapshot build of IGV.

Download IGV to run on Linux / MacOS command line

Download and unzip the Archive. See the downloaded readme.txt for further instructions.



Launch IGV using Java Web Start



1.2 GB

Max memory for Windows with 32-bit Java



2 GB

Only for large memory machines with 64-bit Java



10 GB

Mac users note: If you are notified of security errors that prevent launching IGV, try right-clicking on the downloaded .jnlp file and select Open With > Java Web Start.

Windows users note: Most Windows installs do not include 64-bit Java by default, even if the operating system is 64-bit. Attempting to use the 2GB or greater launch options with 32-bit Java will result in the error "could not create virtual machine".

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
```

登録されていない生物種・配列でも、自分でimportすればOK

Buchneraゲノムを使う

Genome -> Load Genome from File
buc.genome.fasta
を読み込む

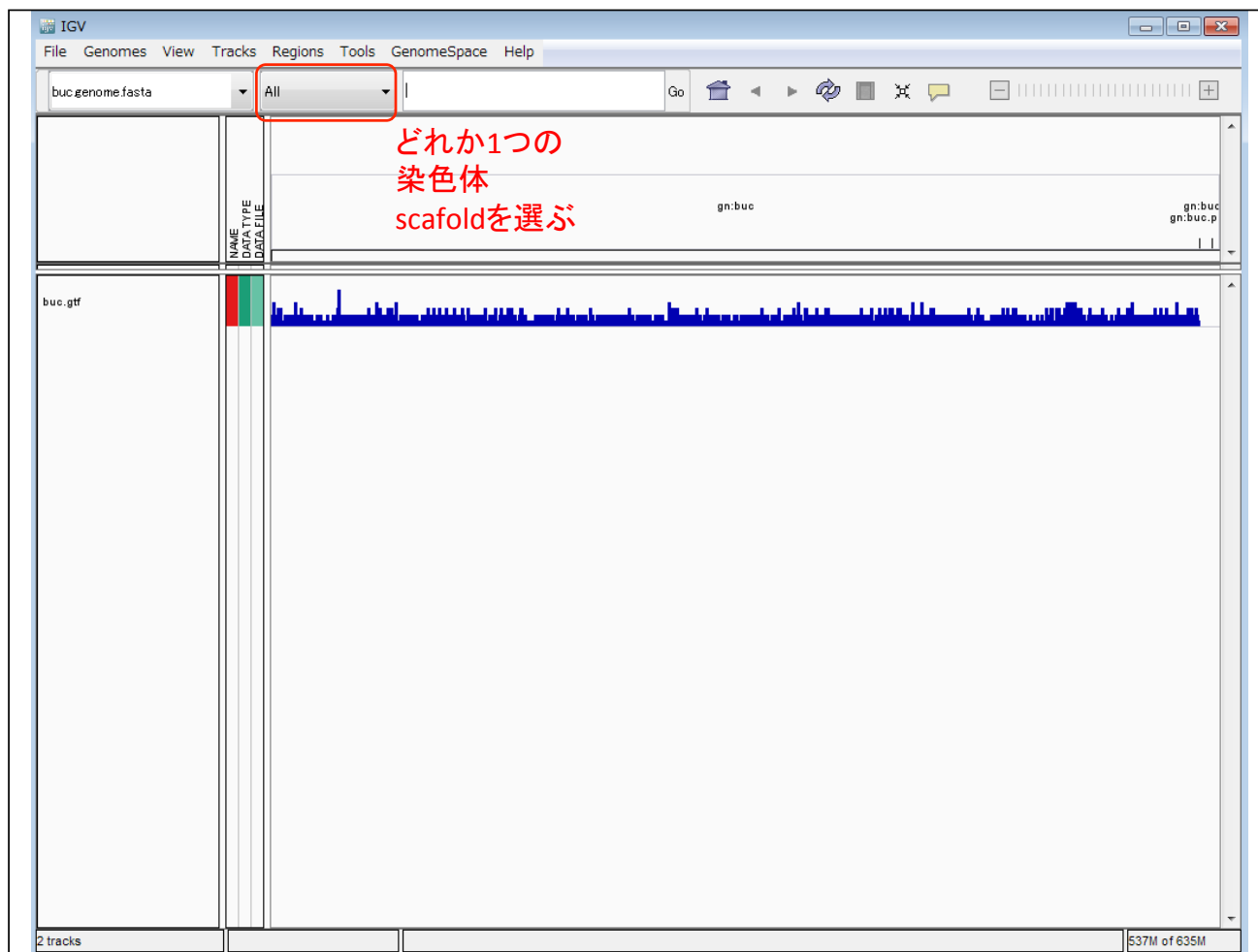
buc.genome.fasta genome rem... 478M of 635M

ゲノム構造を記述したgtfファイルを読み込む

File-> Load from File
buc.gtf ファイルを読み込む

buc.gtf						
gn:buc	KEGG:GENES	CDS	197	2083	+	geneid "BU001"; transcript_id "BU001-RA"; gene_name "gidA";
gn:buc	KEGG:GENES	CDS	2278	3102	+	geneid "BU002"; transcript_id "BU002-RA"; gene_name "atpB";
gn:buc	KEGG:GENES	CDS	3139	3378	+	geneid "BU003"; transcript_id "BU003-RA"; gene_name "atpE";
gn:buc	KEGG:GENES	CDS	3497	3982	+	geneid "BU004"; transcript_id "BU004-RA"; gene_name "atpF";
gn:buc	KEGG:GENES	CDS	3982	4515	+	geneid "BU005"; transcript_id "BU005-RA"; gene_name "atpH";
gn:buc	KEGG:GENES	CDS	4530	6068	+	geneid "BU006"; transcript_id "BU006-RA"; gene_name "atpA";
gn:buc	KEGG:GENES	CDS	6101	6973	+	geneid "BU007"; transcript_id "BU007-RA"; gene_name "atpG";
gn:buc	KEGG:GENES	CDS	6997	8394	+	geneid "BU008"; transcript_id "BU008-RA"; gene_name "atpD";
gn:buc	KEGG:GENES	CDS	8421	8837	+	geneid "BU009"; transcript_id "BU009-RA"; gene_name "atpC";
gn:buc	KEGG:GENES	CDS	8911	11322	-	geneid "BU010"; transcript_id "BU010-RA"; gene_name "gyrB";
gn:buc	KEGG:GENES	CDS	11449	12549	-	geneid "BU011"; transcript_id "BU011-RA"; gene_name "dnaN";
gn:buc	KEGG:GENES	CDS	12554	13918	-	geneid "BU012"; transcript_id "BU012-RA"; gene_name "dnaA";
gn:buc	KEGG:GENES	CDS	14369	14512	+	geneid "BU013"; transcript_id "BU013-RA"; gene_name "rpmH";
gn:buc	KEGG:GENES	CDS	14525	14872	+	geneid "BU014"; transcript_id "BU014-RA"; gene_name "rnpA";
gn:buc	KEGG:GENES	CDS	15011	16609	+	geneid "BU015"; transcript_id "BU015-RA"; gene_name "yidC";

buc.genome.fasta genome rem... 511M of 635M



ビューイングの拡大縮小

カーソルのgene modelに
持っていくことで、gtfファイル
の記述がpop upされる。

トラックの移動
マウスのドラッグ移動が可能
← →による移動も可能
Feature trackを指定して、
Ctrl+F gene model単位で右に移動
Ctrl+B gene model単位で左に移動
Shift+Ctrl+F exon単位で右に移動

シーケンスデータを読み込ませる
File -> Load from File

illumina_ex_B2_Read_bowtie2.mate.sort.bam

同ファイル名のindexファイル
illumina_ex_B2_Read_bowtie2.mate.sort.bam.bai
が同じフォルダーにないといけない。

2 tracks | gn:buc:177,977 | 312M of 646M

読み込まれた

もう一つ別の系統のシーケンスデータも読み込ませる
illumina_ex_B4_Read_bowtie2.mate.sort.bam

3 tracks loaded | gn:buc:189,361 | 476M of 646M

各Trackはドラッグアンドドロップで移動できる

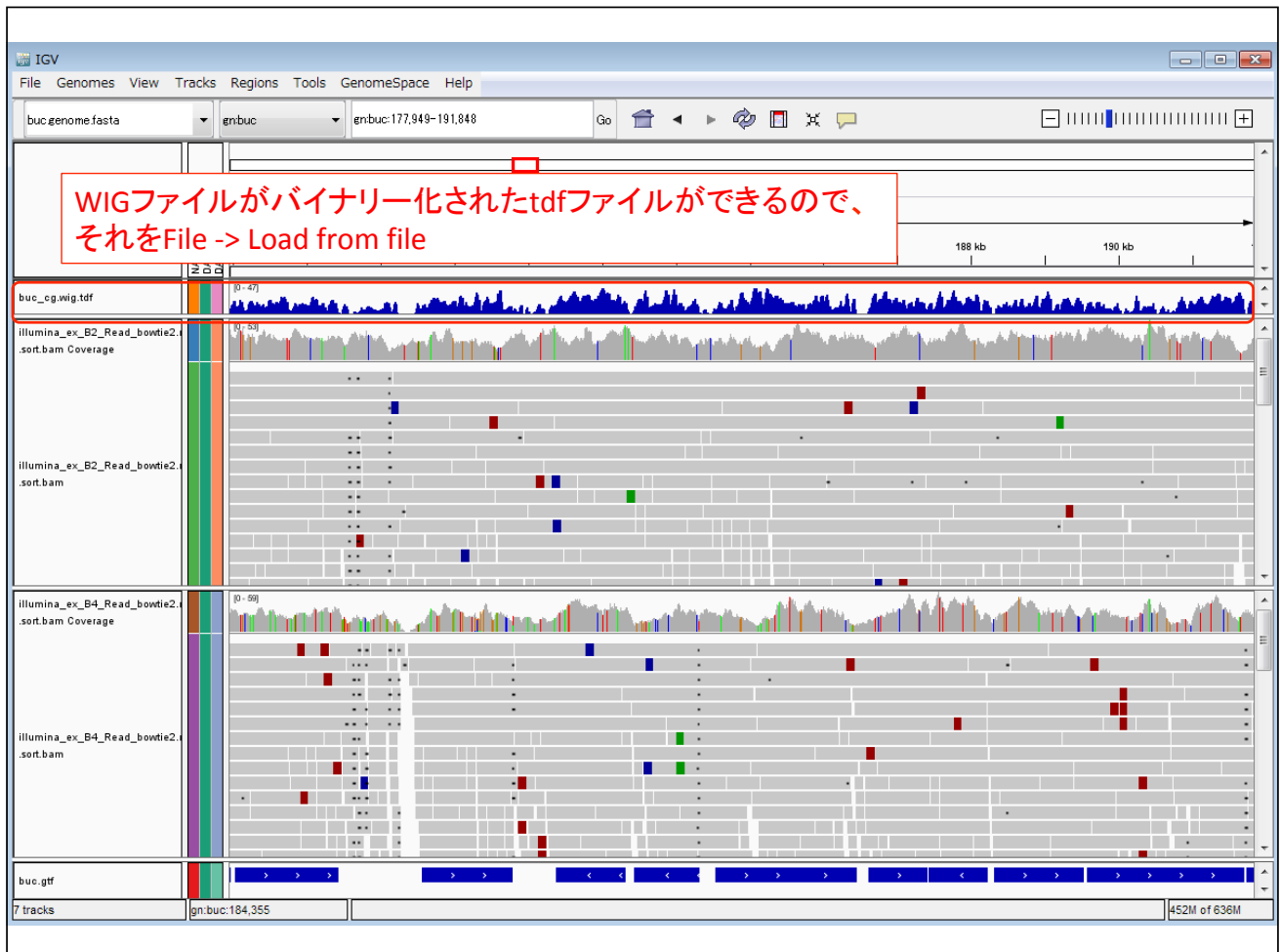
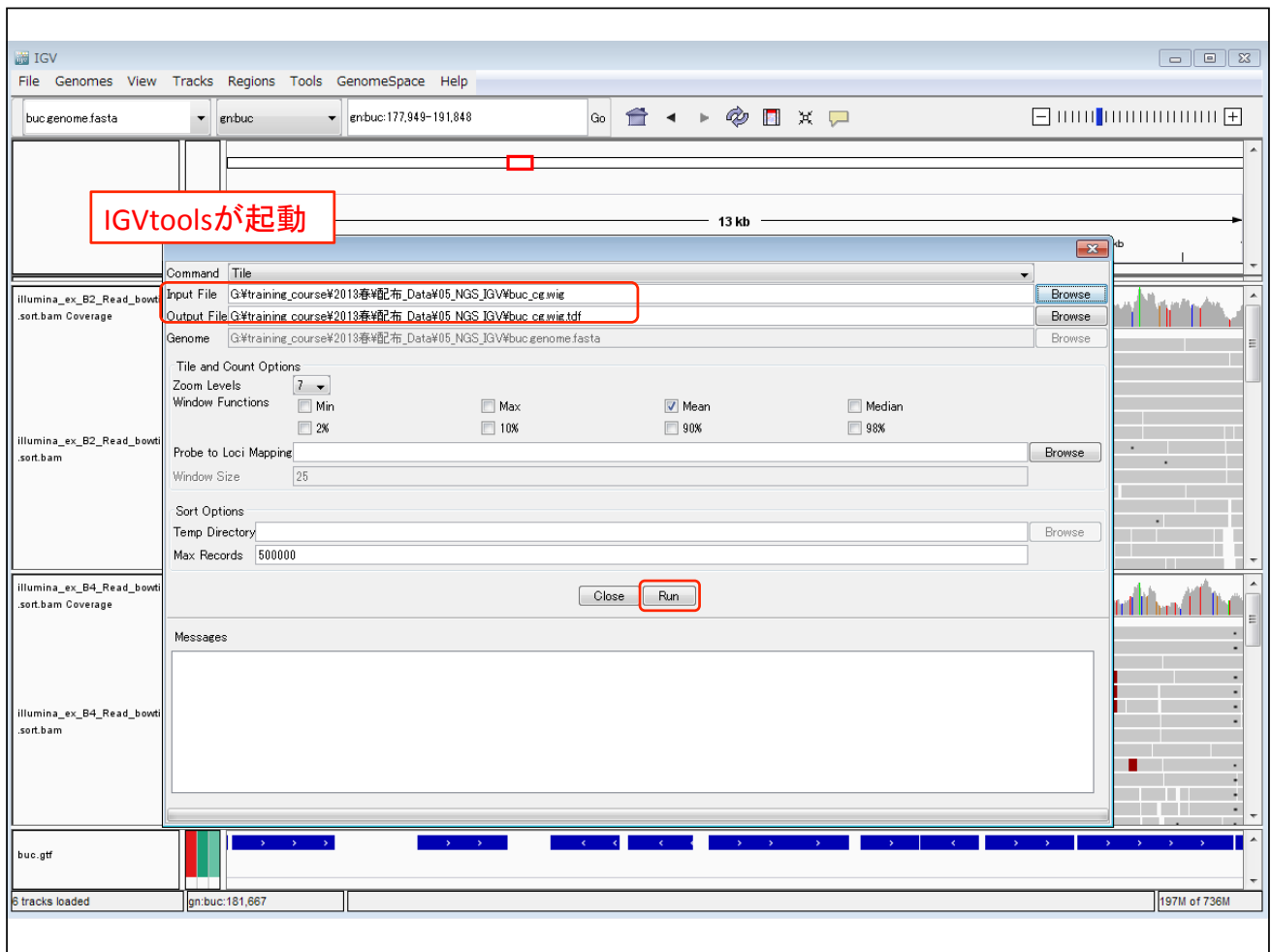
こういう所をクリックしたまま移動させることで幅を変更

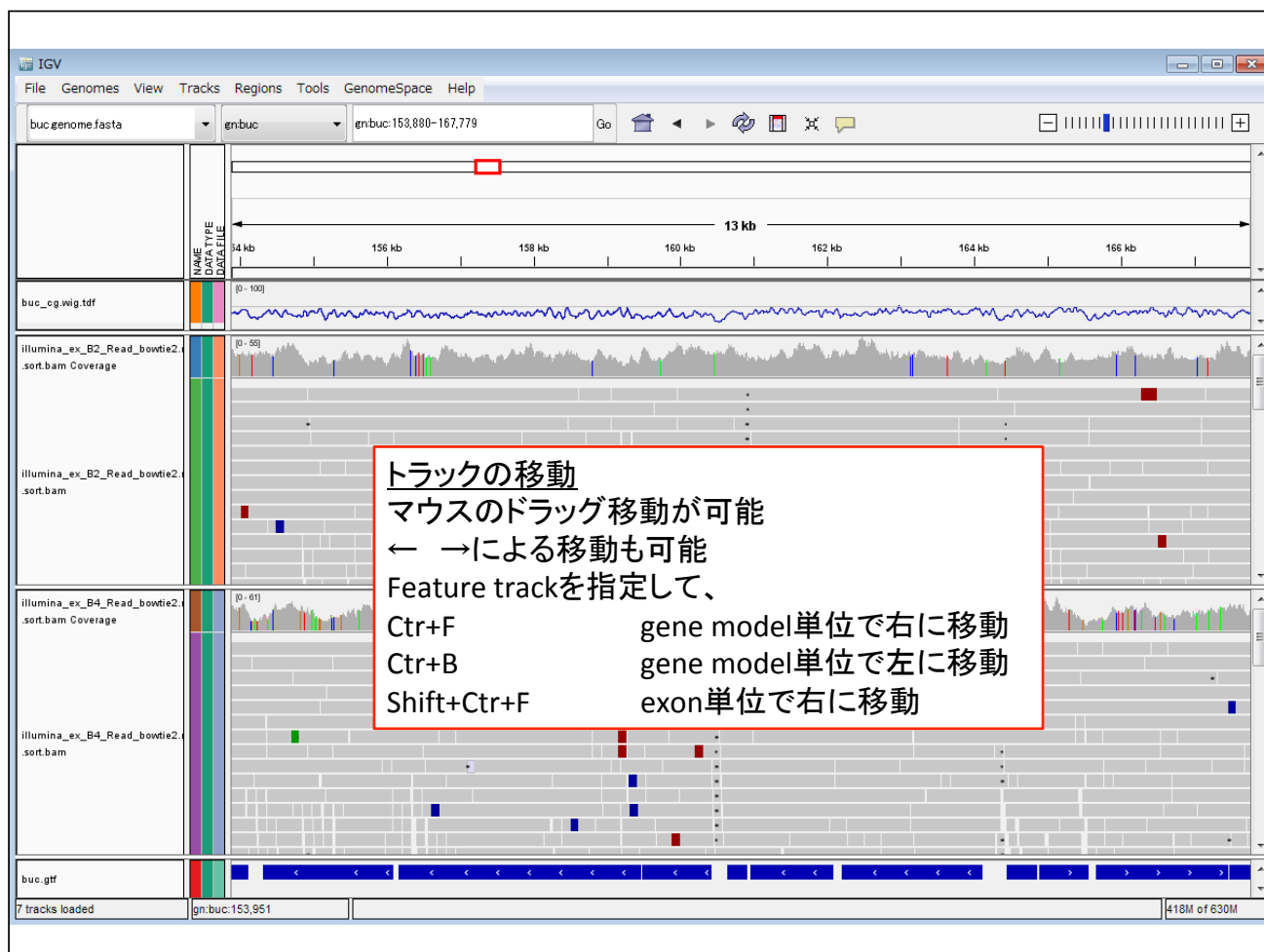
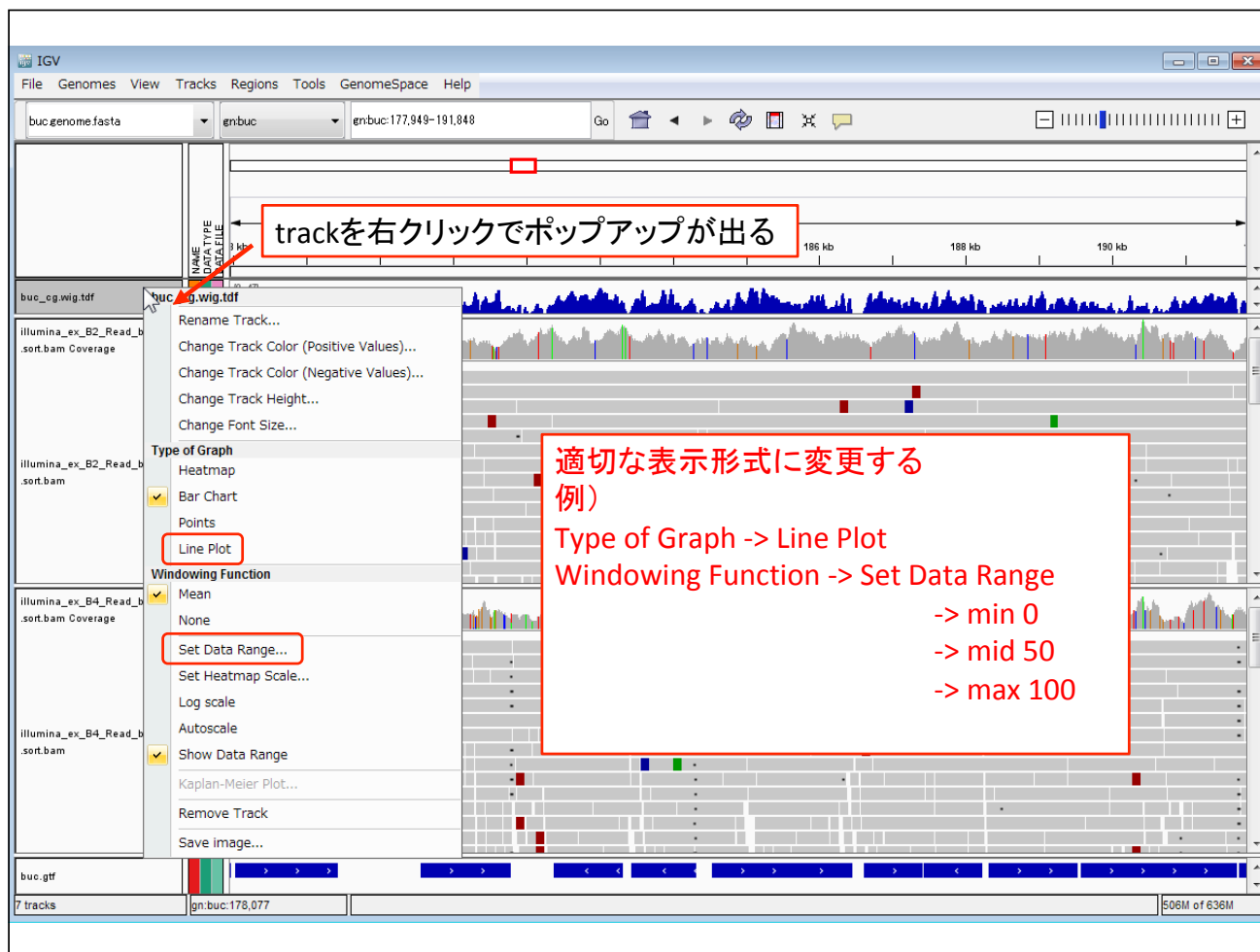
WIGファイルを読み込む

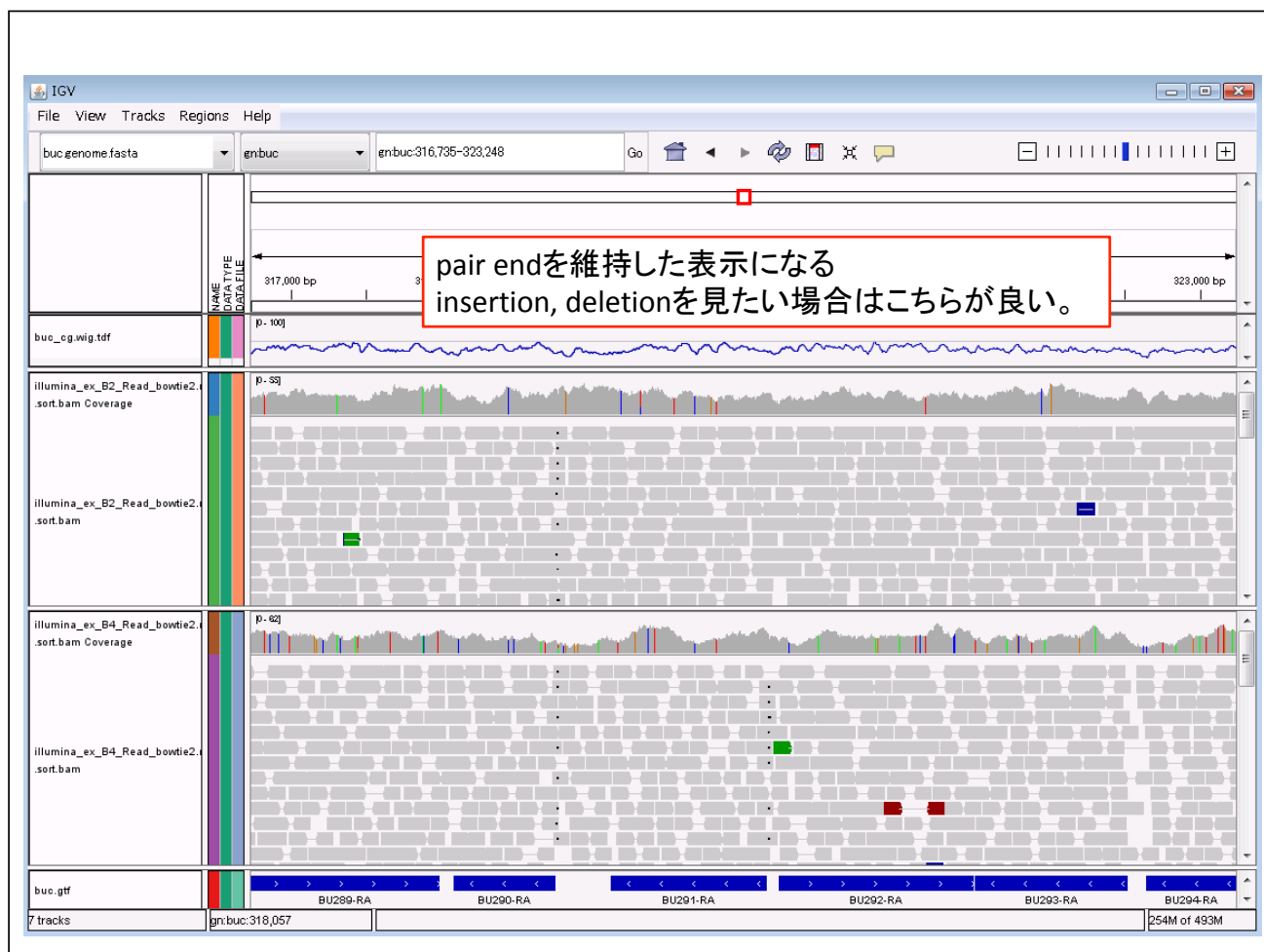
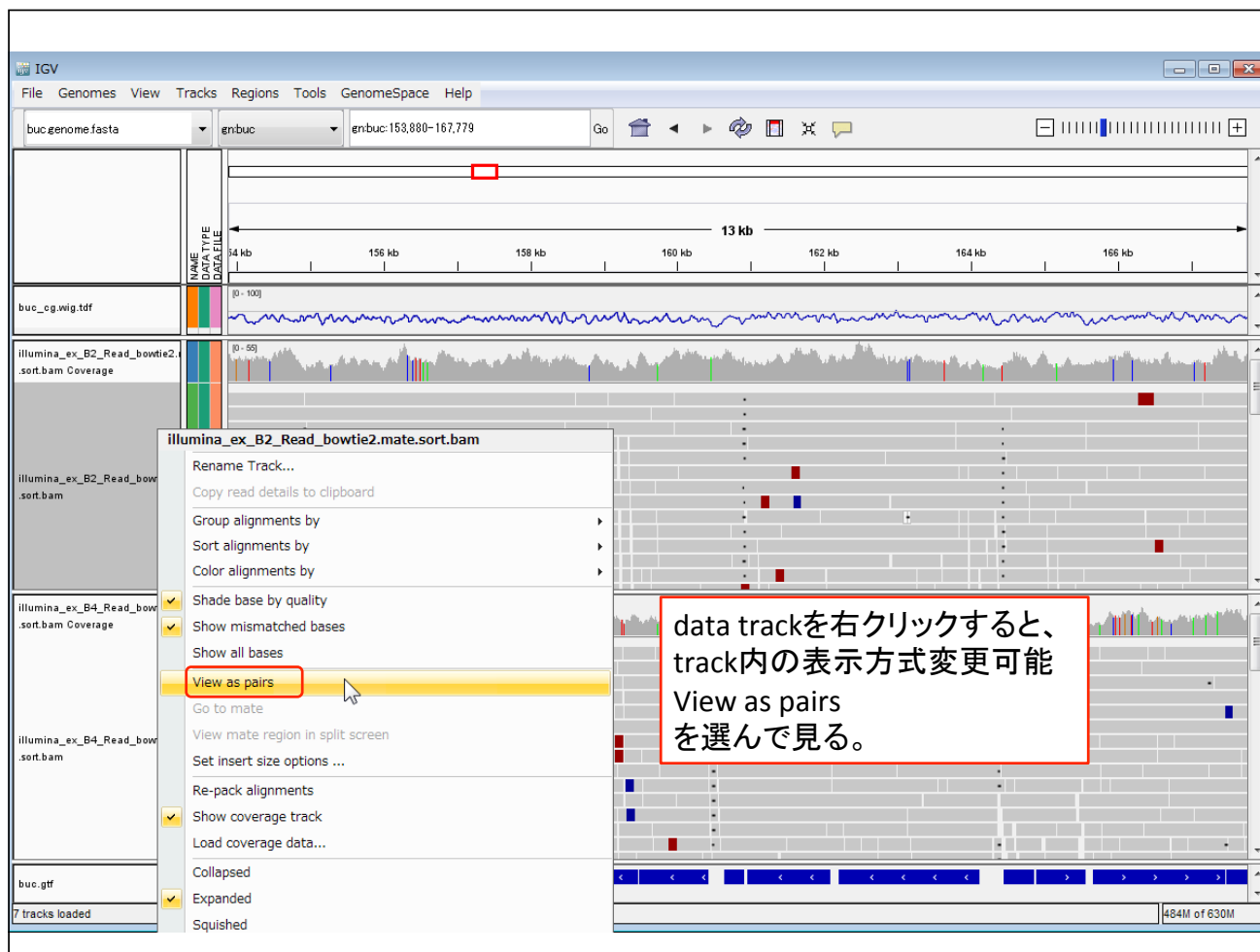
ウィンドウサイズ50baseでCG ratioを計算したWIGファイルを作成してある
buc_cg.wig

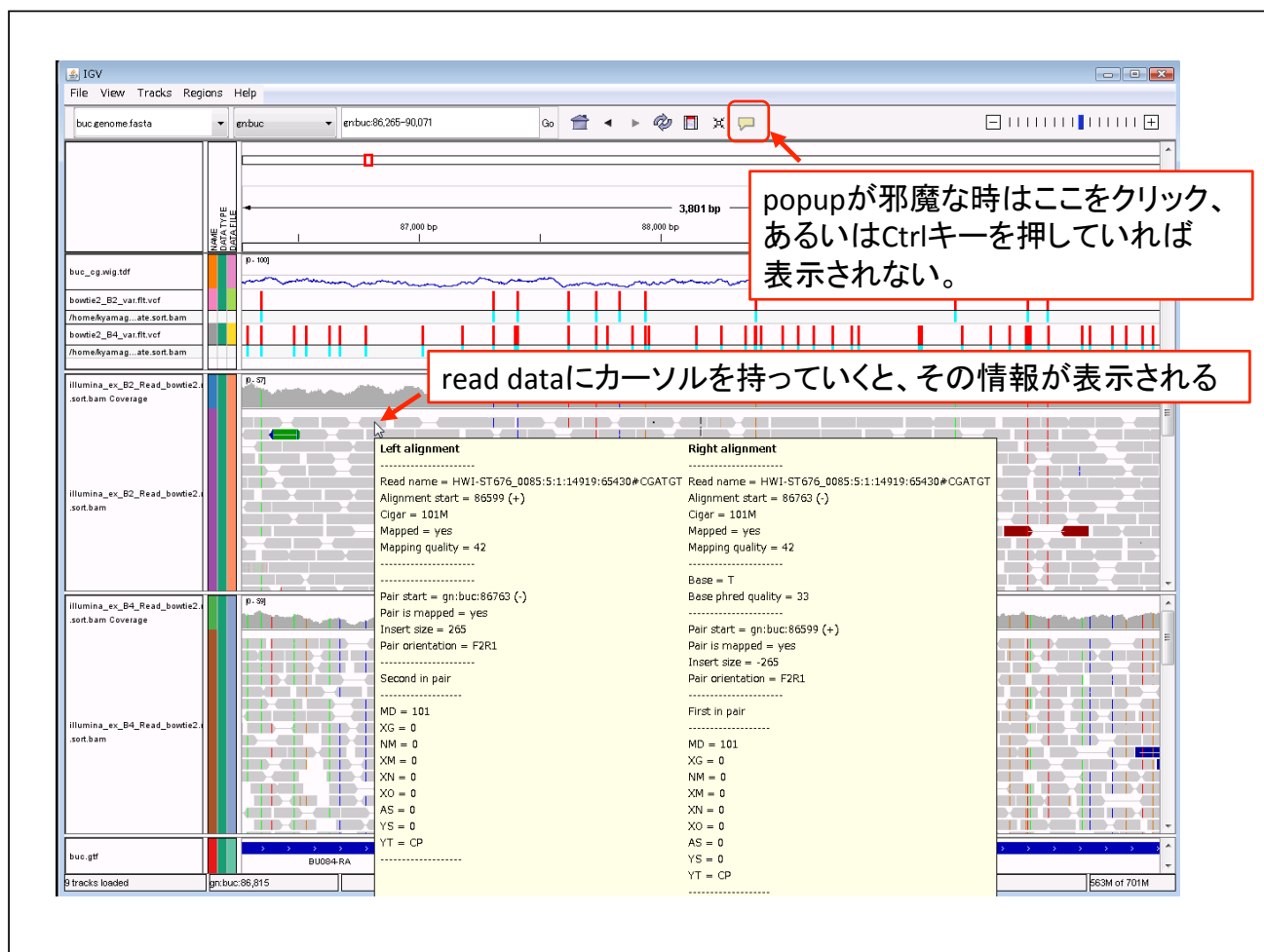
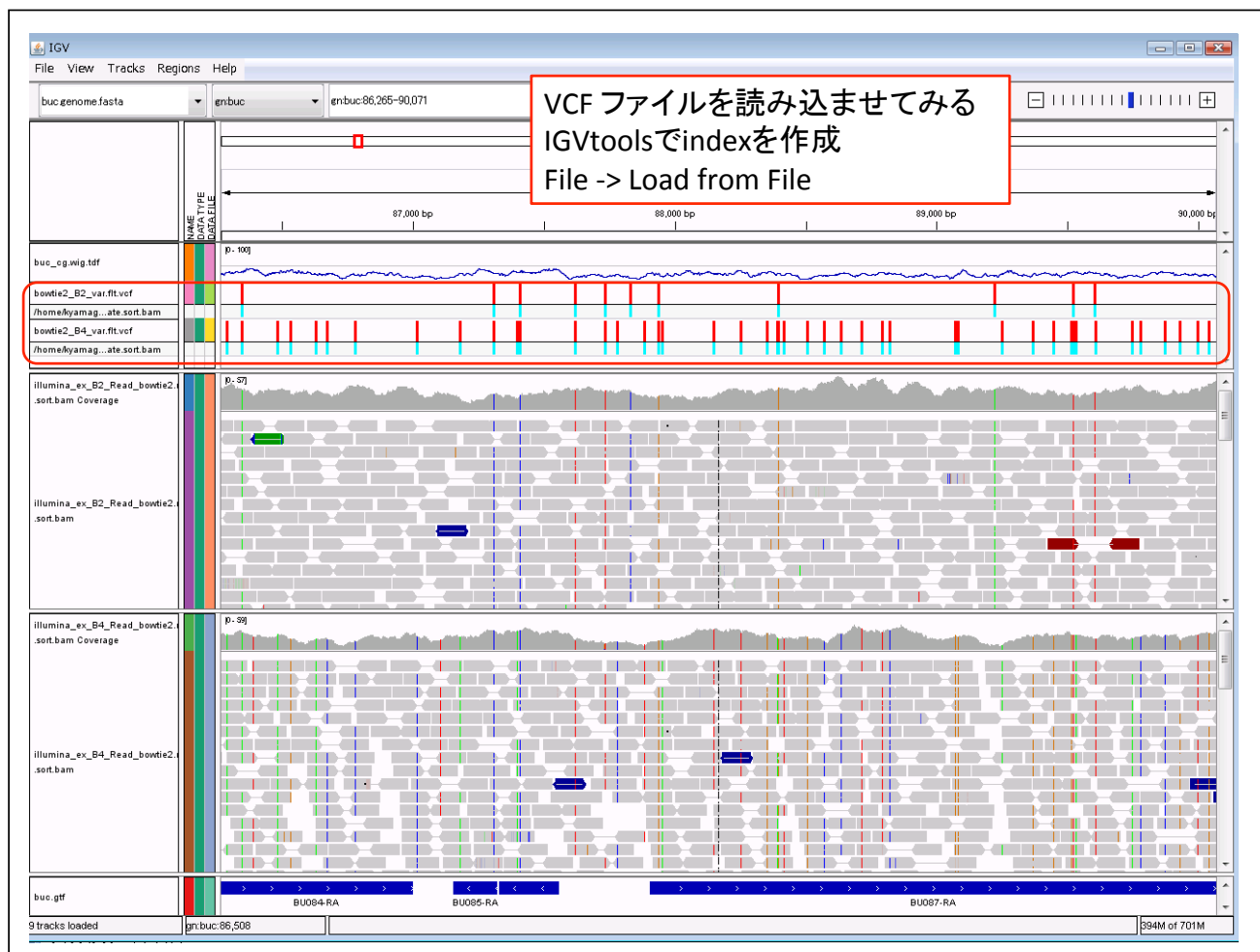
variableStep chrom=gn:buc	
25	16
26	16
27	16
28	16
29	16
30	14
31	12
32	12
33	10
34	10
35	8
36	8
37	8
38	8

File -> Load from File でそのまま読み込ませることも可能だが、
テキストファイルで膨大なサイズだと非常に時間がかかる。
Tools -> Run igvtools









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
- ・自分で見るとにもよし、人に見せるためにもよし
- ・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。