Contents available at ScienceDirect

Diabetes Research and Clinical Practice

journal homepage: www.elsevier.com/locate/diabres

International Diabetes Federation

Type 2 diabetes mellitus disease risk genes identified by genome wide copy number variation scan in normal populations

Manasa Prabhanjan [a,1,2], *Raviraj V. Suresh* [b,1,3], *Megha N. Murthy* [b,1,4], *Nallur B. Ramachandra* [b,*]

[a] *Department of Studies in Genetics and Genomics, University of Mysore, Manasagangotri, Mysore 570006, Karnataka, India*

[b] *Genetics and Genomics Lab, Department of Studies in Genetics and Genomics, University of Mysore, Manasagangotri, Mysore 570006, Karnataka, India*

ARTICLE INFO

Article history:
Received 5 August 2015
Received in revised form
3 December 2015
Accepted 26 December 2015
Available online 13 January 2016

Keywords:
Type 2 diabetes mellitus
Copy number variations
Enrichment
Molecular interaction and network

ABSTRACT

Aims: To identify the role of copy number variations (CNVs) on disease risk genes and its effect on disease phenotypes in type 2 diabetes mellitus (T2DM) in 12 random populations using high throughput arrays.

Methods: CNV analysis was carried out on a total of 1715 individuals from 12 populations, from ArrayExpress Archive of the European Bioinformatics Institute along with our subjects using Affymetrix Genome Wide SNP 6.0 array. CNV effect on T2DM genes were analyzed using several bioinformatics tools and a molecular protein interaction network was constructed to identify the disease mechanism altered by the CNVs.

Results: Analysis showed 34.4% of the total population to be under CNV burden for T2DM, with 83 disease causal and associated genes being under CNV influence. Hotspots were identified on chromosomes 22, 12, 6, 19 and 11.Overlap studies with case cohorts revealed significant disease risk genes such as EGFR, E2F1, PPP1R3A, HLA and TSPAN8.

Conclusions: CNVs play a significant role in predisposing T2DM in normal cohorts and contribute to the phenotypic effects. Thus, CNVs should be considered as one of the major contributors in predisposition of the disease.

1. Introduction

Diabetes is a complex and heterogeneous disease with a staggering global impact and the most recent estimates indicate 346 million people worldwide suffer from this disease (WHO Diabetes Fact sheet No. 312, 2011). Type 2 diabetes mellitus (T2DM) is the most common form of diabetes, accounting for >90% of cases, and occurs when peripheral tissue insulin resistance accompanies insufficient β-cell

* *Corresponding author*. Tel.: +91 821 2419781x888; fax: +91 821 2516056.
E-mail addresses: manasa.prabhanjan@gmail.com (M. Prabhanjan), rajrocksvs@gmail.com (R.V. Suresh), megha.murthy.n@gmail.com (M.N. Murthy), nallurbr@gmail.com, nbruom@gmail.com (N.B. Ramachandra).
[1] These authors contributed equally to the work.
[2] Tel.: +91 9035953408.
[3] Tel.: +91 9535692143.
[4] Tel.: +91 9738401432.

http://dx.doi.org/10.1016/j.diabres.2015.12.015
0168-8227/

insulin production. While >80% of diabetes deaths occur in low- and middle-income countries [1]. India and China have the highest reported prevalence of diabetes with 65 and 98 million in 2013, respectively [2].

Diabetes is caused due to complex interaction between genetic and environmental factors, like poor life style, diet, physical inactivity and overweight. Genetic factors play a major role in causal of T2DM; however, identification and understanding of genetic factors were of great challenge. Genetic variation in the human genome exists in different forms; from single base pair to large structural variation. In recent times, as the technology has improved; SNP studies, large scale association studies, and next generation sequencing were carried out which helped in the better understanding of T2DM [3]. Comparative genomic hybridization (CGH) technique has helped us know about copy number variation (CNVs) and its effect on human genome [4]. Understanding the CNVs is critical for the proper study of disease-associated changes because segmental CNVs have been demonstrated in developmental disorders and susceptibility to disease [5,6]. Therefore, analysis of CNVs at the whole-genome level is required to create a baseline of human genomic variation [7].

Large studies on T2DM individuals have resulted in identification of several SNPs and CNVs linked directly or associated with the disease [8]. As diabetes occurs in general population; studies should be performed in normal population genomes as they might be susceptible to the disease or in the process of manifestation. Therefore, the present investigation is focused on genome wide CNV association on T2DM susceptible/associated genes in 12 normal populations to identify the effect of these T2DM genes when associated CNVs on normal populations.

2. Materials and methods

The present study involves Genome wide CNV analysis of 1774 individuals from twelve random populations using Affymetrix Genome Wide SNP 6.0 array. Raw unprocessed data of 11 populations except Indian population was obtained from International HapMap Consortium [9] and ArrayExpress Archive. The HapMap Consortium covers four populations namely CEU (CEPH collection)—90 samples, CHB (Han Chinese in Beijing, China)—45 samples, JPT (Japanese in Tokyo, Japan)—45 samples and YRI (Yoruba in Ibadan, Nigeria)—90 samples. Samples for 31 Tibetan samples, 155 Chinese samples, 472 of Ashkenazi Jews replicate 1, 480 of Ashkenazi Jews replicate 2, 204 individuals from Taiwan, 55 from Australia and 64 from New World population (Totonacs and Bolivians) (Table 1) were obtained from the ArrayExpress Archive of the European Bioinformatics Institute. The accession numbers of populations obtained from ArrayExpress are E-GEOD-21661, E-GEOD-29851, E-GEOD-30481, E-GEOD-15826, E-GEOD-23636, E-GEOD-23201, E-GEOD-33355 and E-GEOD-33356. Lastly, 43 Indian samples obtained from randomly selected 12 families residing in Karnataka, India. The age of all subjects under the study falls in the range 13–73 years. Fifty-nine individuals were removed from our analysis from the total of 1774 individuals of 12 populations as they did not pass QC thresholds making the final count to 1715. Data has also been made publicly accessible through the University of Mysore Genome Centre Database (URL: http://umgc.uni-mysore.ac.in/index.php/search/cnv). Five milliliters EDTA blood was collected from each member of the Indian study group and genomic DNA was extracted using Promega Wizard Genomic DNA purification kit. The isolated DNA was quantified by Bio-photometer and gel electrophoresis. This research was approved by the University of Mysore Institutional Human Ethics review committee (IHEC). Written informed consent was obtained from all sample donors and the IHEC approved the sample consent procedure. Written informed consent was obtained from parents/guardians in the cases of participants being minors.

2.1. *Genotyping*

Genome-wide genotyping was performed using an Affymetrix Genome-wide Human SNP Array 6.0 chip and Affymetrix CytoScan High-Density (HD) Array having 1.8 million and 2.6

Table 1 – Total CNV presence in 12 populations and its nature of occurrence with their impact on different populations.

Populations	Individuals assessed	Total CNV count	Duplication of CNV's	% DUP	Size (kb)	Deletion of CNV's	% DEL	Size (kb)	Genes overlapped	
									Intact	Partial
Hap Map-YRI-Africa	90	35	2	5.71	381	33	94.28	17,038	31	4
Hap Map-CEU-Europe	90	69	11	15.94	2719	58	84.05	40,091	61	8
Ashkenazi Jews I	464	260	211	81.15	59047	49	23.22	12,283	215	45
Ashkenazi Jews II	480	240	200	90.90	55419	40	18.18	4995	188	52
Hap Map-CHB-China	44	14	2	14.28	378	12	85.71	4076	12	2
CHINA	155	37	23	62.16	6081	14	38.88	2160	26	11
TIBET	31	19	13	68.42	3015	6	31.57	609	13	6
INDIA	38	19	17	89.47	5917	2	10.52	375	16	3
Hap Map-JPT-JAPAN	45	22	3	13.63	512	19	86.36	3163	20	2
AUSTRALIA	53	85	67	78.82	13583	18	21.17	2176	57	28
New world	41	57	30	52.63	8119	27	47.36	3405	48	9
TAIWAN	184	122	114	93.44	29657	8	6.55	1100	102	20
TOTAL	1715	979	692	70.68	184828	285	29.17	91,471	789	190