ARTICLE

A study of Kibbutzim in Israel reveals risk factors for cardiometabolic traits and subtle population structure

Einat Granot-HersHKovitz¹ · David Karasik² · Yechiel Friedlander¹ · Laura Rodriguez-Murillo³ · Rajkumar Dorajoo⁴ · Jianjun Liu^{4,5} · Anshuman Sewda³ · Inga Peter³ · Shai Carmi¹ · Hagit Hochner¹

Received: 15 January 2018 / Revised: 24 June 2018 / Accepted: 17 July 2018 / Published online: 14 August 2018
© European Society of Human Genetics 2018

Abstract

Genetic studies in isolated populations often increase power for identifying loci associated with complex diseases and traits. We present here the Kibbutzim Family Study (KFS), aimed at investigating the genetic basis of cardiometabolic traits in extended Israeli families characterized by long-term social stability and a homogeneous environment. Extensive information on cardiometabolic traits, as well as genome-wide genotypes, were collected on 901 individuals. We observed that most KFS participants were of Ashkenazi Jewish (AJ) genetic origin, confirmed a recent severe bottleneck in the AJ recent history, and detected a subtle within-AJ population structure. Focusing on genetic variants relatively common in the KFS but very rare in Europeans, we observed that AJ-enriched variants appear in cancer-related pathways more than expected by chance. We conducted an association study of the AJ-enriched variants against 16 cardiometabolic traits, and found seven loci (24 variants) to be significantly associated. The strongest association, which we also replicated in an independent study, was between a variant upstream of *MSRA* (frequency $\approx 1\%$ in the KFS and nearly absent in Europeans) and weight ($P = 3.6 \cdot 10^{-8}$). In conclusion, the KFS is a valuable resource for the study of the population genetics of Israel as well as the genetics of cardiometabolic traits.

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41431-018-0230-3>) contains supplementary material, which is available to authorized users.

✉ Shai Carmi
shai.carmi@mail.huji.ac.il

✉ Hagit Hochner
hagith@ekmd.huji.ac.il

¹ Braun School of Public Health, Hebrew University-Hadassah Medical Center, Jerusalem, Israel

² Faculty of Medicine in the Galilee, Bar-Ilan University, Safed, Israel

³ Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁴ Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore

⁵ Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Introduction

Genetic association studies of complex traits in isolated populations are advantageous in identifying risk loci that are rare in the general population but enriched in the isolate [1–3]. The Ashkenazi Jewish (AJ) population has been attractive for genetic studies, because of its unique demographic history of a recent severe bottleneck followed by a rapid expansion and endogamy [4]. AJ were found to carry unique mutations for several Mendelian disorders, as well as risk factors for complex diseases [5–13]. Importantly, while these mutations may be unique or nearly-unique to AJ, they often highlight pathways of broad significance.

Cardiovascular diseases (CVD) are a common cause of death worldwide [14]. Genome-wide association studies (GWAS) in unrelated individuals have identified thousands of genetic variants associated with CVD and their risk factors [15, 16] but the genetic risk is not fully explained. The Kibbutzim Family Study (KFS) was established in 1992 to investigate the environmental and genetic basis of cardiometabolic risk factors [17–20]. The participants belonged, at the time of recruitment, to large families living in close-knit communities, called “Kibbutzim”, in Northern

Israel. Kibbutzim have been communal settlements that have created a relatively homogeneous environment for their members. For example, earnings were uniformly distributed, and Kibbutzim members typically dined jointly. Kibbutzim members are mostly of Ashkenazi Jewish ancestry, with the remaining members belonging to other Jewish subgroups. The KFS is thus expected to be a useful resource for the study of cardiometabolic genetic risk factors.

While most association studies so far were conducted on unrelated individuals, the extended family design of the KFS has the advantages of a reduced sensitivity to population stratification bias and the ability to detect Mendelian inconsistencies [21]. Family-based studies also have the ability to enrich for genetic loci containing rare variants. Familial aggregation, segregation analyses, and linkage and candidate gene association studies were previously conducted in the KFS [17–19, 22–25], focusing on outcomes such as Low-density lipoprotein peak particle diameter [24], fibrinogen variability [23], and red blood cell membrane fatty acid composition [19]. Here, we present results for genome-wide genotyping of 901 KFS participants. We aimed to [1] characterize the population genetics of the KFS population and [2] assess the contribution of genetic variants enriched in the KFS to anthropometric and cardiometabolic traits and other health-related phenotypes.

Methods

Recruitment

The KFS participants were recruited in two phases in 1992–1993 and 1999–2000 [17, 24] from six Kibbutzim in Northern Israel [23]. The first recruitment phase of the study (1992–1993) included 80 extended families, ranging in size from 2 to 43 individuals [26]. During the second phase (1999–2000), participants from the first phase were all invited for repeat examinations (80% response rate) and new participants were recruited, giving a total of 150 extended families ranging in size from 2 to 55 individuals [17]. Families were invited to participate if they consisted of at least four individuals who (i) lived in the Kibbutz, (ii) spanned at least two generations, and (iii) were at least 15 years old. Families were retained if at least two family members consented to participate. Overall, 1033 participants were recruited; 111 were examined only in the first phase, 533 only in the second phase, and 389 in both. Participants completed a self-administered socio-demographic and health questionnaire, including questions on medical and family history and lifestyle [17, 26]. Psychosocial and dietary information was collected as well. Anthropometric and blood pressure traits (described below) were measured in

both phases [19, 20], and peripheral blood samples were collected following a 12-hour fast. All subjects signed an informed consent and the study was approved by the Institutional Review Board of the Hadassah-Hebrew University Medical Center.

Genotyping and quality control

Of the 1033 participants recruited, 938 had high-quality DNA samples ($A_{260}/_{280} > 1.8$, concentration > 50 ng/ μ l). Genotyping was performed using Illumina HumanCoreExome BeadChip, consisting of $\approx 240,000$ tag single-nucleotide polymorphisms (SNPs) and $\approx 240,000$ exome variants. Standard quality control (QC) procedures were applied to filter variants and individuals using Plink 1.90 [27]; for details, please see Supplementary Note 1. A total of 901 individuals and 323,708 variants (281,586 variants with minor allele frequency (MAF) $> 1\%$) passed QC and were used in downstream analyses. The data reported in this paper was deposited at the European Genome-phenome Archive (EGA) under accession number EGAS00001002782.

Principal component analysis

Principal component analysis (PCA) was performed using PC-AiR [28], which is robust to known or cryptic relatedness. Our reference panel included West-Eurasian populations (covering Europe, West-Asia (the Middle East) and the Caucasus, $n = 922$) [29] and the Jewish groups listed in Supplementary Table 1 ($n = 174$) [29]. These samples served as the “unrelated subset” for PC-AiR. We additionally ran PCA using only the Jewish groups ($n = 174$) as the reference population. In that analysis, we also included a panel of AJ recruited in the United States by The Ashkenazi Genome Consortium (TAGC) ($n = 128$) [12], which allowed us to examine differences in ancestry between AJ from Israel (KFS) and the US. Another PCA was run using only the AJ samples from Behar et al. ($n = 29$) as the reference population, to focus on differences between Western and Eastern AJ. Variants used in the PC-AiR analysis were restricted to MAF $> 1\%$ and were pruned to eliminate linkage disequilibrium (LD), using the `--indep-pairwise` command in Plink (window size 50 kb, a shift of ten variants at each step, and LD between variants (r^2) < 0.1).

IBD sharing and demographic reconstruction

We phased the KFS genotypes using SHAPEIT v2 [30], and detected IBD segments using GERMLINE [31] and with additional filtering by HaploScore [32] and SNP density. See Supplementary Note 2 for details. An evaluation of

Table 1 Socio-demographic characteristics of the KFS ($n = 901$)

	All				Men				Women			
	<i>N</i>	mean	sd	%	<i>N</i>	mean	sd	%	<i>N</i>	mean	sd	%
Age	901	44.59	19.55	100	423	45.02	19.46	47	478	44.21	19.65	53
Age-categories		<i>N</i>	%			<i>N</i>	%			<i>N</i>	%	
Age ≤ 29		249	27.7			130	26.80			177	32.30	
Age > 29 and < 59		468	52.0			255	52.58			257	46.90	
Age ≥ 59		183	20.3			100	20.62			114	20.80	
Birthplace												
Israel		649	72.35			308	72.99			341	71.79	
Europe		179	19.96			83	19.67			96	20.21	
America		47	5.24			21	4.98			26	5.47	
North-Africa		12	1.34			6	1.42			6	1.26	
Asia		10	1.11			4	0.95			6	1.26	
Religiosity												
Religious		483	53.8			229	54.27			255	53.46	
Secular		414	46.2			193	45.73			222	46.54	
Education (years)												
< 8		16	1.78			7	1.66			9	1.89	
8–12		403	44.93			197	46.68			206	43.37	
> 12		478	53.29			218	51.66			260	54.74	
Smoking												
Current		87	9.72			46	10.95			41	8.63	
Past Smokers		155	17.32			95	22.62			60	12.63	
No		653	72.96			279	66.43			374	78.74	
Physical activity (hours per week)												
0		497	57.99			253	63.41			244	53.28	
1–2		274	31.97			102	25.56			172	37.55	
≥ 3		86	10.04			44	11.03			42	9.17	
Physical labor at work (sweat or accelerated breathing)												
No		493	55.46			189	45.22			304	64.54	
Occasionally		196	22.05			127	30.38			69	14.65	
Frequently		200	22.50			102	24.40			98	20.81	

the improvement in the accuracy of detected IBD segments due to the pedigree-based phasing is described in Supplementary Note 3. For the demographic inference analysis, we retained only IBD segments shared between Ashkenazi founders ($n = 303$), as identified by the first PC in the PCA of the Jewish populations (Results). The method we used to estimate the population size history of AJ is described in Supplementary Note 4. We calculated runs of homozygosity (ROH) using plink (--homozyg-kb 5000).

Imputation

We imputed the phased genotypes using IMPUTE2 [33]. For the reference panel, we initially used either an Ashkenazi-only reference panel (TAGC; $n = 128$) [12], the 1000 Genomes reference panel phase 1 version 3 ($n = 1092$), or a combined Ashkenazi + 1000 Genomes

panel ($n = 1220$). The estimates provided by IMPUTE2 for the concordance between the true array genotypes and their imputed values were highest when using the combined reference panel, and we thus used that panel for downstream analyses. Imputed genotypes were initially available for 82,328,870 variants. For most analyses, we only considered the 6,858,900 variants with MAF $\geq 1\%$ and imputation quality score ≥ 0.9 .

Identification of rare variants that are relatively common in the KFS

In populations that have undergone recent strong genetic drift (such as Ashkenazi Jews [12, 34]), it is expected that some risk variants of large effects have risen in frequency compared to the general population [35]. We thus focused on variants with a substantially higher frequency in the

KFS compared to the general population, which we take as non-Finnish Europeans (NFE) from The Genome Aggregation Database (gnomAD). We observed that a naive search for variants with a large frequency difference led to numerous artifacts. We thus implemented a stringent QC pipeline. First, we filtered out variants with >10% MAF difference between the KFS (founders only, $n = 393$) and AJ in gnomAD ($n \approx 150$) [36]. Second, we filtered out variants with >10% MAF difference between NFE in the 1000 Genomes Project (phase 3; CEU + GBR + TSI + IBS; $n = 404$) [37] and NFE in gnomAD ($n \approx 7500$) [36]. After applying the two above-mentioned filters, we extracted variants that were very rare ($\text{MAF} < 0.1\%$) in gnomAD NFE but relatively common ($\text{MAF} \geq 1\%$) in the KFS. This resulted in a total of $n = 212,505$ enriched variants.

To determine if the MAF ratio (KFS/NFE) correlated with the functional consequence of the enriched variants, we annotated these variants using SnpEff version 4.3q [38]. We performed gene-set enrichment analysis (GSEA) using the Molecular Signatures Database (MSigDB) on variants present in gnomAD NFE with KFS/NFE MAF ratio > 10 and high/moderate predicted functional impact, totaling 190,598 variants [39].

Association analysis

We performed association tests with BOLT-LMM v2.2 [40]. BOLT-LMM accounts for relationships between individuals and population structure using a linear mixed-model, as well as handles imputed “dosage” data. For building the mixed-model, we used 299,509 genotyped variants ($\text{MAF} > 0.1\%$). We used the 1000 Genomes LD-Score table provided with BOLT-LMM. We only tested the 212,505 enriched variants with >10x higher KFS/NFE MAF ratio (see above). P -value threshold for significance was set at $1.61 \cdot 10^{-6}$ (See Supplementary Note 5 for detailed description).

Supplementary Table 3 lists the 16 anthropometric and cardiometabolic traits we analyzed and their corresponding heritability estimates (based on all imputed genetic variants), as calculated by BOLT-REML [41]. All models were adjusted for age, gender, and phase. Lipid-lowering medication was accounted for by introducing a dichotomous covariate for medication use and blood pressure lowering medication was adjusted by adding 10 and 5 mm Hg to systolic (SBP) and diastolic (DBP) blood pressures, respectively [42]. Lipoprotein (a), C-reactive protein, and triglycerides variables were inverse normal transformed. We observed no improvement in association results when using the non-infinitesimal mixed-model test in BOLT-LMM, and thus all reported results are for the standard infinitesimal model.

To determine the number of independently associated loci, we first excluded associated variants in high LD ($r^2 > 0.95$) with the index SNP (lowest P -value) in each chromosomal region, followed by a conditional analysis using the index SNP as a covariate.

Results

Samples

Our study included 1033 participants (47% male, 53% female) who were recruited during two phases (1992–1993 and 1999–2000) from 150 families (445 founders). The majority of families spanned three (57.3%) and two (28.0%) generations. The mean family size was 6.89 individuals (range 2–55). Participants' characteristics by gender are given in Table 1. The 16 anthropometric and cardiometabolic traits used in the association analysis are summarized in Supplementary Table 4 by gender and age group.

Population genetics

Principal component analysis (PCA)

To study the genetic ancestry of the KFS participants, we ran PCA (Methods) on the genotyped KFS samples ($n = 901$), along with worldwide ($n = 922$) and Jewish ($n = 174$) reference populations [29] (Supplementary Table 1). The first two principal components (Fig. 1) distinguish three main non-Jewish population groups: European, Caucasian, and West-Asian (Middle-Eastern), and six Jewish

Fig. 1 A principal components analysis (PCA) of the KFS samples ($n = 901$, blue cross marks), along with reference samples from Jewish ($n = 174$) and non-Jewish ($n = 922$) populations

Table 2 Significant associations for enriched variants in the KFS (MAF > 1% and 10x higher compared to Europeans) and replication in the JPS cohort

Genomic variant (hg19)	rsID	Effect allele	MAF		KFS (<i>n</i> = 901)				JPS children (<i>n</i> = 857)				JPS mothers (<i>n</i> = 763)				Meta-analysis (KFS + JPS children + JPS mothers)				Nearby genes	GWAS catalog/ PhenGen relevant -associated traits
			KFS (<i>n</i> = 393)	Ashkenazi Jews gnomAD (<i>n</i> = 150)	NFE gnomAD (<i>n</i> = 7,500)	MAF ratio (KFS vs. gnomAD NFE)	Trait	Beta	SE	<i>P</i> -value	Beta	SE	<i>P</i> -value	Beta	SE	<i>P</i> -value	Beta	SE	<i>P</i> -value			
chr5: g:171016625 A > G	rs759883128	G	0.01	0.01	1.30E-04	79.08	Pulse	6.67	1.359	9.20E-07	-1.718	1.561	2.72E-01	-2.932	1.302	2.46E-02	0.764	0.805	3.43E-01			
chr6: g:160195050 A > G	rs754054303	G	0.026	0.02	2.70E-04	94.48	LPA	-0.76	0.158	1.40E-06										ACAT2	HDL	
chr6: g:161011320 T > A	rs185882981	A	0.022	0.02	5.30E-04	40.81	LPA	-0.779	0.161	1.40E-06										LPA	LPA, cholesterol	
chr8: g:9887880 T > G		G	0.01	0.013	6.70E-05	153.13	BMI	4.809	0.985	1.10E-06	7.260	4.051	7.35E-02	3.312	0.945	4.80E-04	4.119	0.673	9.12E-10	MSRA	Obesity, SBP, triglycerides	
chr8: g:9887880 T > G		G	0.01	0.013	6.70E-05	153.13	Weight	17.411	3.159	3.60E-08	26.496	15.724	9.24E-02	8.758	2.121	4.06E-05	11.634	1.750	2.99E-11	MSRA		
chr8: g:9887880 T > G		G	0.01	0.013	6.70E-05	153.13	Waist C.	14.244	2.692	1.20E-07	18.537	11.207	9.85E-02	12.001	3.175	1.70E-04	13.476	2.020	2.52E-11	MSRA		
chr8: g:9924288 G > A	rs759188048	A	0.01	0.013	6.70E-05	152.69	Weight	16.517	3.297	5.50E-07	26.496	15.724	9.24E-02	8.765	2.117	3.87E-05	11.225	1.770	2.29E-10	MSRA		
chr8: g:10171408 T > C		C	0.015	0.02	6.70E-05	227.91	Weight	15.398	3.118	7.90E-07	22.949	9.494	1.58E-02	-5.927	4.851	2.22E-01	10.142	2.528	6.04E-05	MSRA		
chr8: g:10194935 A > C	rs181491777	C	0.015	0.023	1.30E-04	117.46	Weight	15.549	3.117	6.10E-07	16.840	9.504	7.68E-02	-4.956	4.909	3.13E-01	10.169	2.536	6.08E-05	MSRA		
chr8: g:10216348 C > T		T	0.015	0.023	2.70E-04	56.56	Weight	15.425	3.118	7.50E-07	15.872	7.128	2.62E-02	-3.652	3.517	2.99E-01	7.884	2.217	3.77E-04	MSRA		
chr8: g:10260592 C > A	rs779572353	A	0.015	0.02	6.70E-05	227.91	Weight	15.413	3.118	7.70E-07	23.463	9.642	1.52E-02	-6.603	4.842	1.73E-01	9.959	2.530	8.26E-05	MSRA		
chr8: g:10341650 G > A		A	0.015	0.02	1.30E-04	117.46	Weight	15.423	3.118	7.60E-07	26.068	10.719	1.52E-02	-11.028	3.367	1.10E-03	4.207	2.237	6.00E-02	MSRA		
chr8: g:17880544 G > C		C	0.01	0.007	6.70E-05	151.94	WHR	0.093	0.019	1.10E-06	-0.022	0.026	3.96E-01	0.038	0.009	3.29E-05	0.042	0.008	8.75E-08	PCMI		
chr13: g:50074153 A > T	rs756877701	T	0.013	0.013	6.70E-05	190.45	LPA	1.223	0.254	1.50E-06										PHF11, SETDB2, PHF11	Cardiomegaly	
chr13: g:50262586 T > A	rs535598832	A	0.013	0.013	6.70E-05	190.45	LPA	1.224	0.254	1.40E-06										EBPL		
chr13: g:50629665 T > C	rs749307626	C	0.013	0.013	0.00E+00	Inf	LPA	1.225	0.254	1.40E-06										DLEU2	Anthropometry	
chr13: g:50913656 G > C	rs767873499	C	0.013	0.017	0.00E+00	Inf	LPA	1.235	0.254	1.20E-06										DLEU1	Height	

Table 2 (continued)

Genomic variant (hg19)	rsID	Effect allele	MAF		KFS (<i>n</i> = 901)				JPS children (<i>n</i> = 857)				JPS mothers (<i>n</i> = 763)				Meta-analysis (KFS + JPS children + JPS mothers)				Nearby genes	GWAS catalog/ PhenGen relevant -associated traits
			KFS (<i>n</i> = 393)	Ashkenazi Jews gnomAD (<i>n</i> = 150)	NFE gnomAD (<i>n</i> = 7,500)	MAF ratio (KFS vs. gnomAD NFE)	Trait	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	P-value			
chr13: g:50915811 G>C	rs768336927	C	0.013	0.017	0.00E+00	Inf	LPA	1.23	0.254	1.30E-06											<i>DLEU1</i>	
chr13: g:51065795 G>C		C	0.013	0.017	0.00E+00	Inf	LPA	1.23	0.254	1.30E-06											<i>DLEU1</i>	
chr13: g:51216271 G>A	rs751733865	A	0.013	0.023	6.70E-05	190.45	LPA	1.242	0.254	1.00E-06											<i>BCMS</i>	
chr13: g:51384738 C>T	rs780360029	T	0.013	0.026	6.70E-05	189.85	LPA	1.253	0.247	3.80E-07											<i>DLEU7</i> , <i>DLEU7-AS1</i>	Height, HDL
chr13: g:51871870 A>C	rs766345636	C	0.01	0.017	0.00E+00	Inf	LPA	1.275	0.254	4.90E-07											<i>FA-M124A</i>	
chr13: g:51933072 T>C	rs756245132	C	0.01	0.017	0.00E+00	Inf	LPA	1.263	0.253	6.30E-07											<i>SER-PINE3</i> , <i>INTS6</i>	
chr17: g:7290447 C>T	rs566833653	T	0.015	0.007	0.00E+00	Inf	Height	-0.073	0.015	4.70E-07	-0.013	0.020	5.20E-01	-0.163	1.612	9.19E-01	-0.051	0.012	1.99E-05		<i>CDR2L</i>	
chr17: g:73052400 A>G	rs759145164	G	0.015	0.007	0.00E+00	Inf	Height	-0.074	0.015	4.20E-07	-0.013	0.020	5.20E-01	0.295	1.612	8.55E-01	-0.052	0.012	1.58E-05		<i>KCTD2</i>	Triglycerides, stroke, lung volume measurements
chr22: g:36733002 A>G	rs776420285	G	0.01	0.013	5.30E-04	19.30	Hip C.	10.076	2.097	1.50E-06	3.643	3.549	3.05E-01	5.624	2.828	4.71E-02	7.604	1.522	5.82E-07		<i>MYH9</i>	

All genomic coordinates are in hg19. Effect sizes are in units of the phenotypes (Supplementary Table 3). We used a genome-wide significance threshold of $P = 1.61 \times 10^{-6}$ (Methods). The MAF in the KFS was computed in founders only. None of the variants was found in the 1000 Genomes NFE population. The most strongly associated variant, which is associated with BMI, weight, and waist circumference, is highlighted. Horizontal lines distinguish between seven independently associated loci.

C circumference, *WHR* waist-to-hip ratio, *LPA* lipoprotein(a), *SBP* systolic blood pressure

populations: Ashkenazi, Sephardi, North-African, Yemenite, West-Asian, and Caucasian. A partial overlap is observed between AJ and European non-Jews, as well between West-Asian and Caucasian Jewish and non-Jewish populations.

The KFS samples largely overlapped with the AJ reference samples [29]. To study the non-Ashkenazi ancestries in the KFS, we ran PCA with the KFS samples and the Jewish reference populations only (Supplementary Fig. 1). The number of individuals with exclusive AJ ancestry, as distinguished by the first PC ($PC1 \leq 0$), was $n = 733$ (81.4%). The majority of the remaining individuals overlapped with the Sephardi and North-African Jewish clusters, but the Middle-Eastern, Caucasian, and Yemenite Jewish populations were also represented. Some individuals seemed to have a mixed Ashkenazi and other Jewish ancestry, although quantifying their exact number is difficult with PCA.

Self-reported country of birth allowed us to compare the PCA-based and self-reported Jewish ancestry for 247 individuals born outside Israel (Supplementary Fig. 2). Among 140 individuals self-reported as AJ (born in Northern and Central Europe), 136 (97%) met the defined genetic criterion ($PC1 \leq 0$). Among the 11 individuals self-reported as North-African Jewish, 9 (82%) met a pre-defined genetic criterion ($PC1 > 0.03$ and $PC2 > 0.05$).

Next, we asked whether AJ with recent origins in Eastern vs. Western Europe are genetically distinct. We designated KFS individuals born in Germany as Western AJ, and individuals born in Poland, Russia, Hungary, and Romania as Eastern AJ. A PCA plot revealed that Eastern and Western AJ can be distinguished in PC space, albeit imperfectly (Supplementary Fig. 3). We observed the same pattern in the samples of Behar et al. [29].

We observed no differences in PCA between the KFS AJ samples and 128 US-based AJ [12] (Supplementary Figs. 1 and 3), indicating no difference in genetic ancestry between Israel- and US-based AJ. This result, which agrees with the IBD-based analysis of Gusev et al. [43], is expected based on the short time since the migrations of AJ out of Europe and suggests that the source population for these migrations was relatively homogeneous.

IBD sharing and demographic reconstruction

We detected IBD segments longer than 3 cM shared between AJ founders in the KFS (Methods). Using the number and lengths of observed segments, we confirmed a recent severe bottleneck in the AJ recent history (point estimates: effective size ≈ 450 individuals, 23 generations ago) [12]. See Supplementary Note 6 for complete details. IBD sharing also revealed differences in ancestry between Eastern and Western AJ. The mean number of segments

shared within Western AJ was 1.4x larger than within Eastern AJ (8.4 vs. 5.9, $P < 10^{-7}$; Supplementary Table 5), but the mean segment length was similar (≈ 5.5 cM, $P = 0.28$). Sharing levels were particularly high in the group of Western AJ that was distinct by PCA (Supplementary Table 5). The number and lengths of long runs of homozygosity (Methods) did not significantly differ between Eastern and Western AJ (Supplementary Table 6).

Functional annotation of variants enriched in the KFS

We annotated the function of 212,505 variants of $MAF > 1\%$ in the KFS and $< 0.1\%$ in Europeans (Methods). We identified 62 (0.03%) high impact and 291 (0.13%) moderate impact variants, with the remaining predicted to have low or no functional significance (“modifiers”) according to SnpEff (Supplementary Table 7). We observed no correlation between the MAF ratio (KFS/Europeans) and the putative functional significance (Supplementary Fig. 7). Gene-set enrichment analysis (GSEA) on the 201 genes that contained at least one variant with MAF ratio > 10 and a high/moderate functional impact (Methods) identified highly significant enrichment (false discovery rate q -value $< 10^{-5}$) in 25 gene sets (pathways) in the molecular signature database (MSigDB). The top pathways were mostly related to cancers (breast cancer, prostate cancer, skin cancer, and sarcomas, among others) (Supplementary Table 8). There was no enrichment for cancer-related pathways when random sets of genes with variants of no functional significance were analyzed.

Association of variants enriched in the KFS with anthropometric and cardiometabolic traits

We considered the 212,505 enriched variants and used BOLT-LMM to test for an association of these variants with 16 anthropometric and cardiometabolic traits (Methods; qq-plots and Manhattan plots are shown in Supplementary Figs. 8 and 9, respectively). We set the P -value threshold for significance to $1.61 \cdot 10^{-6}$ (Methods). At this significance level, 24 variants were significantly associated (Table 2), comprising seven independent loci. We report gender-specific results for these variants in Supplementary Table 9, and locus zoom plots in Supplementary Fig. 10.

Our main finding is a region spanning seven variants (453 kb, KFS/European MAF ratio between 56 and 228), located in chr8p23.1, and associated with body-weight, waist circumference, and body mass index (BMI). The most significant association was with body weight for (hg19) chr8:g.9887880 T $>$ G ($P = 3.6 \cdot 10^{-8}$), an imputed variant located upstream of *MSRA* (methionine sulfoxide reductase A). This is the only variant with a study-wide significant association ($P < 1.61 \cdot 10^{-6}/16$).

In other chromosomes, a large region (1.9 Mb) in chr13q14.3 showed a significant association with lipoprotein(a) (LPA). The region contained ten variants, all with > 189-fold MAF ratio, with the most significant result at rs780360029 ($P = 3.8 \cdot 10^{-7}$). These variants span eight genes (Table 2), three of which belong to a region that is frequently deleted in B-cell chronic lymphocytic leukemia (*DLEU*) [44]. Two intronic variants in chr6q25.3-26, a known locus for LPA, were significantly associated with LPA; rs754054303 at *ACAT2* gene and rs185882981 at the *LPA* gene (Table 2). Two additional intronic variants in chr17q25.1—rs566833653 in *CDR2L* and rs759145164 in *KCTD2*—were both associated with height ($P = 4.7 \cdot 10^{-7}$ and $P = 4.2 \cdot 10^{-7}$, respectively; Table 2), and are absent in Europeans. Some of the top hits show suggestive differences in P -values between the sexes (Supplementary Table 9).

We pursued replication of our findings in another Israeli cohort, the Jerusalem Perinatal Study (JPS) [45], which consists of parents and their children who were born in the 1970s in Jerusalem. We ran linear regression analyses separately for children ($n = 857$, mean age ≈ 32) and mothers ($n = 763$, mean age ≈ 60) for 12 of the 24 associated variants (the remaining were associated with LPA, which was not available in the JPS) and meta-analyzed the results in all three groups (Table 2). The top hit for weight, BMI, and waist circumference at chr8: g.9887880 T > G had the same direction and magnitude of effect in the KFS and the JPS, with P -values around 10^{-4} in the JPS mothers and 10^{-10} overall (Table 2). The nearby variant rs759188048 similarly replicated, but the results for more distant variants in that locus were mixed. Among the other loci, the association of chr8: g.17880544 G > C with waist-to-hip ratio and that of rs776420285 with hip circumference replicated in the JPS mothers.

Discussion

We analyzed the genotypes of 901 individuals from extended families living in Kibbutzim in Israel, who had detailed records on anthropometric traits and cardiometabolic risk factors. The data enabled us to refine population-genetic patterns of Israeli Jews, as well as study genetic associations with 16 traits.

Ashkenazi Jewish population genetics

PCA confirmed self-reported ancestries and allowed precise assignment of ethnic origins for most KFS individuals. Participants were mostly of AJ origin (81.4%), with the remaining having various other Jewish ancestries. It was previously estimated that AJ have experienced a founder event ≈ 25 – 35 generations ago with an effective population

size of ≈ 300 – 400 individuals [12, 34]. We established that these estimates hold for our independent AJ sample.

A popular theory of Ashkenazi origins is an initial settlement in Western Europe (Northern France and Germany), followed by migration to Poland and an expansion there and in the rest of Eastern Europe [46]. An open question is whether AJ with recent origins in Eastern Europe are genetically distinct from Western European AJ. Early mtDNA and disease mutation studies have identified differences between AJ from different origins [10, 47], and a recent study of mtDNA diversity in AJ has found large differences in haplotype frequencies between Western and Eastern AJ [48, 49]. With genome-wide data, a previous study of ≈ 1300 AJ [4] did not find a correlation (on a PCA plot) between genetic ancestry and a country of origin. A study of IBD sharing across the US did find three AJ sub-clusters, but could not assign the clusters to specific locations [50]. A later study of 29 AJ [29], which is part of the Jewish reference panel used here, did not identify genetic differences between Eastern and Western AJ, except for a minute East-Asian component in the ADMIXTURE analysis that was present in Eastern but not Western AJ. Our analysis of the KFS individuals who reported their country of origin showed that many Western AJ cluster separately from Eastern AJ, and the same pattern was observed in our re-analysis of the data of Behar et. al. [29]. IBD sharing analysis showed 1.4x more shared segments within Western AJ compared to Eastern AJ (and an even higher levels of sharing ($2.1 \times$) for those Western AJ who were distinct on PCA; Supplementary Table 5; Supplementary Fig. 3), but with no difference in mean segment length. An explanation consistent with these observations is that Western AJ consist of two slightly distinct groups: one that descends from a subset of the original founders (represented by those who are distinct on the PCA plot), and another that has migrated there back from Eastern Europe, possibly after absorbing a limited degree of gene flow. We note, however, that we cannot exclude the possibility that the results reflect, at least partly, biased sampling of Western AJ in the KFS.

Analysis of rare European variants that are relatively common in AJ

Studying isolated or founder populations such as AJ is expected to increase power to discover disease-associated genes, due to the rise in frequency of rare or unique risk alleles [35, 51, 52]. Here, we did not observe a correlation between variants enriched in AJ (the KFS) and putative functional significance. Nevertheless, for enriched variants with a functional impact, we identified a significant overlap with several cancer-related gene-sets, including breast

cancer. AJ women have a high risk of familial breast cancer, mostly due to founder mutations in the *BRCA1* and *BRCA2* genes [53]. While no functional enriched variants were observed in *BRCA1* or *BRCA2* in the KFS, a number of genes with functional enriched variants were found to interact with *BRCA1* (Supplementary Table 8). We note that whether cancer is more prevalent in Ashkenazi Jews compared to the general Western population is debated, and possibly limited to colorectal and prostate cancers, if at all [54–56].

We detected seven loci with AJ-enriched variants that were associated with anthropometric and cardiometabolic traits. The most strongly associated locus included seven variants surrounding the *MSRA* gene in chr8p23.1 that were associated with body weight, waist circumference, and BMI. The association of the index SNP in this locus (chr8:g.9887880 T > G) was replicated in another Israeli cohort (Table 2). Variants near this region (100 kb upstream), located between the genes *TNKS* and *MSRA*, were found to be associated with extreme obesity in children and adolescents [57] and with adult waist circumference [58] in individuals of European ancestry. Our findings may implicate *MSRA* as a candidate gene for these observed associations. This gene encodes a ubiquitous and highly conserved protein that carries out the enzymatic reduction of methionine sulfoxide to methionine.

Another region of interest is chr13q14.3, showing significant associations of ten variants with LPA. This region includes the *DLEU* genes, which are frequently deleted in B-cell chronic lymphocytic leukemia, suggesting a role of one or more tumor suppressors [44]. The variant rs749307626 is located in an intronic region of the *DLEU2* gene, which was previously associated with waist-to-hip ratio in a meta-analysis of African and European populations [49]. Three variants are located in the *DLEU1* gene, previously associated with anthropometric traits in another isolate (Korčula Island, Croatia [59]). One variant is located in the *DLEU7* gene, previously associated with height in Europeans and Africans [60]. The variant rs756877701 is located in an intronic region of the *PHF11* gene, which was previously associated with cardiomegaly in the Amish population [61]. Finally, two AJ-enriched variants in the known *LPA* locus on chr6 [62, 63] were associated with LPA in the KFS, providing evidence for the generalizability of our results.

Outlook

We report here the first genetic association study of enriched AJ variants with cardiometabolic traits in the Israeli Jewish population. In this study, we have identified a number of suggestive associations and also refined the understanding of the population genetics of Ashkenazi and other Jewish groups. Current limitations of our study include its relatively small size and its focus on Ashkenazi Jews. Thus, additional

analyses will be required in larger Jewish samples, as well as in other populations, to replicate the findings and elucidate the mechanisms underlying the observed associations. We conclude that the KFS is a valuable source for studying genetics of complex traits as well as Jewish genetics in the setting of a longitudinal family study.

Acknowledgements We are grateful to the study participants, recruiters, interviewers, and nurses. This study was supported by Israeli Science Foundation grants 201/98-1 and 407/17 and partially by National Institutes of Health research grant R01HL088884. Genotyping was also supported in part by a generous gift from the Samson Family (South Africa) to DK.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Kristiansson K, Naukkarinen J, Peltonen L. Isolated populations and complex disease gene identification. *Genome Biol.* 2008;9:109.
- Zeggini E, Gloyn AL, Hansen T. Insights into metabolic disease from studying genetics in isolated populations: stories from Greece to Greenland. *Diabetologia.* 2016;59:938–41.
- Fang S, Zhang S, Sha Q. Literature reviews on methods for rare variant association studies. *Hum Genet Embryol.* 2016;6:1–5.
- Guha S, Rosenfeld JA, Malhotra AK, et al. Implications for health and disease in the genetic signature of the Ashkenazi Jewish population. *Genome Biol.* 2012;13:R2.
- Kenny EE, Pe'er I, Karban A, et al. A genome-wide scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility loci. *PLoS Genet.* 2012;8:1–10. [cited 26 Oct 2017]
- Charrow J. Ashkenazi Jewish genetic disorders. *Fam Cancer.* 2004;3:201–6. [cited 26 Oct 2017]
- Vacic V, Ozelius LJ, Clark LN, et al. Genome-wide mapping of IBD segments in an Ashkenazi PD cohort identifies associated haplotypes. *Hum Mol Genet.* 2014;23:4693–702. [cited 3 Dec 2017]
- Lencz T, Guha S, Liu C, et al. Genome-wide association study implicates *NDST3* in schizophrenia and bipolar disorder. *Nat Commun.* 2013;4:2739.
- Slatkin M. A population-genetic test of founder effects and implications for Ashkenazi Jewish diseases. *Am J Hum Genet.* 2004;75:282–93. [cited 3 Dec 2017]
- Risch N, Tang H, Katzenstein H, Ekstein J. Geographic distribution of disease mutations in the Ashkenazi Jewish population supports genetic drift over selection. *Am J Hum Genet.* 2003;72:812–22. [cited 3 Dec 2017]
- Rivas MA, Avila BE, Koskela J, et al. Insights into the genetic epidemiology of Crohn's and rare diseases in the Ashkenazi Jewish population, T2D-GENES Consortium. *PLoS Genet.* 2018;14:e1007329. [cited 22 Jun 2018]
- Carmi S, Hui KY, Kochav E, et al. Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. *Nat Commun.* 2014;5:4835.
- Zeevi D, Bloom JS, Sadhu MJ, et al. Analysis of the genetic basis of height in large Jewish nuclear families. *bioRxiv.* 1–17. [cited 20 Apr 2018]

14. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med.* 2006;3:2011–30. [cited 2 Nov 2017]
15. Fall T, Ingelsson E. Genome-wide association studies of obesity and metabolic syndrome. *Mol Cell Endocrinol.* 2014;382:740–57. [cited 26 Oct 2017].
16. Atanasovska B, Kumar V, Fu J, Wijmenga C, Hofker MH. GWAS as a driver of gene discovery in cardiometabolic diseases. *Trends Endocrinol Metab.* 2015;26:722–32. [cited 26 Oct 2017]
17. Friedlander Y, Kark JD, Sinnreich R, Tracy RP, Siscovick DS. Fibrinogen and CRP in Israeli families: genetic and environmental sources of concentrations and longitudinal changes. *Atherosclerosis.* 2006;189:169–77. [cited 18 Dec 2012]
18. Friedlander Y, Vatta M, Sotoodehnia N, et al. Possible association of the human KCNE1 (minK) gene and QT interval in healthy subjects: evidence from association and linkage analyses in Israeli families. *Ann Hum Genet.* 2005;69(Pt 6):645–56. [cited 18 Dec 2012]
19. Lemaitre RN, Siscovick DS, Berry EM, Kark JD, Friedlander Y. Familial aggregation of red blood cell membrane fatty acid composition: the Kibbutzim Family Study. *Metabolism.* 2008;57:662–8. [cited 18 Dec 2012]
20. Friedlander Y, Elkana Y, Sinnreich R, Kark JD. Genetic and environmental sources of fibrinogen variability in Israeli families: the Kibbutzim Family Study. *Am J Hum Genet.* 1995;56:1194–206.
21. Ott J, Kamatani Y, Lathrop M. Family-based designs for genome-wide association studies. *Nat Rev Genet.* 2011;12:465–74.
22. Friedlander Y, Lapidus T, Sinnreich R, Kark JD. Genetic and environmental sources of QT interval variability in Israeli families: the kibbutz settlements family study. *Clin Genet.* 1999;56:200–9.
23. Friedlander Y, Kark JD, Sinnreich R, Basso F, Humphries SE. Combined segregation and linkage analysis of fibrinogen variability in Israeli families: evidence for two quantitative-trait loci, one of which is linked to a functional variant (-58G>A) in the promoter of the alpha-fibrinogen gene. *Ann Hum Genet.* 2003;67(Pt 3):228–41.
24. Friedlander Y, Kark JD, Sinnreich R, Edwards KL, Austin MA. Inheritance of LDL peak particle diameter: results from a segregation analysis in Israeli families. *Genet Epidemiol.* 1999;16:382–96.
25. Sinnreich R, Friedlander Y, Luria MH, Sapoznikov D, Kark JD. Inheritance of heart rate variability: the kibbutzim family study. *Hum Genet.* 1999;105:654–61.
26. Sinnreich R, Friedlander Y, Sapoznikov D, Kark JD. Familial aggregation of heart rate variability based on short recordings--the kibbutzim family study. *Hum Genet.* 1998;103:34–40.
27. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.
28. Conomos MP, Reiner AP, Weir BS, Thornton TA. Model-free estimation of recent genetic relatedness. *Am J Hum Genet.* 2016;98:127–48.
29. Behar DM, Metspalu M, Baran Y, et al. No evidence from genome-wide data of a Khazar origin for the Ashkenazi Jews. *Hum Biol.* 2013;85:859–900.
30. O'Connell J, Gurdasani D, Delaneau O, et al. A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet.* 2014;10:e1004234.
31. Gusev A, Lowe JK, Stoffel M, et al. Whole population, genome-wide mapping of hidden relatedness. *Genome Res.* 2009;19:318–26.
32. Durand EY, Eriksson N, Mclean CY. Reducing pervasive false-positive identical-by-descent segments detected by large-scale pedigree analysis. *Mol Biol Evol.* 2014;31:2212–22.
33. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet.* 2012;44:955–9.
34. Palamara PF, Lencz T, Darvasi A, Pe'er I. Length distributions of identity by descent reveal fine-scale demographic history. *Am J Hum Genet.* 2012;91:809–22.
35. Hatzikotoulas K, Gilly A, Zeggini E. Using population isolates in genetic association studies. *Brief Funct Genom.* 2014;13:371–7.
36. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536:285–91.
37. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature.* 2015;526:68–74.
38. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w¹¹¹⁸; iso-2; iso-3. *Fly (Austin).* 2012; 6(2):80–92.
39. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *PNAS.* 2005;102:15545–50. [cited 23 Nov 2017]
40. Loh P-R, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet.* 2015;47:284–90.
41. Loh P-R, Bhatia G, Gusev A, et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat Genet.* 2015;47:1385–92.
42. Cui JS, Hopper JL, Harrap SB. Antihypertensive treatments obscure familial contributions to blood pressure variation. *Hypertension.* 2003;41:207–10.
43. Gusev A, Palamara PF, Aponte G, et al. The architecture of long-range haplotypes shared within and across populations. *Mol Biol Evol.* 2012;29:473–86.
44. Rowntree C, Duke V, Panayiotidis P, et al. Deletion analysis of chromosome 13q14.3 and characterisation of an alternative splice form of LEU1 in B cell chronic lymphocytic leukemia. *Leukemia.* 2002;16:1267–75. [cited 26 Oct 2017]
45. Lawrence GM, Siscovick DS, Calderon-Margalit R, et al. Cohort profile: The Jerusalem perinatal family follow-up study. *Int J Epidemiol.* 2015;45:343–52.
46. Weinryb BD. The Jews of Poland; a social and economic history of the Jewish community in Poland from 1100 to 1800. Jewish Publication Society of America; 1972. Chapter 1.
47. Feder J, Ovadia O, Glaser B, Mishmar D. Ashkenazi Jewish mtDNA haplogroup distribution varies among distinct subpopulations: lessons of population substructure in a closed group. *Eur J Hum Genet.* 2007;15:498–500. [cited 25 Jan 2018]
48. Costa MD, Pereira JB, Pala M, et al. A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. *Nat Commun.* 2013;4:1–10. [cited 25 Jan 2018]
49. Ng MCY, Graff M, Lu Y, et al. Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of African ancestry: African Ancestry Anthropometry Genetics Consortium. Copenhagen GP, editor. *PLOS Genet.* 2017;13:e1006719. Apr 21 [cited 23 Oct 2017]
50. Han E, Carbonetto P, Curtis RE, et al. Clustering of 770,000 genomes reveals post-colonial population structure of North America. *Nat Commun.* 2017;8:14238. [cited 22 Jun 2018]
51. Zeggini E. Europe PMC funders group next-generation association studies for complex traits. *Nat Genet.* 2012;43:287–8.
52. Peltonen L, Palotie A, Lange K. Use of population isolates for mapping complex traits. *Nat Rev Genet.* 2000;1:182–90.
53. Levy-Lahad E, Catane R, Eisenberg S, Kaufman B, et al. Founder BRCA1 and BRCA2 mutations in Ashkenazi Jews in Israel: frequency and differential penetrance in ovarian cancer and in breast-ovarian cancer families. *Am J Hum Genet.* 1997;60:1059–67.

54. Streicher SA, Klein AP, Olson SH, et al. Impact of sixteen established pancreatic cancer susceptibility loci in American jews. *Cancer Epidemiol Biomark Prev.* 2017;10:1540–8. [cited 7 Dec 2017]
55. Lynch HT, Rubinstein WS, Locker GY. Cancer in Jews: introduction and overview. *Fam Cancer.* 2004;3:177–92. [cited 7 Dec 2017].
56. Feldman GE. Do Ashkenazi Jews have a higher than expected cancer burden? Implications for cancer control prioritization efforts. *Isr Med Assoc J.* 2001;3:341–6.
57. Scherag A, Dina C, Hinney A, et al. Two new loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and German study groups. *PLoS Genet.* 2010;6:2–11.
58. Lindgren CM, Heid IM, Randall JC, et al. Genome-wide association scan meta-analysis identifies three loci influencing adiposity and fat distribution. *PLoS Genet.* 2009;5:e1000508.
59. Polasek O, Marusić A, Rotim K, et al. Genome-wide association study of anthropometric traits in Korcula Island, Croatia. *Croat Med J.* 2009;50:7–16. [cited 26 Oct 2017]
60. Kang SJ, Chiang CWK, Palmer CD, et al. Genome-wide association of anthropometric traits in African- and African-derived populations. *Hum Mol Genet.* 2010;19:2725–38. [cited 26 Oct 2017]
61. Parsa A, Chang YPC, Kelly RJ, et al. Hypertrophy-associated polymorphisms ascertained in a founder cohort applied to heart failure risk and mortality. *Clin Transl Sci.* 2011;4:17–23. [cited 26 Oct 2017]
62. Ober C, Nord AS, Thompson EE, et al. Genome-wide association study of plasma lipoprotein(a) levels identifies multiple genes on chromosome 6q. *J Lipid Res.* 2009;50:798–806.
63. Kettunen J, Demirkan A, Würtz P, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun.* 2016;7:11122.