

Differential analysis of mutations in the Jewish population and their implications for diseases

YARON EINHORN¹, DAPHNA WEISSGLAS-VOLKOV¹, SHAI CARMİ²,
HARRY OSTRER³, EITAN FRIEDMAN^{1,4} AND NOAM SHOMRON^{1*}

¹Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

²Braun School of Public Health and Community Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel

³Department of Pathology, Albert Einstein College of Medicine, Bronx, NY, USA

⁴Susanne Levy Gertner Oncogenetics Unit, Sheba Medical Center, Tel-Hashomer, Israel

(Received 2 September 2016; revised 10 January 2017; accepted 31 January 2017)

Abstract

Sequencing large cohorts of ethnically homogeneous individuals yields genetic insights with implications for the entire population rather than a single individual. In order to evaluate the genetic basis of certain diseases encountered at high frequency in the Ashkenazi Jewish population (AJP), as well as to improve variant annotation among the AJP, we examined the entire exome, focusing on specific genes with known clinical implications in 128 Ashkenazi Jews and compared these data to other non-Jewish populations (European, African, South Asian and East Asian). We targeted American College of Medical Genetics incidental finding recommended genes and the Catalogue of Somatic Mutations in Cancer (COSMIC) germline cancer-related genes. We identified previously known disease-causing variants and discovered potentially deleterious variants in known disease-causing genes that are population specific or substantially more prevalent in the AJP, such as in the *ATP* and *HGFAC* genes associated with colorectal cancer and pancreatic cancer, respectively. Additionally, we tested the advantage of utilizing the database of the AJP when assigning pathogenicity to rare variants of independent whole-exome sequencing data of 49 Ashkenazi Jew early-onset breast cancer (BC) patients. Importantly, population-based filtering using our AJP database enabled a reduction in the number of potential causal variants in the BC cohort by 36%. Taken together, population-specific sequencing of the AJP offers valuable, clinically applicable information and improves AJP filter annotation.

1. Introduction

High-throughput sequencing, also known as next-generation sequencing (NGS), reduced the cost and increased the yield of DNA sequencing. As whole-exome sequencing (WES) and whole-genome sequencing (WGS) are increasingly integrated into practical medical care, the importance of studying the genetic structure of ethnically diverse populations using NGS rises. Although most of the variant sites in the human genome are shared among individuals, allele frequencies vary substantially between populations (The International HapMap Consortium, 2005; 1000 Genomes Project Consortium *et al.*, 2012; Visscher

et al., 2012; Carmi *et al.*, 2014; The Genome of the Netherlands Consortium, 2014; Gudbjartsson *et al.*, 2015; Nagasaki *et al.*, 2015). The value and advantages of sequencing diverse populations has already been shown in: genome-wide association studies (Visscher *et al.*, 2012); discovering rare and *de novo* variants; improving variant calling sensitivity and specificity; and improving the accuracy of curating pathogenic variants (Carmi *et al.*, 2014; The Genome of the Netherlands Consortium, 2014; Gudbjartsson *et al.*, 2015). Substantial efforts have been devoted to sequencing large number of individuals from diverse populations in order to create public databases that can assist human genetic studies such as the 1000 Genomes Project (1KG) (1000 Genomes Project Consortium *et al.*, 2012), the Exome Sequencing Project (ESP; <http://evs.gs.washington.edu/EVS/>) and the Exome Aggregation Consortium (ExAC; <http://exac.broadinstitute.org/>).

* Corresponding author: Dr Noam Shomron, Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. Tel +972 36406594. Fax +972 36407432. E-mail: nshomron@post.tau.ac.il

The Ashkenazi Jewish population (AJP) is known to have a high rate of several diseases affecting individuals of that ethnic origin compared with other world ethnicities (Rosner *et al.*, 2009). These include both autosomal recessive disorders due to the founder effect (Slatkin, 2004; Bray *et al.*, 2010; Carmi *et al.*, 2014), such as Gaucher disease (Beutler *et al.*, 1993), cystic fibrosis (Abeliovich *et al.*, 1992) and Tay–Sachs disease (Myerowitz & Costigan, 1988), as well as more common, adult-onset autosomal dominant diseases such as Parkinson's disease (PD) (Ozelius *et al.*, 2006) and hereditary BC and ovarian cancer (Struewing *et al.*, 1997). Notably, the AJP has not been included as part of large-scale international sequencing projects. A recent NGS study of an AJP cohort demonstrated an improvement in imputation accuracy and modelling of Jewish history (Carmi *et al.*, 2014). However, further research is warranted in order to elucidate the possible clinical implications of the AJP allelic architecture and to improve the curation and accuracy of pathogenic variant screening in current and future AJP studies.

Recently, new recommendations for the AJP screening panel were published based on the same dataset as ours (Baskovich *et al.*, 2016). However, that study focused only on the identification of pathogenic variants for the purpose of clinical screening in the AJP, whereas the current study takes a more global view by focusing on the genome and gene-level trends, rather than particular genetic variants, examining the utility of using an AJP-specific reference panel in interpreting clinical sequencing projects involving AJP individuals.

In this study, we focused on the clinical utility and practical implications resulting from WES analysis of 128 Ashkenazi Jews, of whom 74 individuals had no discernible disease and 54 were controls in a PD study. We examined the genetic differences between the AJP and other non-Jewish populations (NJPs) and searched for genes that are more likely to carry pathogenic variants among the AJP than in NJPs. Finally, we applied our findings to 49 independent Ashkenazi Jewish BC patients in order to evaluate the value of utilising an Ashkenazi Jew-specific database as a filtering tool.

2. Methods

Ashkenazi Jew variants

We used an unfiltered variant calling file (VCF) of 128 verified Ashkenazi Jewish individuals who underwent WGS as a part of a population genetic study of the AJP (Carmi *et al.*, 2014). WGS was conducted by Complete Genomics with a high coverage (average coverage >50×). Seventy-four of the individuals were considered healthy and 54 were controls in a PD study. We extracted variants from the whole-exome

region only, based on Illumina's TruSeq Exome Enrichment Kit targets (<https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/truseq-exome-data-sheet-770-2015-007.pdf>), and did not include areas outside this region in our bioinformatics analysis. The target region size was 62 Mb, which targets 20,794 genes and 96.4% of RefSeq43-coding exons. We performed quality check (QC) and applied different filtrations (see Supplementary Methods; available online), which resulted in 222,179 high-quality single-nucleotide variants (SNVs).

BC patient variants

The VCF of 49 Ashkenazi Jewish BC patients, suspected to be hereditary, was obtained using the Genome Analysis Toolkit (GATK) best practice pipeline (McKenna *et al.* 2010), followed by QC (see Supplementary Methods), which resulted in 173,300 variants for the same exome region as the 128 Ashkenazi Jews.

1KG control groups

As control groups, and in order to compare the AJP with other populations, we used the European, African, East Asian (EAS) and South Asian (SAS) populations from the 1KG Project version 3 database (1000 Genomes Project Consortium *et al.*, 2012). The data for these datasets were generated using the Illumina platform, and the variants were called by combining different variant callers, among them GATK's variant caller (<http://www.1000genomes.org/analysis>). For each population, 128 individuals were selected randomly, and the same region that was examined for the AJP was extracted.

3. Results

In this study, we analysed the whole-exome data of 128 Ashkenazi Jewish individuals. We detected 222,179 SNVs, of which 30.6% (68,139) were singletons and 81.7% were shared and were annotated in other European population databases, including the European samples of ESP, ExAC and 1KG. Although this rate of overlap between the AJP and the European population is in line with the known relatedness and genetic similarity between the European population and the AJP (Behar *et al.*, 2003; Costa *et al.*, 2013), approximately 20% of the detected variants were unique to the AJP. The overlap rates between AJP variation and genetically more distant populations including African, EAS and SAS populations (inferred from ExAC and 1KG databases) were significantly smaller, as expected (68–49%,

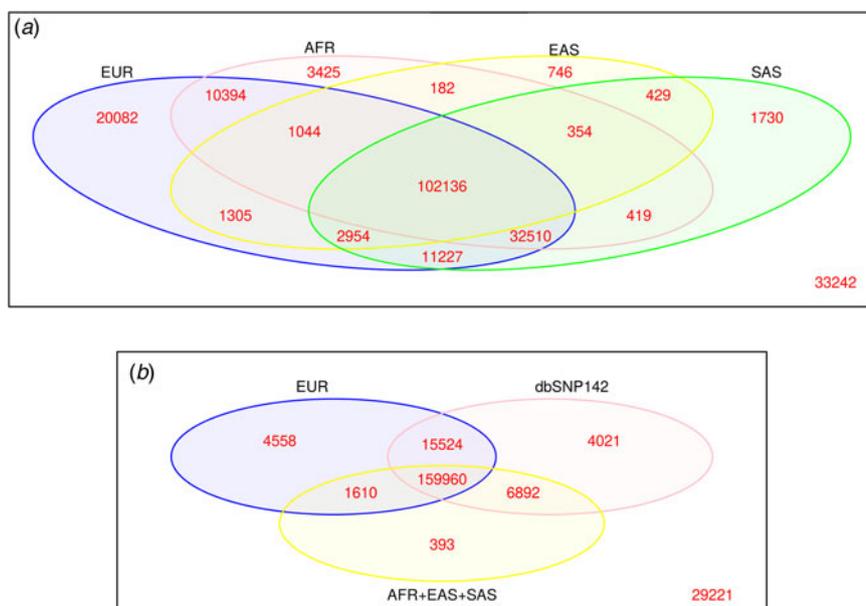


Fig. 1. (a) Overlap of the Ashkenazi Jewish population (AJP) variants with the European (EUR), African (AFR), East Asian (EAS) and South Asian (SAS) populations' variants. (b) Only 3.2% of the variants overlap with one of the non-EUR distal populations. Crossing with the dbSNP142 database resulted in 29,221 novel variants unique to the AJP.

Fig. 1(a), further strengthening the validity of our data. Only 3.2% of the AJP variants were present in one of these distantly related populations but not in the European dataset, resulting in 13.3% (29,221) AJP-unique (i.e. novel) variants not reported in any of the population databases or in dbSNP142 (Fig. 1(b)).

Next, we functionally annotated the coding variants and classified the exonic variants into three categories by severity: (i) 'high impact' including stop-gain or stop-loss variants and variants within 2-bp of a splicing junction; (ii) 'moderate impact' included exonic missense variants; and (iii) 'low impact' included synonymous variants and exonic variants of unknown type due to incomplete gene structure information. Using this classification scheme, 831 variants (19 splice site variants) with high impact were identified, 54,585 were moderate-impact variants and 45,876 were low-impact variants. A similar distribution of variant severity was observed in the 128 European individuals (Supplementary Fig. S1).

Evaluating ACMG and COSMIC set of genes

We evaluated the clinical implications of the high-impact, very rare variants by comparing the existence of these variants in two gene sets: the Catalogue of Somatic Mutations in Cancer (COSMIC; <http://cancer.sanger.ac.uk/cosmic>) and the American College of Medical Genetics and Genomics (ACMG; <https://www.acmg.net/>). COSMIC's Cancer Genes Census catalogues genes that exhibit mutations that are causally implicated in cancer pathogenesis (see Supplementary Material for the complete list). Of all

COSMIC genes harbouring germline cancer mutations ($n = 87$) associated with cancer predisposition, six high-impact variants in five cancer predisposition genes were noted (Supplementary Table S2). Five of the variants were singletons, and one was a doubleton: rs34295337 in *ERCC3*, a gene associated with xeroderma pigmentosum type B (Ma *et al.*, 1994), which is a rare autosomal recessive disease that is associated with skin cancer (Paszowska-Szczur *et al.*, 2013). One variant, rs11571833, in the *BRCA2* gene, was described previously as being associated with an increased risk of developing a variety of cancer types including lung, breast, prostate, gastric and aerodigestive tract cancer (Wang *et al.*, 2014; Delahaye-Sourdeix *et al.*, 2015; Thompson *et al.*, 2015; Meeks *et al.*, 2016; Vijai *et al.*, 2016). Two variants, one in the *DICER1* gene and one in the *NFI* gene, were novel. The *NFI* gene harboured one additional high-impact variant. Notably, *NFI* germline mutations underlie the neurofibromatosis type 1 phenotype, a disease that is reportedly diagnosed at higher rates in the AJP than in the European population (Garty *et al.*, 1994).

The ACMG recommendation for reporting incidental findings in clinical sequencing includes 56 genes (22 genes intersect with COSMIC genes; see Supplementary Material for the complete list). High-impact variants were noted in two ACMG genes. The first variant (rs11571833) in the *BRCA2* gene was already described and discussed above. The second variant, rs200563280, results in a premature stop codon in the *RYR1* gene, a gene that is associated with malignant hyperthermia (Robinson *et al.*, 2006). Thus, the rate of actionable

incidental findings in the AJP is 1.56%, similar to the estimate for Europeans at approximately 2% (Amendola *et al.*, 2015). None of the above variants were mentioned in a recent study, based on the same dataset that expanded the recommendations for an AJP screening panel (Baskovich *et al.*, 2016).

AJP-specific variants

We next examined AJP-specific variants. We defined variants as AJP specific if they were unique (i.e. novel) or very rare (minor allele frequency (MAF) <1%) in the NJPs, but more prevalent in the AJP (MAF >1%). Of the total AJP variants, 17,977 (8%) were AJP specific. To confirm that our dataset is enriched with variants that are unique to the AJP, we performed the same analysis on 128 verified Europeans from the Personal Genome Project (PGP) (Church, 2005; see Supplementary methods). Only 8748 variants (3.6% of the PGP dataset) were more than 1% in the PGP dataset but not in NJPs (both European and non-European populations).

We then looked at genes that are enriched for moderate- to high-impact variant groups that are AJP specific. This analysis yielded 5142 variants. Most genes harboured up to one such variant, 840 genes exhibited two variants and 196 genes displayed three or more moderate- to high-impact variants (Supplementary Fig. S2). After QC (see Supplementary Methods), three outlier genes were filtered out (Supplementary Fig. S2). In this analysis, virtually no correlation between the number of variants and the genomic length of the gene was observed (Pearson's correlation = 0.1). Next, we examined the residual variation intolerance score (RVIS) (Petrovski *et al.*, 2013) in order to identify genes under purifying selection that harbour unique or prevalent mutations in the AJP. Briefly, RVIS measures the tolerance of a gene to contain damaging variation. Genes with a low RVIS are predicted to be less tolerant to variation, and hence are more likely to exhibit a phenotype due to non-synonymous variants. The *APC* gene harboured a high number of AJP-specific variants ($n = 7$) and is in the lowest 0.2 percentile of RVIS (Fig. 2(a)). Mutations in the *APC* gene are associated with a specific form of inherited predisposition to colorectal cancer. Overall, colorectal cancer is more prevalent in the AJP than in NJPs (Feldman, 2001). Notably, the p.I1307K missense mutation in *APC* (rs1801155), which has been previously shown to moderately increase colorectal cancer risk in the AJP (Woodage *et al.*, 1998), was among the identified variants (MAF = 0.047), and was recommended for inclusion in AJP screening (Baskovich *et al.*, 2016). However, additional susceptibility variants were detected in the *APC* gene, suggesting that other variants may contribute to the increased prevalence of colorectal cancer in the AJP. Other genes with low RVIS and

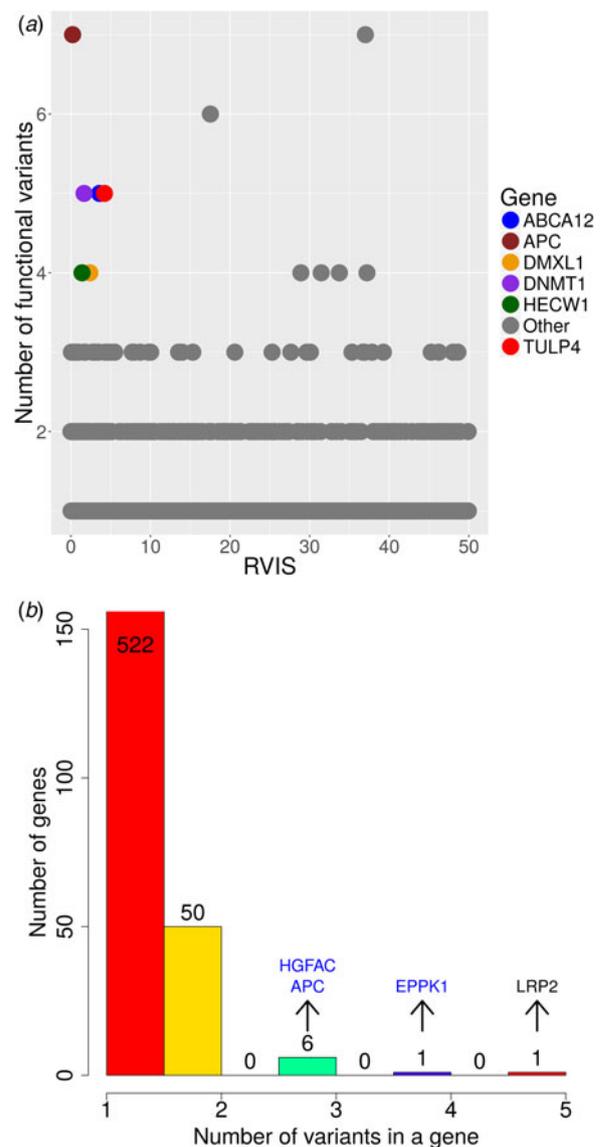


Fig. 2. (a) Genes with a low residual variation intolerance score (RVIS) are less tolerant to rare functional variants. Only six genes had a very low RVIS and four or more high to moderate Ashkenazi Jewish population (AJP)-specific variants, including the *APC* gene, which had the lowest RVIS and highest number of variants at seven. (b) Histogram of the number of AJP-specific deleterious variants in a gene. While most of the genes had two or fewer of these variants, eight genes had three to five variants.

harbouring four AJP-specific damaging variants are *ABCA12*, *TULP4*, *DNMT1*, *DMXL1* and *HECW1*. To the best of our knowledge, the prevalence of the phenotypes associated with these genes (Supplementary Table S3) is not significantly higher in the AJP compared with other NJPs. Hence, the clinical implications and significance of this seemingly high rate of damaging variants in these genes warrant further investigation in additional extended Ashkenazi Jewish studies.

To assess the effect of the AJP-specific variants on protein function, we used the MetaLR (Dong *et al.*,

2015) ensemble tool, which integrates different prediction tools using logistic regression to predict whether a variant is deleterious (see Supplementary Methods). Overall, we obtained 649 AJP-specific deleterious variants in 580 different genes. Only eight genes had at least three AJP-specific deleterious variants (Fig. 2(b) and Supplementary Table S3): *APC*, *ABCA12*, *LRP2*, *EPPKI*, *HGFAC*, *ACAD11*, *HLCS* and *NOX1*. *APC* and *ABCA12* were discussed; the *HGFAC* (three variants) gene is a member of the peptidase S1 protein family and is associated with pancreatic cancer (Kitajima *et al.*, 2008), a cancer type that is known to be more frequent among the AJP (Feldman, 2001). The *EPPKI* gene (four variants) encodes a protein that belongs to the plakin family and is related to ‘vacterl association’ disorder (Hilger *et al.*, 2013). The phenotype of this disorder encompasses Fanconi anaemia, a phenotype that is diagnosed at a higher frequency in the AJP compared with NJPs (Kutler & Auerbach, 2004), and hence, these variants may contribute to these higher occurrence rates. The other genes are associated with different types of rare diseases, but to the best of our knowledge, these conditions are not diagnosed at an increased rate in the AJP (Supplementary Table S3).

Furthermore, to examine whether the genes harbouring AJP-specific deleterious variants were previously implicated as AJP-prevalent phenotypes, we queried VarElect (<http://varelect.genecards.org/>) using the term ‘Ashkenazi’. VarElect can prioritise genotype–phenotype associations based on various databases. Of the 580 queried genes, 14 genes harbouring 17 variants (Table 1) were found to be directly related to the ‘Ashkenazi’ term, denoting conditions that are common to the AJP. Five of the 17 variants are considered to be pathogenic by the Clinvar database, four of the variants were also included in the recent recommendation for the AJP screening panel (Baskovich *et al.*, 2016) and four of the genes are included in the AJP screening panel, but for different variants. To verify our results, we did the same for the 128 European individuals looking at European-specific variants, meaning genes with variants that were very rare in the non-European population but not in the European population (423 genes), and tried to find genes that were related to the ‘Ashkenazi’ phenotype. Although 20 genes were found to be related, none of the variants in them was found to be pathogenic by Clinvar, which further supports our results. Taken together, these results suggest that additional variants, among these 17 variants, are plausibly causal and hence should be further investigated.

Using the Ashkenazi Jewish database in an analysis of Ashkenazi Jewish early BC patients

The major objective of clinical sequencing is to identify the causative mutation from amongst numerous

detected variants. To that end, non-synonymous variants with rare allele frequencies are considered initially as plausible causative mutations. Since the AJP is not included in any of the public databases of international sequencing efforts, the MAFs of closely related populations such as Europeans (Haas *et al.*, 2012; Lee *et al.*, 2012; Rees *et al.*, 2012) are often utilised as surrogates. We evaluated the advantages of using AJP-specific MAFs when screening the WES data of Ashkenazi Jewish samples. Of the 55,416 high- and moderate-impact mutations, 57.7% were classified as very rare based on the general European MAF versus 50.6% based on the AJP MAF, leading to out-filtration of approximately 3900 variants (Fig. 3(a)). Likewise, based on the maximum MAF (MMAF) of all NJPs, 50.1% of the variants were classified as very rare, compared to 40.8% when including the AJP. These results are in line with Carmi *et al.* (2014). For rare variants (MAF <5%), the advantage of using AJP-specific MAFs is somewhat less significant (1.2% difference), in line with the notion that population-specific variants are predominantly very rare (1000 Genomes Project Consortium *et al.*, 2015).

Similarly, potentially deleterious variants are prioritised in clinical NGS applications. Based on the AJP MAF, 79.0% of deleterious variants, based on MetaLR, were considered very rare, whereas 89.6% were considered very rare based on the European MAF (Fig. 3(b)). Furthermore, combining the AJP MAF with the NJP MMAF substantially improved filtering from 85.9% of the variants classified as very rare to just 72.9%. Since the MAFs of numerous populations, but not the AJP, are included in the MetaLR model, adding the AJP MAF can significantly improve the filtering of deleterious variants. Taken together, these significant population-specific differences in rare variants indicate that by utilising AJP-specific MAFs, finer filtration and lower false-positive rates can be achieved in Ashkenazi Jewish sequencing studies.

Importantly, we evaluated the utility of the AJP-specific screening approach using the independent WES data of 49 Ashkenazi Jewish samples derived from high-risk BC cases who do not harbour mutations in the predominant underlying genes – *BRCAl* and *BRCa2*. Of the 2638 predicted deleterious variants, 81.3% were very rare according to the European MAF, compared to 77.5% using the AJP MAF. Similarly, combining the AJP with the NJP MMAF improved filtering by approximately 10% from 75.9% to 64.5% (Supplementary Fig. S3).

In our actual disease gene analysis of the Ashkenazi Jewish BC sample, we screened for very rare variants that are potentially deleterious by MetaLR and are present in at least three BC cases, resulting in 450 potentially deleterious variants. Filtering by using the European MAF resulted in 189 variants in 148 genes,

Table 1. Ashkenazi Jewish population-specific deleterious predicted variants in genes that relate to Ashkenazi Jews according to VarElect.

Gene	Number of AJP-specific deleterious variants	rs number	RVIS percentile	VarElect score ^a	Diseases related to gene by VarElect ^b	Diseases related to variant by Clinvar ^c	Gene/variant mentioned in Baskovich <i>et al.</i> (2016) ^d
<i>APC</i>	3	rs140493115, chr5:112175676, rs200756935	0.22	5.36	Colorectal cancer	–	Gene
<i>WFS1</i>	2	rs71524353, rs35218685	98.91	1.39	Type 2 diabetes	–	–
<i>AK3</i>	1	rs141090350	26.48	1.39	Crohn's disease	–	–
<i>ASPA</i>	1	rs28940279	63.46	2.4	Canavan disease	Canavan disease	Variant
<i>F11</i>	1	rs121965064	14.91	3.1	Factor XI deficiency	Factor XI deficiency	Variant
<i>COL4A3</i>	1	rs55816283	48.42	1.39	Alport syndrome	–	–
<i>CFTR</i>	1	rs1800073	81.9	1.39	Cystic fibrosis	–	Gene
<i>DLD</i>	1	rs121964990	20.72	2.77	Lipoamide dehydrogenase deficiency	Dihydrolipoyl dehydrogenase deficiency	Variant
<i>LDLR</i>	1	rs137853964	9.94	1.96	Familial hypercholesterolaemia	Familial hypercholesterolaemia	Gene
<i>NDUFS4</i>	1	chr5:52942219	44.18	1.39	Leigh syndrome	–	Gene
<i>NOD2</i>	1	rs104895438	91.35	3.39	Crohn's disease	Sarcoidosis	–
<i>KCNH6</i>	1	rs139719681	92.33	1.39	Crohn's disease	–	–
<i>TIMP1</i>	1	rs149986790	94.82	1.39	Refractive error	–	–
<i>VKORC1</i>	1	rs61742245	88.61	1.96	Warfarin pharmacogenetics	Warfarin pharmacogenetics	Variant

^a A score given by VarElect to show how much the gene was found to be related to the Ashkenazi phenotype.

^b The diseases that were related to this gene in the context of the Ashkenazi phenotype according to VarElect.

^c Diseases that were related to the variant according to Clinvar.

^d Is the gene or the variant also found in a recent recommended AJP screening panel?

AJP = Ashkenazi Jewish population; RVIS = residual variation intolerance score.

deleterious variants. Homozygous mutations in this gene are associated with Fanconi anaemia, a disorder that is more commonly encountered in AJP (Kutler & Auerbach, 2004). Moreover, heterozygous mutations in Fanconi anaemia genes are associated with increased cancer risk, primarily BC (Mathew, 2006; Alan & D'Andrea, 2010), and indeed, two of the four AJP-specific deleterious variants in the *EPPK1* gene were also detected in the high-risk BC cohort. Among the observed AJP-specific deleterious variants, five were known to be pathogenic variants that increase the risk of five different diseases that are common to the AJP, and three of them were included in a new recommended screening panel for the AJP (Baskovich *et al.*, 2016). These overlaps confirm the effectiveness of the methodology applied in the present study for finding population-based pathogenic variants, as well as supporting the potential of population screening using NGS. Additionally, by examining specific genes with known and valuable clinical implications and consequences (i.e. ACMG incidental findings genes and COSMIC germline mutation-harboring genes), a number of variants were identified in genes that lead to a phenotype that is seen at a higher occurrence in the AJP than in other populations (e.g. the *NFI* gene).

Based on the results of the present study and the current ACMG incidental findings recommendations, in approximately 3/200 (1.56%) members of the AJP who undergo WES, an incidental finding will emerge. As information about the role of each variant in the exome/genome accumulates and the pathogenicity prediction tools and functional analyses continue to evolve, some of the moderate-impact variants of these genes might also be reclassified as pathogenic, so that the rate of incidental findings may still be altered.

The present study also illustrated the importance of using the Ashkenazi Jewish-specific database in the course of analysing the genetic basis of inherited cancer in the AJP. Using the dataset and analysis tools, the number of potential causal sequence variants underlying an inherited predisposition to BC was reduced by 36%. Such a filtering step is critical to defining a bona fide causal mutation. Therefore, this provides further support for the importance of creating and using a population-specific database when investigating the genetic basis of inherited diseases, rather than using genetically related but not identical populations.

While a recent study of 5685 Ashkenazi Jewish exomes has been published (Rivas *et al.*, 2016), the current study provides evidence that by using whole-exome data from a relatively small number ($n = 128$) of Ashkenazi Jewish individuals, clinically relevant information and improvements in filter annotation are feasible. Thus, the research potential value and

clinical benefits of using NGS technology at a population level are further emphasised.

The Shomron laboratory is supported by the Israel Cancer Research Fund (ICRF), Research Career Development Award (RCDA); Wolfson Family Charitable Fund; Earlier.org – Friends for an Earlier Breast Cancer Test; Claire and Amedee Maratier Institute for the Study of Blindness and Visual Disorders; I-CORE Program of the Planning and Budgeting Committee, The Israel Science Foundation (grant number 41/11); the Israeli Ministry of Defense, Office of Assistant Minister of Defense for Chemical, Biological, Radiological and Nuclear (CBRN) Defense; Foundation Fighting Blindness; Saban Family Foundation, Melanoma Research Alliance; Binational Science Foundation (BSF); Israel Cancer Research Fund (ICRF) Acceleration Grant; Israel Cancer Association (ICA); Donation from the Kateznik K. Association Holocaust; Margot Stoltz Foundation through the Faculty of Medicine grants of Tel Aviv University; The Varda and Boaz Dotan Research Center in Hemato-Oncology, Idea Grant; 'Lirot' Association and the Consortium for Mapping Retinal Degeneration Disorders in Israel; Interdisciplinary grant of the Israeli Ministry of Science, Technology and Space on the Science, Technology and Innovation for the Third Age; The Edmond J. Safra Center for Ethics at Tel Aviv University; Check Point Institute for Information Security; Joint Core Program of Research on the Molecular Basis of Human Disease, Shabbetai Donnolo Fellowships supported by the Italian Ministry of Foreign Affairs; Israel Science Foundation (ISF, 1852/16); and the Edmond J. Safra Center for Bioinformatics at Tel Aviv University.

Supplementary material

For supplementary material accompanying this paper visit <https://doi.org/10.1017/S0016672317000015>.

References

- 1000 Genomes Project Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., Handsaker, R. E., Kang, H. M., Marth, G. T., McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**(7422), 56–65.
- 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature* **526**(7571), 68–74.
- Abeliovich, D., Lavon, I. P., Lerer, I., Cohen, T., Springer, C., Avital, A. & Cutting, G. R. (1992). Screening for five mutations detects 97% of cystic fibrosis (CF) chromosomes and predicts a carrier frequency of 1:29 in the Jewish Ashkenazi population. *American Journal of Human Genetics* **51**(5), 951–956.
- Alan, D. & D'Andrea, M. D. (2010). The Fanconi anemia and breast cancer susceptibility pathways. *The New England Journal of Medicine* **362**(20), 1909–1919.
- Amendola, L. M., Dorschner, M. O., Robertson, P. D., Salama, J. S., Hart, R., Shirts, B. H., Murray, M. L., Tokita, M. J., Gallego, C. J., Kim, D. S., Bennett, J. T., Crosslin, D. R., Ranchalis, J., Jones, K. L., Rosenthal, E. A., Jarvik, E. R., Itsara, A., Turner, E. H., Herman,

- D. S., Schleit, J., Burt, A., Jamal, S. M., Abrudan, J. L., Johnson, A. D., Conlin, L. K., Dulik, M. C., Santani, A., Metterville, D. R., Kelly, M., Foreman, A. K. M., Lee, K., Taylor, K. D., Guo, X., Crooks, K., Kiedrowski, L. A., Raffel, L. J., Gordon, O., Machini, K., Desnick, R. J., Biesecker, L. G., Lubitz, S. A., Mulchandani, S., Cooper, G. M., Joffe, S., Richards, C. S., Yang, Y., Rotter, J. I., Rich, S. S., O'Donnell, C. J., Berg, J. S., Spinner, N. B., Evans, J. P., Fullerton, S. M., Leppig, K. A., Bennett, R. L., Bird, T., Sybert, V. P., Grady, W. M., Tabor, H. K., Kim, J. H., Bamshad, M. J., Wilfond, B., Motulsky, A. G., Scott, C. R., Pritchard, C. C., Walsh, T. D., Burke, W., Raskind, W. H., Byers, P., Hisama, F. M., Rehm, H., Nickerson, D. A. & Jarvik, G. P. (2015). Actionable exomic incidental findings in 6503 participants: challenges of variant classification. *Genome Research* **25**(3), 305–315.
- Baglietto, L., Lindor, N. M., Dowty, J. G., White, D. M., Wagner, A., Garcia, E. B. G., Vriends, A. H. J. T, Dutch Lynch Syndrome Study Group, Cartwright, N. R., Barnetson, R. A., Farrington, S. M., Tenesa, A., Hampel, H., Buchanan, D., Arnold, S., Young, J., Walsh, M. D., Jass, J., Macrae, F., Antill, Y., Winship, I. M., Giles, G. G., Goldblatt, J., Parry, S., Suthers, G., Leggett, B., Butz, M., Aronson, M., Poynter, J. N., Baron, J. A., Le Marchand, L., Haile, R., Gallinger, S., Hopper, J. L., Potter, J., de la Chapelle, A., Vasen, H. F., Dunlop, M. G., Thibodeau, S. N. & Jenkins, M. A. (2010). Risks of Lynch syndrome cancers for *MSH6* mutation carriers. *Journal of the National Cancer Institute* **102**(3), 193–201.
- Baskovich, B., Hiraki, S., Upadhyay, K., Meyer, P., Carmi, S., Barzilai, N., Darvasi, A., Ozelius, L., Peter, I., Cho, J. H., Atzmon, G., Clark, L., Yu, J., Lencz, T., Pe'er, I., Ostrer, H. & Oddoux, C. (2016). Expanded genetic screening panel for the Ashkenazi Jewish population. *Genetics in Medicine* **18**(5), 522–528.
- Behar, D. M., Thomas, M. G., Skorecki, K., Hammer, M. F., Bulygina, E., Rosengarten, D., Jones, A. L., Held, K., Moses, V., Goldstein, D., Bradman, N. & Weale, M. E. (2003). Multiple origins of Ashkenazi Levites: Y chromosome evidence for both Near Eastern and European ancestries. *American Journal of Human Genetics* **73**(4), 768–779.
- Beutler, E., Nguyen, N. J., Henneberger, M. W., Smolec, J. M., McPherson, R. A., West, C. & Gelbart, T. (1993). Gaucher disease: gene frequencies in the Ashkenazi Jewish population. *American Journal of Human Genetics* **52**(1), 85–88.
- Bray, S. M., Mulle, J. G., Dodd, A. F., Pulver, A. E., Wooding, S. & Warren, S. T. (2010). Signatures of founder effects, admixture, and selection in the Ashkenazi Jewish population. *Proceedings of the National Academy of Sciences of the United States of America* **37**, 16222–16227.
- Carmi, S., Hui, K. Y., Kochav, E., Liu, X., Xue, J., Grady, F., Guha, S., Upadhyay, K., Ben-Avraham, D., Mukherjee, S., Bowen, B. M., Thomas, T., Vijai, J., Cruts, M., Froyen, G., Lambrechts, D., Plaisance, S., Van Broeckhoven, C., Van Damme, P., Van Marck, H., Barzilai, N., Darvasi, A., Offit, K., Bressman, S., Ozelius, L. J., Peter, I., Cho, J. H., Ostrer, H., Atzmon, G., Clark, L. N., Lencz, T. & Pe'er, I. (2014). Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. *Nature Communications* **5**, 4835.
- Church, G. M. (2005). The Personal Genome Project. *Molecular Systems Biology* **1**, 2005-0030.
- Costa, M. D., Pereira, J. B., Pala, M., Fernandes, V., Olivieri, A., Achilli, A., Perego, U. A., Rychkov, S., Naumova, O., Hatina, J., Woodward, S. R., Eng, K. K., Macaulay, V., Carr, M., Soares, P., Pereira, L. & Richards, M. B. (2013). A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. *Nature Communications* **4**, 2543.
- Delahaye-Sourdeix, M., Anantharaman, D., Timofeeva, M. N., Gaborieau, V., Chabrier, A., Vallée, M. P., Lagiou, P., Holcátová, I., Richiardi, L., Kjaerheim, K., Agudo, A., Castellsagué, X., Macfarlane, T. V., Barzan, L., Canova, C., Thakker, N. S., Conway, D. I., Znaor, A., Healy, C. M., Ahrens, W., Zaridze, D., Szeszenia-Dabrowska, N., Lissowska, J., Fabianova, E., Mates, I. N., Bencko, V., Foretova, L., Janout, V., Curado, M. P., Koifman, S., Menezes, A., Wünsch-Filho, V., Eluf-Neto, J., Boffetta, P., Fernández Garrote, L., Polesel, J., Lener, M., Jaworowska, E., Lubiński, J., Boccia, S., Rajkumar, T., Samant, T. A., Mahimkar, M. B., Matsuo, K., Franceschi, S., Byrnes, G., Brennan, P. & McKay, J. D. (2015). A rare truncating *BRCA2* variant and genetic susceptibility to upper aerodigestive tract cancer. *Journal of the National Cancer Institute* **107**(5), djv037.
- Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K. & Liu, X. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Human Molecular Genetics* **24**(8), 2125–2137.
- Feldman, G. E. (2001). Do Ashkenazi Jews have a higher than expected cancer burden? Implications for cancer control prioritization efforts. *The Israel Medical Association journal: IMAJ* **3**(5), 341–346.
- Garty, B. Z., Laor, A. & Danon, Y. L. (1994). Neurofibromatosis type 1 in Israel: survey of young adults. *Journal of Medical Genetics* **31**(11), 853–857.
- Gudbjartsson, D. F., Helgason, H., Gudjonsson, S. A., Zink, F., Oddson, A., Gylfason, A., Besenbacher, S., Magnusson, G., Halldorsson, B. V., Hjartarson, E., Sigurdsson, G. T., Stacey, S. N., Frigge, M. L., Holm, H., Saemundsdottir, J., Helgadóttir, H. T., Johannsdóttir, H., Sigfusson, G., Thorgerisson, G., Sverrisson, J. T., Gretarsdóttir, S., Walters, G. B., Rafnar, T., Thjodleifsson, B., Bjornsson, E. S., Olafsson, S., Thorarinsdóttir, H., Steingrimsdóttir, T., Gudmundsdóttir, T. S., Theodors, A., Jonasson, J. G., Sigurdsson, A., Bjornsdóttir, G., Jonsson, J. J., Thoraransen, O., Ludvigsson, P., Gudbjartsson, H., Eyjolfsson, G. I., Sigurdardóttir, O., Olafsson, I., Arnar, D. O., Magnusson, O. T., Kong, A., Masson, G., Thorsteinsdóttir, U., Helgason, A., Sulem, P. & Stefansson, K. (2015). Large-scale whole-genome sequencing of the Icelandic population. *Nature Genetics* **47**(5), 435–444.
- Haas, J. T., Winter, H. S., Lim, E., Kirby, A., Blumenstiel, B., DeFelice, M., Gabriel, S., Jalas, C., Branski, D., Grueter, C. A., Toporovski, M. S., Walther, T. C., Daly, M. J. & Farese Jr, R. V. (2012). *DGAT1* mutation is linked to a congenital diarrheal disorder. *The Journal of Clinical Investigation* **122**(12), 4680–4684.
- Hilger, A., Schramm, C., Pennimpede, T., Wittler, L., Dworschak, G. C., Bartels, E., Engels, H., Zink, A. M., Degenhardt, F., Müller, A. M., Schmiedeke, E., Grasshoff-Derr, S., Märzheuser, S., Hosie, S., Holland-Cunz, S., Wijers, C. H., Marcelis, C. L., van Rooij, I. A., Hildebrandt, F., Herrmann, B. G., Nöthen, M. M., Ludwig, M., Reutter, H. & Draaken, M. (2013). *De novo* microduplications at 1q41, 2q37.3, and 8q24.3 in patients with VATER/VACTERL association. *European Journal of Human Genetics: EJHG* **21**(12), 1377–1382.
- Kitajima, Y., Ide, T., Ohtsuka, T. & Miyazaki, K. (2008). Induction of hepatocyte growth factor activator gene

- expression under hypoxia activates the hepatocyte growth factor/c-Met system via hypoxia inducible factor-1 in pancreatic cancer. *Cancer Science* **99**(7), 1341–1347.
- Kutler, D. I. & Auerbach, A. D. (2004). Fanconi anemia in Ashkenazi Jews. *Familial Cancer* **3**(3–4), 241–248.
- Lee, Y. C., Durr, A., Majczenko, K., Huang, Y. H., Liu, Y. C., Lien, C. C., Tsai, P. C., Ichikawa, Y., Goto, J., Monin, M. L., Li, J. Z., Chung, M. Y., Mundwiller, E., Shakkottai, V., Liu, T. T., Tesson, C., Lu, Y. C., Brice, A., Tsuji, S., Burmeister, M., Stevanin, G. & Soong, B. W. (2012). Mutations in *KCND3* cause spinocerebellar ataxia type 22. *Annals of Neurology* **72**(6), 859–869.
- Ma, L., Siemssen, E. D., Noteborn, M. H. M. & van der Eb, A. J. (1994). The xeroderma pigmentosum group B protein ERCC3 produced in the baculovirus system exhibits DNA helicase activity. *Nucleic Acids Research* **22**(20), 4095–4102.
- Mathew, C. G. (2006). Fanconi anaemia genes and susceptibility to cancer. *Oncogene* **25**(43), 5875–5884.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**(9), 1297–1303.
- Meeks, H. D., Song, H., Michailidou, K., Bolla, M. K., Dennis, J., Wang, Q., Barrowdale, D., Frost, D.; EMBRACE, McGuffog, L., Ellis, S., Feng, B., Buys, S. S., Hopper, J. L., Southey, M. C., Tesoriero, A.; kConFab Investigators, James, P. A., Bruinsma, F., Campbell, I. G.; Australia Ovarian Cancer Study Group, Broeks, A., Schmidt, M. K., Hogervorst, F. B.; HEBON, Beckman, M. W., Fasching, P. A., Fletcher, O., Johnson, N., Sawyer, E. J., Riboli, E., Banerjee, S., Menon, U., Tomlinson, I., Burwinkel, B., Hamann, U., Marme, F., Rudolph, A., Janavicius, R., Tihomirova, L., Tung, N., Garber, J., Cramer, D., Terry, K. L., Poole, E. M., Tworoger, S. S., Dorfling, C. M., van Rensburg, E. J., Godwin, A. K., Guénel, P., Truong, T.; GEMO Study Collaborators, Stoppa-Lyonnet, D., Damiola, F., Mazoyer, S., Sinilnikova, O. M., Isaacs, C., Maugard, C., Bojesen, S. E., Flyger, H., Gerdes, A. M., Hansen, T. V., Jensen, A., Kjaer, S. K., Hogdall, C., Hogdall, E., Pedersen, I. S., Thomassen, M., Benitez, J., González-Neira, A., Osorio, A., Hoya Mde, L., Segura, P. P., Diez, O., Lazaro, C., Brunet, J., Anton-Culver, H., Eunjung, L., John, E. M., Neuhausen, S. L., Ding, Y. C., Castillo, D., Weitzel, J. N., Ganz, P. A., Nussbaum, R. L., Chan, S. B., Karlan, B. Y., Lester, J., Wu, A., Gayther, S., Ramus, S. J., Sieh, W., Whittemore, A. S., Monteiro, A. N., Phelan, C. M., Terry, M. B., Piedmonte, M., Offit, K., Robson, M., Levine, D., Moysich, K. B., Cannioto, R., Olson, S. H., Daly, M. B., Nathanson, K. L., Domchek, S. M., Lu, K. H., Liang, D., Hildebrandt, M. A., Ness, R., Modugno, F., Pearce, L., Goodman, M. T., Thompson, P. J., Brenner, H., Butterbach, K., Meindl, A., Hahnen, E., Wappenschmidt, B., Brauch, H., Brüning, T., Blomqvist, C., Khan, S., Nevanlinna, H., Pelttari, L. M., Aittomäki, K., Butzow, R., Bogdanova, N. V., Dörk, T., Lindblom, A., Margolin, S., Rantala, J., Kosma, V. M., Mannermaa, A., Lambrechts, D., Neven, P., Claes, K. B., Maerken, T. V., Chang-Claude, J., Flesch-Janys, D., Heitz, F., Varon-Mateeva, R., Peterlongo, P., Radice, P., Viel, A., Barile, M., Peissel, B., Manoukian, S., Montagna, M., Oliani, C., Peixoto, A., Teixeira, M. R., Collavoli, A., Hallberg, E., Olson, J. E., Goode, E. L., Hart, S. N., Shimelis, H., Cunningham, J. M., Giles, G. G., Milne, R. L., Healey, S., Tucker, K., Haiman, C. A., Henderson, B. E., Goldberg, M. S., Tischkowitz, M., Simard, J., Soucy, P., Eccles, D. M., Le, N., Borresen-Dale, A. L., Kristensen, V., Salvesen, H. B., Bjorge, L., Bandera, E. V., Risch, H., Zheng, W., Beeghly-Fadiel, A., Cai, H., Pylkäs, K., Tollenaar, R. A., Ouweland, A. M., Andrulis, I. L., Knight, J. A.; OCGN, Narod, S., Devilee, P., Winqvist, R., Figueroa, J., Greene, M. H., Mai, P. L., Loud, J. T., García-Closas, M., Schoemaker, M. J., Czene, K., Darabi, H., McNeish, I., Siddiquil, N., Glasspool, R., Kwong, A., Park, S. K., Teo, S. H., Yoon, S. Y., Matsuo, K., Hosono, S., Woo, Y. L., Gao, Y. T., Foretova, L., Singer, C. F., Rappaport-Furhauer, C., Friedman, E., Laitman, Y., Rennert, G., Imyanitov, E. N., Hulick, P. J., Olopade, O. I., Senter, L., Olah, E., Doherty, J. A., Schildkraut, J., Koppert, L. B., Kiemeny, L. A., Massuger, L. F., Cook, L. S., Pejovic, T., Li, J., Borg, A., Öfverholm, A., Rosing, M. A., Wentzensen, N., Henriksson, K., Cox, A., Cross, S. S., Pasini, B. J., Shah, M., Kabisch, M., Torres, D., Jakubowska, A., Lubinski, J., Gronwald, J., Agnarsson, B. A., Kupryjanczyk, J., Moes-Sosnowska, J., Fostira, F., Konstantopoulou, I., Slager, S., Jones, M.; PRostate cancer AssOCIation group To Investigate Cancer Associated aLterations in the genome, Antoniou, A. C., Berchuck, A., Swerdlow, A., Chenevix-Trench, G., Dunning, A. M., Pharoah, P. D., Hall, P., Easton, D. F., Couch, F. J., Spurdle, A. B. & Goldgar, D. E. (2016). *BRCA2* polymorphic stop codon K3326X and the risk of breast, prostate, and ovarian cancers. *Journal of the National Cancer Institute* **108**(2), djv315.
- Myerowitz, R. & Costigan, F. C. (1988). The major defect in Ashkenazi Jews with Tay–Sachs disease is an insertion in the gene for the alpha-chain of beta-hexosaminidase. *Journal of Biological Chemistry* **263**(35), 18587–18589.
- Nagasaki, M., Yasuda, J., Katsuoka, F., Nariai, N., Kojima, K., Kawai, Y., Yamaguchi-Kabata, Y., Yokozawa, J., Danjoh, I., Saito, S., Sato, Y., Mimori, T., Tsuda, K., Saito, R., Pan, X., Nishikawa, S., Ito, S., Kuroki, Y., Tanabe, O., Fuse, N., Kuriyama, S., Kiyomoto, H., Hozawa, A., Minegishi, N., Douglas Engel, J., Kinoshita, K., Kure, S., Yaegashi, N.; ToMMo Japanese Reference Panel Project & Yamamoto, M. (2015). Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. *Nature Communications* **6**, 8018.
- Ozelius, L. J., Senthil, G., Saunders-Pullman, R., Ohmann, E., Deligtisch, A., Tagliati, M., Hunt, A. L., Klein, C., Henick, B., Hailpern, S. M., Lipton, R. B., Soto-Valencia, J., Risch, N. & Bressman, S. B. (2006). *LRRK2* G2019S as a cause of Parkinson's disease in Ashkenazi Jews. *New England Journal of Medicine* **354**(4), 424–425.
- Paszowska-Szczur, K., Scott, R. J., Serrano-Fernandez, P., Mirecka, A., Gapska, P., Górski, B., Cybulski, C., Maleszka, R., Sulikowski, M., Nagay, L., Lubinski, J. & Dębniak, T. (2013). Xeroderma pigmentosum genes and melanoma risk. *International Journal of Cancer* **133**(5), 1094–1100.
- Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S. & Goldstein, D. B. (2013). Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genetics* **9**(8), e1003709.
- Rees, M. G., Ng, D., Ruppert, S., Turner, C., Beer, N. L., Swift, A. J., Morken, M. A., Below, J. E., Blech, I.; NISC Comparative Sequencing Program, Mullikin, J. C., McCarthy, M. I., Biasecker, L. G., Gloyn, A. L. & Collins, F. S. (2012). Correlation of rare coding variants in the gene encoding human glucokinase regulatory

- protein with phenotypic, cellular, and kinetic outcomes. *The Journal of Clinical Investigation* **122**(1), 205–217.
- Rivas, M. A., Koskela, J., Huang, H., Stevens, C., Avila, B. E., Haritunians, T., Neale, B., Kurki, M., Ganna, A., Graham, D., Glaser, B., Peter, I., Atzmon, G., Barzilai, N., Levine, A., Schiff, E., Pontikos, N., Weisburd, B., Karczewski, K. J., Minikel, E., Petersen, B.-S., Beaugerie, L., Seksik, P., Cosnes, J., Schreiber, S., Bokemeyer, B., Bethge, J.; NIDDK IBD Genetics Consortium, T2D-GENES Consortium, Heap, G., Ahmad, T., Plagnol, V., Segal, A. W., Targan, S., Turner, D., Saavalainen, P., Farkkila, M., Kontula, K., Pirinen, M., Palotie, A., Brant, S. R., Duerr, R. H., Silverberg, M. S., Rioux, J. D., Weersma, R. K., Franke, A., MacArthur, D. G., Jalas, C., Sokol, H., Xavier, R. J., Pulver, A., Cho, J. H., McGovern, D. P. B. & Daly, M. J. (2016). Insights into the genetic epidemiology of Crohn's and rare diseases in the Ashkenazi Jewish population. *bioRxiv* 077180.
- Robinson, R., Carpenter, D., Shaw, M.-A., Halsall, J. & Hopkins, P. (2006). Mutations in *RYR1* in malignant hyperthermia and central core disease. *Human Mutation* **27**(10), 977–989.
- Rosner, G., Rosner, S. & Orr-Urtreger, A. (2009). Genetic testing in Israel: an overview. *Annual Review of Genomics and Human Genetics* **10**(1), 175–192.
- Slatkin, M. (2004). A population-genetic test of founder effects and implications for Ashkenazi Jewish diseases. *American Journal of Human Genetics* **75**(2), 282–293.
- Struewing, J. P., Hartge, P., Wacholder, S., Baker, S. M., Berlin, M., McAdams, M., Timmerman, M. M., Brody, L. C. & Tucker, M. A. (1997). The risk of cancer associated with specific mutations of *BRCA1* and *BRCA2* among Ashkenazi Jews. *New England Journal of Medicine* **336**(20), 1401–1408.
- The Genome of the Netherlands Consortium (2014). Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nature Genetics* **46**(8), 818–825.
- The International HapMap Consortium (2005). A haplotype map of the human genome. *Nature* **437**(7063), 1299–1320.
- Thompson, E. R., Goringe, K. L., Rowley, S. M., Li, N., McNerny, S., Wong-Brown, M. W., Devereux, L., Li, J.; Lifepool Investigators, Trainer, A. H., Mitchell, G., Scott, R. J., James, P. A. & Campbell, I. G. (2015). Reevaluation of the *BRCA2* truncating allele c.9976A>T (p.Lys3326Ter) in a familial breast cancer context. *Scientific Reports* **5**, 14800.
- Vijai, J., Topka, S., Villano, D., Ravichandran, V., Maxwell, K. N., Maria, A., Thomas, T., Gaddam, P., Lincoln, A., Kazzaz, S., Wenz, B., Carmi, S., Schrader, K. A., Hart, S. N., Lipkin, S. M., Neuhausen, S. L., Walsh, M. F., Zhang, L., Lejbkowitz, F., Rennert, H., Stadler, Z. K., Robson, M., Weitzel, J. N., Domchek, S., Daly, M. J., Couch, F. J., Nathanson, K. L., Norton, L., Rennert, G. & Offit, K. (2016). A recurrent *ERCC3* truncating mutation confers moderate risk for breast cancer. *Cancer Discovery* **6**(11), 1267–1275.
- Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. (2012). Five years of GWAS discovery. *American Journal of Human Genetics* **90**(1), 7–24.
- Wang, Y., McKay, J. D., Rafnar, T., Wang, Z., Timofeeva, M. N., Broderick, P., Zong, X., Laplana, M., Wei, Y., Han, Y., Lloyd, A., Delahaye-Sourdeix, M., Chubb, D., Gaboricau, V., Wheeler, W., Chatterjee, N., Thorleifsson, G., Sulem, P., Liu, G., Kaaks, R., Henrion, M., Kinnersley, B., Vallée, M., LeCalvez-Kelm, F., Stevens, V. L., Gapstur, S. M., Chen, W. V., Zaridze, D., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., Mates, D., Bencko, V., Foretova, L., Janout, V., Krokán, H. E., Gabrielsen, M. E., Skorpen, F., Vatten, L., Njølstad, I., Chen, C., Goodman, G., Benhamou, S., Voeder, T., Vålk, K., Nelis, M., Metspalu, A., Lener, M., Lubiński, J., Johansson, M., Vineis, P., Agudo, A., Clavel-Chapelon, F., Bueno-de-Mesquita, H. B., Trichopoulos, D., Khaw, K. T., Johansson, M., Weiderpass, E., Tjønneland, A., Riboli, E., Lathrop, M., Scelo, G., Albanes, D., Caporaso, N. E., Ye, Y., Gu, J., Wu, X., Spitz, M. R., Dienemann, H., Rosenberger, A., Su, L., Matakidou, A., Eisen, T., Stefansson, K., Risch, A., Chanock, S. J., Christiani, D. C., Hung, R. J., Brennan, P., Landi, M. T., Houlston, R. S. & Amos, C. I. (2014). Rare variants of large effect in *BRCA2* and *CHEK2* affect risk of lung cancer. *Nature Genetics* **46**(7), 736–741.
- Wasielewski, M., Riaz, M., Vermeulen, J., van den Ouweland, A., Labrijn-Marks, I., Olmer, R., van der Spaa, L., Klijn, J. G., Meijers-Heijboer, H., Dooijes, D. & Schutte, M. (2010). Association of rare *MSH6* variants with familial breast cancer. *Breast Cancer Research and Treatment* **123**(2), 315–320.
- Win, A. K., Lindor, N. M. & Jenkins, M. A. (2013). Risk of breast cancer in Lynch syndrome: a systematic review. *Breast Cancer Research* **15**(2), R27.
- Woodage, T., King, S. M., Wacholder, S., Hartge, P., Struewing, J. P., McAdams, M., Laken, S. J., Tucker, M. A. & Brody, L. C. (1998). The *APC* I1307K allele and cancer risk in a community-based study of Ashkenazi Jews. *Nature Genetics* **20**(1), 62–65.