ORIGINAL ARTICLE

DOI: 10.4274/jcrpe.galenos.2022.2022-3-11

J Clin Res Pediatr Endocrinol 2022;14(3):293-301

Comparative Analyses of Turkish Variome and Widely Used Genomic Variation Databases for the Evaluation of Rare Sequence Variants in Turkish Individuals: Idiopathic Hypogonadotropic Hypogonadism as a Disease Model

Leman Damla Kotan

Çukurova University Faculty of Medicine, Department of Pediatric Endocrinology, Adana, Turkey

What is already known on this topic?

The absence of population-specific genetic variation reference databases causes misleading results in rare variant evaluations.

What this study adds?

This study confirmed that Turkish Variome could represent the Turkish population for rare genomic variant evaluation.

Abstract

Objective: With the increasing use of whole-exome sequencing, one of the challenges in identifying the causal allele for a Mendelian disease is the lack of availability of population-specific human genetic variation reference databases. The people of Turkey were not represented in GnomAD or other publicly available large databases until recently, when the first comprehensive genomic variation database, Turkish Variome (TRV), was published. The aim of this study was to evaluate whether TRV or other publicly available large genomic variation databases can reliably be used for rare disease variant evaluation in Turkish individuals.

Methods: Sixty non-disease-causing, non-synonymous variants (minor allele frequencies >1%) were identified in 58 genes that are known to be associated with idiopathic hypogonadotropic hypogonadism from a large Turkish patient cohort. The allelic frequencies of these variants were then compared with those in various public genomic variation databases, including TRV.

Results: Our cohort variants showed the highest correlations with those in the TRV, Iranome, and The Greater Middle East Variome, in decreasing order.

Conclusion: These results suggest that the TRV is the appropriate database to use for rare genomic variant evaluations in the Turkish population. Our data also suggest that variomes from geographic neighborhoods may serve as substitute references for populations devoid of their own genomic variation databases.

Keywords: Allele frequency, Turkish Variome, variant evaluation, genomic variation databases

Introduction

The widespread use of next-generation sequencing (NGS), particularly whole-exome sequencing (WES), in medical practice, has resulted in massive data accumulation (1). In order to accurately interpret the differences in the DNA sequences of individuals, criteria based on specific parameters are used. One of the essential parameters is allele frequency (AF), which represents the prevalence of a gene variant in a given population. Variants with minor AFs less than 1% are considered rare and can play a causative role in Mendelian and complex disorders. Genetic alterations observed with a much higher frequency than expected for the disease in a population are generally interpreted as benign (2,3). As many variants are proven to be population-specific, large databases evolved into a comprehensive body of data comprising of datasets from individual subpopulations

Address for Correspondence: Leman Damla Kotan MD, Çukurova University Faculty of Medicine, Department of Pediatric Endocrinology, Adana, Turkey
E-mail: damlakotan@gmail.com, dkotan@cu.edu.tr **ORCID:** orcid.org/0000-0001-6176-8986

Conflict of interest: None declared
Received: 14.03.2022
Accepted: 06.04.2022

(4). Failure to use population-specific databases can lead to unreliable or even misleading results in variant evaluation.

The people of Turkey live in the Anatolian peninsula, which is geographically at the crossroads of three major continents, through which major population movements have occurred during all periods of human history. Therefore, it was thought that this geographic region might have a genetic admixture. The genetic structure of Turkish people has been investigated in small scale studies using different methods (5,6,7,8). In a recent study, Kars et al. (9) published the first comprehensive genomic variation database, Turkish Variome (TRV), which compiles whole genome and whole exome data from 3362 individuals from various regions of Turkey.

The aim of this study was to evaluate whether any large population variomes, including TRV, can reliably be used in variant evaluations for the population of Turkey. Therefore, 60 non-disease-causing non-synonymous variants (minor AFs greater than 1%) in 58 genes, known to be associated with idiopathic hypogonadotropic hypogonadism (IHH), were compared with the AFs in various population databases worldwide.

Methods

Patient Cohort

The study used genetic variants from a large, rare disease cohort. The cohort included a total of 290 independent patients (112 female and 178 male) from seven geographic regions of Turkey (the Marmara region, Black Sea, Aegean, Mediterranean, Central Anatolia, Eastern Anatolia, and Southeastern Anatolia), roughly representing the population of Turkey.

Disease Model

IHH is a rare disease characterized by pubertal failure and infertility, with a prevalence of 1/10-100.000 (10). Mutations detected in patients with IHH, based on the Mendelian inheritance model, are currently thought to be responsible for approximately 50% of all cases (11).

Genetic Analyses

A total of 290 WES data sets were screened for potentially pathogenic nucleotide changes [frameshifts, in-frame changes (insertion and deletion), nonsense (stop-loss and stop-gain), two-base splice-sites (donor/acceptor), and missense] located in the exons of 58 genes known to be associated with IHH. Intronic areas, distant regions, and synonymous changes were excluded. Currently known-IHH-associated genes are listed in Table 1 (11).

Selection of the Study Variants

Based on the prevalence of IHH (1/10.100.000), those variants with an AF lower than 0.0001 were excluded from the study as they can be of high pathogenicity. We also excluded those that can be potentially pathogenic with an AF of 0.01-0.0001. In this study, we only included those with AF greater than 0.01, which are extremely unlikely to be disease-causing for IHH.

WES Analyses

Briefly, the genomic DNA samples from each patient were prepared as an Illumina sequencing library. Afterward, sequencing libraries were enriched for proper targets with the Illumina Exome Enrichment protocol. Captured libraries were sequenced with Illumina HiSeq 2000 Sequencer (Macrogen, Seoul, South Korea). The reads were mapped to UCSC hg19.

Databases

The seven established databases used for the AF correlations with our cohort were: GnomAD, which includes European Finnish, Europen Non-Finnish (ENF), Ashkenazi Jewish, East Asian, South Asian, Latino/Admixed American, and African/African-American subcategories (12); The NHLBI Trans-Omics for Precision Medicine representing a diverse population around the world with multi-ethnic data content (European, Hispanic/Latino, African, Asian) (13); The Greater Middle East (GME) Variome Project, which includes the GME world population, from Morocco in the west to Pakistan in the East including 163 alleles from the Turkish peninsula (14); Iranome, which includes Iranian Arabs, Kurds, Persians, Persian Gulf Islanders, Azeris, and Turkmen ethnic groups (15); GenomeAsia, which includes South East Asian, Oceania, North East Asian, African, West Eurasia, South Asian, and American subpopulations (16); the 4.7KJPN, which represents the overall Japanese population (17); and Online Archive of Brazilian Mutations, which includes Brazilian population (18). The GnomAD ENF category includes Southern European, Bulgarian, North-Western European, Swedish, and Estonian subpopulations. Categories named as "others: were not included in the study. The AFs were collected from the databases in February 2022. URLs of databases are provided in the web resources.

Statistical Analysis

Statistical analyses were performed using the Statistical Package for Social Sciences, version 20.0 (IBM Inc., Armonk, NY, USA), and a p value of < 0.001 was considered

Table 1. Genes known to be associated with idiopathic hypogonadotropic hypogonadism

Approved symbol	Approved name	HGNC ID	Chromosomal location
AMH	anti-Mullerian hormone	464	19p13.3
AMHR2	anti-Mullerian hormone receptor type 2	465	12q13.13
ANOS1	anosmin 1	6211	Xp22.31
AXL	AXL receptor tyrosine kinase	905	19q13.2
CCDC141	coiled-coil domain containing 141	26821	2q31.2
CHD7	chromodomain helicase DNA binding protein 7	20626	8q12.2
DCC	DCC netrin 1 receptor	2701	18q21.2
DLG2	discs large MAGUK scaffold protein 2	2901	11q14.1
DMXL2	Dmx like 2	2938	15q21.2
DUSP6	dual specificity phosphatase 6	3072	12q21.33
FEZF1	FEZ family zinc finger 1	22788	7q31.32
FGF17	fibroblast growth factor 17	3673	8p21.3
FGF8	fibroblast growth factor 8	3686	10q24.32
FGFR1	fibroblast growth factor receptor 1	3688	8p11.23
FLRT3	fibronectin leucine rich transmembrane protein 3	3762	20p12.1
FSHB	follicle stimulating hormone subunit beta	3964	11p14.1
GNRH1	gonadotropin releasing hormone 1	4419	8p21.2
GNRHR	gonadotropin releasing hormone receptor	4421	4q13.2
HESX1	HESX homeobox 1	4877	3p14.3
HS6ST1	heparan sulfate 6-O-sulfotransferase 1	5201	2q14.3
IGSF10	immunoglobulin superfamily member 10	26384	3q25.1
IL17RD	interleukin 17 receptor D	17616	3p14.3
KISS1	KiSS-1 metastasis suppressor	6341	1q32.1
KISS1R	KISS1 receptor	4510	19p13.3
KLB	klotho beta	15527	4p14
LEP	leptin	6553	7q32.1
LEPR	leptin receptor	6554	1p31.3
LHB	luteinizing hormone subunit beta	6584	19q13.33
NDNF	neuron derived neurotrophic factor	26256	4q27
NR0B1	nuclear receptor subfamily 0 group B member 1	7960	Xp21.2
NSMF	NMDA receptor synaptonuclear signaling and neuronal migration factor	29843	9q34.3
NTN1	netrin 1	8029	17p13.1
OTUD4	OTU deubiquitinase 4	24949	4q31.21
PCSK1	proprotein convertase subtilisin/kexin type 1	8743	5q15
PLXNA1	plexin A1	9099	3q21.3
PLXNA3	plexin A3	9101	Xq28
PNPLA6	patatin like phospholipase domain containing 6	16268	19p13.2
POLR3A	RNA polymerase III subunit A	30074	10q22.3
POLR3B	RNA polymerase III subunit B	30348	12q23.3
PROK2	prokineticin 2	18455	3p13
PROKR2	prokineticin receptor 2	15836	20p12.3
RAB18	RAB18, member RAS oncogene family	14244	10p12.1
RAB3GAP1	RAB3 GTPase activating protein catalytic subunit 1	17063	2q21.3
RAB3GAP2	RAB3 GTPase activating non-catalytic protein subunit 2	17168	1q41
RNF216	ring finger protein 216	21698	7p22.1
SEMA3A	semaphorin 3A	10723	7q21.11
SEMA3E	semaphorin 3E	10727	7q21.11

J Clin Res Pediatr Endocrinol
2022;14(3):293-301

Table 1. Continued

Approved symbol	Approved name	HGNC ID	Chromosomal location
SEMA3F	semaphorin 3F	10728	3p21.31
SMCHD1	structural maintenance of chromosomes flexible hinge domain containing 1	29090	18p11.32
SOX10	SRY-box transcription factor 10	11190	22q13.1
SPRY4	sprouty RTK signaling antagonist 4	15533	5q31.3
SRA1	steroid receptor RNA activator 1	11281	5q31.3
STUB1	STIP1 homology and U-box containing protein 1	11427	16p13.3
TAC3	tachykinin precursor 3	11521	12q13.3
TACR3	tachykinin receptor 3	11528	4q24
TBC1D20	TBC1 domain family member 20	16133	20p13
TUBB3	tubulin beta 3 class III	20772	16q24.3
WDR11	WD repeat domain 11	13831	10q26.12

HGNC: https://www.genenames.org, HGNC: HUGO Gene Nomenclature Committee

Table 2. The allelic frequencies (AF) of the variants in the IHH cohort with greater than 1% and AF values in selected databases

Gene	dbSNP	Cohort AF	TRV AF	Genome Asia AF	GnomAD AF	GME AF	ToPMed Bravo AF	Iranome AF	4.7KJPN AF	ABraOM
AXL	rs7249222	0.9844	1.0000	1.0000	1.0000	-	1.0000	1.0000	0.0000	1.0000
DCC	rs9951523	0.9810	0.9886	0.9980	0.9853	0.9914	0.9913	0.9919	1.0000	0.9885
AMH	rs10417628	0.9542	0.9782	0.9890	0.9817	0.9777	0.9834	0.9794	0.9999	0.9926
AMH	rs10407022	0.8224	0.8091	0.6480	0.7629	0.7773	0.6976	0.7956	0.6523	0.7775
IGSF10	rs7619322	0.7948	0.7848	0.7390	0.7196	0.8145	0.6976	0.7937	0.6613	0.7471
SEMA3F	rs1046956	0.7000	0.7164	0.9150	0.7347	0.6491	0.6479	0.7406	0.9999	0.6789
IL17RD	rs6780995	0.6931	0.6630	0.4160	0.6201	0.6500	0.6218	0.6981	0.1611	0.6362
KISS1R	rs350132	0.6775	0.7823	0.7630	0.7942	0.8127	0.8204	0.7432	0.7679	0.7830
SMCHD1	rs2276092	0.6396	0.6456	0.6540	0.6960	0.7134	0.7388	0.6625	0.5547	0.7348
ANOS1	rs808119	0.5970	0.6157	-	0.5590	0.7189	0.5325	0.5381	0.6755	0.5450
SRA1	rs5871740	0.4879	0.5120	0.4660	0.1400	0.4506	-	0.5331	0.5113	-
DMXL2	rs12102203	0.4586	0.4670	0.4180	0.4936	0.4526	0.4850	0.4544	0.4661	0.5090
DUSP6	rs2279574	0.4344	0.4683	0.5590	0.5177	0.4558	0.4200	0.4719	0.5175	0.4367
SRA1	rs3085220	0.4224	0.5120	0.4660	0.1221	-	-	0.5331	0.5113	0.1976
DCC	rs2229080	0.4034	0.4611	-	0.4416	0.4879	0.3782	0.4888	0.6595	0.4293
RAB3GAP1	rs10445686	0.3862	0.3843	0.3210	0.1973	0.3524	0.1264	0.3919	0.4118	0.2348
HS6ST1	rs200979099	0.3672	-	-	0.3773	0.4030	0.0907	0.4381	0.0144	0.3801
FLRT3	rs6079391	0.3568	0.3694	0.2380	0.4098	0.2754	0.3375	0.3325	0.1492	0.3924
PLXNA3	rs5945430	0.3383	0.3038	-	0.2148	0.3668	0.6666	0.4113	0.0524	0.2784
LEPR	rs1137101	0.3241	0.3532	0.6800	0.5089	0.3487	0.4871	0.3394	0.8621	0.4384
SMCHD1	rs633422	0.3068	0.3236	-	0.3383	-	0.2502	0.2781	0.2417	0.3481
KISS1	rs4889	0.3051	0.3186	0.3080	0.2874	0.3132	0.3238	0.2932	0.4662	0.3001
KLB	rs4975017	0.2862	0.2897	0.4350	0.3428	0.2275	0.2525	0.2938	0.3801	0.2889
PCSK1	rs6234	0.2517	0.2725	0.2800	0.2633	0.2573	0.2263	0.2662	0.2093	0.2430
PCSK1	rs6235	0.2466	0.2719	0.2780	0.2595	0.2532	0.2110	0.2631	0.2094	0.2356
PNPLA6	rs17854645	0.2362	0.2403	0.0820	0.1620	0.2152	0.1479	0.2459	0.0001	0.1674
KISS1	rs71745629	0.2292	0.2794	0.2750	0.2230	-	0.1705	0.2214	0.4625	0.2059
DMXL2	rs17524906	0.2224	0.2151	0.0760	0.1819	0.2064	0.1806	0.2156	0.0119	0.1962
LEPR	rs1805094	0.2000	0.2496	0.0970	0.1584	-	0.1822	0.2662	0.1010	0.1937
KLB	rs17618244	0.1965	0.2182	0.1750	0.1793	0.2139	0.1357	0.1994	0.1838	0.1929
HS6ST1	rs3958533	0.1931	-	-	0.1193	0.3171	0.1031	0.4206	0.0214	0.2529
CCDC141	rs61397643	0.1896	0.1642	0.1250	0.1310	0.1614	0.1165	0.1950	0.1172	0.1403

Table 2. Continued

Gene	dbSNP	Cohort AF	TRV AF	Genome Asia AF	GnomAD AF	GME AF	ToPMed Bravo AF	Iranome AF	4.7KJPN AF	ABraOM
GNRH1	rs6185	0.1793	0.1926	0.2920	0.2299	0.1596	0.1844	0.1725	0.5192	0.1945
HS6ST1	rs199993343	0.1727	-	-	0.0898	0.1554	0.0222	0.3302	0.0081	0.0371
IL17RD	rs17057718	0.1706	0.1848	0.3050	0.1983	0.1898	0.1500	0.1931	0.3382	0.1724
LEPR	rs1137100	0.1379	0.1771	0.4140	0.2959	0.1077	0.2148	0.1556	0.7813	0.2298
DUSP6	rs770087	0.1327	0.1613	0.1230	0.1960	0.1930	0.2890	0.1537	0.0388	0.2578
SEMA3E	rs61729612	0.1224	0.1215	0.1066	0.1161	0.1406	0.1022	0.1144	0.2172	0.1215
RAB3GAP2	rs2289189	0.1120	0.1376	0.0840	0.0840	0.0936	0.0603	0.1206	0.0809	0.0853
CCDC141	rs34883828	0.0931	0.1254	0.0560	0.1070	0.1168	0.1294	0.1044	0.0097	0.1124
CCDC141	rs12988301	0.0862	0.1097	0.1010	0.0832	0.1072	0.1194	0.0887	0.1557	0.1100
IGSF10	rs12487205	0.0810	0.0792	0.0320	0.0341	0.0694	0.0358	0.0818	0.0413	0.0517
IGSF10	rs17204557	0.0810	0.0772	0.0690	0.0416	0.0720	0.0310	0.0806	0.0409	0.0476
CCDC141	rs17362588	0.0637	0.0546	0.0110	0.0604	0.0629	0.0452	0.0618	-	0.0689
RAB3GAP2	rs12045447	0.0586	0.0463	0.0470	0.0390	0.0808	0.0587	0.0481	0.0326	0.0418
PCSK1	rs6232	0.0534	0.0547	0.0210	0.0381	0.0523	0.0290	0.0518	0.0003	0.0287
POLR3B	rs17038460	0.0517	0.0612	0.0100	0.0460	0.0634	0.0410	0.0550	-	0.0870
LHB	rs1800447	0.0465	0.0621	-	0.0665	0.0888	0.0768	0.0381	0.0452	0.0343
LHB	rs34349826	0.0431	0.0610	-	0.0522	0.0833	-	0.0375	0.0452	0.0269
CCDC141	rs13419085	0.0431	0.0412	0.0900	0.0606	0.0453	0.0266	0.0600	0.0166	0.0492
HS6ST1	rs201154532	0.0413	0.0442	-	0.0541	0.0229	0.0443	0.0538	0.0060	0.0345
KISS1	rs12998	0.0327	0.0365	0.0360	0.0325	0.0483	0.0241	0.0393	0.0375	0.0377
KISS1	rs35431622	0.0275	0.0422	0.0021	0.0518	0.0724	0.1207	0.0487	-	0.0812
KLB	rs35372803	0.0275	0.0213	0.0066	0.0362	0.0186	0.0292	0.0150	0.0028	0.0213
SEMA3A	rs147436181	0.0275	0.0180	0.0057	0.0137	0.0176	0.0106	0.0275	-	0.0188
DUSP6	rs61734372	0.0241	0.0200	0.0046	0.0238	0.0121	0.0201	0.0156	-	0.0246
FLRT3	rs35253731	0.0206	0.0209	0.0232	0.0319	0.0382	0.0669	0.0218	0.0124	0.0402
OTUD4	rs36225838	0.0206	0.0261	0.0054	0.0256	0.0282	0.0221	0.0243	-	0.0336
STUB1	rs148553428	0.0120	0.0064	0.0046	0.0037	0.0065	0.0024	0.0093	-	0.0016
IGSF10	rs34114908	0.0189	0.0110	0.0017	0.0127	0.0120	0.0084	0.0081	-	0.0147

Variants are arranged by allelic frequency from high to low.
dbSNP: Database of Single Nucleotide Polymorphisms, TRV: Turkish Variome, GME: The Greater Middle East Variom Project, GnomAD: genome aggregation database, TopMeD Bravo: Trans-Omics for Precision Medicine, ToMMo-4.7KJPN: Tohoku Medical Megabank Organization, ABraOM: Online Archive of Brazilian Mutations, IHH: idiopathic hypogonadotropic hypogonadism, -: absent, GME: http://igm.ucsd.edu/gme/, TopMED Bravo: https://bravo.sph.umich.edu/freeze5/hg38/, ToMMo-4.7KJPN: https://jmorp.megabank.tohoku.ac.jp/202102/, ABraOM: https://abraom.ib.usp.br

statistically significant. The Spearman's correlation method was used as the variables in the comparison of the groups were non-normally distributed. The correlation coefficients (CCs) between the study cohort and each of the databases/subgroups were analyzed separately. All correlation analysis results were found to be statistically significant. Next, we compared the CCs based on the concept of comparison of correlations from independent samples.

Results

In this study, a total of 60 variants with an AF greater than 1% were detected in 30 of 58 IHH-associated genes in the WES data from the cohort of 290 independent Turkish IHH patients (Table 2). No variants above the cut-off were observed in 26 of the listed IHH genes, while 17 genes had more than one (maximum five) variant and 13 genes had only one. The great majority of the changes (95.0%) were missense, and 5.0% were frameshift (two insertions and one deletion). Each of the variants in the study cohort was observed only in the Iranome and GnomAD.

A statistically significant correlation was observed between the study cohort and each one of the databases analyzed (Table 3). The highest CCs were observed between the study cohort and the following databases, in decreasing order: TRV (0.994), Iranome (0.983), and GME (0.981). Comparison of correlations from independent samples indicated that the

CCs of these three databases with our study cohort were not statistically different from each other (Table 3, shown in bold). The remaining 32 CCs were significantly different. Thus, the comparison results were interpreted as such that the three databases (TRV, Iranome, and GME) can be used as the reference databases for Turkish individuals.

Discussion

Population studies have repeatedly revealed the importance of local datasets in research and clinical practice, rather than using comprehensive databases with a wide-ranging sample size (4,19,20). Knowing the AF differences between

Table 3. The correlation coefficients (CCs) between the IHH cohort and various databases

Databases	Total number of 60 variants encountered	Correlation coefficients
TurkishVariome	**57**	**0.994**
Iranome	**60**	**0.983**
The Greater Middle East Variom Project (GME)	**55**	**0.981**
Arab#	60	0.978
Azeri#	60	0.976
Persian#	60	0.975
West Eurasia*	50	0.974
Lur#	60	0.969
Kurd#	60	0.968
Turkmen#	60	0.963
Persian Gulf Islander#	60	0.958
Southern European +, ~	60	0.955
Online Archive of Brazilian Mutations (ABraOM)	59	0.951
South Asian*	50	0.949
Baloch#	60	0.948
Bulgarian +, ~	60	0.943
GenomeAsia	50	0.940
GnomAD	60	0.935
European Non-Finnish +	60	0.923
South Asian +	60	0.920
Ashkenazi Jewish +	60	0.920
Latino/Admixed American +	60	0.913
North-Western European +, ~	60	0.911
Swedish +, ~	60	0.907
Trans-Omics for Precision Medicine (TopMed)	57	0.906
North East Asian*	50	0.892
European Finnish +	60	0.886
American*	50	0.868
Estonian +, ~	60	0.867
South East Asian*	50	0.861
East Asian +	60	0.849
Oceania*	50	0.841
African/African-American +	60	0.817
African*	50	0.802
Tohoku Medical Megabank Organization (4.7KJPN)	52	0.639

The Spearman's Correlation method was used for non-normally distributed variables in the comparison of the groups. Statistical analyses were performed using SPSS 20.0 and a p-value of < 0.001 was considered statistically significant. The correlation coefficients (CCs) between the study cohort and each of the databases/subgroups were analyzed separately. All correlation analysis results were found to be statistically significant. The databases/subgroups were arranged by CCs from high to low. Comparison of correlations from independent samples indicated that the CC of these three databases with our study cohort were not statistically different from each other (shown in bold). Symbols indicate from which database the subgroups were collected; *: GenomeAsia, +: GnomAD (Genome aggregation database), ~: GnomAD ENF (European Non- Finnish), #: Iranome.
gnomAD: https://gnomad.broadinstitute.org, GenomeAsia: https://browser.genomeasia100k.org, Iranome: http://www.iranome.ir, ABraOM: https://abraom.ib.usp.br, IHH: idiopathic hypogonadotropic hypogonadism

populations is also essential for developing machine-learning-based methods that use clustering scores for pathogenicity classifications (21). Disease genetics studies in a given population may also provide information for community characteristics, such as mutation history, local adaptations, and avoiding false-positive genetic diagnoses of Mendelian disorders. In this way, identifying and labeling population-specific genetic changes, such as individual/family-specific variants, will significantly reduce the burden of variants of uncertain significance (22,23). In our study, the common variants of Turkish IHH patients were observed at varying frequencies in different populations, supporting the hypothesis that a population-specific reference database should be used to facilitate the selection of pathogenic variants.

It is essential to understand that common and rare alleles have different characteristics. A rare variation is needed to survive many generations to rise to a moderate frequency, while common ones tend to be inherited over long periods due to negligible effects and are most likely classified as benign. Thus, they are excellent candidates for determining demographic histories or periodical features, such as ancestral origins and migration routes (24,25,26). Blekhman et al. (27) observed that Mendelian-disease gene variants, in general, are under purifying selection pressures. The IHH-associated gene variants should be expected to be subjected to additional negative selection pressures as pathogenic variants in these genes result in infertility. This reproductive disadvantage causes them to be rapidly purged from the population (27). Consequently, the AFs of the IHH gene variants are expected to be more skewed compared to most of the Mendelian disease genes except for those with very high mortality. However, common variants (minor allele frequencies $>1\%$) are free of such distortions. Based on the foregoing argument, we selectively compared the common variants in the IHH-related genes with those of the TRV and other publically available databases. Our cohort results showed a nearly one-to-one correlation (0.994) with TRV, which is comprised of NGS data from individuals participating in genetic studies of various diseases, such as obesity, amyotrophic lateral sclerosis, and Parkinson's disease. Our study using a rare disease model, IHH, which is not represented in the TRV patient subpopulations, confirms that TRV is well representative of the Turkish population. Kars et al. (9) also reported the close genetic relationship between Balkan and Caucasian populations and those of Turkey. Previously, similar to our methods paradigm, Alkan et al. (5) studied the 16 genomes from various regions of Turkey and compared them to those in the 1000 Genomes Project, and showed that the genetic structure of the people of Turkey is similar to those of Europe, particularly the Southern Europe/Mediterranean region, compared to other gene pools. Similarly, in our study, a close correlation, albeit to a lesser extent, was also observed with those of West Eurasia, including Caucasia (0.974) and Southern Europe (0.955).

It is well-known that consanguineous unions increase the incidence of recessively inherited diseases (28). Our study included 290 independent IHH patients, and consanguinity was present in 56.0%. This rate is higher than the general Turkish population (21.1%), probably due to a rare disease that could be recessively inherited (29). Studies have reported that high consanguineous marriage is common in many regions, including Turkey, Iran, and Pakistan (29,30,31). The kinship union is influenced by culture, religion, geographical conditions, or socioeconomic boundaries. The AFs in our study cohort did not show remarkable similarity for those in different databases in distant geographies. However, the close correlations with the non-European neighbors of the Anatolian peninsula, Iranome (0.983) and GME (0.981) suggest our genetic similarity for alleles that are relatively difficult to spread due to this social structure (28).

Study Limitations

The use of WES analyses performed at different periods in the study may have resulted in differences between reads that confidently support alleles.

Conclusion

Our findings confirm that TRV can be reliably used for variant evaluations from the Turkish population. Our results also indicate that variomes from geographic neighborhoods may serve as substitute references in variant evaluation for populations devoid of representative databases.

Acknowledgements

I am grateful to Prof. Dr. A. Kemal Topaloğlu for his valuable comments and continuing support of this project. I want to thank Prof. Dr. Ayşe Serin and Dr. Hüsniye Canan for useful discussions. I also thank Assoc. Prof. Dr. İhsan Turan for helping with statistical analyses.

Ethics

Ethics Committee Approval: The Ethics Committee of the Çukurova University Faculty of Medicine approved this study (decision no: 47, date: 02.11.2018).

Informed Consent: Informed consent form was obtained from all patients and/or their representative.

Peer-review: Externally and internally peer-reviewed.

Financial Disclosure: This work was supported by the Çukurova University Scientific Research project number: 11364.

References

1. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, Hardison M, Person R, Bekheirnia MR, Leduc MS, Kirby A, Pham P, Scull J, Wang M, Ding Y, Plon SE, Lupski JR, Beaudet AL, Gibbs RA, Eng CM. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med 2013;369:1502-1511. Epub 2013/10/04
2. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, Committee ALQA. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405-424. Epub 2015/03/06
3. Ku CS, Loy EY, Salim A, Pawitan Y, Chia KS. The discovery of human genetic variations and their use as disease markers: past, present and future. J Hum Genet 2010;55:403-415. Epub 2010/05/21
4. van Rooij JGJ, Jhamai M, Arp PP, Nouwens SCA, Verkerk M, Hofman A, Ikram MA, Verkerk AJ, van Meurs JBJ, Rivadeneira F, Uitterlinden AG, Kraaij R. Population-specific genetic variation in large sequencing data sets: why more data is still better. Eur J Hum Genet 2017;25:1173-1175. Epub 2017/09/15
5. Alkan C, Kavak P, Somel M, Gokcumen O, Ugurlu S, Saygi C, Dal E, Bugra K, Gungor T, Sahinalp SC, Ozoren N, Bekpen C. Whole genome sequencing of Turkish genomes reveals functional private alleles and impact of genetic interactions with Europe, Asia and Africa. BMC Genomics 2014;15:963. Epub 2014/11/08
6. Hodoglugil U, Mahley RW. Turkish population structure and genetic ancestry reveal relatedness among Eurasian populations. Ann Hum Genet 2012;76:128-141. Epub 2012/02/16
7. Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA. Excavating Y-chromosome haplotype strata in Anatolia. Hum Genet 2004;114:127-148. Epub 2003/10/31
8. Di Benedetto G, Erguven A, Stenico M, Castri L, Bertorelle G, Togan I, Barbujani G. DNA diversity and population admixture in Anatolia. Am J Phys Anthropol 2001;115:144-156. Epub 2001/06/01
9. Kars ME, Basak AN, Onat OE, Bilguvar K, Choi J, Itan Y, Caglar C, Palvadeau R, Casanova JL, Cooper DN, Stenson PD, Yavuz A, Bulus H, Gunel M, Friedman JM, Ozcelik T. The genetic structure of the Turkish population reveals high levels of variation and admixture. Proc Natl Acad Sci U S A 2021;118. Epub 2021/08/25
10. Bianco SD, Kaiser UB. The genetic and molecular basis of idiopathic hypogonadotropic hypogonadism. Nat Rev Endocrinol 2009;5:569-576. Epub 2009/08/27
11. Topaloglu AK, Turan I. Genetic Etiology of Idiopathic Hypogonadotropic Hypogonadism. Endocrines 2022;3:1-15.
12. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation C. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285-291. Epub 2016/08/19
13. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Taliun SAG, Corvelo A, Gogarten SM, Kang HM, Pitsillides AN, LeFaive J, Lee SB, Tian X, Browning BL, Das S, Emde AK, Clarke WE, Loesch DP, Shetty AC, Blackwell TW, Smith AV, Wong Q, Liu X, Conomos MP, Bobo DM, Aguet F, Albert C, Alonso A, Ardlie KG, Arking DE, Aslibekyan S, Auer PL, Barnard J, Barr RG, Barwick L, Becker LC, Beer RL, Benjamin EJ, Bielak LF, Blangero J, Boehnke M, Bowden DW, Brody JA, Burchard EG, Cade BE, Casella JF, Chalazan B, Chasman DI, Chen YI, Cho MH, Choi SH, Chung MK, Clish CB, Correa A, Curran JE, Custer B, Darbar D, Daya M, de Andrade M, DeMeo DL, Dutcher SK, Ellinor PT, Emery LS, Eng C, Fatkin D, Fingerlin T, Forer L, Fornage M, Franceschini N, Fuchsberger C, Fullerton SM, Germer S, Gladwin MT, Gottlieb DJ, Guo X, Hall ME, He J, Heard-Costa NL, Heckbert SR, Irvin MR, Johnsen JM, Johnson AD, Kaplan R, Kardia SLR, Kelly T, Kelly S, Kenny EE, Kiel DP, Klemmer R, Konkle BA, Kooperberg C, Kottgen A, Lange LA, Lasky-Su J, Levy D, Lin X, Lin KH, Liu C, Loos RJF, Garman L, Gerszten R, Lubitz SA, Lunetta KL, Mak ACY, Manichaikul A, Manning AK, Mathias RA, McManus DD, McGarvey ST, Meigs JB, Meyers DA, Mikulla JL, Minear MA, Mitchell BD, Mohanty S, Montasser ME, Montgomery C, Morrison AC, Murabito JM, Natale A, Natarajan P, Nelson SC, North KE, O'Connell JR, Palmer ND, Pankratz N, Peloso GM, Peyser PA, Pleiness J, Post WS, Psaty BM, Rao DC, Redline S, Reiner AP, Roden D, Rotter JI, Ruczinski I, Sarnowski C, Schoenherr S, Schwartz DA, Seo JS, Seshadri S, Sheehan VA, Sheu WH, Shoemaker MB, Smith NL, Smith JA, Sotoodehnia N, Stilp AM, Tang W, Taylor KD, Telen M, Thornton TA, Tracy RP, Van Den Berg DJ, Vasan RS, Viaud-Martinez KA, Vrieze S, Weeks DE, Weir BS, Weiss ST, Weng LC, Willer CJ, Zhang Y, Zhao X, Arnett DK, Ashley-Koch AE, Barnes KC, Boerwinkle E, Gabriel S, Gibbs R, Rice KM, Rich SS, Silverman EK, Qasba P, Gan W, Consortium NT-OfPM, Papanicolaou GJ, Nickerson DA, Browning SR, Zody MC, Zollner S, Wilson JG, Cupples LA, Laurie CC, Jaquish CE, Hernandez RD, O'Connor TD, Abecasis GR. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature 2021;590:290-299. Epub 2021/02/12
14. Scott EM, Halees A, Itan Y, Spencer EG, He Y, Azab MA, Gabriel SB, Belkadi A, Boisson B, Abel L, Clark AG, Greater Middle East Variome C, Alkuraya FS, Casanova JL, Gleeson JG. Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. Nat Genet 2016;48:1071-1076. Epub 2016/07/19
15. Fattahi Z, Beheshtian M, Mohseni M, Poustchi H, Sellars E, Nezhadi SH, Amini A, Arzhangi S, Jalalvand K, Jamali P, Mohammadi Z, Davarnia B, Nikuei P, Oladnabi M, Mohammadzadeh A, Zohrehvand E, Nejatizadeh A, Shekari M, Bagherzadeh M, Shamsi-Gooshki E, Borno S, Timmermann B, Haghdoost A, Najafipour R, Khorram Khorshid HR, Kahrizi K, Malekzadeh R, Akbari MR, Najmabadi H. Iranome: A catalog of genomic variations in the Iranian population. Hum Mutat 2019;40:1968-1984. Epub 2019/07/26
16. GenomeAsia KC. The GenomeAsia 100K Project enables genetic discoveries across Asia. Nature 2019;576:106-111. Epub 2019/12/06
17. Tadaka S, Saigusa D, Motoike IN, Inoue J, Aoki Y, Shirota M, Koshiba S, Yamamoto M, Kinoshita K. jMorp: Japanese Multi Omics Reference Panel. Nucleic Acids Res 2018;46:D551-D557. Epub 2017/10/27
18. Naslavsky MS, Yamamoto GL, de Almeida TF, Ezquina SAM, Sunaga DY, Pho N, Bozoklian D, Sandberg TOM, Brito LA, Lazar M, Bernardo DV, Amaro E, Jr., Duarte YAO, Lebrao ML, Passos-Bueno MR, Zatz M. Exomic variants of an elderly cohort of Brazilians in the ABraOM database. Hum Mutat 2017;38:751-763. Epub 2017/03/24

19. Lim ET, Wurtz P, Havulinna AS, Palta P, Tukiainen T, Rehnstrom K, Esko T, Magi R, Inouye M, Lappalainen T, Chan Y, Salem RM, Lek M, Flannick J, Sim X, Manning A, Ladenvall C, Bumpstead S, Hamalainen E, Aalto K, Maksimow M, Salmi M, Blankenberg S, Ardissino D, Shah S, Horne B, McPherson R, Hovingh GK, Reilly MP, Watkins H, Goel A, Farrall M, Girelli D, Reiner AP, Stitziel NO, Kathiresan S, Gabriel S, Barrett JC, Lehtimaki T, Laakso M, Groop L, Kaprio J, Perola M, McCarthy MI, Boehnke M, Altshuler DM, Lindgren CM, Hirschhorn JN, Metspalu A, Freimer NB, Zeller T, Jalkanen S, Koskinen S, Raitakari O, Durbin R, MacArthur DG, Salomaa V, Ripatti S, Daly MJ, Palotie A, Sequencing Initiative Suomi P. Distribution and medical impact of loss-of-function variants in the Finnish founder population. PLoS Genet 2014;10:e1004494. Epub 2014/08/01

20. Gudbjartsson DF, Helgason H, Gudjonsson SA, Zink F, Oddson A, Gylfason A, Besenbacher S, Magnusson G, Halldorsson BV, Hjartarson E, Sigurdsson GT, Stacey SN, Frigge ML, Holm H, Saemundsdottir J, Helgadottir HT, Johannsdottir H, Sigfusson G, Thorgeirsson G, Sverrisson JT, Gretarsdottir S, Walters GB, Rafnar T, Thjodleifsson B, Bjornsson ES, Olafsson S, Thorarinsdottir H, Steingrimsdottir T, Gudmundsdottir TS, Theodors A, Jonasson JG, Sigurdsson A, Bjornsdottir G, Jonsson JJ, Thorarensen O, Ludvigsson P, Gudbjartsson H, Eyjolfsson GI, Sigurdardottir O, Olafsson I, Arnar DO, Magnusson OT, Kong A, Masson G, Thorsteinsdottir U, Helgason A, Sulem P, Stefansson K. Large-scale whole-genome sequencing of the Icelandic population. Nat Genet 2015;47:435-444. Epub 2015/03/26

21. Maffucci P, Bigio B, Rapaport F, Cobat A, Borghesi A, Lopez M, Patin E, Bolze A, Shang L, Bendavid M, Scott EM, Stenson PD, Cunningham-Rundles C, Cooper DN, Gleeson JG, Fellay J, Quintana-Murci L, Casanova JL, Abel L, Boisson B, Itan Y. Blacklisting variants common in private cohorts but not in public databases optimizes human exome analysis. Proc Natl Acad Sci U S A 2019;116:950-959. Epub 2018/12/29

22. Shirts BH, Pritchard CC, Walsh T. Family-Specific Variants and the Limits of Human Genetics. Trends Mol Med 2016;22:925-934. Epub 2016/10/16

23. He L, Pang X, Chen P, Wang X, Yang T, Wu H. Carrier re-sequencing reveals rare but benign variants in recessive deafness genes. Sci Rep 2017;7:11355. Epub 2017/09/14

24. Slatkin M. Allele age and a test for selection on rare alleles. Philos Trans R Soc Lond B Biol Sci 2000;355:1663-1668. Epub 2000/12/29

25. Raychaudhuri S. Mapping rare and common causal alleles for complex human diseases. Cell 2011;147:57-69. Epub 2011/10/04

26. Niroula A, Vihinen M. How good are pathogenicity predictors in detecting benign variants? PLoS Comput Biol 2019;15:e1006481. Epub 2019/02/12

27. Blekhman R, Man O, Herrmann L, Boyko AR, Indap A, Kosiol C, Bustamante CD, Teshima KM, Przeworski M. Natural selection on genes that underlie human disease susceptibility. Curr Biol 2008;18:883-889. Epub 2008/06/24

28. Bittles A. Consanguinity and its relevance to clinical genetics. Clin Genet 2001;60:89-98. Epub 2001/09/13

29. Tuncbilek E, Koc I. Consanguineous marriage in Turkey and its impact on fertility and mortality. Ann Hum Genet 1994;58:321-329. Epub 1994/10/01

30. Saadat M, Ansari-Lari M, Farhud DD. Consanguineous marriage in Iran. Ann Hum Biol 2004;31:263-269. Epub 2004/06/19

31. Afzal M, Ali SM, Siyal HB. Consanguineous marriages in Pakistan. Pak Dev Rev 1994;33:663-674. Epub 1994/01/01