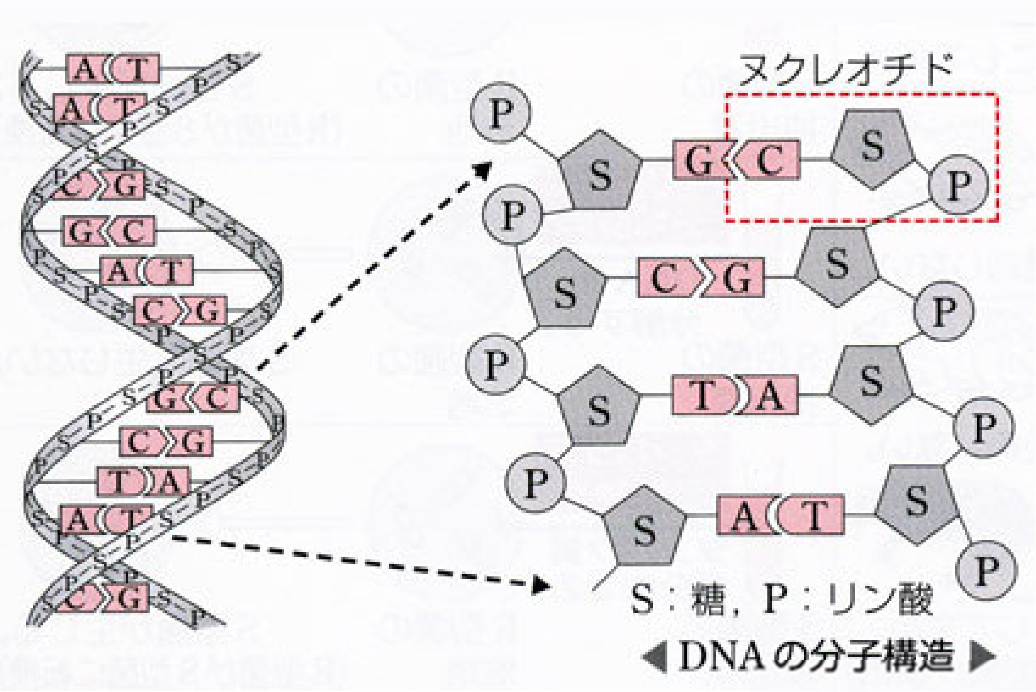
epigenome解析

東京科学大学　ILA国府台
中林　潤

JH人材育成課　バイオインフォマティシャン育成講座　⑦

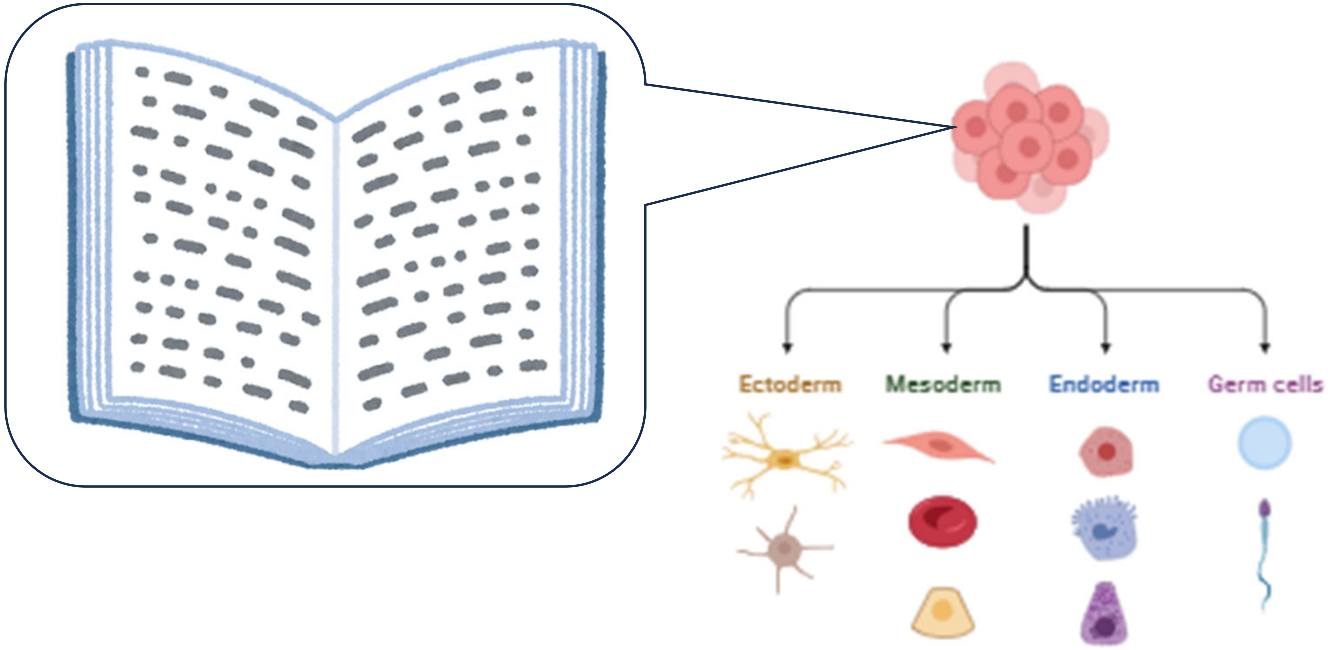
epigenomeとは？

- gene：遺伝子
- genome：遺伝子の全体
- epigenome
  epi-：上の、超えて、間に、後のなどの意味を表す接頭辞
  遺伝子の塩基配列以外の要素



◀DNAの分子構造▶

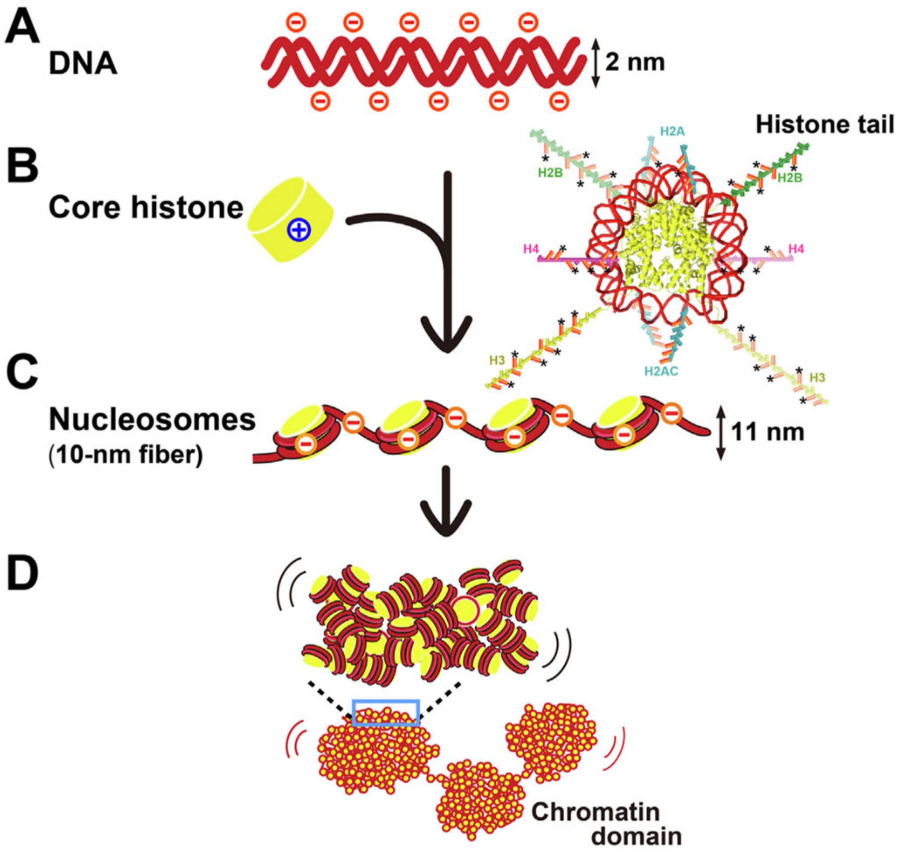
epigenomeとは？



同じ遺伝子を共有する細胞から
様々な細胞が分化する。
↓
遺伝子の使い分け

- 遺伝情報＝塩基配列：文字
- epigenome：何ページの何行目をいつ読むか

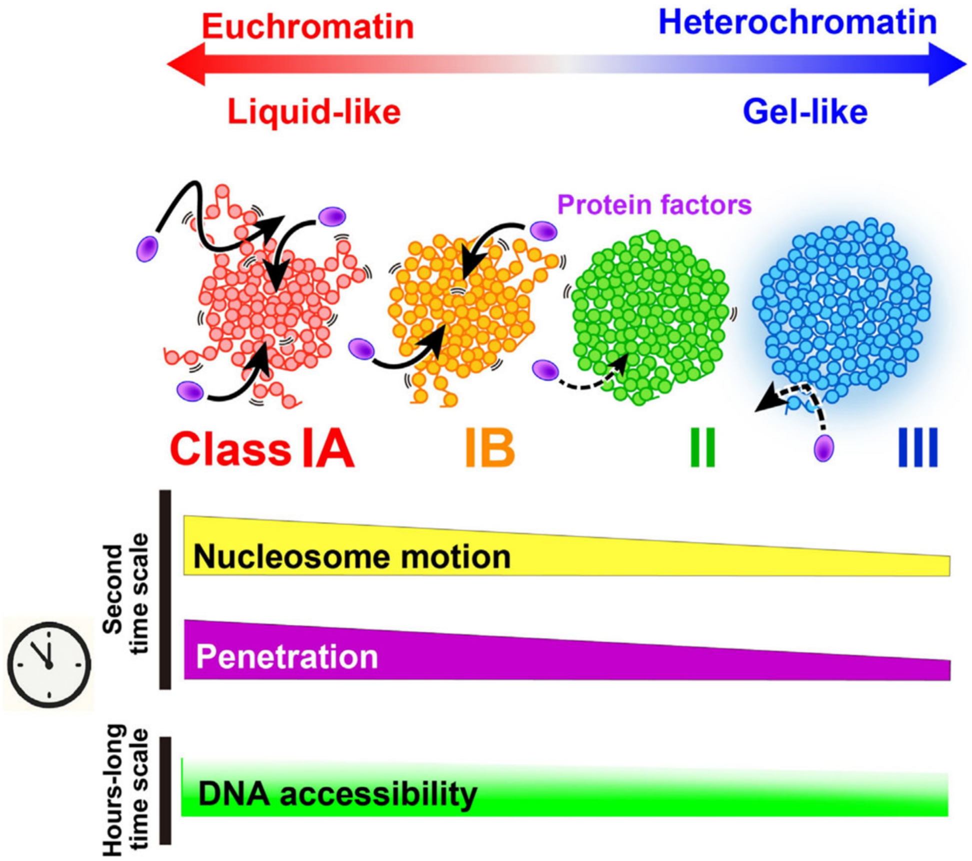
chromatin構造



- マイナスにチャージしているDNAがプラスにチャージしているヒストンに巻き付き、chromatinが形成される。
- 複数のchromatinが集まってさらに高次の構造が形成される。

Euchromatin and Heterochromatin : Implications for DNA accessibility and transcription
Minami K, Semeigazin A, Nakazato K, Maeshima K. ***J Mol Biol.*** 2026 438:169270

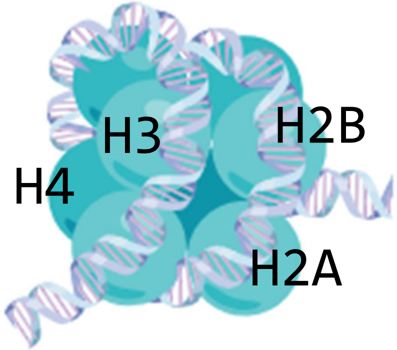
EuchromatinとHeterochromatin



chromatinの状態によってDNAへのアクセスのしやすさが変わる。

Euchromatin and Heterochromatin : Implications for DNA accessibility and transcription
Minami K, Semeigazin A, Nakazato K, Maeshima K. ***J Mol Biol.*** 2026 438:169270

構造と機能の連関



chromatin
ヒストンの8量体に
DNAが巻き付いた構造

ヒストン修飾：DNAの状態を示す
H3K4me1：primed enhancer
H3K4me3：active promoter
H3K27ac：super enhancer

転写領域：DNAがほどける

nucleosome
複数のchromatinにより形成される
高次構造の基本単位

Created in BioRender.com bio

chromatinの状態は遺伝子の状態を知る手掛かりとなる

epigenome解析

- 遺伝子の状態をゲノムワイドに解析する。
- DNA断片の配列が遺伝子のどの部分か解析する。
  Peak call、Annotation、Tag count
- 塩基配列自体が解析の対象ではない。

epigenome解析

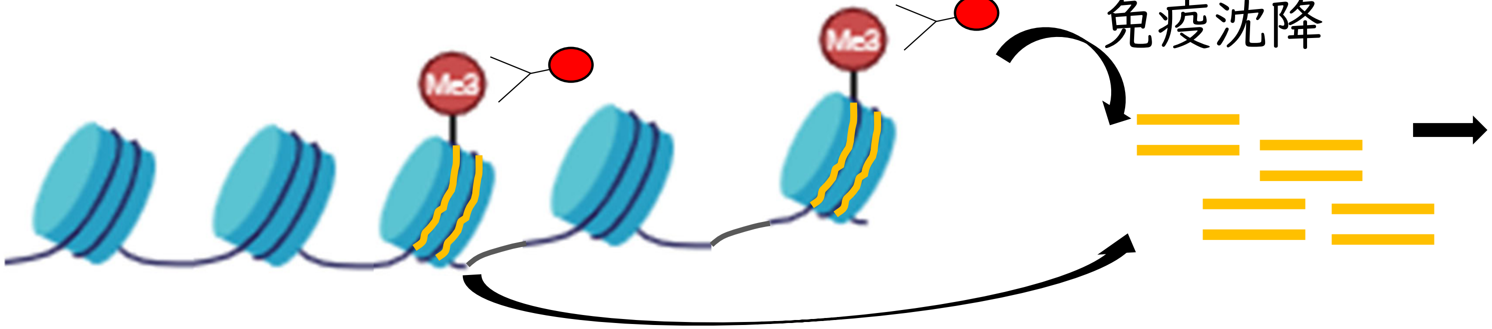
- ChIP-seq
- ATAC-seq
- Hi-C

ChIP-seq

- ヒストン修飾部位をゲノムワイドに探索する手法
- ヒストン修飾の特異的抗体を用いた免疫沈降

ヒストン修飾

- H3K4me1
- H3K4me3
- H3K27ac



抗体

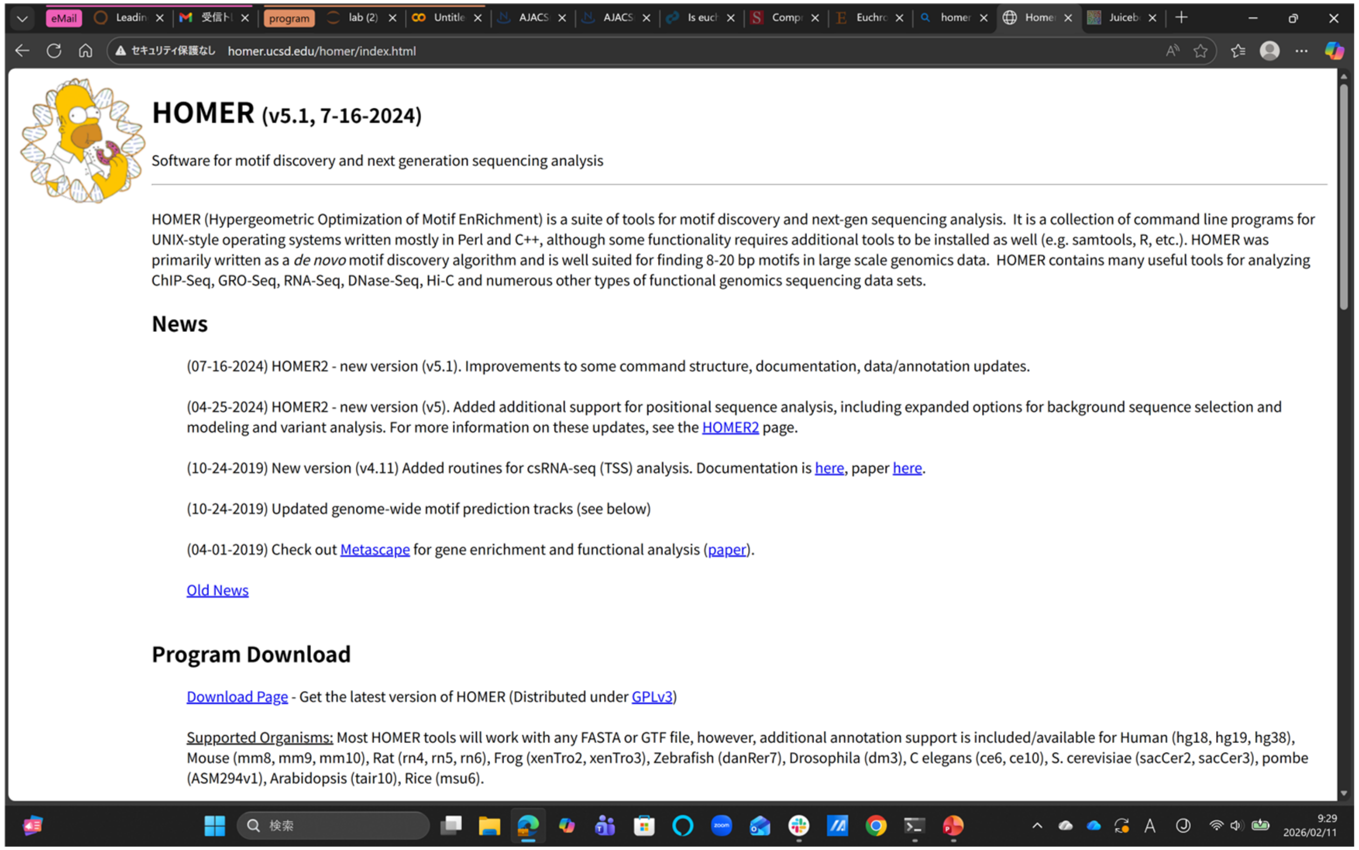
免疫沈降

次世代シーケンサーでシーケンス

⬇

mapping
Peak call
Annotation
Tag count

HOMER



HOMER (v5.1, 7-16-2024)

Software for motif discovery and next generation sequencing analysis

HOMER (Hypergeometric Optimization of Motif EnRichment) is a suite of tools for motif discovery and next-gen sequencing analysis. It is a collection of command line programs for UNIX-style operating systems written mostly in Perl and C++, although some functionality requires additional tools to be installed as well (e.g. samtools, R, etc.). HOMER was primarily written as a *de novo* motif discovery algorithm and is well suited for finding 8-20 bp motifs in large scale genomics data. HOMER contains many useful tools for analyzing ChIP-Seq, GRO-Seq, RNA-Seq, DNase-Seq, Hi-C and numerous other types of functional genomics sequencing data sets.

News

(07-16-2024) HOMER2 - new version (v5.1). Improvements to some command structure, documentation, data/annotation updates.

(04-25-2024) HOMER2 - new version (v5). Added additional support for positional sequence analysis, including expanded options for background sequence selection and modeling and variant analysis. For more information on these updates, see the HOMER2 page.

(10-24-2019) New version (v4.11) Added routines for csRNA-seq (TSS) analysis. Documentation is here, paper here.

(10-24-2019) Updated genome-wide motif prediction tracks (see below)

(04-01-2019) Check out Metascape for gene enrichment and functional analysis (paper).

Old News

Program Download

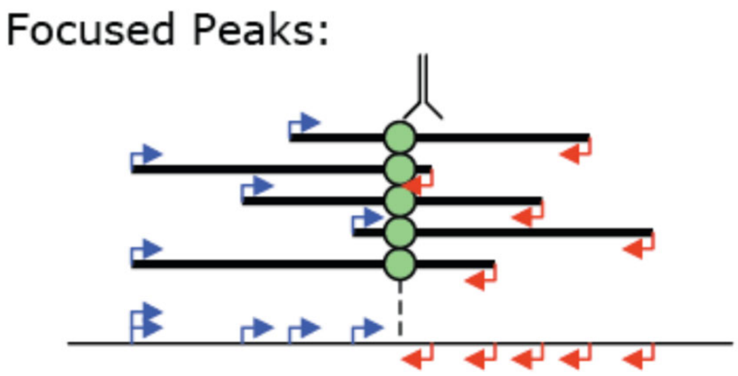
Download Page - Get the latest version of HOMER (Distributed under GPLv3)

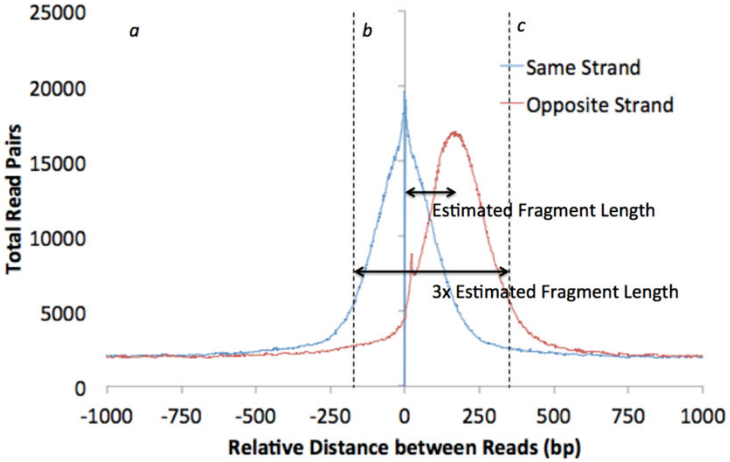
Supported Organisms: Most HOMER tools will work with any FASTA or GTF file, however, additional annotation support is included/available for Human (hg18, hg19, hg38), Mouse (mm8, mm9, mm10), Rat (rn4, rn5, rn6), Frog (xenTro2, xenTro3), Zebrafish (danRer7), Drosophila (dm3), C elegans (ce6, ce10), S. cerevisiae (sacCer2, sacCer3), pombe (ASM294v1), Arabidopsis (tair10), Rice (msu6).

- Peak call
- Annotation
- motif検索

http://homer.ucsd.edu/homer/index.html

Peak call





Same strand and Diff Strand Fold Enrichment: $\frac{\text{Density } b}{\text{Density } a \text{ \& } c}$

Same/Diff Fold Enrichment: $\frac{\text{Density } b \text{ (same strand)}}{\text{Density } b \text{ (Opposite Strand)}}$

抗体の場所を中心として二峰性の分布を示す。
↓
タグカウントと分布からピークを同定する。

BEDファイルフォーマット

- ゲノムの位置情報

  染色体番号　開始点　終点

chr1	4496414	4496640
chr1	4748129	4748571
chr1	4768512	4768805

UCSC genome browserにアップロードして表示できる

UCSC Genome Browser

Tools

- **Genome Browser** - Interactively visualize genomic data
- **BLAT** - Rapidly align sequences to the genome
- **In-Silico PCR** - Rapidly align PCR primer pairs to the genome
- **Table Browser** - Download and filter data from the Genome Browser
- **LiftOver** - Convert genome coordinates between assemblies
- **REST API** - Returns data requested in JSON format
- **Variant Annotation Integrator** - Annotate genomic variants
- **More tools...**

News

July 23, 2025 - **Bionano DLE-1 CTTAAG sites for human, hg38 and hg19**

July 15, 2025 - **New pathogenicity predictor tracks MCAP and MutScore**

July 15, 2025 - **ENCODE4 Long-read Transcripts for hg38/mm10**

July 9, 2025 - **Unusually Conserved Regions track for hg38**

July 7, 2025 - **UCSC Genome Browser turns 25**

June 26, 2025 - **New Genome Browser tutorials**

Meetings and Workshops: Come see us in person!

- **McKusick Short Course in Mammalian and Medical Genetics** -- Bar Harbor, ME. July 22, 2025
- **Variant Effect Prediction Training Course (VEPTC)** -- Porto, Portugal. September 28 - October 1, 2025
- **ASHG Pre-Conference Course** -- Boston, MA. Oct 13-14, 2025

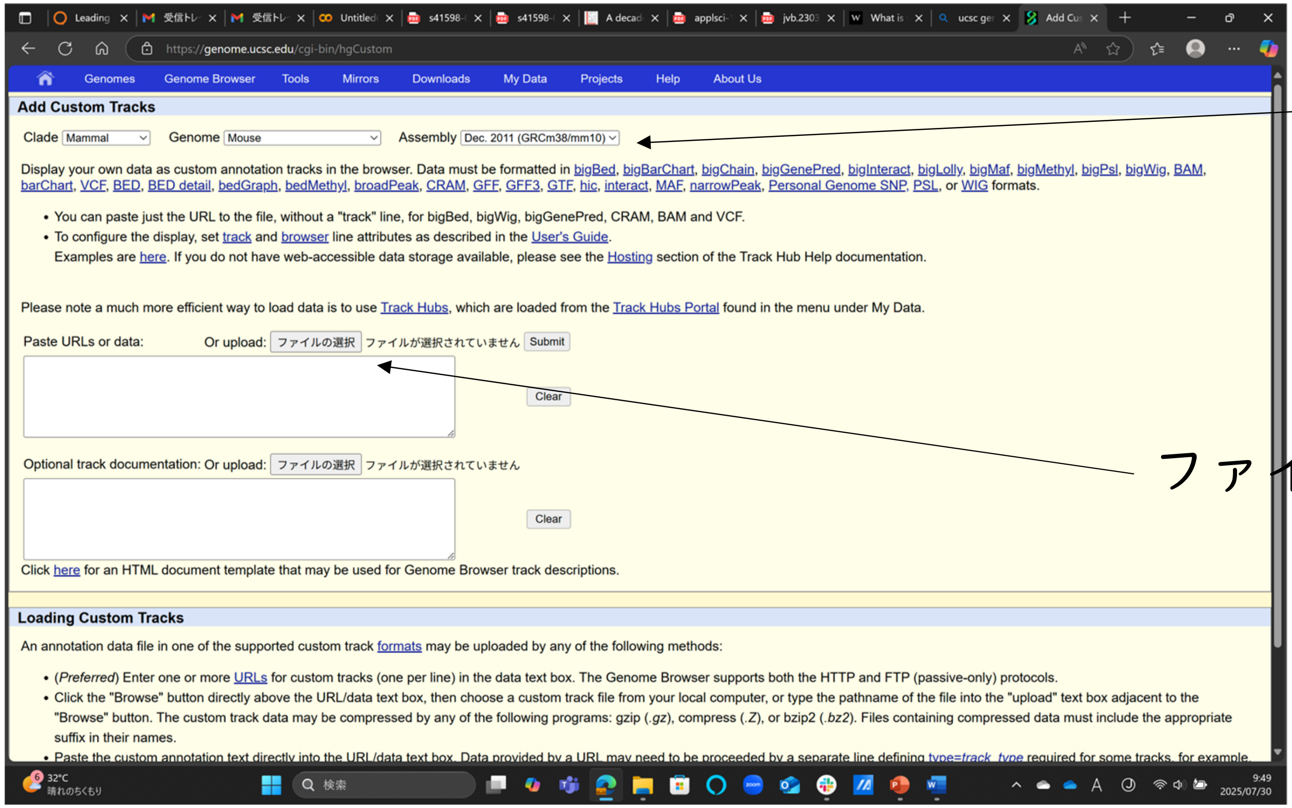
Feel free to contact us if you are interested in attending a workshop, or meeting someone from the team to collaborate, get help, or ask any questions at the meetings.

Sharing data

Learning

https://genome.ucsc.edu/

Custom Track



Add Custom Tracks

Clade Mammal Genome Mouse Assembly Dec. 2011 (GRCm38/mm10)

Display your own data as custom annotation tracks in the browser. Data must be formatted in bigBed, bigBarChart, bigChain, bigGenePred, bigInteract, bigLolly, bigMaf, bigMethyl, bigPsl, bigWig, BAM, barChart, VCF, BED, BED detail, bedGraph, bedMethyl, broadPeak, CRAM, GFF, GFF3, GTF, hic, interact, MAF, narrowPeak, Personal Genome SNP, PSL, or WIG formats.

- You can paste just the URL to the file, without a "track" line, for bigBed, bigWig, bigGenePred, CRAM, BAM and VCF.
- To configure the display, set track and browser line attributes as described in the User's Guide.
  Examples are here. If you do not have web-accessible data storage available, please see the Hosting section of the Track Hub Help documentation.

Please note a much more efficient way to load data is to use Track Hubs, which are loaded from the Track Hubs Portal found in the menu under My Data.

Paste URLs or data: Or upload: ファイルの選択 ファイルが選択されていません Submit

Clear

Optional track documentation: Or upload: ファイルの選択 ファイルが選択されていません

Clear

Click here for an HTML document template that may be used for Genome Browser track descriptions.

Loading Custom Tracks

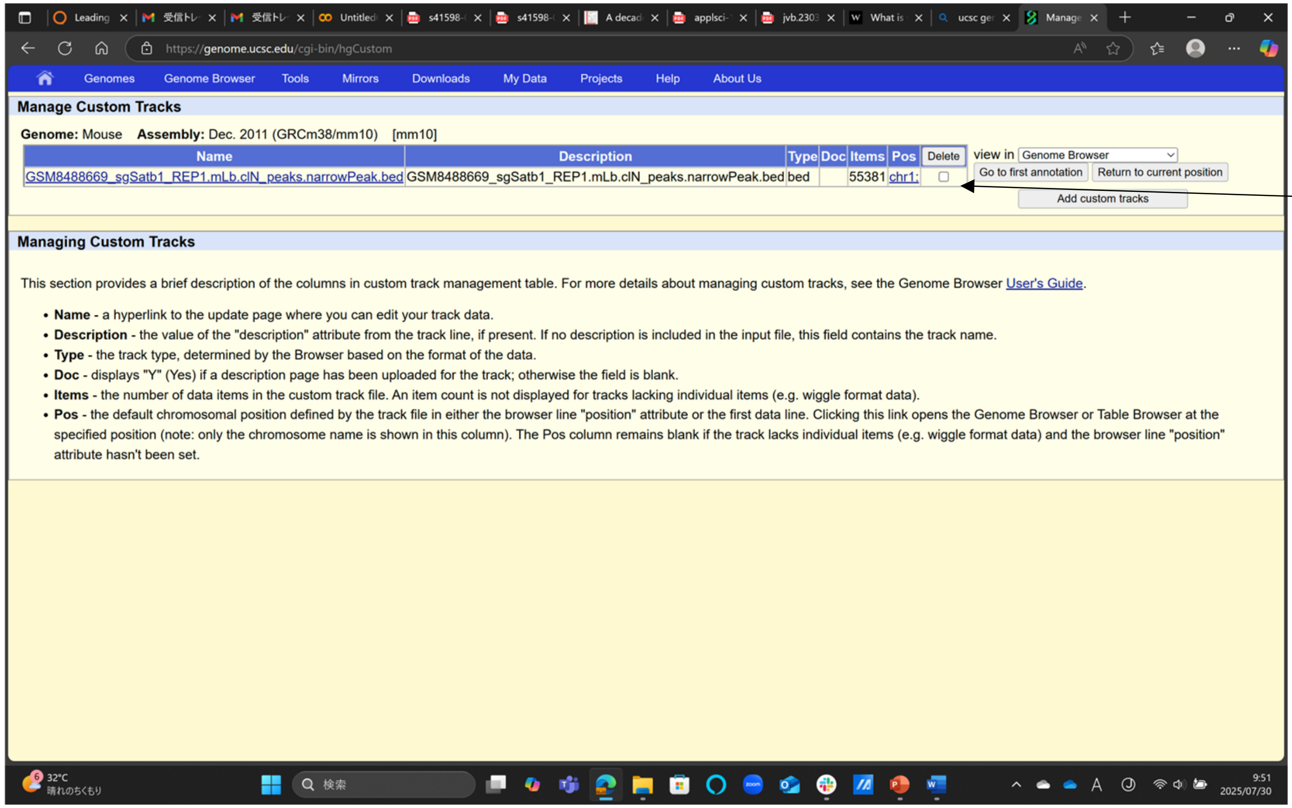
An annotation data file in one of the supported custom track formats may be uploaded by any of the following methods:

- (*Preferred*) Enter one or more URLs for custom tracks (one per line) in the data text box. The Genome Browser supports both the HTTP and FTP (passive-only) protocols.
- Click the "Browse" button directly above the URL/data text box, then choose a custom track file from your local computer, or type the pathname of the file into the "upload" text box adjacent to the "Browse" button. The custom track data may be compressed by any of the following programs: gzip (*.gz*), compress (*.Z*), or bzip2 (*.bz2*). Files containing compressed data must include the appropriate suffix in their names.
- Paste the custom annotation text directly into the URL/data text box. Data provided by a URL may need to be proceeded by a separate line defining *type=track_type* required for some tracks, for example

参照ゲノムを選択

ファイルを選択してsubmit

Custom Track



Manage Custom Tracks

Genome: Mouse **Assembly:** Dec. 2011 (GRCm38/mm10) [mm10]

Name	Description	Type	Doc	Items	Pos	Delete
GSM8488669_sgSatb1_REP1.mLb.clN_peaks.narrowPeak.bed	GSM8488669_sgSatb1_REP1.mLb.clN_peaks.narrowPeak.bed	bed		55381	chr1:	☐

Managing Custom Tracks

This section provides a brief description of the columns in custom track management table. For more details about managing custom tracks, see the Genome Browser User's Guide.

- **Name** - a hyperlink to the update page where you can edit your track data.
- **Description** - the value of the "description" attribute from the track line, if present. If no description is included in the input file, this field contains the track name.
- **Type** - the track type, determined by the Browser based on the format of the data.
- **Doc** - displays "Y" (Yes) if a description page has been uploaded for the track; otherwise the field is blank.
- **Items** - the number of data items in the custom track file. An item count is not displayed for tracks lacking individual items (e.g. wiggle format data).
- **Pos** - the default chromosomal position defined by the track file in either the browser line "position" attribute or the first data line. Clicking this link opens the Genome Browser or Table Browser at the specified position (note: only the chromosome name is shown in this column). The Pos column remains blank if the track lacks individual items (e.g. wiggle format data) and the browser line "position" attribute hasn't been set.

アップロードした
ファイルが表示さ
れる

Custom Track

アップロードしたBEDファイルに記載された位置が黒く表示される。

https://genome.ucsc.edu/cgi-bin/hgTracks?db=mm10&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&posit...

UCSC Genome Browser on Mouse (GRCm38/mm10)

New to the Genome Browser? See our short (2-3 minute) guided tutorial. All tutorials can be found in the top blue bar menu under **Help** > **Interactive Tutorial**.

chr12:56,694,976-56,714,605 19,630 bp.

gene, chromosome range, search terms, help pages, see exam

GSM8488669_sgSatb1_REP1.mLb.clN_peaks.narrowPeak.bed

All gaps of unknown nucleotides (N's), including AGP annotated gaps

GENCODE VM23 Comprehensive Transcript Set (only Basic displayed by default)

RefSeq genes from NCBI

Mouse ESTs That Have Been Spliced

Single Cell RNA-Seq Gene Expression from Tabula Muris

ENCODE Candidate Cis-Regulatory Elements (cCREs) combined from all cell types

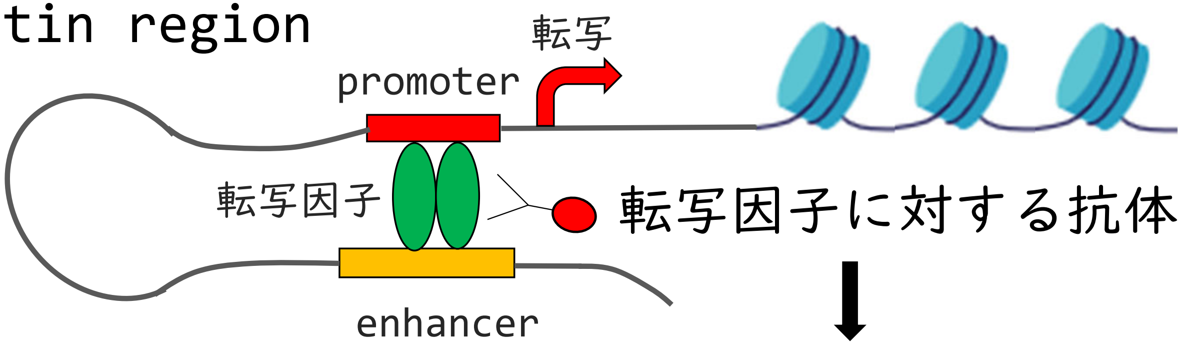
Placental Mammal Basewise Conservation by PhyloP

Multiz Alignments of 60 Vertebrates

https://genome.ucsc.edu/cgi-bin/hgTracks?db=mm10&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr12%3A56694976-56714605&hgsid=2841573970_tg9KeEWD2ty1Gz7M7ZT7agfRLHeU

転写因子抗体でChIP-seq

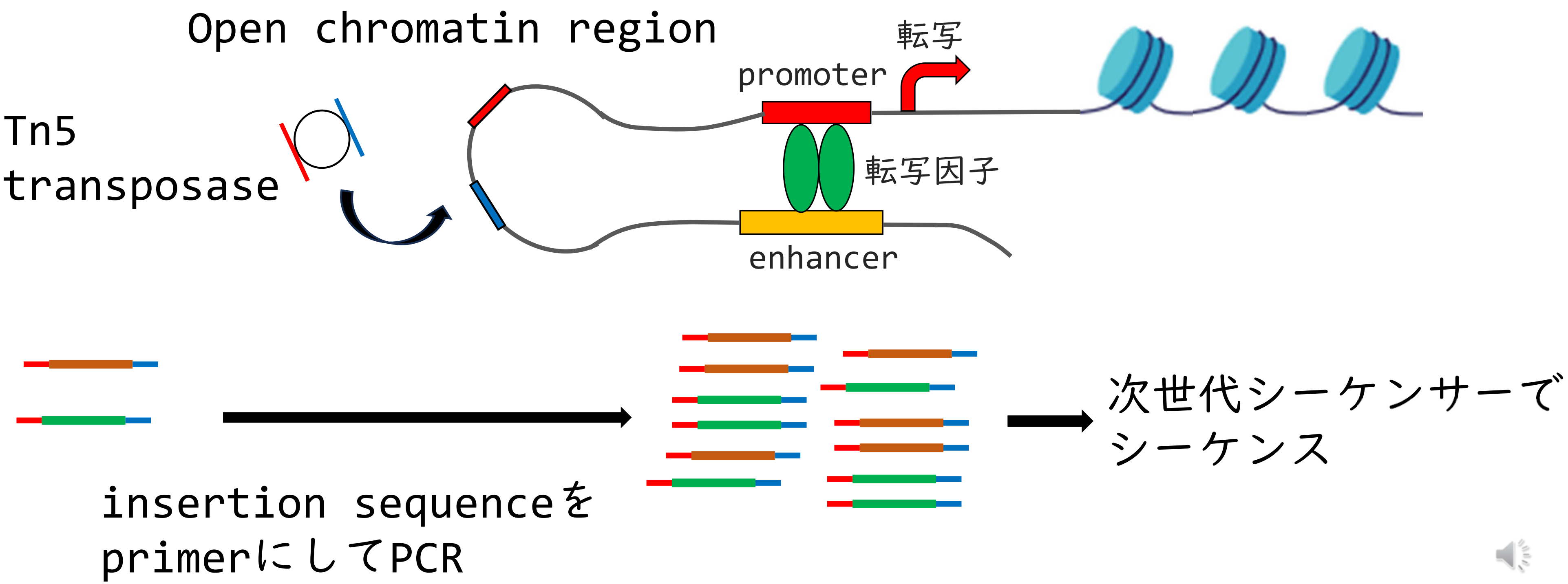
Open chromatin region



ChIP-seq
転写因子結合領域をゲノムワイドに探索できる。

ATAC-seq

- Open Chromatin領域をゲノムワイドに検索する手法
- transposaseがopen chromatin領域にアクセスできることを利用



Open chromatin region

Tn5
transposase

転写

promoter

転写因子

enhancer

insertion sequenceを
primerにしてPCR

次世代シーケンサーで
シーケンス

Hi-C

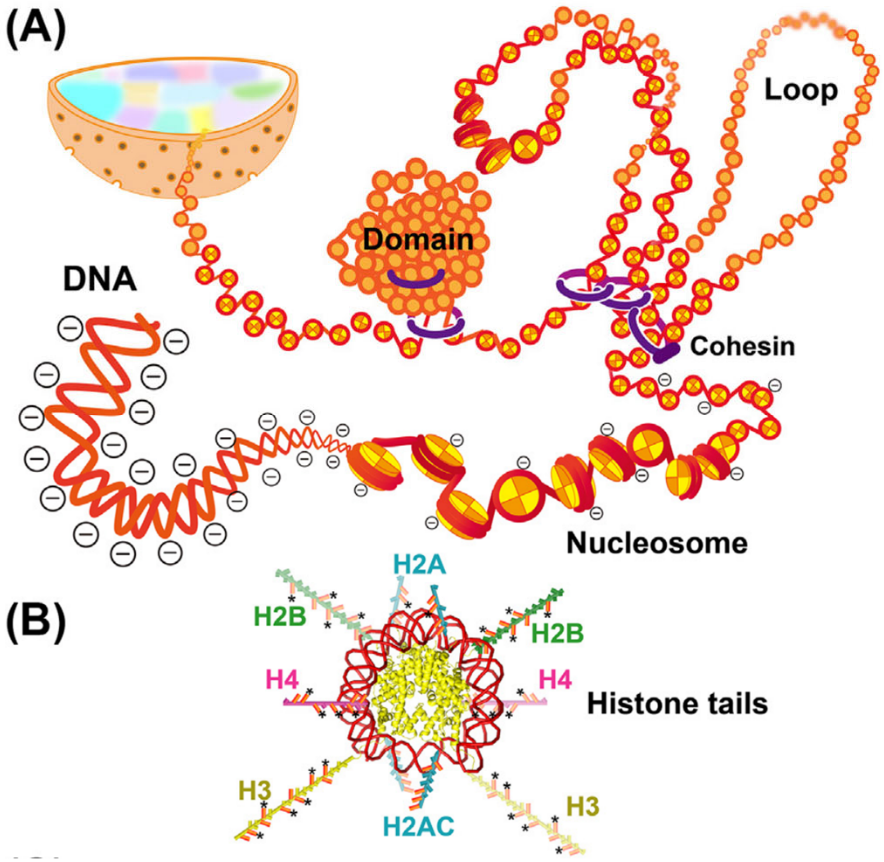
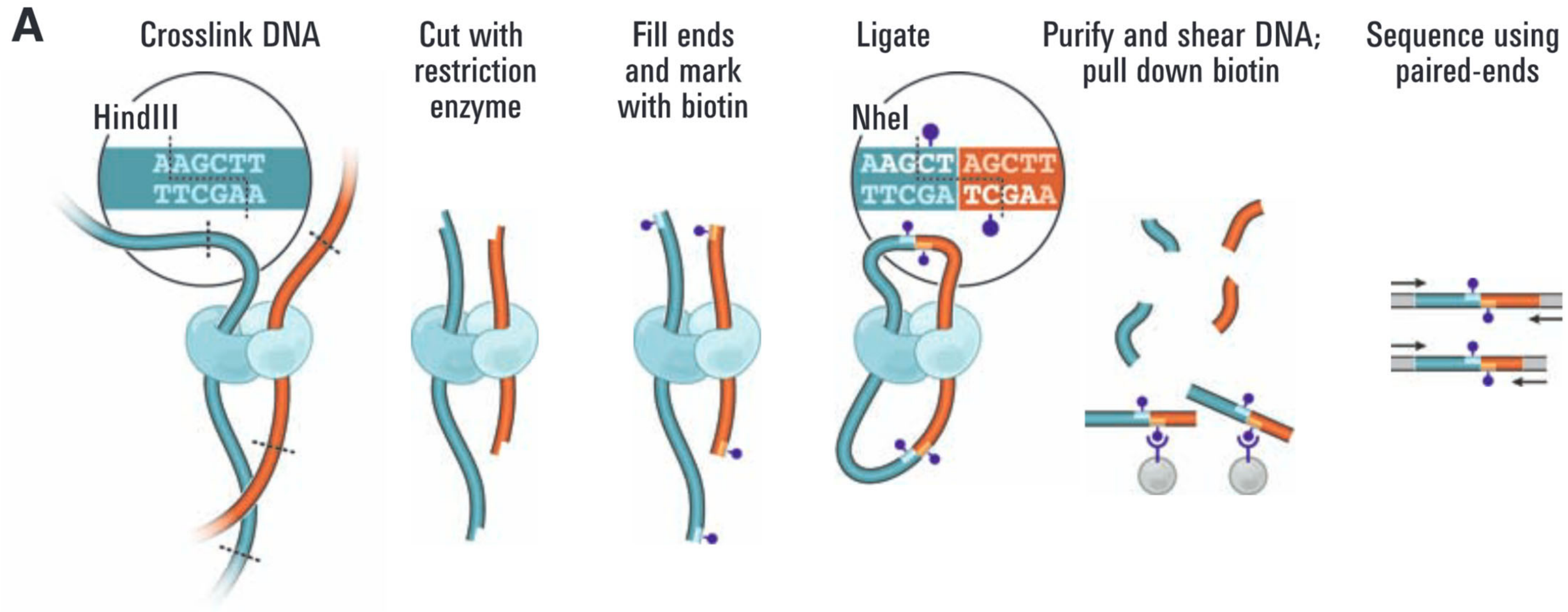


Figure 1. A scheme for hierarchical chromatin organization. (A) Negatively charged DNA is wrapped around a positively charged histone octamer to make a nucleosome. The chain of nucleosomes can form a loop structure held by cohesin or other proteins, and is compacted into chromatin domains [e.g., topologically associating domain (TAD)/contact domain/loop domain]. The number of cohesin bound to the domain may depend on the number of CTCF binding sites at the base of the loop. (B) Nucleosome core structure and histone tails extending away from the nucleosome core. H2A also has a carboxy-terminal tail. Basic residues (lysines and arginines) in the tail domains are colored in orange. Asterisks indicate sites of lysine acetylation. (Illustrations are based on data in [88,89] and reproduced from [90] with permission.) (C) A common model of euchromatin and heterochromatin.

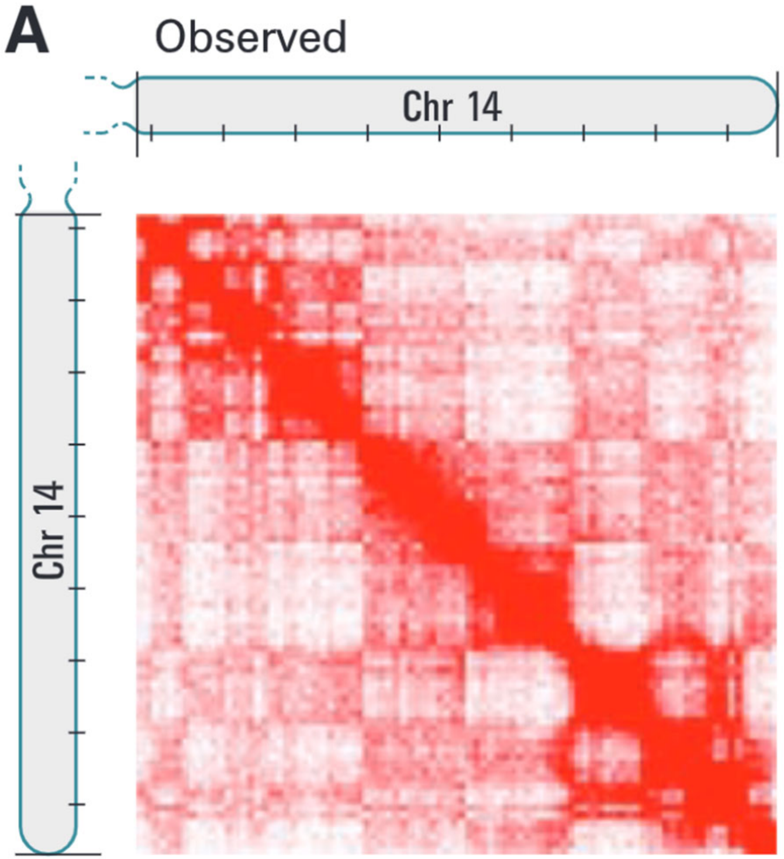
Is euchromatin really open in the cell?
Maeshima K, Iida S, Shimazoe MA, Tamura S and Ide S. ***Trends in Cell Biol.*** 2024 Jan 34(1):7-17

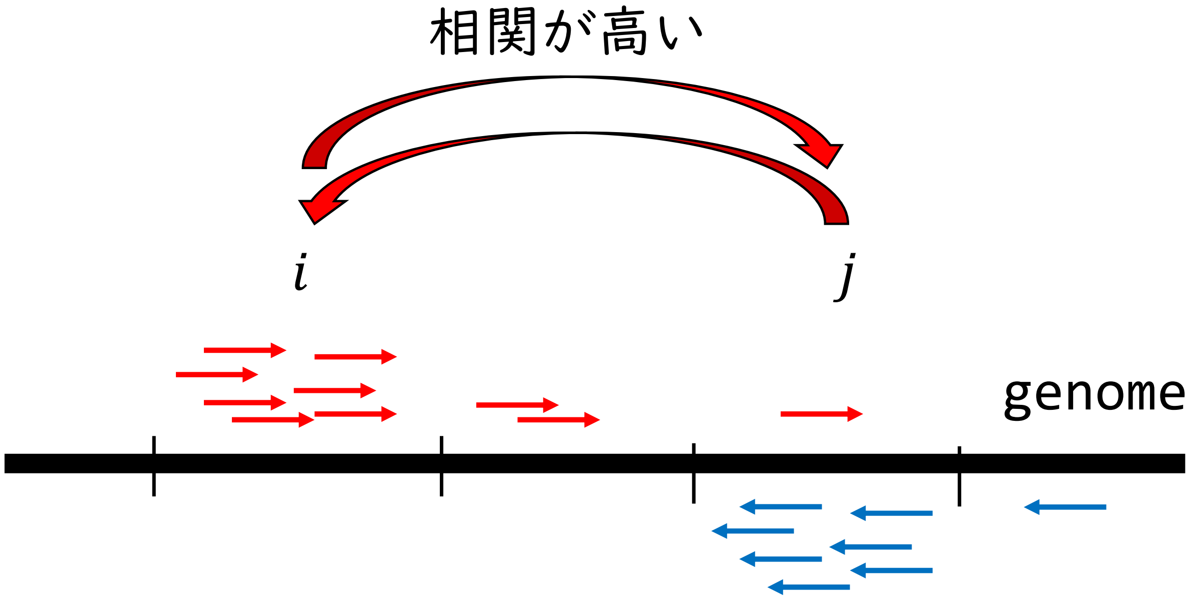
Hi-Cの原理



Comprehensive mapping of long-range interactions reveals folding principles of the human genome.
Liberman-Eiden E. et al ***Science*** 2009 Oct 9;326(5950):289-293.

Contact map

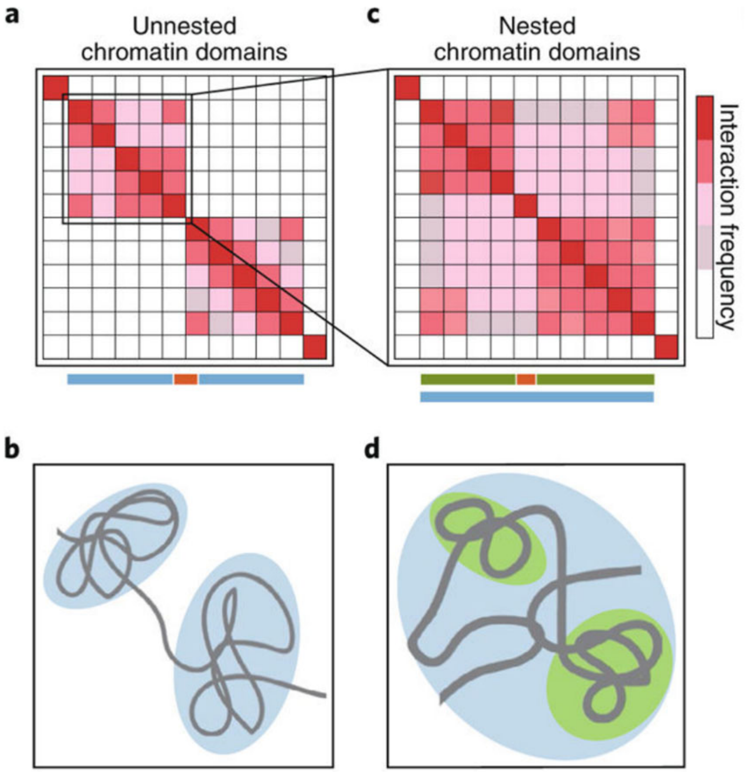


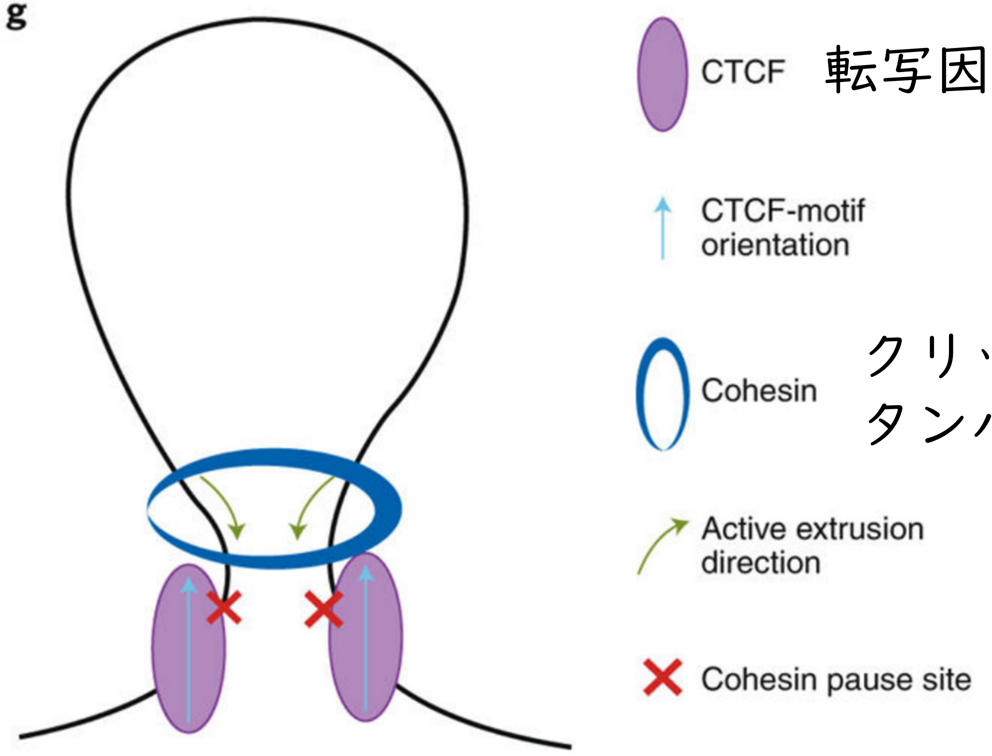


ゲノムを一定の間隔で区切る（bin）
タグカウントのbin間の相関を色で表したheat map

Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Liberman-Eiden E. et al ***Science*** 2009 Oct 9;326(5950):289-293.

Topological Associating Domain（TAD）



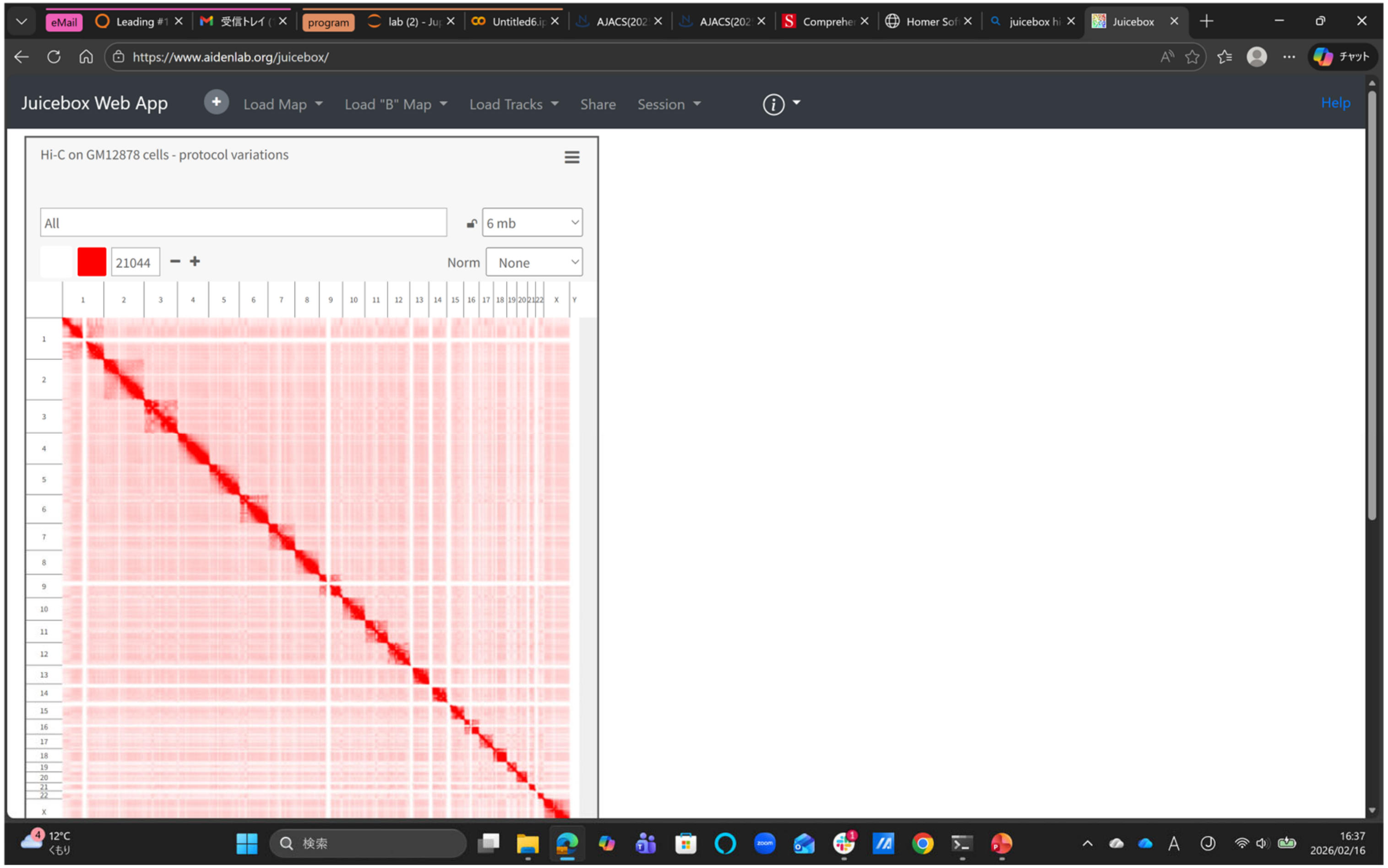


転写因子

クリップ役の
タンパク

DNAが区画化されたメガベーススケールの構造

JuiceBox



Hi-C on GM12878 cells - protocol variations

All

6 mb

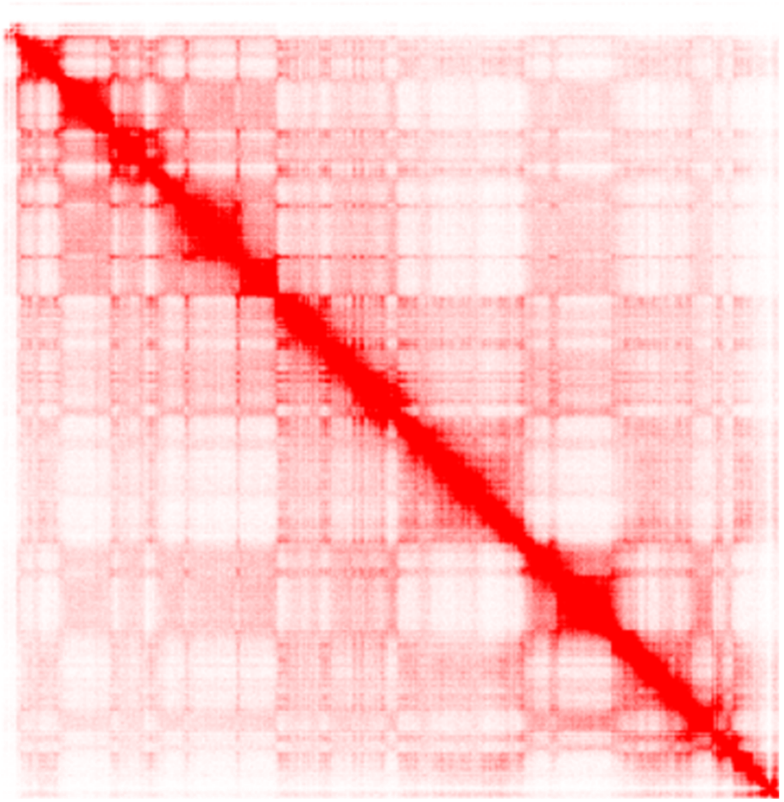
21044

Norm None

Baylor College of Medicineの
Aiden Labが提供する
Hi-C解析用のウェブアプリケーション
ブラウザ上で解析できる。
ENCODEなどに収録されたデータの閲覧もできる。

https://www.aidenlab.org/juicebox/

14番染色体のContact map



Whole-genome haplotype reconstruction using proximity ligation and shotgun sequencing.
Selvaraj S, Dixon JR, Bansal V and Ren B. ***Nat Biotechnol.*** 2013 Dec;31(12):1111-1118.

まとめ

- epigenome
  塩基配列以外の遺伝子の要素

- epigenome解析
  遺伝子の状態を全ゲノム規模で解析する手法
  ChiP-seq、ATAC-seq、Hi-Cなど
  標的となる遺伝子領域を収集→NGSでシーケンス→マッピング→Peak Call→Annotation→タグカウント