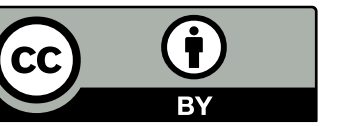




# JoGo: Japanese Open Genome and Omics Platform



## 1.0に向けた取り組みについて

長崎 正朗 (九州大学・生体防御医学研究所) [nagasaki@bioreg.kyushu-u.ac.jp](mailto:nagasaki@bioreg.kyushu-u.ac.jp)

河合洋介 (国立国際医療研究センター) [ykawai-tky@umin.ac.jp](mailto:ykawai-tky@umin.ac.jp)

JoGo (Japanese Open Genome Omics Platform)は、全ゲノムの塩基配列レベルまで研究利用公開されているバイオバンクの試料から、新たに取得した日本人の長鎖型全ゲノム情報、クロマチンアクセシビリティ、トランスクリプトームなどのオミクス情報を統合した情報を提供することを目標としたポータルである。また、これらの情報について、CoLoRSなどの国際連携を通じて、長鎖型全ゲノム情報解析パイプラインと統一することで将来的に数千人規模の海外のデータとの比較ができるデータベースとすることを目標としている。さらに、国内のゲノムポータルデータベースであるMGENDやTogoVarとの連携を進めている。この活動を通じて、塩基配列レベルから解釈できるヒトゲノムオミクス情報基盤ポータルを整備することで、日本のデータサイエンス・バイオインフォマティクスの研究はもとより、教育にも活用できるデータベースとなる。本ポスターでは、現在進めているJoGo第1版公開に向けたポータル整備現状を中心に説明を行う。

### <1> JoGo Genomic DB Portal

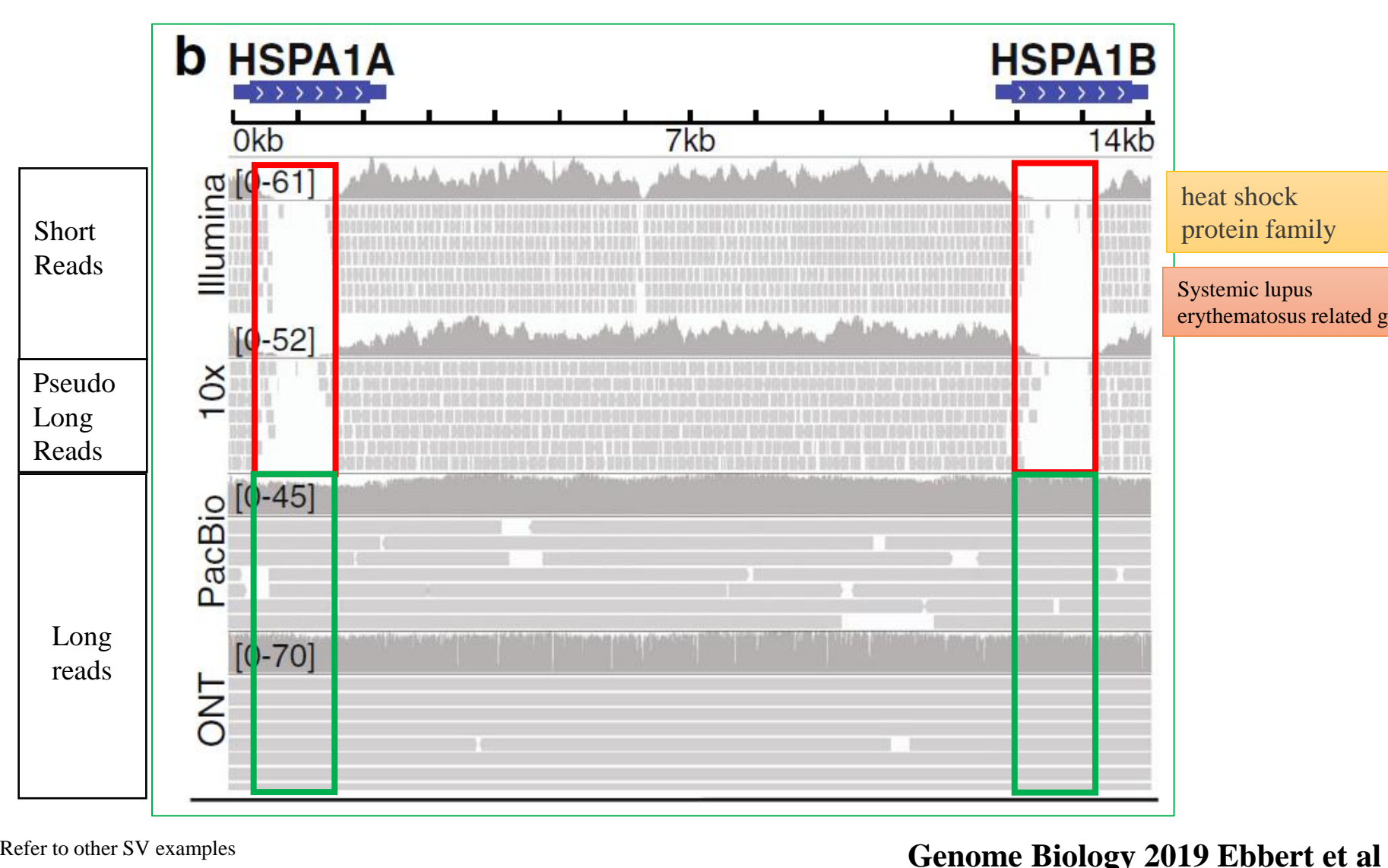
Original data

Development of **open genomic information infrastructure** to accelerate sequence-level understanding of SVs.

- International 258 samples (includes 105 original Japanese samples)
- High-resolution long-read genomic data in whole chromosomal region
- Release base resolution information
- Web portal and downloadable contents

#### ① R&D Concepts: Background

Short read WGS (srWGS) promotes the base level human genome analysis; however, SV regions are still uncovered (Figure below: regions of genes that are difficult to discriminate by srWGS due to too similar sequences (in red box)). The situation is now ready with up-to-date long-range sequencers (in green box).



Refer to other SV examples

Genome Biology 2019 Ebbert et al

### <2> JoGo Omics DB Portal

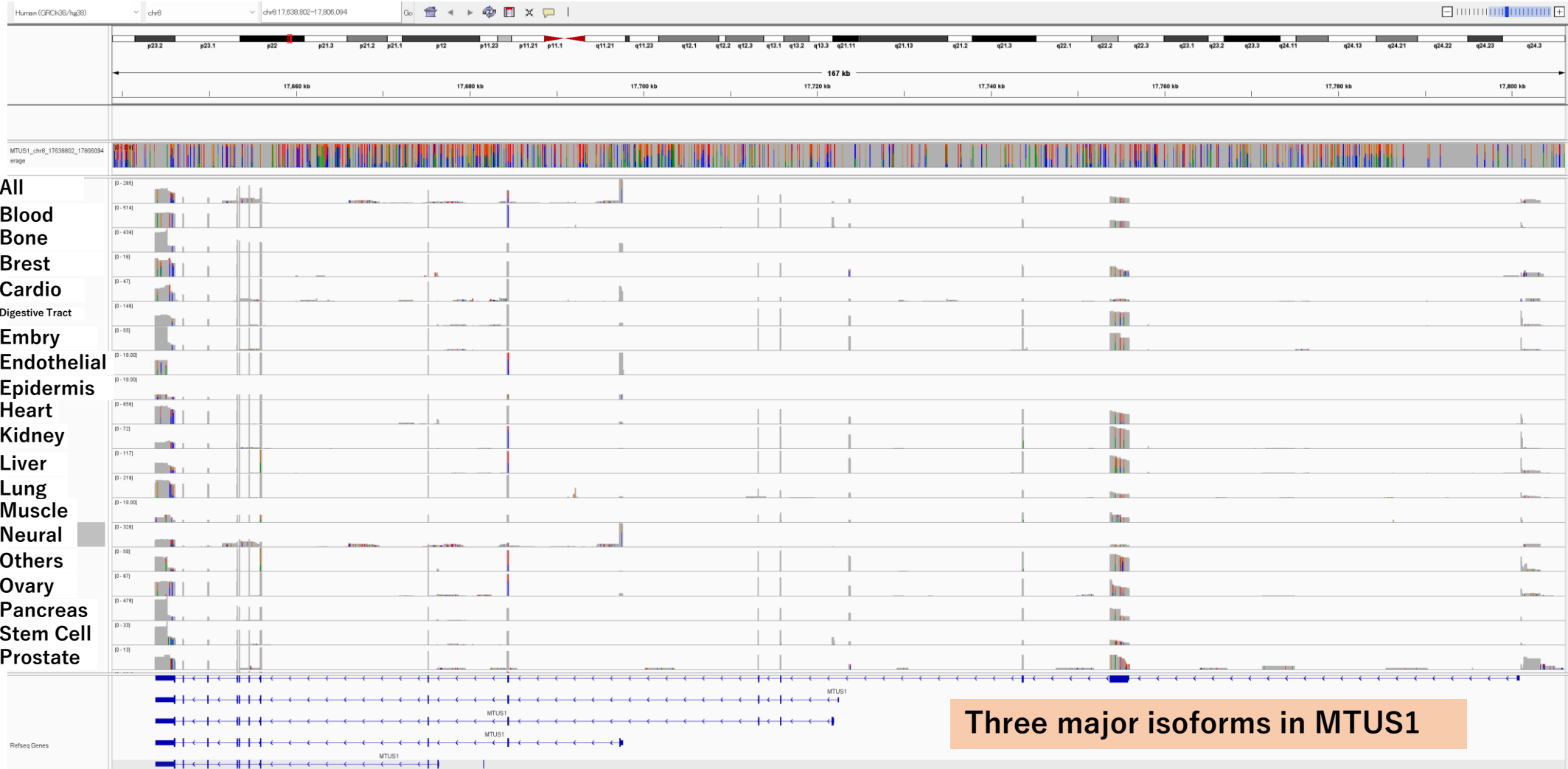
Development of **an omics annotation infrastructure** to accelerate understanding of SVs.

- Original ATAC-Seq / RNA-Seq
- The same samples in (1)
- Catalog public Iso-seq data

Public Data Catalog public ISO-SEQ data from long read analysis

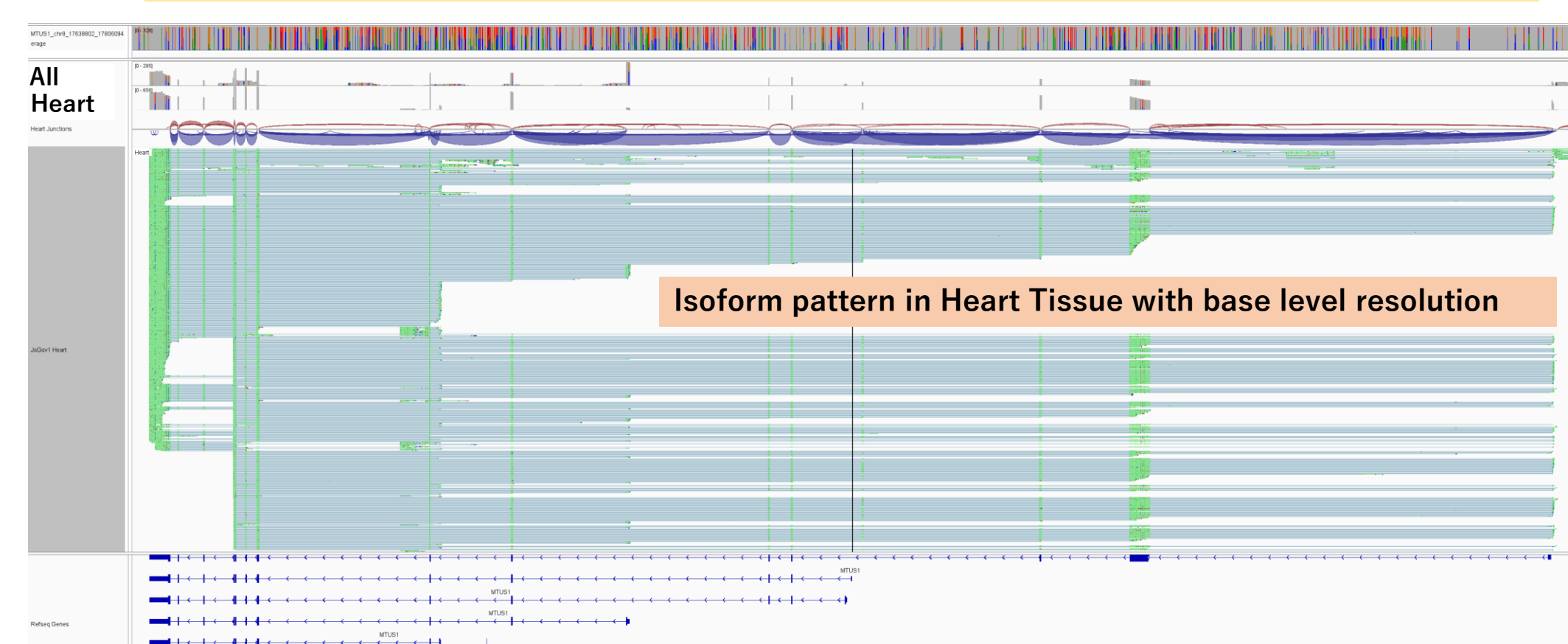
Tissue level expression

MTUS1 chr8:17638802-17806094



Three major isoforms in MTUS1

Base resolution Isoform Expression

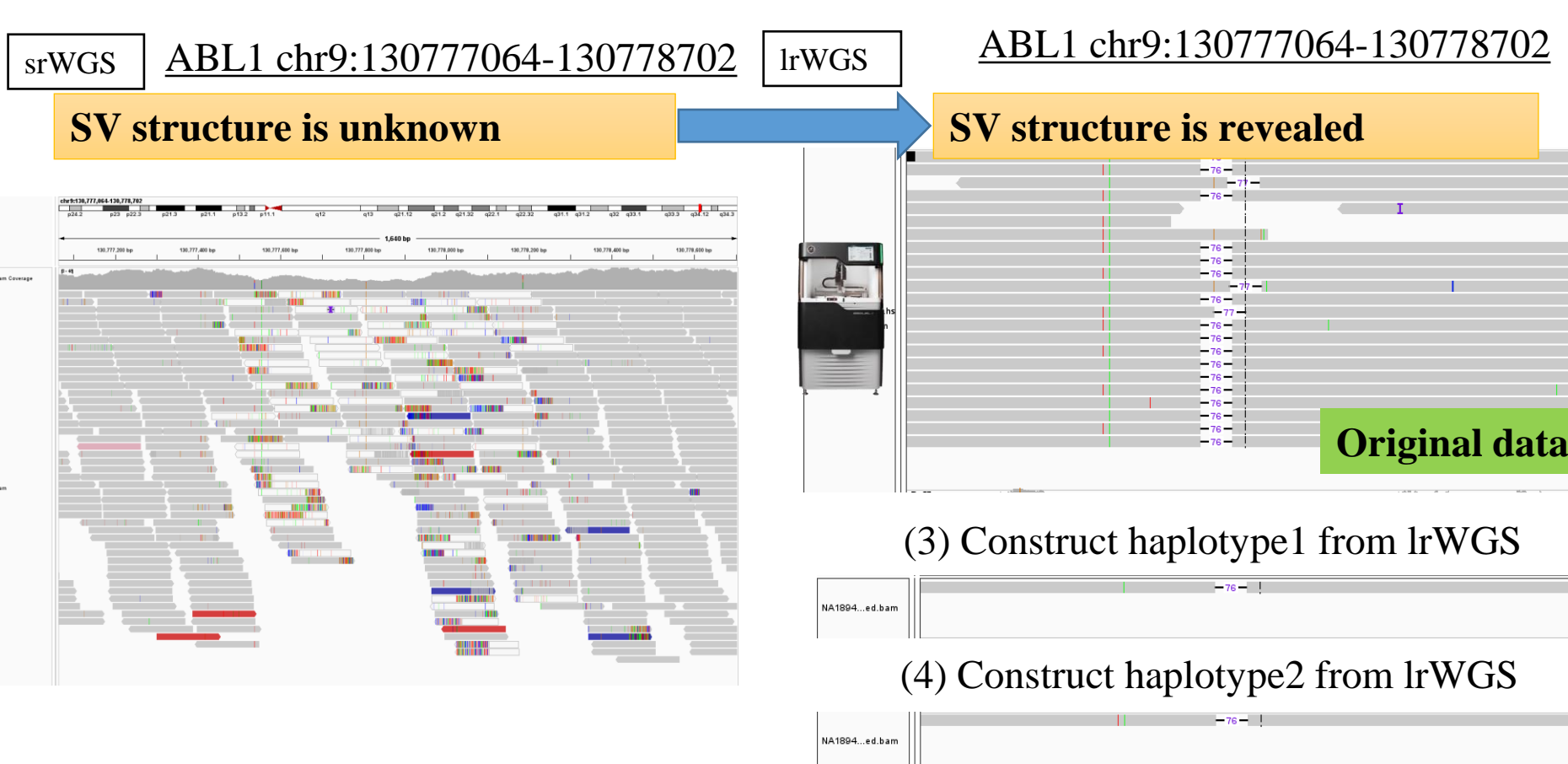


Isoform pattern in Heart Tissue with base level resolution

Construct a web portal to decipher 20% of SVs based on open Japanese whole-genome information(\*) from long-read sequencer.

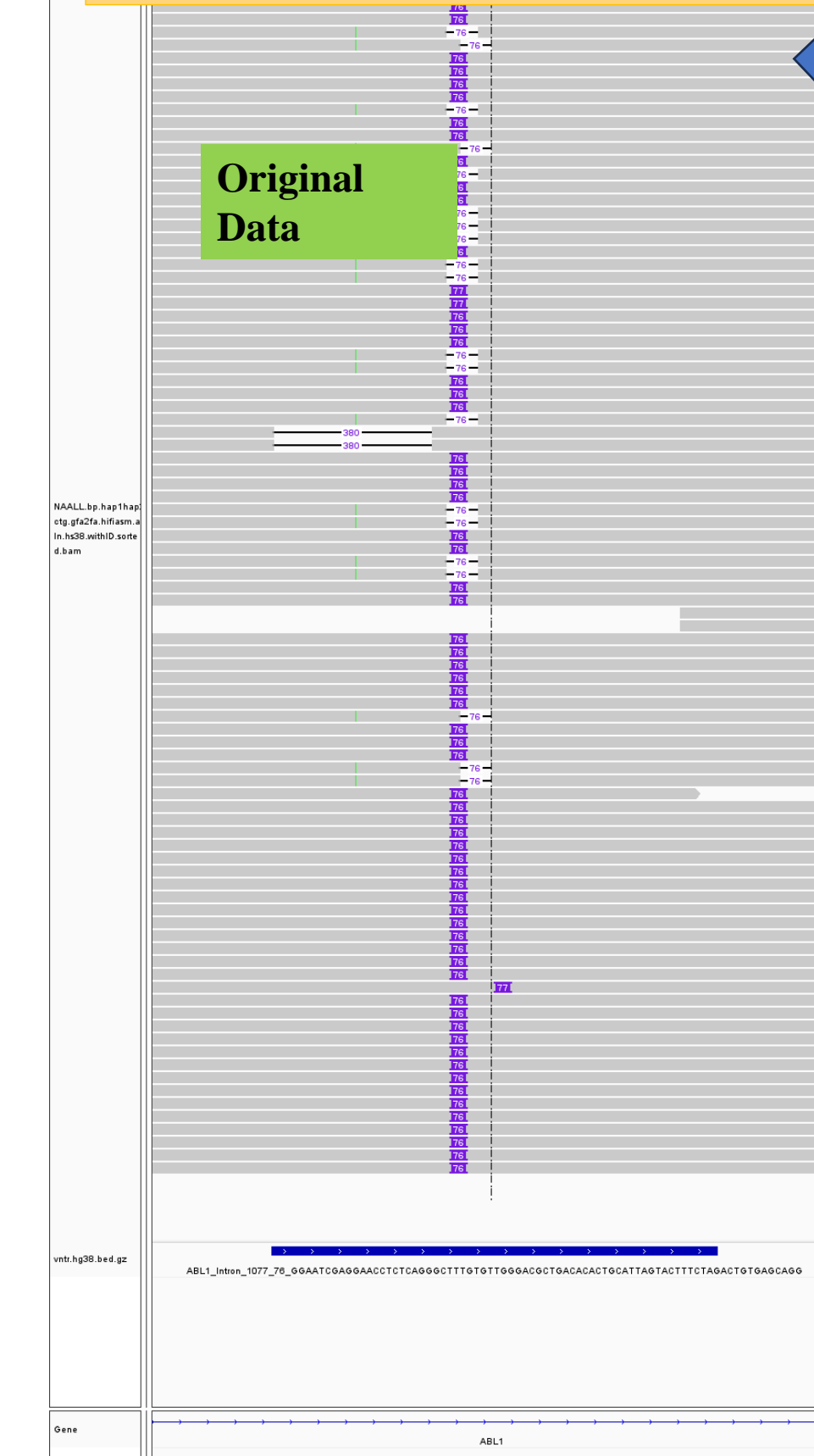
#### Invidial Level Analysis

\* For historical samples and information that have established scientific value, are widely used for research purposes, and are publicly available, see the supplementary material for details.



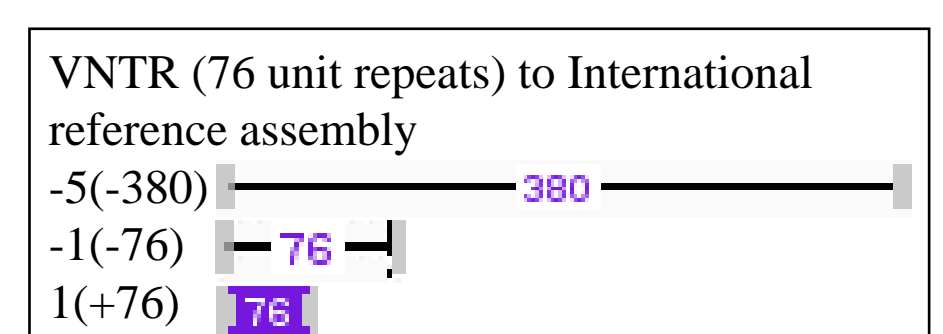
#### Population Level Analysis

What kind of Genomic diversity exists at ABL1 VNTR region among Japanese population?



Public one-hundred Japanese sequence data derived from long-read sequencing technology.

JoGo can inform 3 groups in Japanese population.



① Cover SVs of minor allele frequency > 1% among

Users can capture the haplotype structure in Japanese, e.g., combinations of SNPs and complexity in near chromosomal regions

## JoGo Japanese Open Genome and Omics Platform

<https://jogo.csml.org/>

Release in FY2024  
early access is available  
Please contact us

### <3> RDF and external DB

- <https://github.com/dbcls/visc/wiki/VISC-hackathon3.5>
- <https://mgend.med.kyoto-u.ac.jp/>



- Use format of SVs standards by VISC.
- Collaboration to MGEND

- Members of VISC (Variant Information Standardization Collegium) at 11-13/Mar/2023
- Member of the development of MGEND (Medical Genomics Japan Variant Database)

### <4> International Collaboration to CoLoRS

Target 2000 samples

Public Data <https://colordsb.org/>

- International standards of SVs data analysis
- Integration of international SVs statistics to JoGo portal.

- The main differences to JoGo is two points:
  - JoGo handles the integration of omics dataset.
  - JoGo releases base resolution information.

Genic Haplotype Example in JoGo v1: ALDH2 haplotype catalog and their haplotype counts by population

Translate hapid		regionname	genename	aahapid	aalen	chm13v2	grch38p0	total	AFR	AMR	EAS	EUR	SAS
ALDH2_chr12_111761933_111822532	ALDH2	h0001	h0001	m0001	517	1	1	310	87	52	127	6	36
ALDH2_chr12_111761933_111822532	ALDH2	h0002	h0002	m0002	517	0	0	24	0	0	24	0	0
ALDH2_chr12_111761933_111822532	ALDH2	h0003	h0003	m0003	517	0	0	4	0	0	0	0	4
ALDH2_chr12_111761933_111822532	ALDH2	h0004	h0004	m0004	517	0	0	1	0	0	1	0	0
ALDH2_chr12_111761933_111822532	ALDH2	h0005	h0005	m0005	517	0	0	1	1	0	0	0	0

Transcript hapid		regionname	genename	aahapid	minorhapid	aalen	grch38p0	chm13v2	total	AFR	AMR	EAS	EUR	SAS
ALDH2_chr12_111761933_111822532	ALDH2	h0001	h0001	m0001	m0001	517	0	1	304	84	52	124	6	36
ALDH2_chr12_111761933_111822532	ALDH2	h0001	h0001	m0002	m0002	517	0	0	3	0	0	3	0	0
ALDH2_chr12_111761933_111822532	ALDH2	h0001	h0001	m0003	m0003	517	0	0	1	1	0	0	0	0
ALDH2_chr12_111761933_111822532	ALDH2	h0001	h0001	m0004	m0004	517	0	0	1	1	0	0	0	0

Gene hapid		regionname	genename	aahapid	minorhapid	mminorhapid	aalen	grch38p0	chm13v2	total	AFR	AMR	EAS	EUR	SAS
ALDH2_chr12_111761933_111822532	ALDH2	h0001	h0001	m0001	m0001	m0001	517	0	1	304	84	52	124	6	36
ALDH2_chr12_111761933_111822532	ALDH2	h0001	h0001	m0001	m0001	m0001	517	1	0	1	0	0	0	0	0

Total number of cataloged genic haplotypes in JoGo v1

Region	# of Translate hapid	# of Transcript hapid	# of gene hapid
total	161,840	285,163	2,743,929
AFR	87,599	154,067	881,133
AMR	60,558	97,192	565,526
EAS	73,557	115,283	1,071,792
EUR	35,129	48,749	149,784
SAS	49,218	75,332	377,944

Distribution of GRCh38/CHM13v2 major hapid



Distribution of GRCh38/CHM13v2 transcript hapid of the most frequent translate hapid



Hapid distribution between GRCh38 and CHM13v2



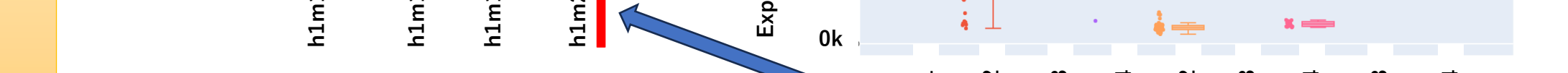
Translate hapid of CHM13v2



Translate hapid of GRCh38



Advanced Analysis for defined hapids in JoGo v1



### <Result・Impact>

The project develops a genomic and omics portal to international samples including original Japanese samples with base pair level resolutions and collaborates with CoLoRS international SV consortium to integrate the global haplotype information from 2000 samples with the standardized data analysis pipelines. The obtained SVs will be shared with the RDF format defined by the VISC. JoGo contributes to Japanese genome information for understanding haplotypes from rare diseases to multifactorial diseases.

当研究に限らず、バイオメディカル情報解析に関わりたい助教・研究員・技術開発を募集中です。興味のあるかたは長崎までご連絡ください。  
We are hiring researchers and technical staff in the laboratory of biomedical information analysis. Feel free to contact to [nagasaki@csml.org](mailto:nagasaki@csml.org). <https://nagasaki.csml.org/>