

Web-Based Database and Viewer of East Asian Copy Number Variations

Ji-Hong Kim[†], Hae-Jin Hu[†] and Yeun-Jun Chung*

Department of Microbiology, Integrated Research Center for Genome Polymorphism, The Catholic University of Korea School of Medicine, Seoul 137-701, Korea

Abstract

We have discovered copy number variations (CNVs) in 3,578 Korean individuals with the Affymetrix Genome-Wide SNP array 5.0, and 4,003 copy number variation regions (CNVRs) were defined in a previous study. To explore the details of the variants easily in related studies, we built a database, cataloging the CNVs and related information. This system helps researchers browsing these variants with gene and structure variant annotations. Users can easily find specific regions with search options and verify them from system-integrated genome browsers with annotations.

Keywords: copy number variation, database, genome browser

Availability: The East Asian CNV database (EACDB) can be accessed at www.irgcp.com/EACNVDB.html. Some configuration and server installation should be done before the system integration. Contact yejun@catholic.ac.kr for detailed information.

Introduction

Copy number variation (CNV) is a common type of structural variation in the human genome. They have been suggested to be related to disease susceptibility or human phenotype diversity [1-3]. Current genome-wide association studies of CNVs are attempting to find how they are related to disease susceptibility or phenotypic diversity. Due to the fact that frequencies of CNVs show ethnic differences [4-7], a finding of disease or phenotype association of a specific CNV in one pop-

ulation is hard to generalize to other populations. However, by comparing the frequencies of target CNVs among different ethnic groups, we could assume the population-specific disease susceptibility or phenotype difference.

In this regard, CNV frequency information of various populations has become a major concern of population or disease association studies.

With increasing interest in CNVs, many CNV projects have been announced recently. Many of them have already established large databases of CNVs from many different ethnic groups (Database of Genomic Variants [DGV], <http://projects.tcag.ca>; The Copy Number Variants Projects, <http://www.sanger.ac.uk/research/areas/humangenetics/cnv/>) [8]. There are also several databases for supporting population-specific studies, including the CNV Control Database for Japanese (<http://gwas.lifesciencedb.jp>), the Singapore Human Mutation and Polymorphism Database (shmed.bii.a-star.edu.sg) [9], and the Thailand Mutation and Variation Database (www4a.biotec.or.th) [10]. The accumulation of CNV information seems promising, in that we can gradually extend the knowledge of CNVs everywhere in the world. However, it will take a long time to fill the gap of CNV databases covering the whole population.

As a step to fill this gap, we discovered a Korean CNV with Affymetrix Genome-Wide SNP array 5.0 (Affymetrix, Santa Clara, CA, USA) in a previous study [11]. In this study, we built a Korean database based on the findings of 4,003 copy number variation regions (CNVRs). To build a tool that is easily accessible via the web for Korean-specific CNV studies, we built a viewer for browsing these CNVs.

Features and Results

System requisites

Java Runtime Environment of Sun Microsystems 1.6.0 (Oracle, Redwood City, CA, USA) or equivalent is required, since the system is written in Java language. We used MySQL database for storing and retrieving CNV data and GBrowse2 [12] for drawing regional information. Both are freely available from their distribution websites.

Since the main search pages are written in Java Server Page (JSP) language, Apache Tomcat is needed

[†]These authors contribute equally to this work.

*Corresponding author: E-mail yejun@catholic.ac.kr

Tel +82-2-2258-7343, Fax +82-2-596-8969

Received 31 January 2012, Revised 16 February 2012,

Accepted 18 February 2012

for the application server and Apache HTTP server is needed for Gbrowse2 viewer pages (Fig. 1).

Data

CNVs of our previous study were retrieved from the Affymetrix Genome-Wide Human SNP array 5.0 of 3,578 Korean individuals. A total of 4,003 CNVRs were defined, and 2,077 CNVRs (51.9%) were potentially novel. The annotation data for genes were collected based on the Human Mar. 2006 NCBI36/hg18 build, and reference structure variants were retrieved from DGV (hg18.v8, aut,2009).

Database and viewer

Our web-based database viewer can display previously

discovered CNVs by their positions (Fig. 2). Users can also filter out CNVs based on the DGV overlapped regions, CNV type (Gain/Loss/Complex), or their frequencies. Each selected region could be diagnosed in detail by clicking on it. DGV and OMIM ID columns are linked with corresponding websites, and CNVR position columns are linked with the genome browser. The genome browser is integrated based on the open source project GBrowse2. Users can seek or zoom in/out of CNVs across the chromosome by entering positions or clicking zoom buttons. GBrowse2 can also display interesting areas by dragging the region bar without reloading the entire page. Gene information of the selected area is also displayed, and details will be given on separate pop-up page by clicking on it.

Discussion

For a fast-paced research environment, a viewer for searching and observing data in one step is very handy. We hope our system can help researchers who are interested not only in our target polymorphism study but also in a viewer for polymorphisms for general purposes.

We chose a web-based platform for the tool because of its usability and ease of maintenance. Since its workload is very small as an input query for a viewer, a web-based platform does not have drawbacks, like other calculation applications with very large data.

Acknowledgements

This study was supported by a grant of the Korea Health 21 R&D Project, Ministry of Health and Welfare, Republic of Korea (A040002). We thank the KARE consortium for providing the original genotyping data.

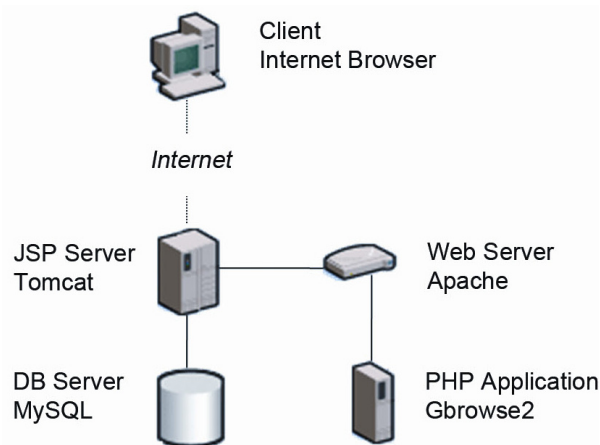


Fig. 1. System Structures of East Asian Copy Number Variation Database (EACDB). JSP: PHP (Hypertext Preprocessor).

EACDB East-Asian CNV Database
 DGV Ver.: hg18.v8.aug.2009
 Genome Data Assembly: Build 36, Mar.2006

HOME About Search Site Map Contact Us

1 1 248911641 ~ 248911641

Chr	Start	End	Locus	Freq.	DGV	Type	Omim ID	Phenotype
1	5129	84845	ip36.33	3.94	Y	Complex		
1	711270	808556	ip36.33	0.06	Y	Gain		
1	1381435	1454183	ip36.33	0.06	Y	Gain	612317	
1	1618027	1670811	ip36.33	0.34	Y	Loss	176873	
1	1819067	1869815	ip36.33	0.48	Y	Loss	610171	
1	2447463	2449428	ip36.32	0.06	Y	Complex	606192	
1	2459035	2475181	ip36.32	0.22	Y	Loss		
1	3130438	3142828	ip36.32	0.31	Y	Loss	605557	
1	4022951	4045475	ip36.32	0.06	N	Loss		
1	4162814	4199587	ip36.32	0.06	N	Loss		
1	5021315	5071724	ip36.32	0.08	Y	Loss		
1	5414088	5443191	ip36.31	0.06	Y	Loss		
1	5705308	5718214	ip36.31	0.17	N	Loss		
1	10631379	10638758	ip36.32	0.06	Y	Gain	609895	
1	12775459	12936641	ip36.21	1.9	Y	Complex		
1	14237550	14242955	ip36.21	0.06	Y	Loss		
1	16219857	16231574	ip36.13	0.08	Y	Loss	602024	*Bartter syndrome, type 4, digenic, 602522(3)
1	16891258	16921744	ip36.13	0.5	Y	Complex		
1	16949054	16987221	ip36.13	0.06	Y	Loss		
1	17067742	17139108	ip36.13	0.78	Y	Complex		

2011 Integrated Research Center for Genome Polymorphism
 The Catholic University of Korea, School of Medicine 205, Bangsi-dong, Seocho-gu, Seoul, 1307-707

Ease Asian CNV Database (chr1): 이 206 kbp 정보는 chr1 의 660,000 - 865,999 사이의 정보입니다

Browser Select Tracks Snapshots Community Tracks Custom Tracks Preferences

Search
 [chr1:660,000-865,999] Search
 Download Track Data | Configure | Go
 Save Snapshot | Load Snapshot

Examples chr1
 Data Source
 [Ease Asian CNV Database (chr1)]
 Scroll/Zoom: [x] [y] [z] [206 kbp] [x] [y] [z] Flip

Region
 [chr1:660,000-865,999] 50 kbp
 660k 665k 670k 675k 680k 685k 690k 695k 700k 705k 710k 715k 720k 725k 730k 735k 740k 745k 750k 755k 760k 765k 770k 775k 780k 785k 790k 795k 800k 805k 810k 815k 820k 825k 830k 835k 840k 845k 850k 855k 860k 865k 870k

Genes
 NR6M0115
 LOC648307
 FIN1C
 FLJ39619
 SPTCL1
 MOC

East Asian CNV
 CNV_2

Select Tracks Clear highlighting

The Generic Genome Browser for questions about the data at this site, please contact its webmaster. For support of the browser software only, send email to gmod-gbrowse@lists.sourceforge.net or visit the GMOD Project web pages.

Fig. 2. Database search page and genome browser.

References

1. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, *et al*. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 2007;315:848-853.
2. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. *Nat Rev Genet* 2009;10:551-564.
3. Beckmann JS, Estivill X, Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet* 2007; 8:639-646.
4. Lam KW, Jeffreys AJ. Processes of copy-number change in human DNA: the dynamics of {alpha}-globin gene deletion. *Proc Natl Acad Sci U S A* 2006;103: 8921-8927.
5. Conrad DF, Hurler ME. The population genetics of structural variation. *Nat Genet* 2007;39(7 Suppl):S30-S36.
6. Hegele RA. Copy-number variations add a new layer of complexity in the human genome. *CMAJ* 2007;176: 441-442.
7. Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, *et al*. A comprehensive analysis of common copy-number variations in the human genome. *Am J Hum Genet* 2007;80:91-104.
8. Feuk L, Marshall CR, Wintle RF, Scherer SW. Structural variants: changing the landscape of chromosomes and design of disease studies. *Hum Mol Genet* 2006;15 Spec No 1:R57-R66.
9. Tan EC, Loh M, Chuon D, Lim YP. Singapore Human Mutation/Polymorphism Database: a country-specific database for mutations and polymorphisms in inherited disorders and candidate gene association studies. *Hum Mutat* 2006;27:232-235.
10. Ruangrit U, Srikumool M, Assawamakin A, Ngamphiw C, Chuechote S, Thaiprasarnsup V, *et al*. Thailand mutation and variation database (ThaiMUT). *Hum Mutat* 2008;29:E68-E75.
11. Yim SH, Kim TM, Hu HJ, Kim JH, Kim BJ, Lee JY, *et al*. Copy number variations in East-Asian population and their evolutionary and functional implications. *Hum Mol Genet* 2010;19:1001-1008.
12. Donlin MJ. Using the generic genome browser (GBrowse). *Curr Protoc Bioinformatics* 2009;Chapter 9:Unit 9.9.