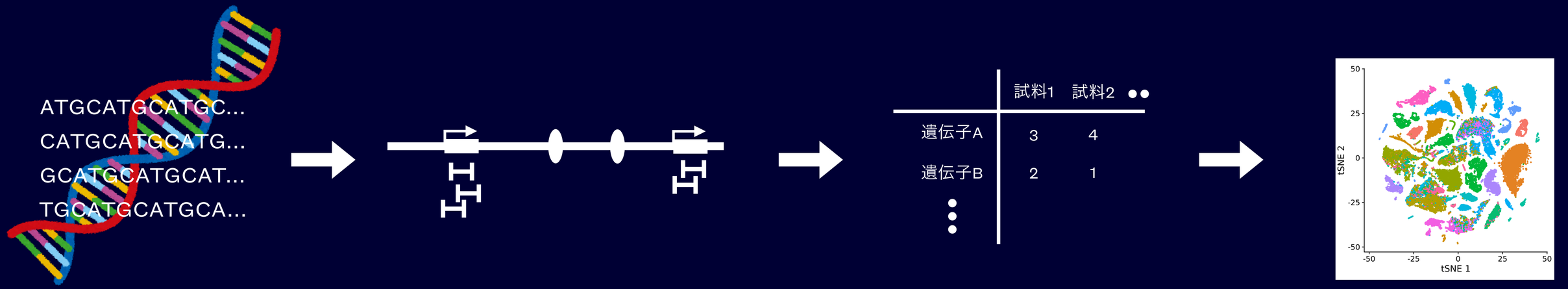
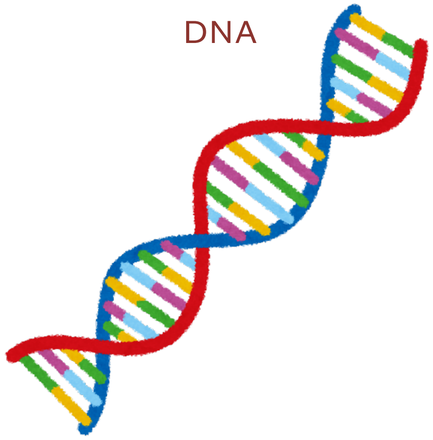
第4回京都大学研究データマネジメントワークショップ



生命科学研究（ゲノム機能解析研究）のデータマネジメント

京都大学高等研究院ヒト生物学高等研究拠点(ASHBI)・単一細胞ゲノム情報解析コア(SIGNAC)　コアマネジャー　辻村太郎

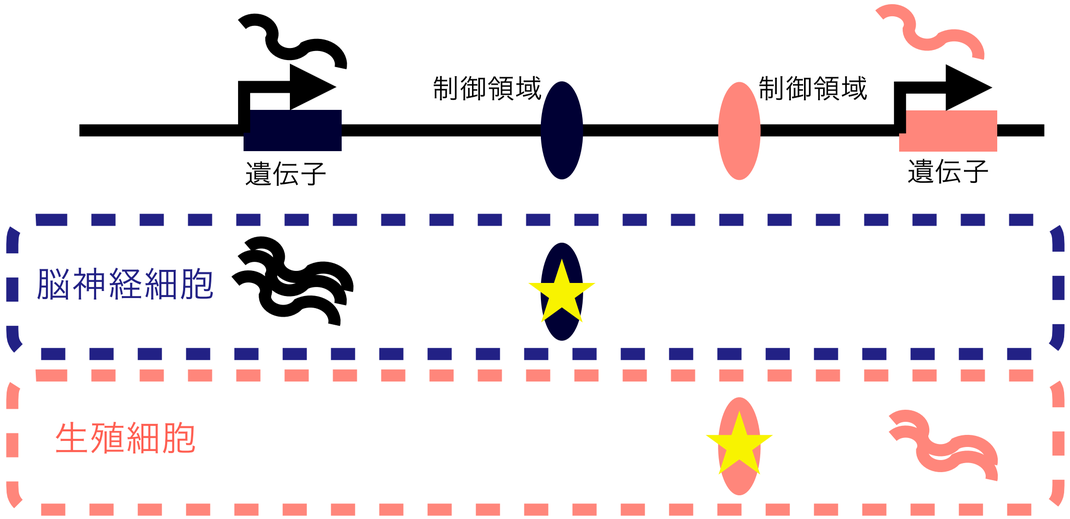
ゲノム



ATGCATGCATGC...

ヒトゲノムは**30億**のA, T, G, Cの並び

ゲノムには遺伝子（~2万）や無数の制御領域がコードされている。



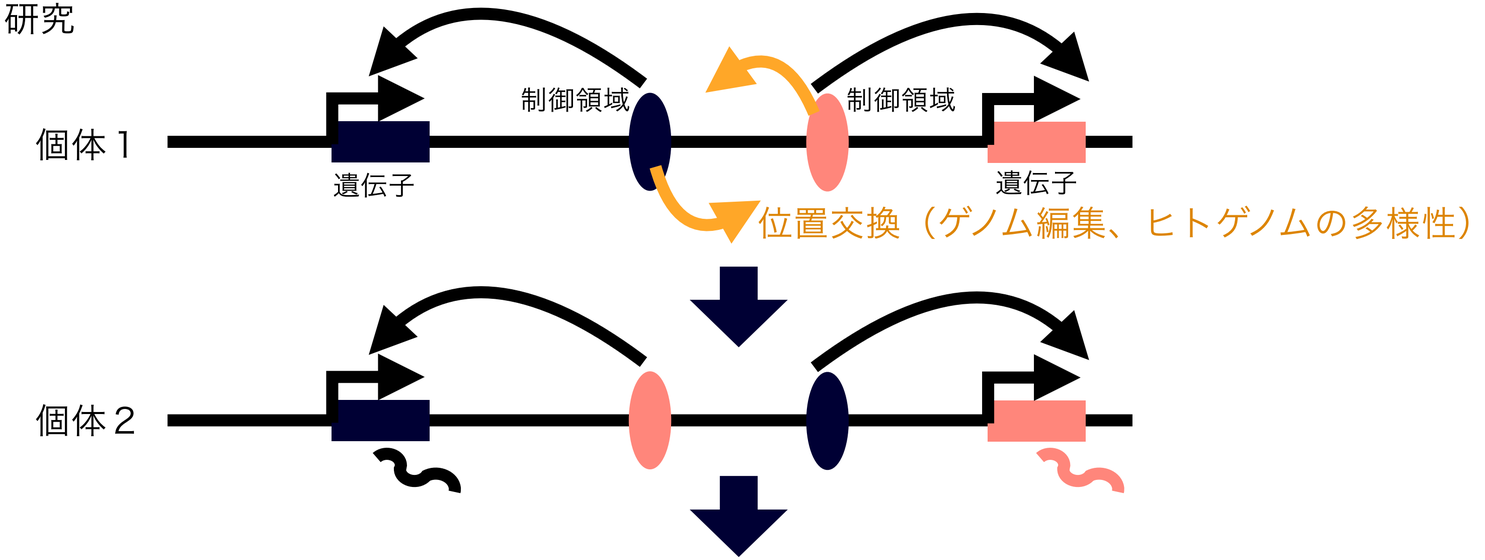
ゲノムの状態は細胞ごとに異なる。

ゲノムの配列は生物種、個体間で異なる。

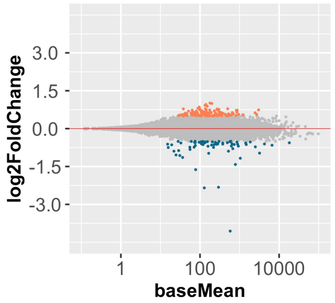
複雑で多様な生き物

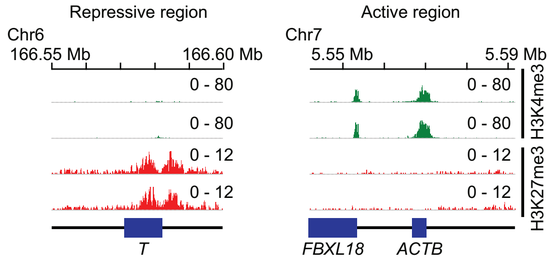
ゲノム機能解析

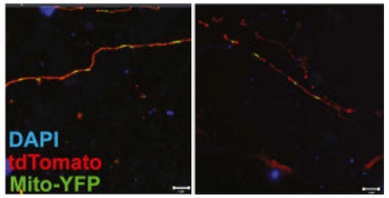
ゲノム配列の機能と進化を研究

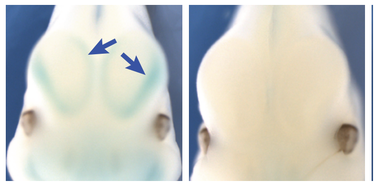


遺伝子の発現状態、**ゲノムの３次元的構造・化学的状態**、細胞の状態、生体システムの状態、を解析









２０年前のシーケンス解析

各DNA断片をそれぞれ個別に増幅

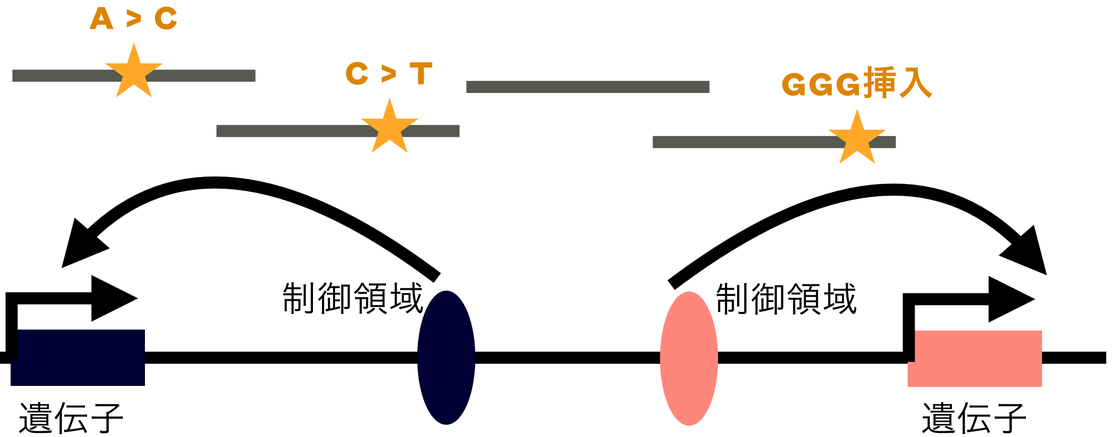
それぞれ個別に配列解析

（１本あたり800bpくらい）

ヒトゲノム計画（１人）：

~10年、~3000億円

全ゲノム配列解析

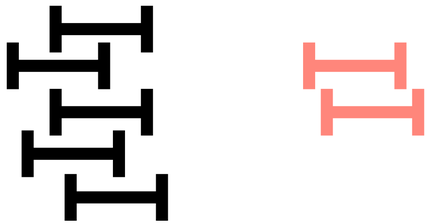


~100Gb

発現解析

~10Gb

エピジェネティクス解析



~10Gb

次世代シーケンス解析

各DNA断片を**まとめて**処理

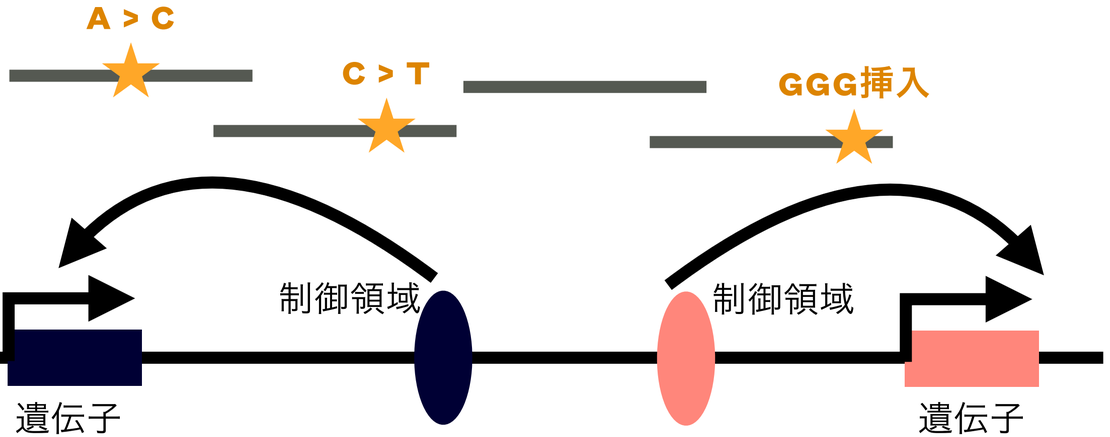
各DNA断片を**まとめて**解析

~3Tb

ヒトゲノム（１人）：

1~2日、数万円

全ゲノム配列解析

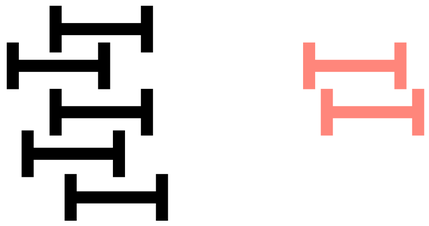


~100Gb

発現解析

~10Gb

エピジェネティクス解析



~10Gb

シーケンスデータの処理プロセス

遺伝子発現解析を例に

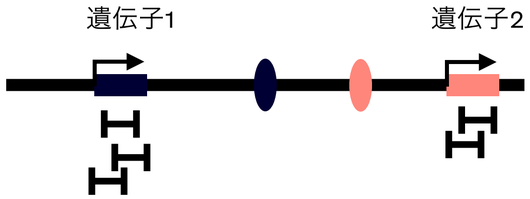
生データ
100Gb~3Tb

配列データ
.fastqファイル

サンプル１
ATGCATGCATGC...
CATGCATGCATG...
GCATGCATGCAT...
TGCATGCATGCA...

サンプル２

ゲノムへのマッピング
.bamファイル

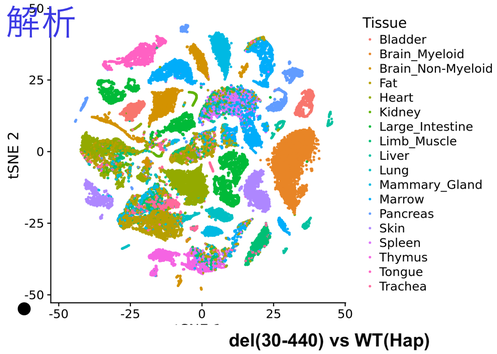


遺伝子との照合
.countファイル

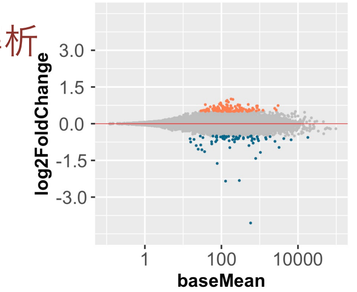
	試料1	試料2	●●
遺伝子1	3	4	
遺伝子2	2	1	
⋮			

データ解析

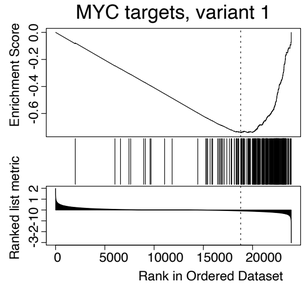
サンプルの分布

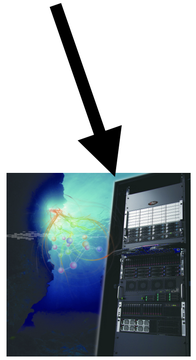


差分解析



エンリッチメント解析





実験・サンプルごとに分割、
配列ファイルに変換

ASHBiファイルサーバーに保存

定期的に自動バックアップ

シーケンスデータの処理プロセス

遺伝子発現解析を例に

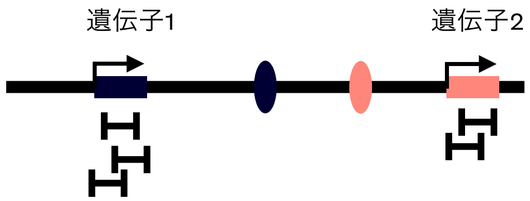
生データ
100Gb~3Tb

配列データ
.fastqファイル

サンプル１
ATGCATGCATGC...
CATGCATGCATG...
GCATGCATGCAT...
TGCATGCATGCA...

サンプル２

ゲノムへのマッピング
.bamファイル

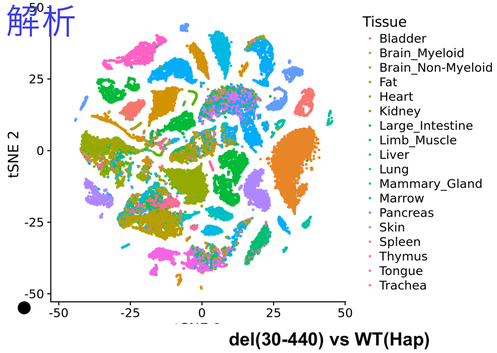


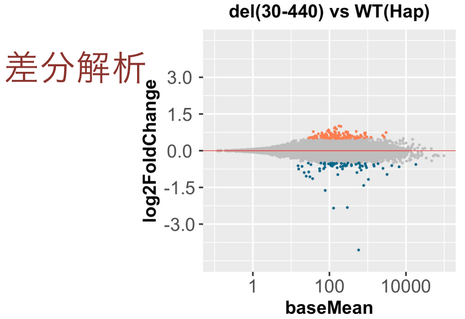
遺伝子との照合
.countファイル

	試料1	試料2	●●
遺伝子A	3	4	
遺伝子B	2	1	
⋮			

データ解析

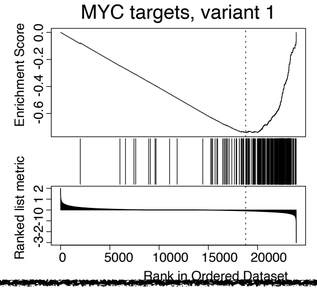
サンプルの分布





差分解析

エンリッチメント解析



ASHBiファイルサーバーに保存
定期的に自動バックアップ

実験・サンプルごとに分割、
配列ファイルに変換

公的データベースに登録

論文で発表

公的データベースGEO (NCBI), ArrayExpress (EBI), DRA (DDBJ)

eLife RESEARCH ARTICLE

Controlling gene activation by enhancers through a drug-inducible topological insulator

Taro Tsujimura[1,2†]*, Osamu Takase[1,2], Masahiro Yoshikawa[1,2], Etsuko Sano[1,2], Matsuhiko Hayashi[3], Kazuto Hoshi[4,5], Tsuyoshi Takato[4,5], Atsushi Toyoda[6], Hideyuki Okano[2], Keiichi Hishikawa[1,2]*

[1]Department of iPS Cell Research & Epigenetic Medicine, Keio University School of Medicine, Tokyo, Japan; [2]Department of Physiology, Keio University School of Medicine, Tokyo, Japan; [3]Apheresis and Dialysis Center, Keio University School of Medicine, Tokyo, Japan; [4]Division of Tissue Engineering, University of Tokyo Hospital, Tokyo, Japan; [5]Department of Oral and Maxillofacial Surgery, University of Tokyo Hospital, Tokyo, Japan; [6]Department of Genomics and Evolutionary Biology, National Institute of Genetics, Mishima, Japan

***For correspondence:** taro.tsujimura@keio.jp (TT); hishikawa-tky@umin.ac.jp (KH)

Present address: †Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University, Kyoto, Japan

Competing interest: See page 30

Funding: See page 31

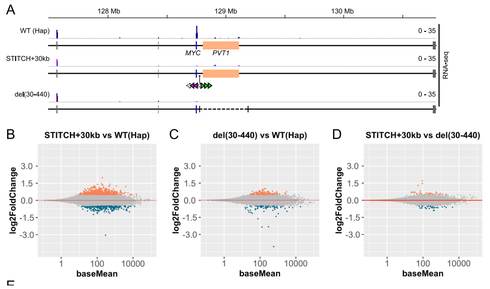
Received: 26 April 2019
Accepted: 06 April 2020
Published: 05 May 2020

Reviewing editor: Job Dekker, University of Massachusetts Medical School, United States

Abstract While regulation of gene-enhancer interaction is intensively studied, its application remains limited. Here, we reconstituted arrays of CTCF-binding sites and devised a synthetic topological insulator with tetO for chromatin-engineering (STITCH). By coupling STITCH with tetR linked to the KRAB domain to induce heterochromatin and disable the insulation, we developed a drug-inducible system to control gene activation by enhancers. In human induced pluripotent stem cells, STITCH inserted between *MYC* and the enhancer down-regulated *MYC*. Progressive mutagenesis of STITCH led to a preferential escalation of the gene-enhancer interaction, corroborating the strong insulation ability of STITCH. STITCH also altered epigenetic states around *MYC*. Time-course analysis by drug induction uncovered deposition and removal of H3K27me3 repressive marks follows and reflects, but does not precede and determine, the expression change. Finally, STITCH inserted near *NEUROG2* impaired the gene activation in differentiating neural progenitor cells. Thus, STITCH should be broadly useful for functional genetic studies.

Introduction

Interaction of genes and enhancers is greatly affected by architectural proteins that bind to chromatin and organize folding of the genome (*Dekker et al., 2017*). Most notably, CTCF mediates loop formation of chromatin in association with a cohesin complex, which physically bundles two distant loci of the genomic DNA (*Parelho et al., 2008*; *Phillips-Cremins et al., 2013*; *Wendt et al., 2008*). The genome-wide contact maps of chromatin show that the CTCF-binding sites often demarcate boundaries of so-called contact domains or topologically associating domains (TADs), where chromatin association takes place more preferentially inside than outside (*Dixon et al., 2012*; *Phillips-Cremins et al., 2013*; *Rao et al., 2014*). The looping between two CTCF-binding sites is mostly established where they are in the converging orientations with each other (*de Wit et al., 2015*; *Guo et al., 2015*; *Rao et al., 2014*; *Vietri Rudan et al., 2015*). Loss of cohesin or CTCF resulted in disappearance of contact domains (*Gassler et al., 2017*; *Nora et al., 2017*; *Rao et al., 2017*; *Schwarzer et al., 2017*; *Wutz et al., 2017*). According to the extrusion model, the cohesin ring extrudes the chromatin fiber from a site of loading and pauses at a CTCF-binding site that is oriented towards the ring (*Fudenberg et al., 2016*; *Sanborn et al., 2015*). This model is widely accepted as the underlying mechanism for the formation of the loops and contact domains.



Search results for taro tsujimura

6 experiments

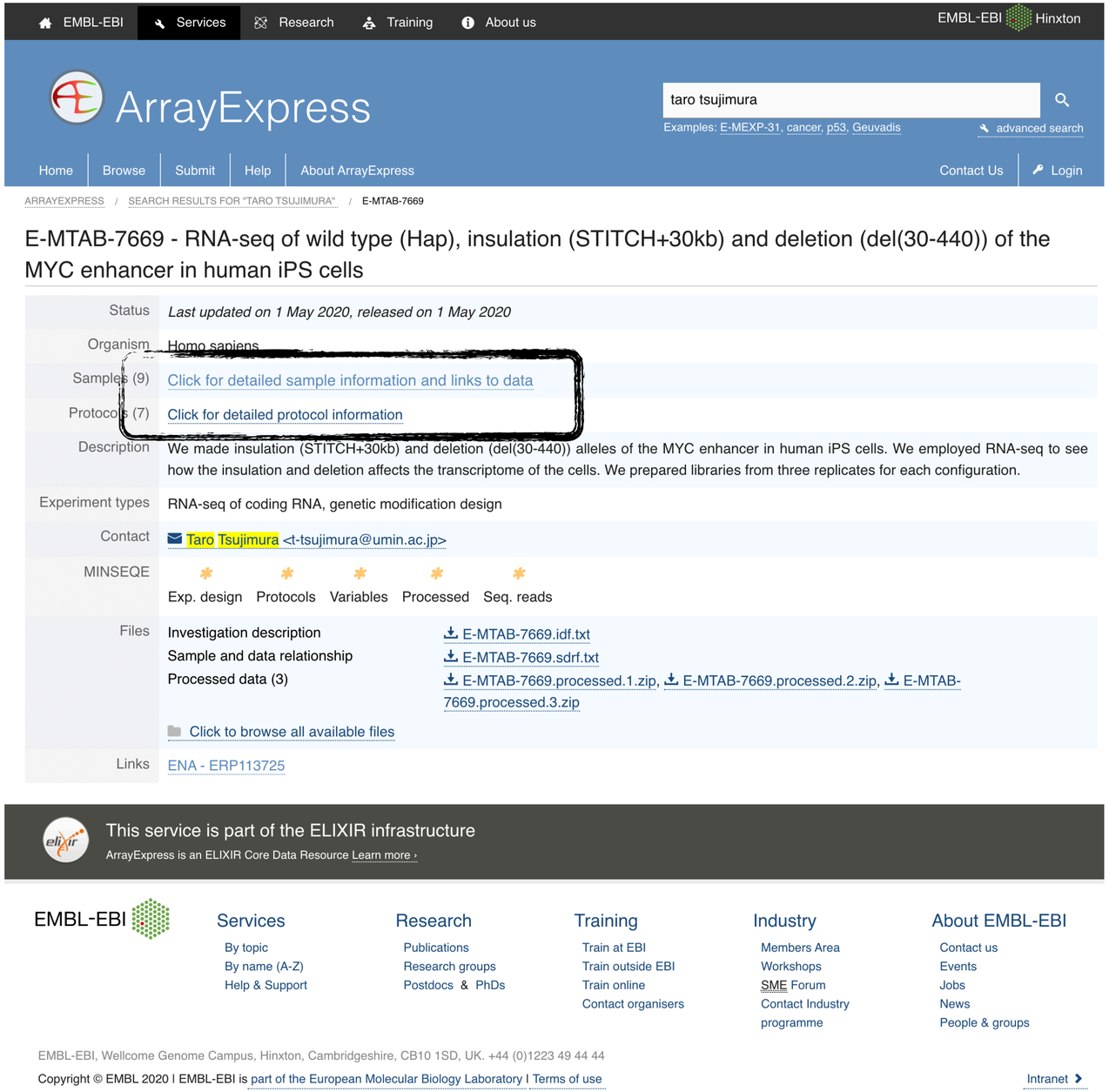
Accession	Title	Type	Organism	Assays	Released	Processed	Raw	Atlas
E-MTAB-8957	4C-seq to show the effects of insertion of STITCH into NEUROG2-65kb positions on the chromatin conformation in neural progenitor cells differentiated from human iPSCs	4C	Homo sapiens	4	01/05/2020	↓	↓	-
E-MTAB-8492	4C-seq to show the effects of insertion of STITCH into MYC+30kb and NEUROG2-65kb positions on the chromatin conformation in human iPSCs and differentiated neural progenitor cells	4C	Homo sapiens	16	01/05/2020	↓	↓	-
E-MTAB-7670	nChIP-seq for CTCF, H3K4me3, H3K27me3 and H3K9me3, in wild type (Hap), STITCH+30kb and STITCH/KRAB clones of human iPS cells	ChIP-seq	Homo sapiens	22	01/05/2020	🔗	↓	-
E-MTAB-7669	RNA-seq of wild type (Hap), insulation (STITCH+30kb) and deletion (del(30-440)) of the MYC enhancer in human iPS cells	RNA-seq of coding RNA	Homo sapiens	9	01/05/2020	🔗	↓	-
E-MTAB-7668	4C-seq from viewpoint at MYC promoter (VP-MYC1 and VP-MYC2), in wild type (Hap) and variously modified alleles around the locus in human iPS cells	4C	Homo sapiens	58	01/05/2020	🔗	↓	-
E-MTAB-2488	4C analysis to capture chromatin conformations of the Tfap2c-Bmp7 locus upon different genomic rearrangements in mice	other	Mus musculus	24	19/01/2015	-	↓	-

Export table in Tab-delimited format | Export matching metadata in XML format | Subscribe to RSS feed matching this search

This service is part of the ELIXIR infrastructure

ArrayExpress is an ELIXIR Core Data Resource Learn more ›

1つのデータセットの概要



データファイルへのリンク

E-MTAB-7669 - RNA-seq of wild type (Hap), insulation (STITCH+30kb) and deletion (del(30-440)) of the MYC enhancer in human iPS cells

Display full sample-data table

Export table in Tab-delimited format

9 rows

	Sample Attributes				Variables	Assay	Links to Data			
Source Name	organism	sex	genotype	cell type	genotype	Assay Name	ENA	FASTQ	Processed	Processed
RNAseq_del(30-440)_rep1	Homo sapiens	female	del(30-440)	induced pluripotent stem cell	del(30-440)	RNAseq_del(30-4				
RNAseq_del(30-440)_rep2	Homo sapiens	female	del(30-440)	induced pluripotent stem cell	del(30-440)	RNAseq_del(30-4				
RNAseq_del(30-440)_rep3	Homo sapiens	female	del(30-440)	induced pluripotent stem cell	del(30-440)	RNAseq_del(30-4				
RNAseq_Hap_rep1	Homo sapiens	female	Hap	induced pluripotent stem cell	Hap	RNAseq_Hap_rep				
RNAseq_Hap_rep2	Homo sapiens	female	Hap	induced pluripotent stem cell	Hap	RNAseq_Hap_rep				
RNAseq_Hap_rep3	Homo sapiens	female	Hap	induced pluripotent stem cell	Hap	RNAseq_Hap_rep				
RNAseq_STITCH+30kb_rep1	Homo sapiens	female	STITCH+30kb	induced pluripotent stem cell	STITCH+30kb	RNAseq_STITCH-				
RNAseq_STITCH+30kb_rep2	Homo sapiens	female	STITCH+30kb	induced pluripotent stem cell	STITCH+30kb	RNAseq_STITCH-				
RNAseq_STITCH+30kb_rep3	Homo sapiens	female	STITCH+30kb	induced pluripotent stem cell	STITCH+30kb	RNAseq_STITCH-				

プロトコルの記載

7 protocols

Accession	Type
P-MTAB-83023	high throughput sequence alignment protocol
We mapped the sequences to the human genome (hg19) with HISAT2 (Kim et al., 2015).	
P-MTAB-83024	normalization data transformation protocol
We made BedGraph tracks with HOMER (Heinz et al., 2010). The data ranges are indicated by counts per 10 million. We assigned the mapped reads to annotated genes with HTSeq (Anders et al., 2015).	
P-MTAB-83021	nucleic acid sequencing protocol
The libraries were sequenced with HiSeq2500 System (Illumina) using HiSeq SR Rapid Cluster Kit v2-HS (Illumina, Cat#GD-402-4002) and HiSeq Rapid SBS Kit v2-HS 50 Cycle (Illumina, Cat#FC-402-4022).	
P-MTAB-83022	growth protocol
We cultured the cells in the StemFit® AK02N medium (ReproCELL, Cat#RCAK02N) on dish coated with iMatrix-511 (ReproCELL, Cat#NP892-012) without feeder cells.	
P-MTAB-83020	nucleic acid library construction protocol
We first enriched mRNA using NEBNext® Poly(A) mRNA Magnetic Isolation (New England Biolabs, Cat#E7490S). Then subsequently, we used NEXTflex Rapid RNA-Seq kit (Bioo Scientific, Cat#NOVA-5238-01) for the library preparation with customly designed oligo DNAs as primers for the PCR reaction.	
P-MTAB-83018	sample collection protocol
The cells were dissociated from dish with TrypLE™ Select (Thermo Fisher Scientific K.K., Cat#12563-011) .	
P-MTAB-83019	nucleic acid extraction protocol
RNA was extracted using the High-pure RNA isolation kit (Roche, Cat#11828665001) in presence of the DNase I included in the kit.	

ゲノムシーケンスデータまとめ

複数のプロジェクト

生データ
100Gb~3Tb

ASHBiのファイルサーバーに保存
定期的に自動バックアップ

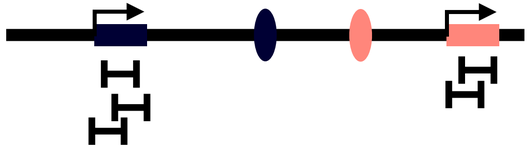
実験・サンプルごとに分割

配列データ
.fastqファイル

サンプル１
ATGCATGCATGC...
CATGCATGCATG...
GCATGCATGCAT...
TGCATGCATGCA...

サンプル２

ゲノムへのマッピング
.bamファイル



遺伝子との照合
.countファイル

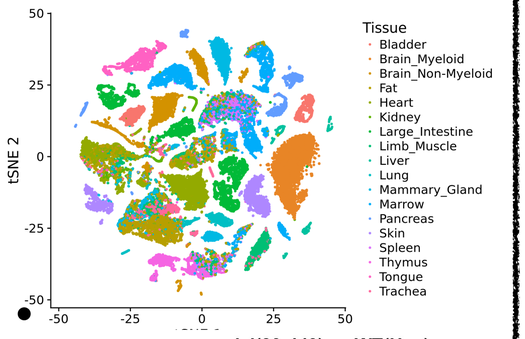
	試料1	試料2	●●
遺伝子A	3	4	
遺伝子B	2	1	
●			

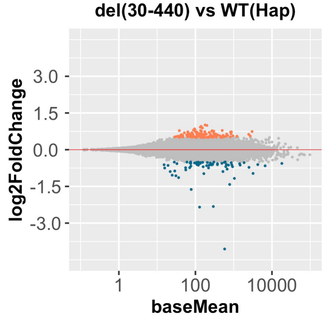
公的データベース

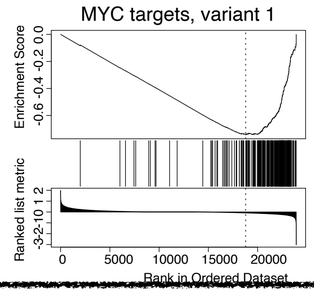
データの概要、取得方法をテキストで記載

論文

データ解析







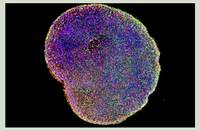
実際は複雑：「実験ノート」「プロジェクト」「リソース」「データ」

2021年8月19日

遺伝子Aのノックアウトマウスの解析

胎児前脳RNA-seqをした(**wt vs KO**)。

認知症疾患iPS細胞由来の**脳オルガノイド**の解析



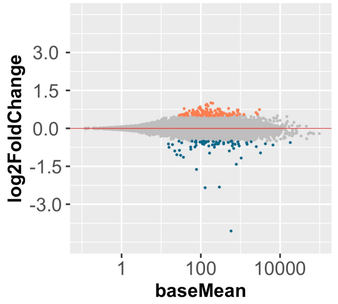
ゲノム上のヒストン修飾のChIP-seqをした（**健常vs疾患**）。

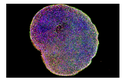
RNA-seqとChIP-seqはまとめて次世代シーケンス解析した。

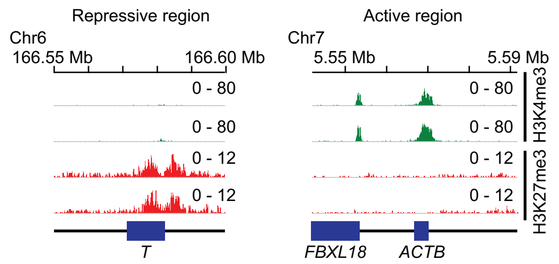
データの場所は「20210819_sequence」フォルダ。

分割

NGSデータ



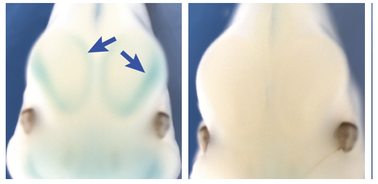




2021年9月19日

遺伝子Aノックアウトマウスの解析

胎児前脳でのレポーター遺伝子の発現を見た(**wt vs KO**)。

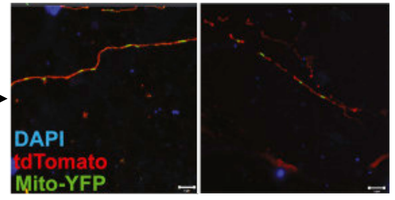


「実体顕微鏡」フォルダ

└「20210919_geneA_WTvsKO」

認知症疾患iPS細胞由来の**神経細胞**の解析

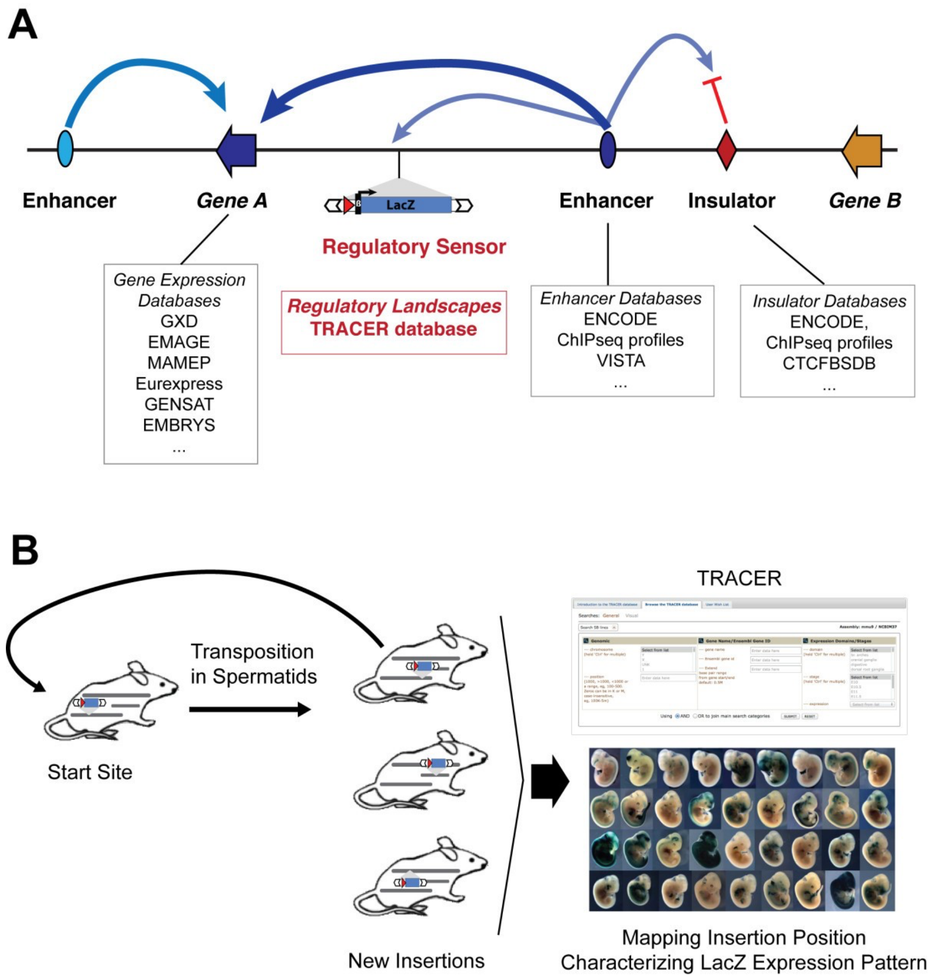
分化させた神経細胞の状態を解析した（**健常vs疾患**）。



「共焦点顕微鏡」フォルダ

└「20210919_organoid」

マウスリソースのデータベースの構築
（欧州分子生物学研所・ポスドク時代の取り組み）

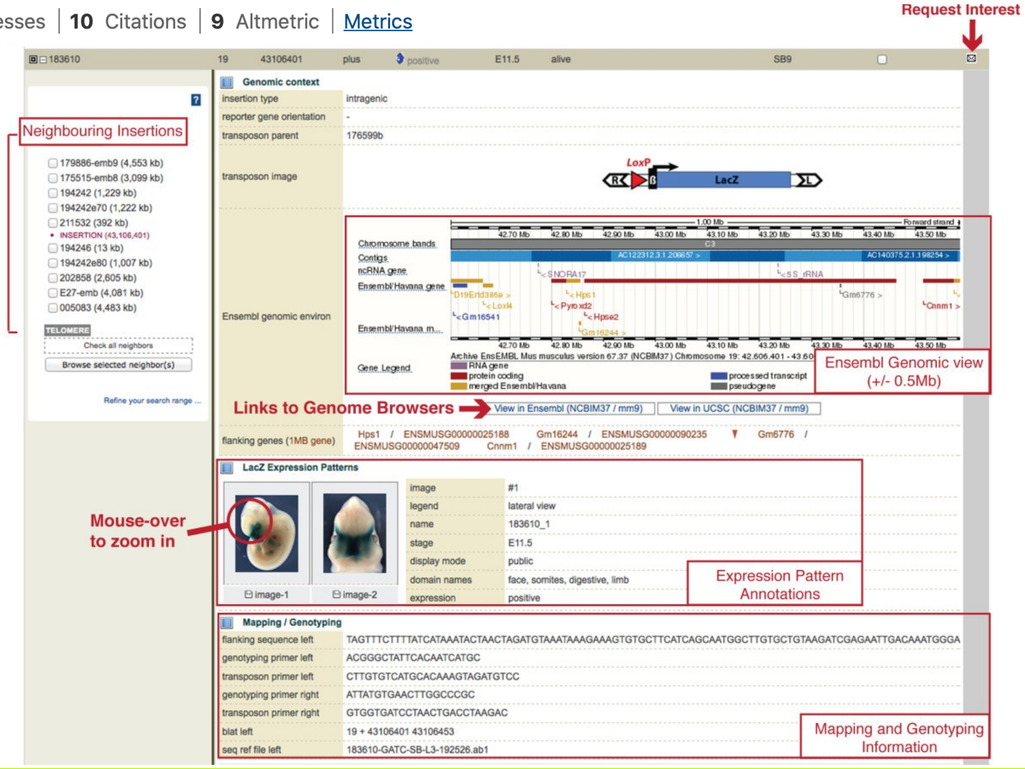


Database | Open Access | Published: 02 April 2013

TRACER: a resource to study the regulatory architecture of the mouse genome

Chao-Kung Chen, Orsolya Symmons, Veli Vural Uslu, Taro Tsujimura, Sandra Ruf, Damian Smedley ✉ & François Spitz ✉

BMC Genomics **14**, Article number: 215 (2013) | Cite this article



個人プロジェクトの派生データの統合→データベース

PLOS GENETICS

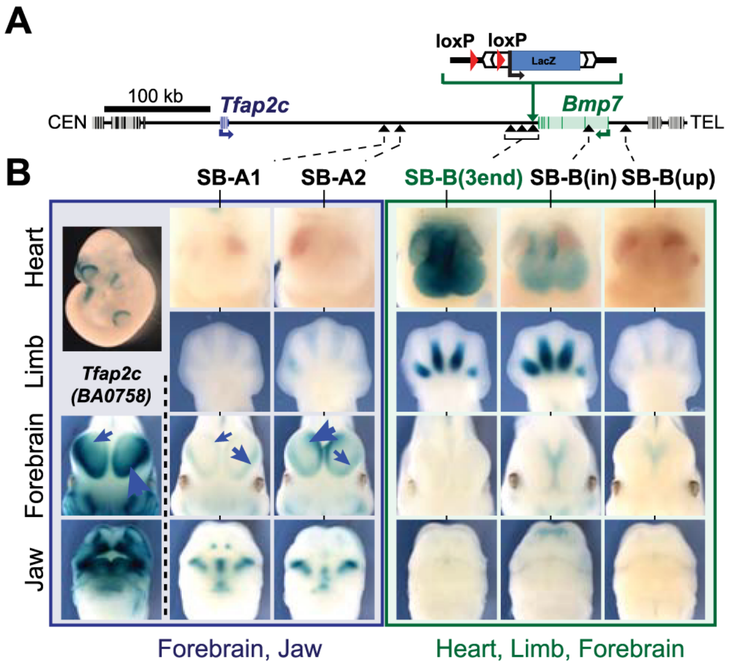
OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

A Discrete Transition Zone Organizes the Topological and Regulatory Autonomy of the Adjacent *Tfap2c* and *Bmp7* Genes

Taro Tsujimura, Felix A. Klein, Katja Langenfeld, Juliane Glaser, Wolfgang Huber, François Spitz

Published: January 8, 2015 • https://doi.org/10.1371/journal.pgen.1004897



nature genetics

Explore our content ⌄ Journal information ⌄

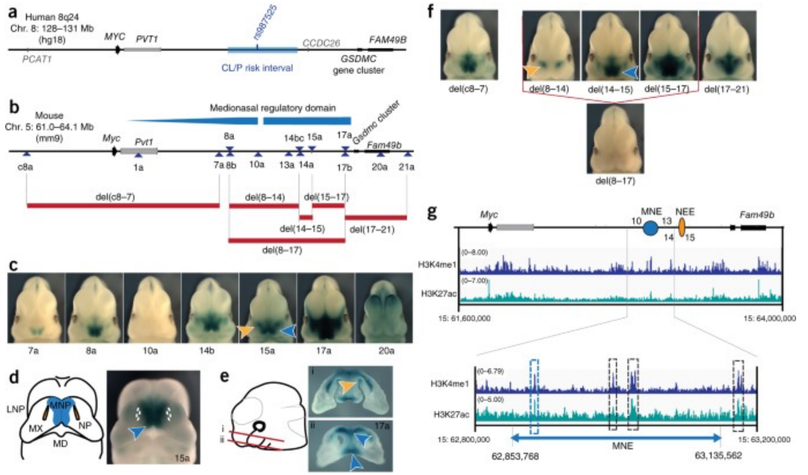
nature › nature genetics › letters › article

Published: 25 May 2014

Long-range enhancers regulating *Myc* expression are required for normal facial morphogenesis

Veli Vural Uslu, Massimo Petretich, Sandra Ruf, Katja Langenfeld, Nuno A Fonseca, John C Marioni & François Spitz ✉

Nature Genetics **46**, 753–758(2014) | Cite this article



Developmental Cell

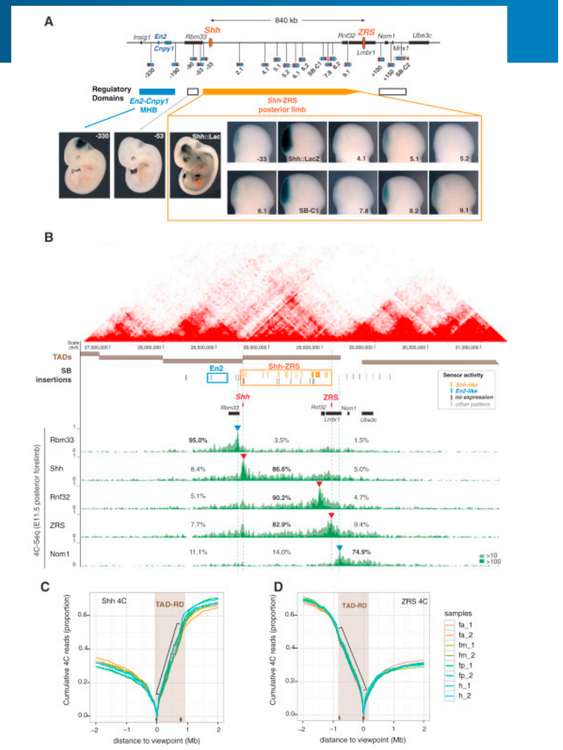
ARTICLE | VOLUME 39, ISSUE 5, P529-543, DECEMBER 05, 2016

The *Shh* Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances

Orsolya Symmons [3] • Leslie Pan • Silvia Remeseiro • ... Felix Klein • Wolfgang Huber • François Spitz [5, 6] ✉ • Show all authors • Show footnotes

Open Access • Published: November 17, 2016 • DOI: https://doi.org/10.1016/j.devcel.2016.10.015 •

Check for updates



データベースを活用した統括的解析研究

Functional and topological characteristics of mammalian regulatory domains

Orsolya Symmons[1], Veli Vural Uslu[1], Taro Tsujimura[1], Sandra Ruf[1], Sonya Nassari[1], Wibke Schwarzer[1], Laurence Ettwiller[2,3] and François Spitz[1,4]

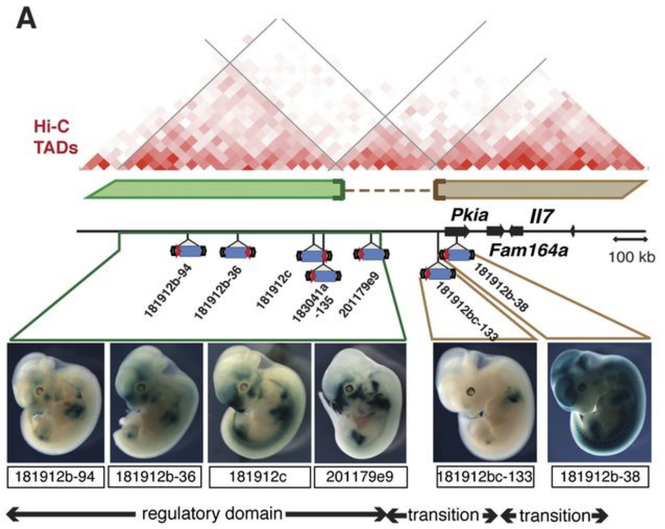
[+] Author Affiliations

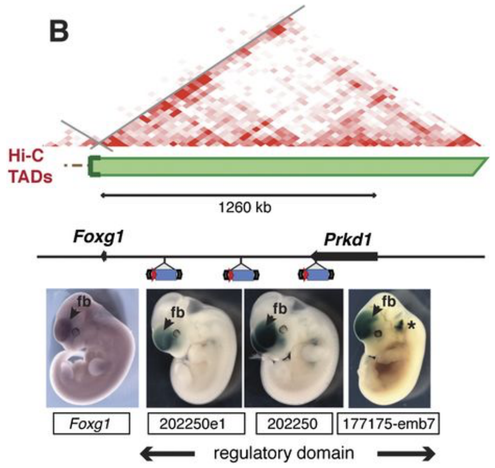
[+] Author Notes

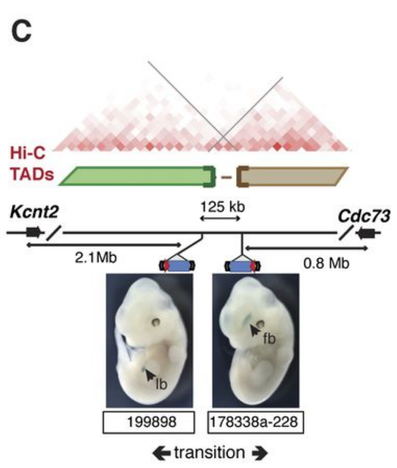
Abstract

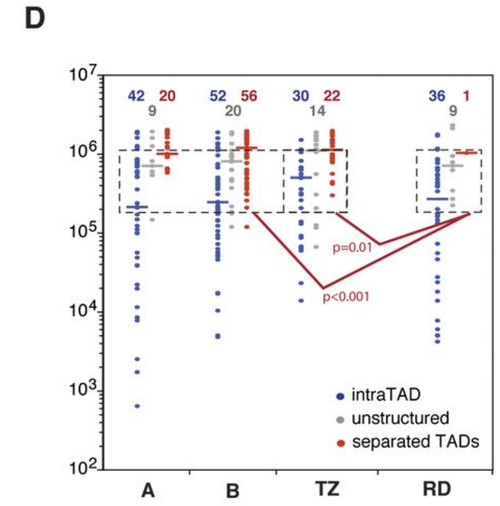
Long-range regulatory interactions play an important role in shaping gene-expression programs. However, the genomic features that organize these activities are still poorly characterized. We conducted a large operational analysis to chart the distribution of gene regulatory activities along the mouse genome, using hundreds of insertions of a regulatory sensor. We found that enhancers distribute their activities along broad regions and not in a gene-centric manner, defining large regulatory domains. Remarkably, these domains correlate strongly with the recently described TADs, which partition the genome into distinct self-interacting blocks. Different features, including specific repeats and CTCF-binding sites, correlate with the transition zones separating regulatory domains, and may help to further organize promiscuously distributed regulatory influences within large domains. These findings support a model of genomic organization where TADs confine regulatory activities to specific but large regulatory domains, contributing to the establishment of specific gene expression profiles.

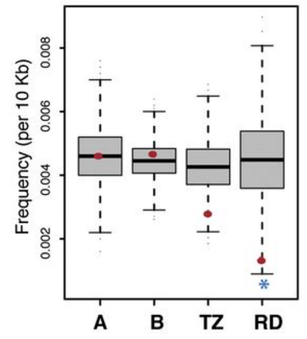
ゲノムの「制御ドメイン」を明らかにした。











ASHBiでのゲノムデータマネジメントとその課題

生データ
100Gb~3Tb

ASHBiのファイルサーバーに保存
定期的に自動バックアップ

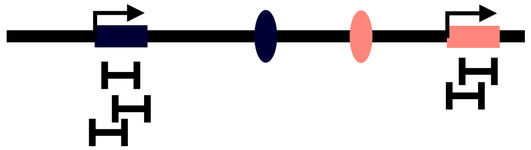
各ラボごとに管理

配列データ
.fastqファイル

サンプル１
ATGCATGCATGC...
CATGCATGCATG...
GCATGCATGCAT...
TGCATGCATGCA...

サンプル２

ゲノムへのマッピング
.bamファイル

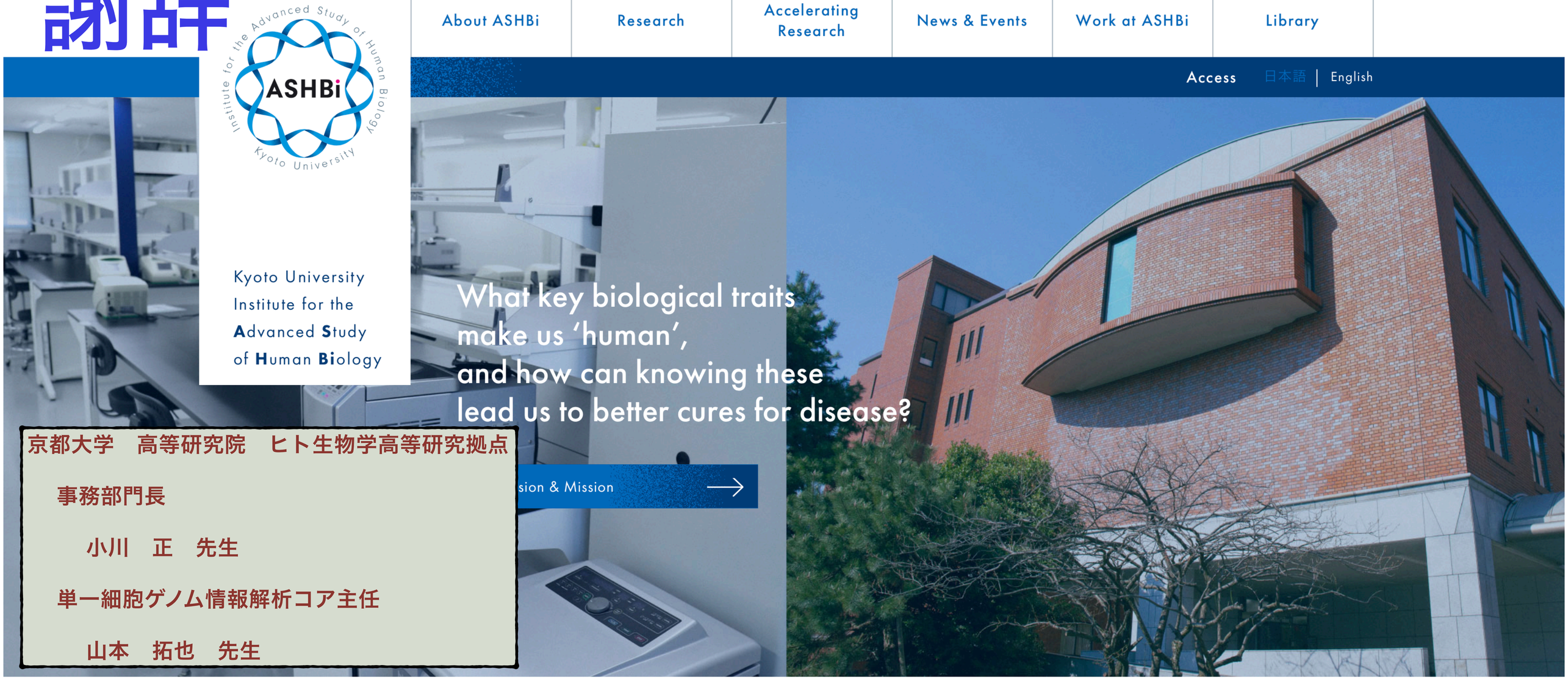


遺伝子との照合
.countファイル

	試料1	試料2
遺伝子A	3	4
遺伝子B	2	1

- 統括的な管理方法？
- データの内容をどう管理？
  - 「ヒト」の個人情報保護との整合性？
  - ブラインド状態でデータ解析の場合は？
- リソース、他の多角的解析データ群との照合をどう実現？

謝辞



京都大学　高等研究院　ヒト生物学高等研究拠点

事務部門長

小川　正　先生

単一細胞ゲノム情報解析コア主任

山本　拓也　先生