December 15, 2014

VaDE

Database manual version 1.1

Biomedical Informatics Laboratory

Department of Molecular Life Science

Division of Basic Medical Science and Molecular Medicine

Tokai University School of Medicine

Contents

1. What's VaDE?

The VarySysDB Disease Edition (VaDE) is a database of human genome polymorphisms involved in traits such as various disease susceptibilities or drug responses, which have been collected from a number of academic papers.

Recently, many genome-wide association studies (GWASs) have been performed and identified various disease-associated genomic polymorphisms. These data are valuable for medical research. However, use of these data has been difficult for general life scientists because the information have been described in numerous academic articles. We therefore started a project to construct a database of human genome polymorphisms involved in various traits from 2013. In principle, the information has been obtained from a large number of collected GWAS articles. The VaDE database was born by integrating with the VarySysDB database of functional information of human genome polymorphism that has been previously built.

Most of the data that has been registered in VaDE is genomic polymorphisms associated with diseases or drug responses. Besides, it contains a number of genomic polymorphisms associated with general traits such as height or weight. VaDE provides a wealth of information about these genomic polymorphisms such as odds ratios, β values, sample populations, *p* values and so on. Furthermore, VaDE evaluates reproducibility of associations in multiple independent studies.

By using VaDE, you can easily search and get the reliable information of genomic polymorphisms associated with disease susceptibility. This information can be used in researches for predicting disease risks, which lead to application to preventive medicine in the future. In addition, data registered in VaDE is available in a wide range of fields such as drug discovery, forensic medicine, and anthropology, so the role of this database will become increasingly important in the future.

The VaDE database address is http://bmi-tokai.jp/VaDE/.

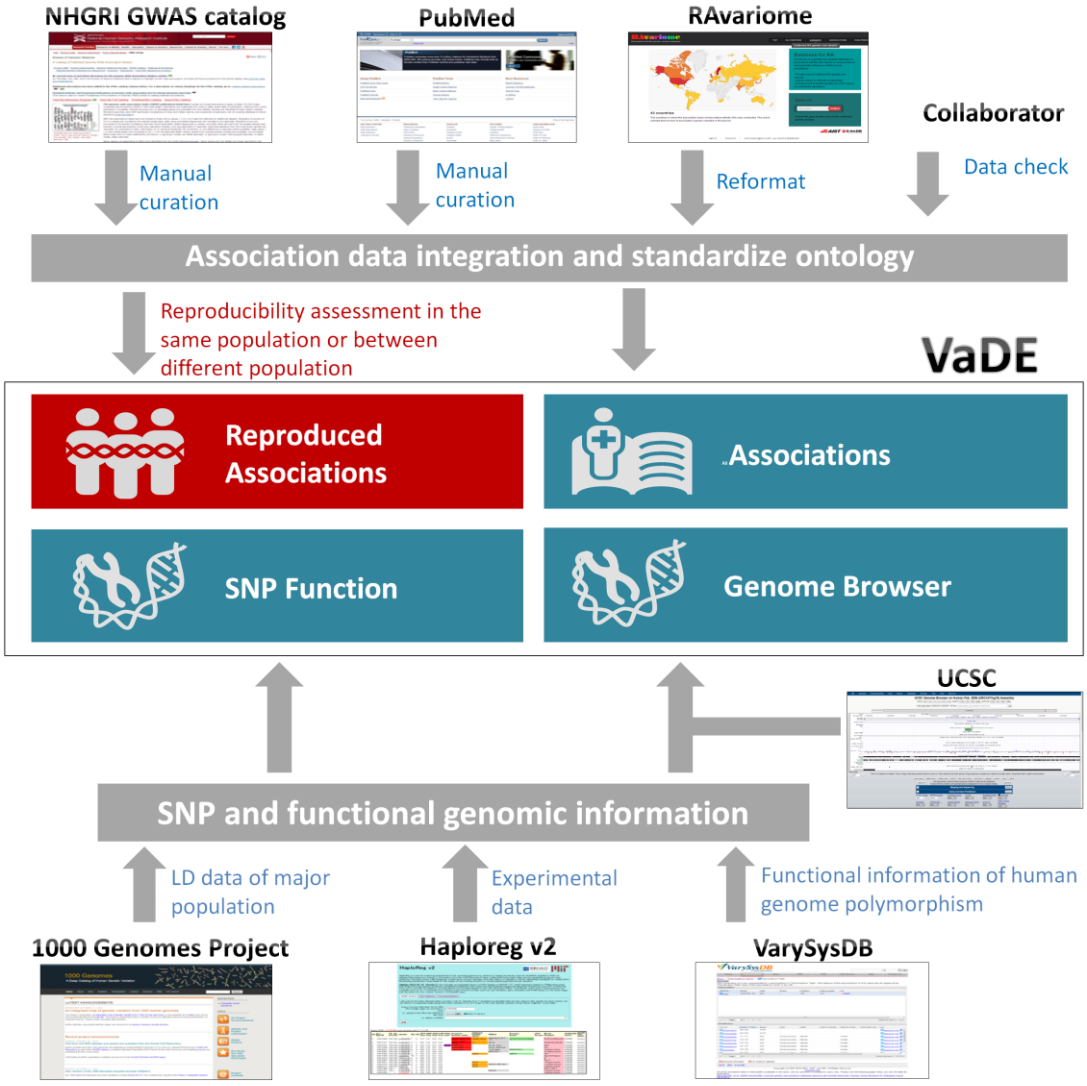

Figure 1. Flow of the VaDE database construction

2. Major pages in VaDE and their links

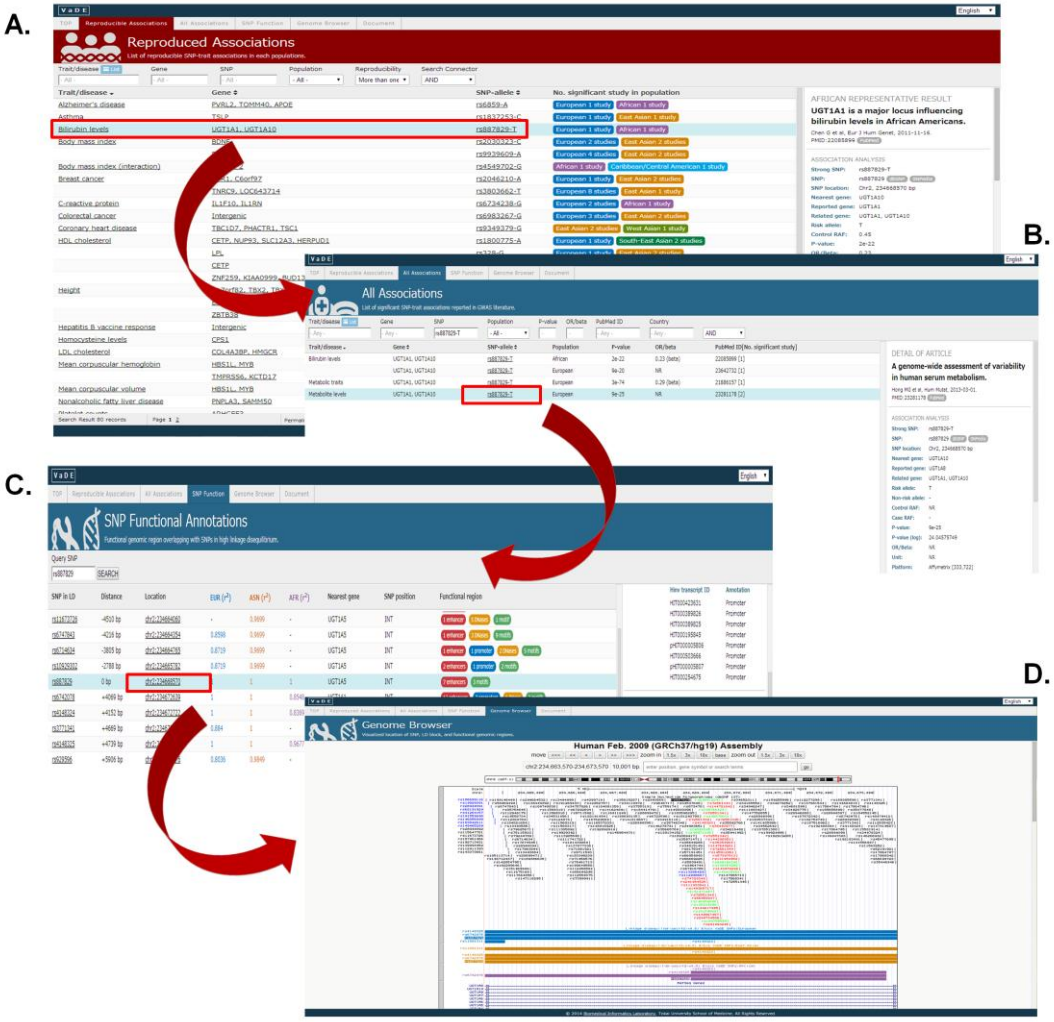

Figure 2. Links among the major pages in VaDE: (A) Reproduced Associations page, (B) All Associations page, (C) SNP Functional Annotations page, (D) Genome Browser page

There are hyperlinks among all major pages of VaDE, so you can move to pages that provide more detailed SNP information in a step-wise manner. Each page provides the following information: (A) a list of reproducible SNP-trait associations in each population, (B) a list of detailed information of SNP-trait associations, (C) a list of SNPs in high linkage disequilibrium with the selected SNP and their functional information, and (D) Genome Browser.

3. Detailed information of pages and search system

3.1 Top page

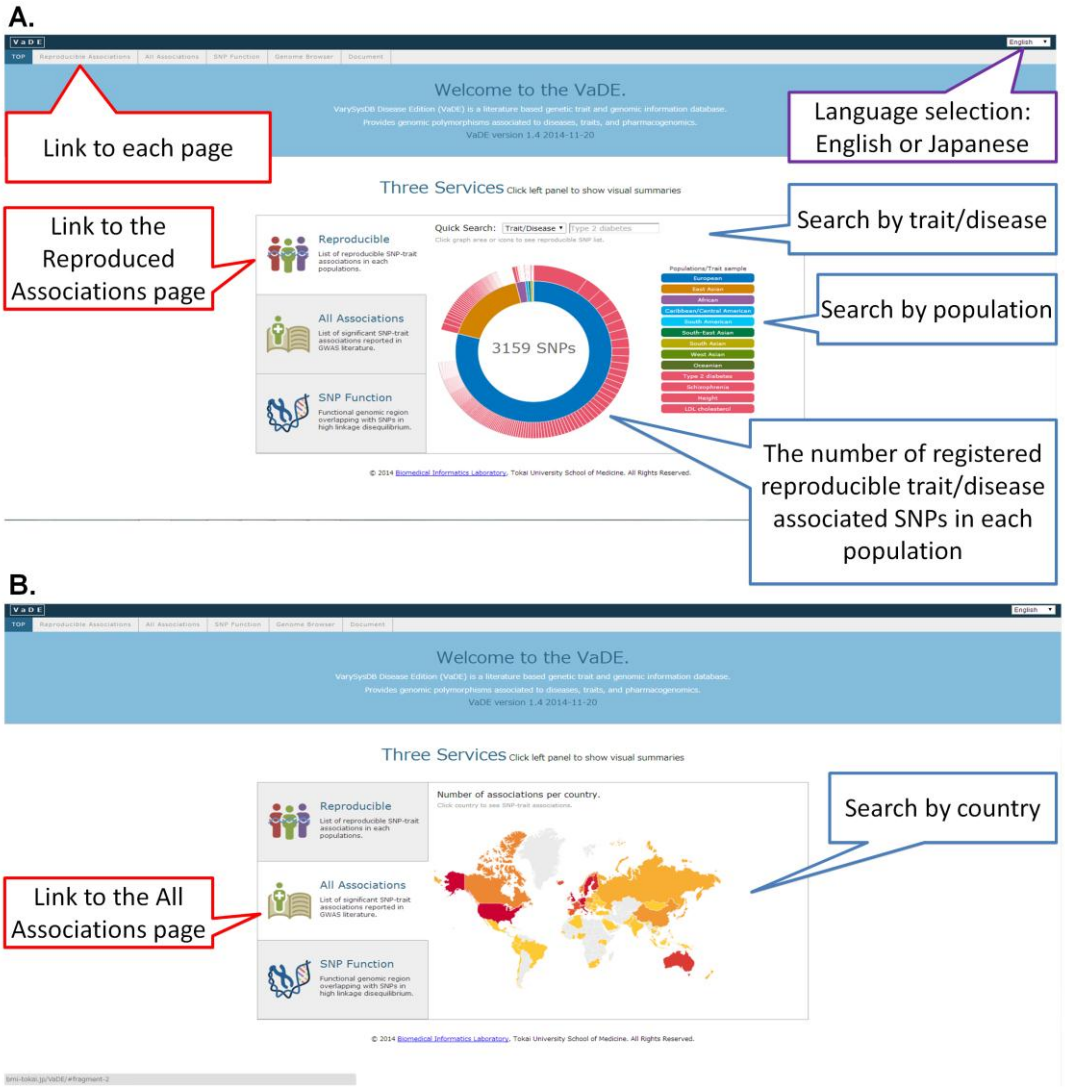

Figure 3-1. Top page: (A) search window for reproduced associations in each trait and population, (B) search window for all SNP-trait associations in a country

[Page description] On the Top page (A), you can search by trait/disease names or population names, and move to the Reproduced Associations page. On the Top page (B), you can search by country names on the world atlas, and move to the All Associations page. Here, the search phrases need to be written in English (The same shall apply hereafter).

3.2 Reproduced Associations page

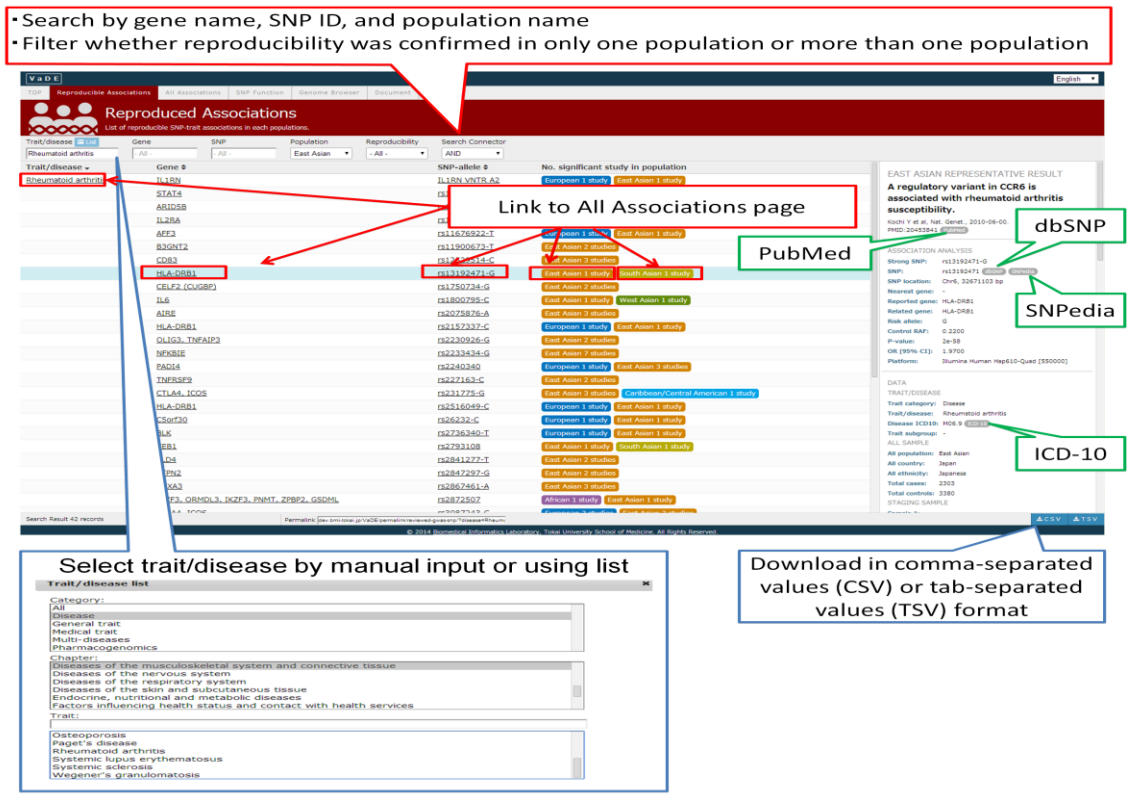

Figure 3-2. Reproduced Associations page

[Page description] The Reproduced Associations page provides information of reproduced trait/disease associated SNPs reported in two or more studies with independent samples for each population. In the left section, a list reproducible SNPs is displayed with trait/disease, reported gene, SNP-allele, population examined, and the number of their significant study (GWAS: P-value <1.0×10^{-5}, replication study: P-value <0.05). When you select an item in the list, you can move to the All Associations page with search by the item (Refer to the next page). In the right section, detailed information of selected SNP-trait associations from the study is shown, using the largest number of cases as a representative result for each population. There are links to PubMed, dbSNP, SNPedia, and ICD-10. You can download all the data by clicking of the CSV or TSV buttons.

[Search method] You can search association data by trait/disease name, gene name (gene symbol), SNP ID (dbSNP rs number), population name, and the condition of reproducibility (in one region or in multiple regions). Figure 3-2 shows a result of search by Rheumatoid arthritis and East Asian.

3.3 All Associations page

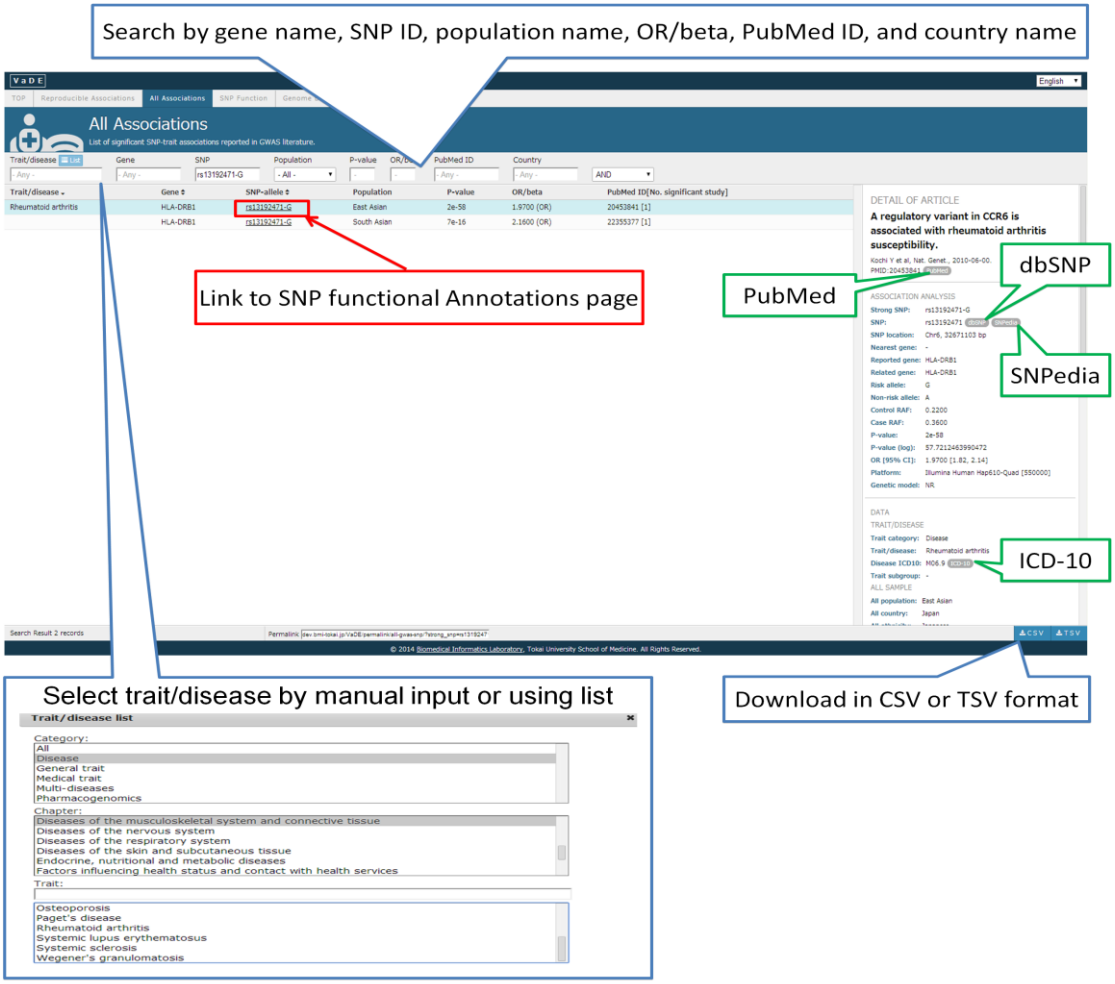

Figure 3-3. All Associations page

[Page description] The All Associations page provides information of all trait/disease associated SNPs registered in VaDE regardless of reproducibility. In the left section, a list of SNPs is displayed with trait/disease, reported gene, SNP-allele, population examined, *P*-value, odds ratio (OR)/beta-value, PubMed ID of original article and the number of their significant study (GWAS: P-value $<1.0\times10^{-5}$, replication study: P-value <0.05). When you select a SNP-allele, you can move to the Functional Annotations page with search by the SNP-allele (Refer to the next page). In the right section, detailed information of each SNP-trait association is shown. There are links to PubMed, dbSNP, SNPedia, and ICD-10. You can download all the data by clicking the CSV or TSV button.

[Search method] You can search association data by trait/disease name, gene name (gene symbol), SNP ID (dbSNP rs number), population name, upper limit of *P*-value, lower limit of OR/beta-value, PubMed ID, and country name.

3.4 SNP Functional Annotations page

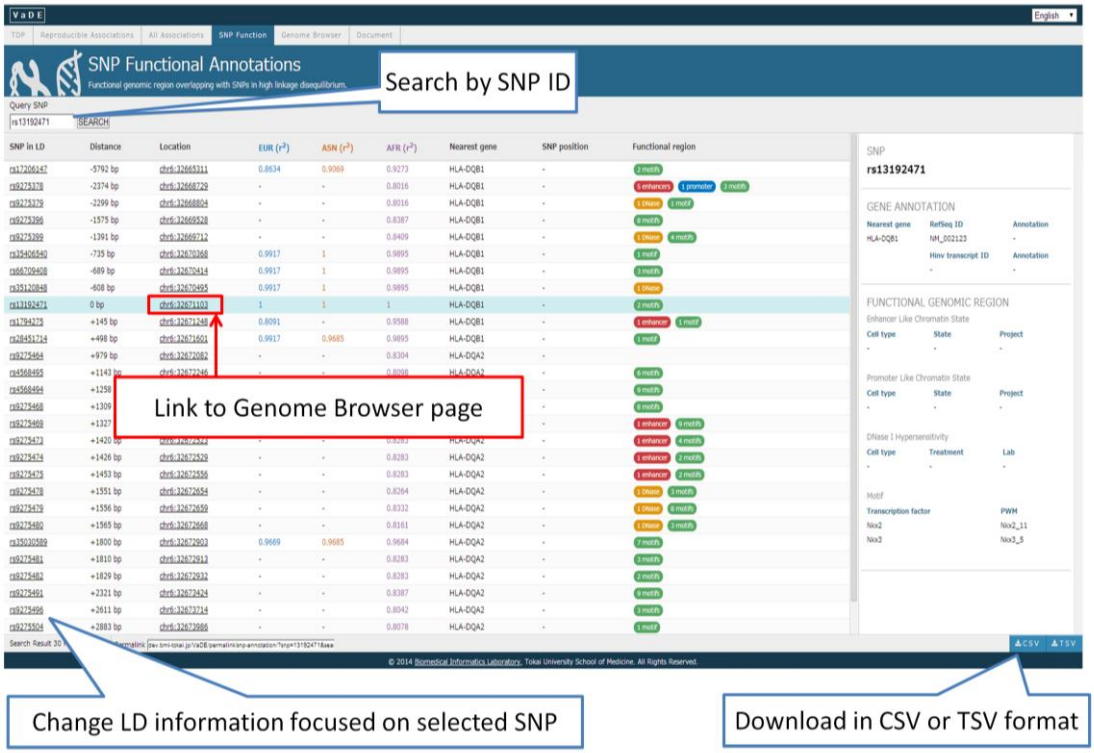

Figure 3-4: SNP Functional Annotations page

[Page description] The SNP Functional Annotations page provides information of query SNP, SNPs in high linkage disequilibrium (LD) with the query SNP, and functional genomic region overlapping with each SNP location. In the left section, a list of SNPs in LD (r^2>0.8) is displayed with their location, r^2 in each major population (European, Asian, and African), nearest gene, SNP position, and functional region (enhancer, promoter, DNase I, or motif). When you select the location of a target SNP, you can move to the Genome Browser page with search by the SNP and the location (Refer to the next page). In the right section, detailed information of each SNP is shown. You can download the list of the data by clicking the CSV or TSV button.

[Search method] You can search these data by SNP ID (dbSNP rs number). Also, you are able to change a query SNP by selecting the other SNP in LD.

3.5 Genome Browser page

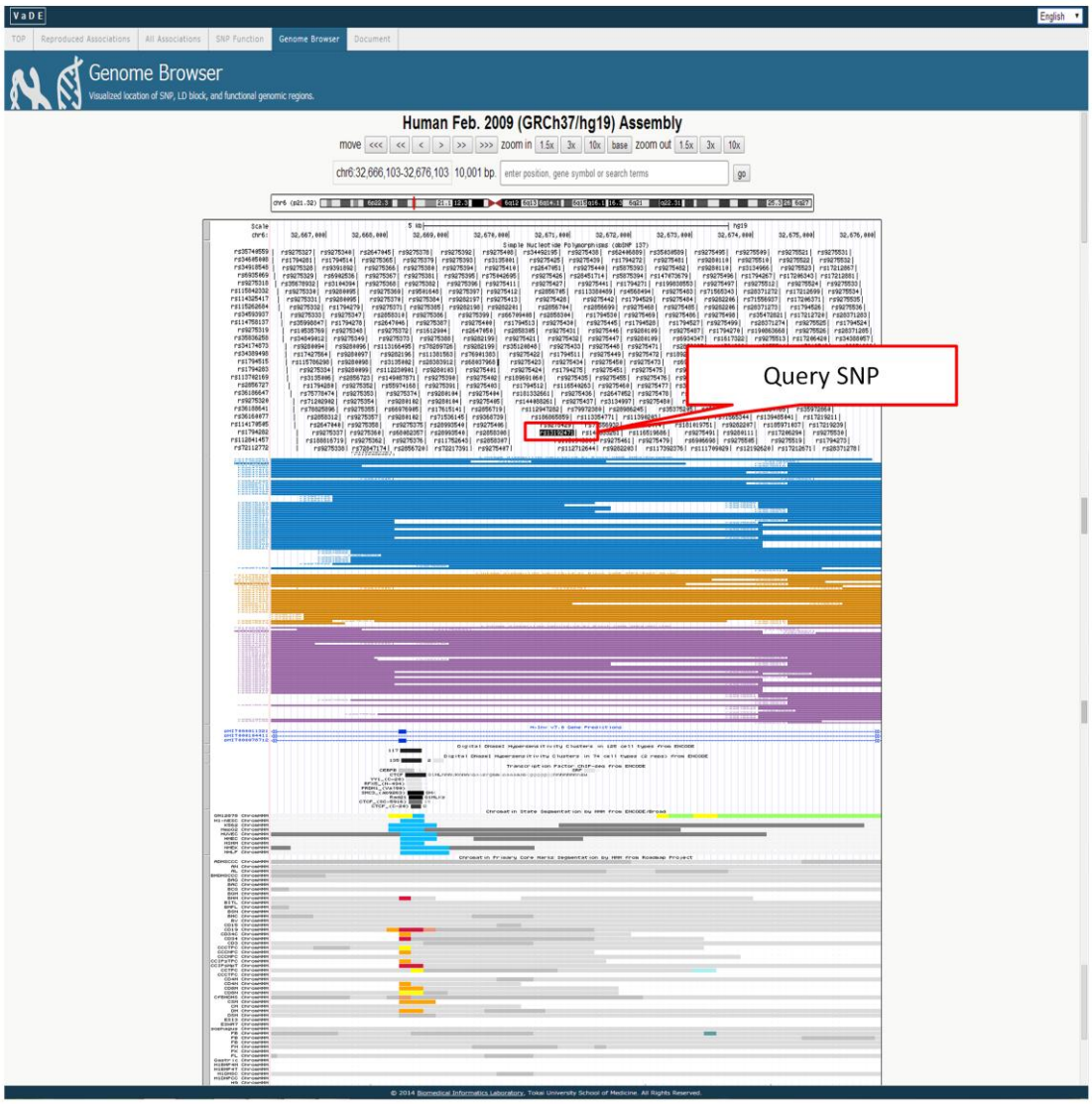

Figure 3-5. Genome Browser page

[Page description] The Genome Browser page that incorporated UCSC Genome Browser provides information of positional relationship on the genome with a focus on a query SNP. You can find information of all registered SNPs near the query SNP, LD blocks in each major population (European, Asian, and African), genes, and functional genomic regions.

Color	State	High frequency chromatin marker (frequency over 50%)
1	Active Promoter	H3K4me2, H3K4me3, H3K27ac, H3K9ac
2	Weak Promoter	H3K4me1, H3K4me2, H3K4me3
3	Inactive/poised Promoter	H3K27me3, H3K4me2
4	Strong enhancer	H3K4me1, H3K4me2, H3K4me3, H3K27ac, H3K9ac
5	Strong enhancer	H3K4me1, H3K4me2, H3K27ac
6	Weak/poised enhancer	H3K4me1, H3K4me2
7	Weak/poised enhancer	H3K4me1
8	Insulator	CTCF
9	Transcriptional transition	H3K36me3(low), H4K20me1(low), H3K4me1(low)
10	Transcriptional elongation	H3K36me3(low)
11	Weak transcribed	H3K36me3(very low), H4K20me1(very low)
12	Polycomb-repressed	H3K27me3(low)
13	Heterochromatin; low signal	(no signal)
14	Repetitive/Copy Number Variation	(low freq. of all chromatin marks)
15	Repetitive/Copy Number Variation	(high freq. of all chromatin marks)

Table 3-5-1. Chromatin in state segmentation by HMM from ENCODE/Broad

Color	State
1	TSS_poised
2	TSS_flanking_more_upstream
3	TSS_active
4	TSS_weak
5	TSS_flanking_downstream
6	TSS_flanking_more_downstream
7	Transcription
8	Transcription_weak
9	Transcription_Enhancer-like
10	Transcription_Enhancer-like_(short_genes)
11	Enhancer_weak_1
12	Enhancer_weak_2
13	Enhancer_active
14	Enhancer_active_with_weakK4me1_strong_K27ac
15	Enhancer_poised
16	Repressed_polycomb_weak
17	Repressed_polycomb
18	H3K9me3_K27me3
19	Zinc_finger_genes_H3K36me3_K9me3
20	Heterochromatin_at_repeats
21	Heterochromatin
22	Quiescent_1
23	Quiescent_2

24	Quiescent_3
25	Quiescent_low_H3K9ac

Table 3-5-2. Chromatin in Core Marks segmentation by HMM from Roadmap Project

Color coding of chromatin segmentation in the Genome Browser is shown in Table 3-5-1 and Table 3-5-2.

4. Additional information

Further information and utilization for the VaDE database is presented in the following paper.

Nagai Y, Takahashi Y, and Imanishi T (2014) VaDE: a manually-curated database of reproducible associations between various traits and human genomic polymorphisms. *Nucleic Acids Research*, Database Issue 2014;doi:10.1093/nar/gku1037.

Also, statistics of the VaDE database is presented in the online Document page (http://bmi-tokai.jp/VaDE/document/).

If you have any questions, please contact us by email to the address below.

vade@ml.tokai-u.jp