

CanVaS: Documenting the genetic variation spectrum of Greek cancer patients

Despoina Kalfakakou¹, Florentia Fostira², Athanasios Papathanasiou², Paraskevi Apostolou², Vasiliki Dellatola², Ioanna Gavra², Ioannis Vlachos³, Zacharias Scouras¹, Eleni Drosopoulou¹, Drakoulis Yannoukakos², and Irene Konstantopoulou²

¹Aristotle University of Thessaloniki

²National Centre for Scientific Research-Demokritos

³Beth Israel Deaconess Medical Center / Harvard Medical School Program in Placebo Studies and the Therapeutic Encounter

June 15, 2021

Abstract

National genetic variation registries vastly increase the level of detail for the relevant population, while directly affecting patient management. Herein, we report CanVaS, a Cancer Variation reSource aiming to document the genetic variation of cancer patients in Greece. CanVaS comprises germline genetic data from 7,363 Greek individuals with a personal and/or family history of malignancy. The dataset incorporates ~24,000 functionally annotated rare variants in 97 established or suspected cancer susceptibility genes. For each variant, allele frequency for the Greek population, interpretation for clinical significance, anonymized family and segregation information, as well as phenotypic traits of the carriers, are included. Moreover, information on the geographic distribution of the variants across the country are provided, enabling the study of Greek population isolates. Direct comparisons between Greek (sub)populations with relevant genetic resources is supported, allowing fine-grain localized adjustment of guidelines and clinical decision-making. Most importantly, anonymized data are available for download, while the Leiden Open Variation Database schema is adopted, enabling integration/interconnection with central resources. CanVaS could become a stepping-stone for a countrywide effort to characterize the cancer genetic variation landscape, concurrently supporting national and international cancer research. The database can be accessed at: <http://ithaka.rrp.demokritos.gr/CanVaS>

DATABASE ARTICLE

Title: *CanVaS: Documenting the genetic variation spectrum of Greek cancer patients*

Authors:

Despoina Kalfakakou^{1,2} (ORCID: 0000-0002-3369-2360), Florentia Fostira² (ORCID: 0000-0003-2751-2332), Athanasios Papathanasiou² (ORCID: 0000-0001-6520-9751), Paraskevi Apostolou², Vasiliki Dellatola², Ioanna E. Gavra² (ORCID: 0000-0002-3803-9168), Ioannis S. Vlachos^{3,4} (ORCID: 0000-0002-8849-808X), Zacharias G. Scouras¹ (ORCID: 0000-0001-7812-3682), Eleni Drosopoulou¹ (ORCID: 0000-0001-9805-0572), Drakoulis Yannoukakos² (ORCID: 0000-0001-7509-3510), Irene Konstantopoulou² (ORCID: 0000-0002-0470-0309)

¹ Department of Genetics, Development & Molecular Biology, School of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece

² Molecular Diagnostics Laboratory, INRaSTES, National Center for Scientific Research “Demokritos”, Athens, Greece

³ Cancer Research Institute, Department of Pathology, Beth Israel Deaconess Medical Center / Harvard Medical School, Boston, MA, USA

⁴ Broad Institute of MIT and Harvard, Cambridge, MA, USA

Corresponding author : Despoina Kalfakakou, Molecular Diagnostics Laboratory, INRaSTES, National Center for Scientific Research “Demokritos”, Patriarchou Gregoriou E’ & Neapoleos Str., Aghia Paraskevi, 15310, Athens, Greece, phone: +302106503811, e-mail: d.kalfakakou@rrp.demokritos.gr, ORCID ID: 0000-0002-3369-2360

Grant sponsor information:

This work was supported by Stavros Niarchos Foundation Industrial Fellowship for PhD candidates.

Abstract:

National genetic variation registries vastly increase the level of detail for the relevant population, while directly affecting patient management. Herein, we report CanVaS, a Cancer Variation reSource aiming to document the genetic variation of cancer patients in Greece. CanVaS comprises germline genetic data from 7,363 Greek individuals with a personal and/or family history of malignancy. The dataset incorporates ~24,000 functionally annotated rare variants in 97 established or suspected cancer susceptibility genes. For each variant, allele frequency for the Greek population, interpretation for clinical significance, anonymized family and segregation information, as well as phenotypic traits of the carriers, are included. Moreover, information on the geographic distribution of the variants across the country are provided, enabling the study of Greek population isolates. Direct comparisons between Greek (sub)populations with relevant genetic resources is supported, allowing fine-grain localized adjustment of guidelines and clinical decision-making. Most importantly, anonymized data are available for download, while the Leiden Open Variation Database schema is adopted, enabling integration/interconnection with central resources. CanVaS could become a stepping-stone for a countrywide effort to characterize the cancer genetic variation landscape, concurrently supporting national and international cancer research. The database can be accessed at: <http://ithaka.rrp.demokritos.gr/CanVaS>

Key Words:

database, hereditary cancer, Greek population, genetic variation, germline variation, founder variants

Introduction:

Recent advances in biotechnology have resulted in an exponential growth of produced genetic data, enabling the genetic characterization of individuals, cohorts, or even populations. However, the fragmentation of sequencing activities and small-scale siloed data resources create a divided genetic data landscape inhibiting open research. To this end, standardized genetic databases have become powerful tools driving progress in the community.

However, very early on, even right after the completion of the 1000 Genomes Project (Genomes Project et al., 2015), it became apparent that numerous rare variants are found exclusively in specific populations. Unfortunately, certain populations are under-represented in large-scale studies, such as the 1000 Genomes Project, Exome Sequencing Project (Fu et al., 2013), Exome Aggregation Consortium (Lek et al., 2016) and Genome Aggregation Consortium (Koch, 2020), as a result of numerous technical or logistics reasons, as well as a recorded preference of studying already well characterized populations and inadequate genomic inclusion of other diverse populations, which are often undersampled (Bentley, Callier, & Rotimi, 2017).

Standardized and harmonized national genetic registries and databases can become powerful tools for data management and access, not only by aggregating and catering invaluable data, but also by increasing the level of detail, context, and information, while facilitating data integration with central genetic resources. Because of the extensive inter- and intra-population genetic data recording, national genetic data registries manage to capture variation between and within populations, while providing context, as well as geographical

or highly specific information to that population, which cannot be captured in a generic resource. Therefore, allele frequencies with high specificity can be calculated while enabling the identification of rare alleles associated with disease risk that can be common in specific populations. Moreover, the determination of the geographic origin of variant carriers facilitates the identification of variant locality, which in turn helps explaining the genetic background and diversity in different (sub)populations and the changes in allele and genotype frequencies over time. To this end, such efforts can serve as the foundation of nation-wide precision medicine, enabling the establishment of population-specific genetic testing protocols, and can potentially lead to cost-effectiveness in surgical and therapeutic treatments.

Another obstacle for progress in the field is the scarcity of publicly available genetic data paired with phenotypic and clinical information. These data are of high importance for a number of applications such as phenotypic heterogeneity determination, genotype-phenotype correlation analysis, variant classification and precision medicine. Most public genetic databases, record either variation data that are not linked to the carrier's phenotype or genotype-phenotype associations, without detailed variation data. OMIM (Hamosh, Scott, Amberger, Bocchini, & McKusick, 2005) contains general information about associations between genes and corresponding phenotypes, without recording genetic variation. dbGAP (Tryka et al., 2014) and NHGRI-EBI GWAS catalogue (Buniello et al., 2019) document variation and phenotype association identified from Genome Wide Association Studies. The central installation of Leiden Open Variation Database (LOVD) (Fokkema et al., 2011) and ClinVar (Landrum et al., 2020) encourage contributors to submit phenotypes of individuals along with variants, but detailed clinical data are not generally recorded. That said, linking genotyping with clinical data can provide significant insights into the genetic variation that contribute to the development of human disease (Deans et al., 2015).

In order to provide a first paradigm tackling the aforementioned open challenges, we established CanVaS, a national database that records germline genetic variation and associated clinical data of Greek cancer patients. With CanVaS, we manage to document the genetic heterogeneity of cancer susceptibility genes in Greeks, a population of unique genetic makeup, exhibiting extended genetic heterogeneity, but also strong founder effects (Apostolou et al., 2020; Apostolou et al., 2017; Konstantopoulou et al., 2014; Pertesi, Konstantopoulou, & Yannoukakis, 2011). Most importantly, all the genetic data are accompanied with detailed phenotypic and clinical traits of tested individuals, whether a pathogenic variant has been identified or not, enabling genotype-phenotype correlation analysis. CanVaS data have been integrated in a LOVD installation to facilitate easy integration and interconnection with other public resources.

Experimental Design, Materials and Methods:

Data were generated by sequencing germline DNA of cancer patients and/or their healthy relatives that have been referred for genetic testing due to personal or family history of malignancy to the reference laboratory of Molecular Diagnostics Laboratory of the National Centre for Scientific Research (NCSR) 'Demokritos', in Athens-Greece, between the years 1999-2019. Written informed consent was obtained from all individuals prior to genetic testing, allowing anonymous use of their data for research purposes. Tested individuals were interrogated on their personal and family history of cancer and geographic origin of their parents. Phenotypic data of tumors were manually extracted from histopathology reports.

Genetic testing was performed on patients' DNA, using either Sanger sequencing, Next Generation Sequencing (NGS)-implemented multigene panel or Multiplex Ligation-dependent Probe Amplification (MLPA) as previously described (Fostira et al., 2020). More specifically, individuals referred for genetic testing between 1999-2012 were tested through Sanger sequencing exclusively. After 2012, genetic testing was performed through gene panel sequencing, unless the individual had a clear clinical diagnosis of a genetic syndrome; in that case targeted sequencing of the candidate gene was performed. Overall, all analyses were complemented with MLPA for -relevant to phenotype- genes, an experimental approach that enables the detection of large genomic rearrangements (deletions/duplications), encompassing one or more exons.

All Variant Call Format (VCF) files were filtered based on variant quality (Genotype Quality > 40, Variant Read Depth > 20x), following selection for rare variants (gnomAD Non Finnish Europeans Allele Frequency

< 0.01). Moreover, known artifacts that are repeatedly identified during experimentation were also filtered. Identified rare variants were classified following the American College of Medical Genetics (ACMG) guidelines for variant classification (Richards et al., 2015).

LOVD open-source database software was used for the data integration, as it enables interconnection with the central LOVD installation. Custom table columns that were created in addition to the already existing columns of LOVD can be found in Table 1.

Results:

Database statistics

CanVaS documents data deriving from germline genetic testing of 1-97 cancer-predisposing genes (Table 2) of 7,363 individuals. Of them, 6,271 (85.17%) are female and 1,092 (14.83%) are male.

Overall, 22,089 rare variants in genes that are known or suspected to predispose to cancer have been identified and imported to the resource, corresponding to 7,968 unique variants. Of them, 2,204 are reported as Pathogenic/Likely pathogenic variants (PV/LPV; 9.98%), 10,696 are reported as variants of unknown significance (VUS; 48.42%), and 9,189 variants are reported as Benign/Likely benign variants (BV/LBV; 41.60%) (Figure 1). The rest of the variants are found in intergenic regions that have been identified in Genome Wide Association Studies and are not studied further herein. Of all the PV/LPVs, 1,817 (82.5%) are positioned in highly penetrant genes (Houweling et al., 2011; Janoueix-Lerosey et al., 2008; Menko et al., 2014; Mosse et al., 2008; Puntervoll et al., 2013; Salpea & Stratakis, 2014; Yang et al., 2016), 266 (12.0%) in moderately penetrant genes (Daly et al., 2020; Provenzale et al., 2020; Stewart et al., 2019), while 121 (5.5%) are found in genes with low or uncertain penetrance (Daly et al., 2020; Li et al., 2020; Provenzale et al., 2020). Most importantly, 1,643 (74.6%) PV/LPVs concern clinically actionable genes that are included in ACMG's list of 59 clinically important genes (Kalia et al., 2017) (Table 2).

A total of 7,061 different disease-phenotypes in 6,001 (81.5%) individuals are reported in CanVaS, while 1,362 individuals (18.5%) are asymptomatic blood relatives of the aforementioned patients. The most prevalent phenotype is breast cancer (BrCa), accounting for 3,970 individuals (53.92%; mean age \pm SD: 46.82 \pm 12.05), of which 103 (1.04% of all individuals) is male BrCa (mean age \pm SD: 61.95 \pm 13.4). Moreover, 1,017 (13.81%) patients have been diagnosed with ovarian cancer (OvCa; mean age \pm SD: 53.08 \pm 13.04), while 510 (6.93%) have a history of colon polyps, polyposis and/or colorectal cancer (CRC; mean age \pm SD of CRC: 50.6 \pm 13.85). Additionally, 158 (2.15%), 131 (1.78%) and 102 (1.39%) individuals have been diagnosed with pancreatic (mean age \pm SD: 60.14 \pm 12.19), thyroid (mean age \pm SD: 46.42 \pm 15.01) and endometrial cancer (mean age \pm SD: 52.08 \pm 12.75), respectively. Although definitely more rare, additional diagnoses involve cancerous manifestations, such as brain or diffuse gastric cancer (a total of 857 or 11.64% of patients with other cancer phenotypes) and non-cancerous diagnoses consistent with cancer syndromes, such as Tuberous Sclerosis and Neurofibromatosis (317 or 4.3% of patients with non-cancerous phenotypes) (Figure 2). A total of 709 (9.63%) individuals in CanVaS are presented with more than one phenotypes.

Among the individuals that had at least one cancer diagnosis, 91 (1.52%) were children or adolescents (0-18 years old) at the time of diagnosis, 360 (6%) individuals were young adults (19-30 years old), 2,158 (35.97%) were diagnosed between ages 31-45, 1,846 (30.77%) were diagnosed between ages 46-60, while 1,018 (16.97%) individuals were diagnosed after the age of 60. For 527 (8.78%) individuals their age at first diagnosis was not available.

Provided information

Any person familiar with LOVD can easily navigate through CanVaS, as the database format is standardized and commonly adopted in genetic settings. That said, in addition to the default information, the tables of the database have been enriched with custom columns (Table 1) so that supplementary information can be provided, consistent with a national variation registry, in order to enable recording of the phenotypes in great detail.

Each “Individual” entry is accompanied with demographic information, i.e. gender and age. Importantly, the particular geographic region/province of the tested individual’s descent is provided. When applicable, this information comes in great detail, i.e. specific Greek village/city of the individual’s origins, enabling the documentation of variant locality and population isolates. In addition to demographic information, each “Individual” entry is supplemented with a link to the published article(s) including corresponding data, where available. A summary of the individual’s phenotype(s), screening(s) and identified variant(s) is provided; in this case the user can get into details by following the relevant links on CanVaS’s interface (Figure 3).

Each entry for an identified variant includes information on its genomic position, description on all transcripts according to Human Genome Variation Society (HGVS) nomenclature (den Dunnen et al., 2016), variant segregation within families and origin (paternal or maternal). Moreover, each unique variant was individually assessed and classified accordingly, following ACMG guidelines. All the aforementioned information is accompanied by custom data, that are crucial to a population database. Thus, for each “Variant” entry, its founder status and locality is provided, where available, enabling the calculation of its average frequency in the Greek population. Apart from the HGVS nomenclature, other “popular” names of the variant are documented for easiest identification, a feature particularly useful for large genomic rearrangements, which are often described only on the genomic level. Nevertheless, information on the absence of a variant is as strong as information on its presence for genetic studies. Thereby, the individual’s “Screening” entry documents not only the type of genetic testing performed, but whether or not any variants were identified as well (Figure 4).

Specific attention was paid on the documentation of the phenotypic data, as they can serve as a powerful tool for each analysis. For this reason, emphasis was placed on describing the data in a structured and detailed way, minimizing free text entries. Therefore, each “Phenotype” entry is coupled with important clinical data, such as the patient’s tumor morphology, grade, stage, hormone receptor status and lymph node status. Alongside, the medical history on any associated phenotype of the individual’s family is provided (Figure 5). Based on the sensitive nature of this information, phenotypic data are controlled-accessed and are available only upon request through the database portal.

Discussion:

Herein, we report CanVaS, a Cancer Variation reSource aiming to document the genetic variation of cancer patients in Greece. The dataset comprises the overall subset of all identified rare germline variants in cancer patients, along with their classification and allele frequency in Greek population, rendering CanVaS a valuable resource for cancer susceptibility genetic variants. The collection of ~2,200 and ~9,200 pathogenic/likely pathogenic and benign/likely benign variants, respectively, can provide useful information to researchers involved in variant interpretation and classification.

CanVaS includes a large number of variants with unknown clinical significance (VUS), for which additional information is needed to enable classification. Importantly the highly granular annotation and the targeted population (cancer patients and their relatives) can be used to support VUS prioritization and gene-specific, functional investigations.

CanVaS provides the variant’s geographic distribution in Greece, while recording Greek founder variants and their locality. Overall, this approach empowers subpopulation analysis, study of Greek population isolates, and investigation of possible founder effects. The strong founder effects observed in Greek subpopulations has been recently highlighted in a genetic analysis performed in Cretan isolates, where it was shown that three pathogenic variants in *BRCA1* and *BRCA2*, unique to the Cretan subpopulation, namely NM_007294.3:c.5492del in *BRCA1* and NM_000059.3:c.7806-2A>T and NM_000059.3:c.6842-2675_7008-5558del, both in *BRCA2*, account for 48% of identified PVs in *BRCA1* and *BRCA2* among breast and ovarian cancer patients of Cretan descent (Apostolou et al., 2020).

Moreover, the in-depth study of the Greek population and the evaluation of founder effects have enabled the determination of the pathogenicity of rare variants, that are confined in Greeks, as in the case of *BRCA1* variant NM_007294.3:c.5212G>A (p.Gly1738Arg). This is a founder pathogenic variant for the Greek popu-

lation, which has been found in 53 families to date. Its pathogenicity was hard to prove, as it is a missense substitution, but its recurrence in Greek breast/ovarian cancer patients made its classification possible, enabling the proper management of patients not only in Greece, but in Greek diaspora as well (Anagnostopoulos et al., 2008).

Notably, data recorded in CanVaS are enriched with phenotypic information, a type of data that is not shared by most genetic repositories. This is a significant addition with high utility to the scientific community, enabling clinical applications through appropriate variant classification. Furthermore, the data provided include patient primary diagnosis, detailed histopathology, phenotypic traits that are consistent with cancer syndromes and information on family history, while this information is provided in a standardized and structured framework, permitting efficient and sophisticated data analyses. The depth and variety of these data enables them to support a great number of downstream applications, such as phenotypic heterogeneity determination, phenotype-genotype correlation analyses, precision medicine and variant pathogenicity interrogations.

A prominent example of a variant classification aided by the available clinical data is *BRCA1* variant NM.007294.3:c.245T>C (p.Leu82Pro). This variant is a rare VUS found in a young woman who has been diagnosed two times with a primary triple negative breast tumor (Supplementary Figure 1) – a phenotype consistent with the presence of a pathogenic variant in *BRCA1*. This information, paired with *in silico* prediction tools that classify the variant as “probably damaging” and existing functional data showing a severe effect of this change in *BRCA1* gene (Findlay et al., 2018), lead to a “Likely Pathogenic” classification of this variant.

Many ambitious efforts for the construction of National Mutation Databases have been made in the past, focusing on clinically relevant variants that have been found in a number of genetic diseases (Charoute et al., 2014; Fakhro et al., 2016; Kleanthous et al., 2006; Patrinos, van Baal, Petersen, & Papadakis, 2005; Pradhan et al., 2011; Ruangrit et al., 2008; Sipila & Aula, 2002; van Baal et al., 2007; Zlotogora, van Baal, & Patrinos, 2007). However, the majority have unfortunately become obsolete with time, a fact that highlights the difficulty of maintaining and updating such resources. The need for national variation databases is crucial for the study of genetic heterogeneity of different populations. Databases that specifically focus on specific disease/gene(s) are easier to manage, through which larger genetic databases can be supported, enabling collaboration and data exchange. As a part of the Brazilian Initiative for Precision Medicine (Rocha, Secolin, Rodrigues, Carvalho, & Lopes-Cendes, 2020), the BRCA database, documenting 180 variants in *BRCA1* and *BRCA2* genes that have been found in the Brazilian population is a first step towards this direction.

To our knowledge, CanVaS is the largest population database for hereditary cancer and genetic susceptibility to cancer, documenting variants in 97 genes coupled with clinical data. Through detailed documentation of paired genotypic and phenotypic data, CanVaS can become a valuable resource for professionals involved in cancer genetics. Our long-term ambition is to organize a coordinated effort along with other public and private diagnostic laboratories in order to centralize genetic and phenotypic data that contribute to genetic cancer susceptibility among Greeks.

Authors’ contributions:

IK, DY, FF and DK designed the study. FF, IK, DY, ISV, PA, DK and VD carried out the experiments that generated the genotypic data. IK, FF, DY, PA and VD collected the clinical and phenotypic data. DK, AP and IEG curated the clinical and phenotypic data. DK, AP, PA and VD curated the classification of variants. DK, IK, and FF validated the curation of the data. DK generated the figures. DK wrote the manuscript that was reviewed and edited by all authors.

Acknowledgements:

The authors are grateful to all the patients and their families that participated in this study. They would also like to thank all Molecular Diagnostic Laboratory alumni that contributed to the generation of these data and Ms Athanasia Kalfakakou for the design of CanVaS’s logo.

CanVaS was supported by Stavros Niarchos Foundation Industrial Fellowship for PhD candidates.

Data Availability Statement:

The data that support the findings of this study are openly available in CanVaS at <http://ithaka.rp.demokritos.gr/CanVaS>

Conflict of Interest:

The authors declare that they have no conflict of interest.

References:

- Anagnostopoulos, T., Pertesi, M., Konstantopoulou, I., Armaou, S., Kamakari, S., Nasioulas, G., . . . Yannoukakis, D. (2008). G1738R is a BRCA1 founder mutation in Greek breast/ovarian cancer patients: evaluation of its pathogenicity and inferences on its genealogical history. *Breast Cancer Res Treat*, *110* (2), 377-385. doi:10.1007/s10549-007-9729-y
- Apostolou, P., Fostira, F., Kouroussis, C., Kalfakakou, D., Delimitsou, A., Agelaki, S., . . . Saloustros, E. (2020). BRCA1 and BRCA2 germline testing in Cretan isolates reveals novel and strong founder effects. *Int J Cancer*, *147* (5), 1334-1342. doi:10.1002/ijc.32903
- Apostolou, P., Pertesi, M., Aleporou-Marinou, V., Dimitrakakis, C., Papadimitriou, C., Razis, E., . . . Fostira, F. (2017). Haplotype analysis reveals that the recurrent BRCA1 deletion of exons 23 and 24 is a Greek founder mutation. *Clin Genet*, *91* (3), 482-487. doi:10.1111/cge.12824
- Bentley, A. R., Callier, S., & Rotimi, C. N. (2017). Diversity and inclusion in genomic research: why the uneven progress? *J Community Genet*, *8* (4), 255-266. doi:10.1007/s12687-017-0316-6
- Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., . . . Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*, *47* (D1), D1005-D1012. doi:10.1093/nar/gky1120
- Charoute, H., Nahili, H., Abidi, O., Gabi, K., Rouba, H., Fakiri, M., & Barakat, A. (2014). The Moroccan Genetic Disease Database (MGDD): a database for DNA variations related to inherited disorders and disease susceptibility. *Eur J Hum Genet*, *22* (3), 322-326. doi:10.1038/ejhg.2013.151
- Daly, M. B., Pilarski, R., Yurgelun, M. B., Berry, M. P., Buys, S. S., Dickson, P., . . . Darlow, S. D. (2020). NCCN Guidelines Insights: Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic, Version 1.2020. *J Natl Compr Canc Netw*, *18* (4), 380-391. doi:10.6004/jnccn.2020.0017
- Deans, A. R., Lewis, S. E., Huala, E., Anzaldo, S. S., Ashburner, M., Balhoff, J. P., . . . Mabee, P. (2015). Finding our way through phenotypes. *PLoS Biol*, *13* (1), e1002033. doi:10.1371/journal.pbio.1002033
- den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., McGowan-Jordan, J., . . . Taschner, P. E. (2016). HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat*, *37* (6), 564-569. doi:10.1002/humu.22981
- Fakhro, K. A., Staudt, M. R., Ramstetter, M. D., Robay, A., Malek, J. A., Badii, R., . . . Rodriguez-Flores, J. L. (2016). The Qatar genome: a population-specific tool for precision medicine in the Middle East. *Hum Genome Var*, *3*, 16016. doi:10.1038/hgv.2016.16
- Findlay, G. M., Daza, R. M., Martin, B., Zhang, M. D., Leith, A. P., Gasperini, M., . . . Shendure, J. (2018). Accurate classification of BRCA1 variants with saturation genome editing. *Nature*, *562* (7726), 217-222. doi:10.1038/s41586-018-0461-z
- Fokkema, I. F., Taschner, P. E., Schaafsma, G. C., Celli, J., Laros, J. F., & den Dunnen, J. T. (2011). LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat*, *32* (5), 557-563. doi:10.1002/humu.21438

- Fostira, F., Kostantopoulou, I., Apostolou, P., Papamentzelopoulou, M. S., Papadimitriou, C., Faliakou, E., . . . Yannoukakos, D. (2020). One in three highly selected Greek patients with breast cancer carries a loss-of-function variant in a cancer susceptibility gene. *J Med Genet*, *57* (1), 53-61. doi:10.1136/jmedgenet-2019-106189
- Fu, W., O'Connor, T. D., Jun, G., Kang, H. M., Abecasis, G., Leal, S. M., . . . Akey, J. M. (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature*, *493* (7431), 216-220. doi:10.1038/nature11690
- Genomes Project, C., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., . . . Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526* (7571), 68-74. doi:10.1038/nature15393
- Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A., & McKusick, V. A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res*, *33* (Database issue), D514-517. doi:10.1093/nar/gki033
- Houweling, A. C., Gijzen, L. M., Jonker, M. A., van Doorn, M. B., Oldenburg, R. A., van Spaendonck-Zwarts, K. Y., . . . Menko, F. H. (2011). Renal cancer and pneumothorax risk in Birt-Hogg-Dube syndrome; an analysis of 115 FLCN mutation carriers from 35 BHD families. *Br J Cancer*, *105* (12), 1912-1919. doi:10.1038/bjc.2011.463
- Janoueix-Lerosey, I., Lequin, D., Brugieres, L., Ribeiro, A., de Pontual, L., Combaret, V., . . . Delattre, O. (2008). Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. *Nature*, *455* (7215), 967-970. doi:10.1038/nature07398
- Kalia, S. S., Adelman, K., Bale, S. J., Chung, W. K., Eng, C., Evans, J. P., . . . Miller, D. T. (2017). Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med*, *19* (2), 249-255. doi:10.1038/gim.2016.190
- Kleanthous, M., Patsalis, P. C., Drousiotou, A., Motazacker, M., Christodoulou, K., Cariolou, M., . . . Patrinos, G. P. (2006). The cypriot and Iranian National Mutation Frequency Databases. *Hum Mutat*, *27* (6), 598-599. doi:10.1002/humu.9422
- Koch, L. (2020). Exploring human genomic diversity with gnomAD. *Nat Rev Genet*, *21* (8), 448. doi:10.1038/s41576-020-0255-7
- Konstantopoulou, I., Tsitlaidou, M., Fostira, F., Pertesi, M., Stavropoulou, A. V., Triantafyllidou, O., . . . Yannoukakos, D. (2014). High prevalence of BRCA1 founder mutations in Greek breast/ovarian families. *Clin Genet*, *85* (1), 36-42. doi:10.1111/cge.12274
- Landrum, M. J., Chitipiralla, S., Brown, G. R., Chen, C., Gu, B., Hart, J., . . . Kattman, B. L. (2020). ClinVar: improvements to accessing data. *Nucleic Acids Res*, *48* (D1), D835-D844. doi:10.1093/nar/gkz972
- Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., . . . Exome Aggregation, C. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, *536* (7616), 285-291. doi:10.1038/nature19057
- Li, Y., Zhang, J., Adikaram, P. R., Welch, J., Guan, B., Weinstein, L. S., . . . Simonds, W. F. (2020). Genotype of CDC73 germline mutation determines risk of parathyroid cancer. *Endocr Relat Cancer*, *27* (9), 483-494. doi:10.1530/ERC-20-0149
- Menko, F. H., Maher, E. R., Schmidt, L. S., Middelton, L. A., Aittomaki, K., Tomlinson, I., . . . Linehan, W. M. (2014). Hereditary leiomyomatosis and renal cell cancer (HLRCC): renal cancer risk, surveillance and treatment. *Fam Cancer*, *13* (4), 637-644. doi:10.1007/s10689-014-9735-2
- Mosse, Y. P., Laudenslager, M., Longo, L., Cole, K. A., Wood, A., Attiyeh, E. F., . . . Maris, J. M. (2008). Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature*, *455* (7215), 930-935.

doi:10.1038/nature07261

- Patrinos, G. P., van Baal, S., Petersen, M. B., & Papadakis, M. N. (2005). Hellenic National Mutation database: a prototype database for mutations leading to inherited disorders in the Hellenic population. *Hum Mutat*, 25 (4), 327-333. doi:10.1002/humu.20157
- Pertesi, M., Konstantopoulou, I., & Yannoukakos, D. (2011). Haplotype analysis of two recurrent genomic rearrangements in the BRCA1 gene suggests they are founder mutations for the Greek population. *Clin Genet*, 80 (4), 375-382. doi:10.1111/j.1399-0004.2010.01532.x
- Pradhan, S., Sengupta, M., Dutta, A., Bhattacharyya, K., Bag, S. K., Dutta, C., & Ray, K. (2011). Indian genetic disease database. *Nucleic Acids Res*, 39 (Database issue), D933-938. doi:10.1093/nar/gkq1025
- Provenzale, D., Ness, R. M., Llor, X., Weiss, J. M., Abbadessa, B., Cooper, G., . . . Ogba, N. (2020). NCCN Guidelines Insights: Colorectal Cancer Screening, Version 2.2020. *J Natl Compr Canc Netw*, 18 (10), 1312-1320. doi:10.6004/jnccn.2020.0048
- Puntervoll, H. E., Yang, X. R., Vetti, H. H., Bachmann, I. M., Avril, M. F., Benfodda, M., . . . Molven, A. (2013). Melanoma prone families with CDK4 germline mutation: phenotypic profile and associations with MC1R variants. *J Med Genet*, 50 (4), 264-270. doi:10.1136/jmedgenet-2012-101455
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Committee, A. L. Q. A. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, 17 (5), 405-424. doi:10.1038/gim.2015.30
- Rocha, C. S., Secolin, R., Rodrigues, M. R., Carvalho, B. S., & Lopes-Cendes, I. (2020). The Brazilian Initiative on Precision Medicine (BIPMed): fostering genomic data-sharing of underrepresented populations. *NPJ Genom Med*, 5 , 42. doi:10.1038/s41525-020-00149-6
- Ruangrit, U., Srikumool, M., Assawamakin, A., Ngamphiw, C., Chuechote, S., Thaiprasarnsup, V., . . . Tongsimma, S. (2008). Thailand mutation and variation database (ThaiMUT). *Hum Mutat*, 29 (8), E68-75. doi:10.1002/humu.20787
- Salpea, P., & Stratakis, C. A. (2014). Carney complex and McCune Albright syndrome: an overview of clinical manifestations and human molecular genetics. *Mol Cell Endocrinol*, 386 (1-2), 85-91. doi:10.1016/j.mce.2013.08.022
- Sipila, K., & Aula, P. (2002). Database for the mutations of the Finnish disease heritage. *Hum Mutat*, 19 (1), 16-22. doi:10.1002/humu.10019
- Stewart, D. R., Best, A. F., Williams, G. M., Harney, L. A., Carr, A. G., Harris, A. K., . . . Schultz, K. A. P. (2019). Neoplasm Risk Among Individuals With a Pathogenic Germline Variant in DICER1. *J Clin Oncol*, 37 (8), 668-676. doi:10.1200/JCO.2018.78.4678
- Tryka, K. A., Hao, L., Sturcke, A., Jin, Y., Wang, Z. Y., Ziyabari, L., . . . Feolo, M. (2014). NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic Acids Res*, 42 (Database issue), D975-979. doi:10.1093/nar/gkt1211
- van Baal, S., Kaimakis, P., Phommarinh, M., Koumbi, D., Cuppens, H., Riccardino, F., . . . Patrinos, G. P. (2007). FINDbase: a relational database recording frequencies of genetic defects leading to inherited disorders worldwide. *Nucleic Acids Res*, 35 (Database issue), D690-695. doi:10.1093/nar/gkl934
- Yang, S. Y., Hsiung, C. N., Li, Y. J., Chang, G. C., Tsai, Y. H., Chen, K. Y., . . . Hsu, H. M. (2016). Fanconi anemia genes in lung adenocarcinoma- a pathway-wide study on cancer susceptibility. *J Biomed Sci*, 23 , 23. doi:10.1186/s12929-016-0240-9
- Zlotogora, J., van Baal, S., & Patrinos, G. P. (2007). Documentation of inherited disorders and mutation frequencies in the different religious communities in Israel in the Israeli National Genetic Database. *Hum*

Mutat, 28 (10), 944-949. doi:10.1002/humu.20551

Figure Legends:

Figure 1: Distribution of unique variant and classification entries in each gene reported in CanVaS.

Figure 2: CanVaS phenotypic entries. Values are plotted in log₂ scale.

Figure 3: Snapshot depicting CanVaS interface of an “Individual” entry.

Figure 4: Snapshot depicting CanVaS interface of a “Variant” entry.

Figure 5: Snapshot depicting CanVaS interface of a “Phenotype” entry.

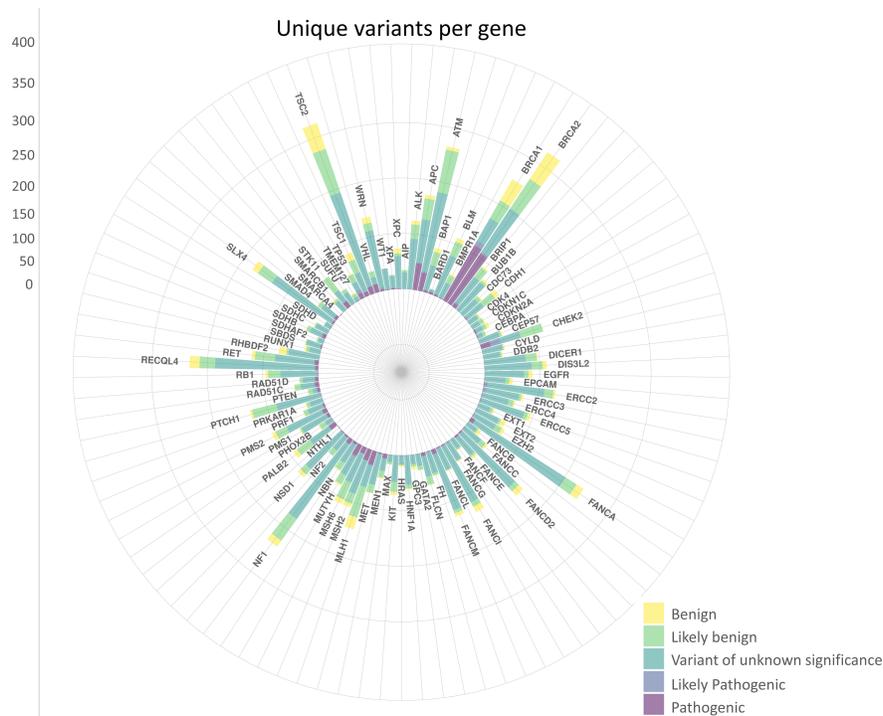
Supplementary Figure 1: Phenotypic traits of a woman carrying the *BRCA1* variant NM_007294.3:c.245T>C (p.Leu82Pro).

Hosted file

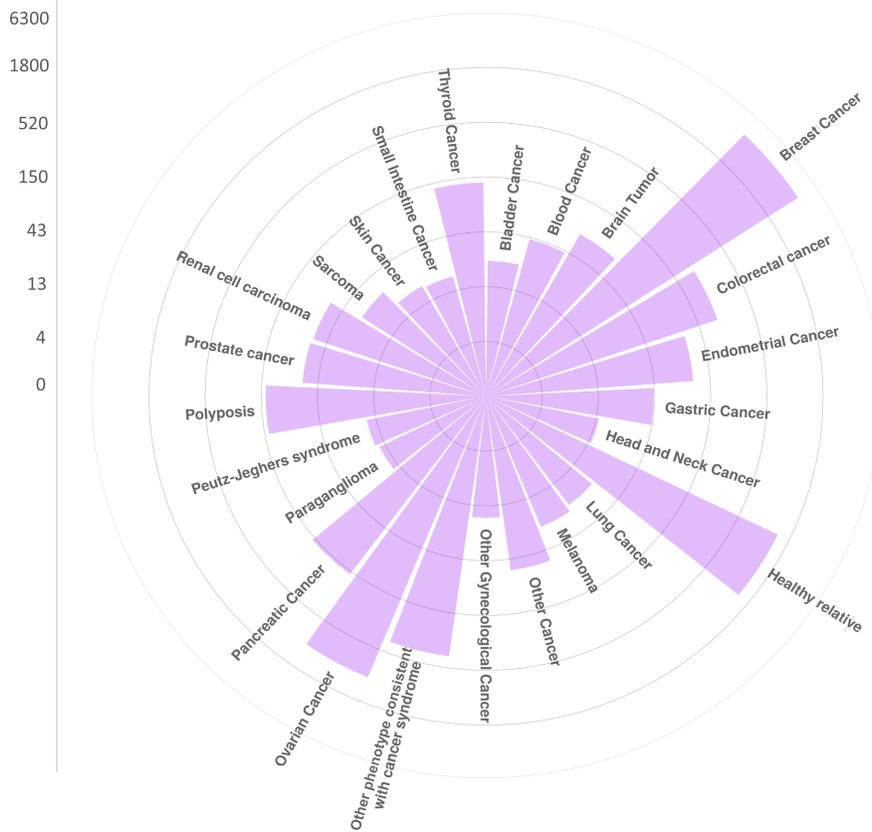
Table1.pdf available at <https://authorea.com/users/404113/articles/515407-canvas-documenting-the-genetic-variation-spectrum-of-greek-cancer-patients>

Hosted file

Table2.pdf available at <https://authorea.com/users/404113/articles/515407-canvas-documenting-the-genetic-variation-spectrum-of-greek-cancer-patients>



Recorded Phenotypes



CanVaS - A Greek Cancer Patient Genetic Variation Resource LOVD v3.0 Build 23 | Current LOVD status | Register as submitter | Log in

Genes | Transcripts | Variants | Individuals | Diseases | Screenings | Submit | Documentation

Individual #00014743

Reference: [Pedra et al. \(2020\)](#) Link to published data

Gender: Female

Geographic origin: Chalkidiki Geographic origin of both parents enables subpopulation analysis and investigation of possible founder effects

Population: Thessaloniki

Consanguinity: Greek

Remarks: no

Panel size: 1

Diseases: BRCA

Owner name: Despoina Kafakakou

Phenotypes Detailed clinical data are not publicly available, but any user has access to the individual's disease and age at onset

Breast Cancer (BRCA)

Phenotype ID	Age of onset	Phenotype details	Inheritance	BRCA/FH	Subtype	Owner
000008424	57y					Despoina Kafakakou

Screenings List of genetic screenings performed

Screening ID	Template	Technique	Type	Genes screened	Variants found	Owner
000001216	DNA	HLPA		1 BRCA1		Despoina Kafakakou

Variants List of identified variants

1 entry on 1 page. Showing entry 1.

[100 per page] Legend

Chr	Allele	Effect	DNA change (genomic) (hg19)	Reference	DB-ID	dbSNP ID	Also Known As	Geographic Origin
17	Maternal (inferred)	+/+	g-41193676_41198104delinsCTGTG	Pedra et al. 2020 Koutroumpis et al. 2013 Lacortou et al. 2009	BRCA1_000102		BRCA1 exon 24 deletion	Euxinus Pontus

[100 per page] Legend

CanVaS - A Greek Cancer Patient Genetic Variation Resource
BRCA1 (BRCA1 DNA repair associated)

Curator: Despoina Kalfakakou

Genes Transcripts Variants Individuals Diseases Screenings Submit Documentation

Genomic variant #0000025246

Individual ID	00014743
Chromosome	17
Allele	Maternal (inferred)
Affects function (as reported)	Affects function
Affects function (by curator)	Affects function
DNA change (genomic) (Relative to hg19 / GRCh37)	g.41193676_41198104delinsCTGTG
Reference	[Pectea et al. 2011] [Constantopoulos et al. 2013] [Armakou et al. 2009]
DB-ID	BRCA1_000102
dBSNP ID	-
Variant remarks	-
Genetic origin	Germline
Segregation	Yes
Also Known As	BRCA1 exon 24 deletion
Is Greek Founder	Yes
Geographic Origin (for founder variants)	Euxinus Pontus
Average frequency (large NGS studies)	Variant not found in online data sets
Average frequency in greek population	0.00152
Owner	Despoina Kalfakakou

Options

Impact on clinical significance as it was evaluated by Molecular Diagnostics Lab, following ACMG guidelines

Segregation in family

Documentation of Greek founder variants and their locality

Greek Allele Frequency calculated from CanVaS data

Variant on transcripts

Gene	Transcript	Affects function	Exon	DNA change (cDNA)	RNA change	Protein
BRCA1	NM_007294.3	+/+	-	c.5468-285_*4019delinsCACAG	r.0?	p.0?
BRCA1	NM_007297.3	+/+	-	c.5327-285_*4019delinsCACAG	r.0?	p.0?
BRCA1	NM_007298.3	+/+	-	c.2156-285_*4019delinsCACAG	r.0?	p.0?
BRCA1	NM_007299.3	+/+	-	c.2082-285_*4125delinsCACAG	r.0?	p.0?
BRCA1	NM_007300.3	+/+	-	c.5531-285_*4019delinsCACAG	r.0?	p.0?

HGVS nomenclature for all transcripts

Screenings Associated screening entry

Screening ID	Template	Technique	Type	Genes screened	Variants found	Owner
0000001216	DNA	MLPA		1 BRCA1	1	Despoina Kalfakakou

CanVaS - A Greek Cancer Patient Genetic Variation Resource

LOVD v.3.0 Build 23 | Current LOVD status | Welcome, Despoina Kalfakakou | Your account | Unfinished submissions | Log out

Genes Transcripts Variants Individuals Diseases Screenings Submit Users Configuration Setup Documentation

Phenotype #000008472

Individual ID	00016110
Associated disease	BRCA
Age of onset	47y
Phenotype details	-
Inheritance	-
Histology	Ductal
Behaviour	Invasive
Grade	III
Stage	III
ER	Negative
PR	Negative
HER2	Negative
BRCA/FH	Strong
Subtype	-
Subcategory	-
Owner name	Despoina Kalfakakou
Phenotype data status	Public
Created by	Despoina Kalfakakou
Date created	2020-05-26 11:51:40 +03:00 (EEST)
Last edited by	Despoina Kalfakakou
Date last edited	2020-05-26 11:51:40 +03:00 (EEST)

Options

Detailed phenotypic information

Information on Family History

Because of the sensitive nature of this information, phenotypic data are controlled accessed only.

Powered by LOVD v.3.0 Build 23
LOVD software ©2004-2020 Leiden University Medical Center

CanVaS - A Greek Cancer Patient Genetic Variation Resource

LOVD v.3.0 Build 23 | Current LOVD status | Welcome, Despoina Kalfakakou | Your account | Unfinished submissions | Log out

Genes Transcripts Variants Individuals Diseases Screenings Submit Users Configuration Setup Documentation

Individual #0014990

Individual ID	00014990
Reference	[Constantopoulos et al. 2013]
Gender	F
Geographic origin	Kalamata
Geographic origin	Greece
Population	no
Consanguinity	-
Remarks	-
Parent IDs	2
Individual	BRCA
Owner name	Despoina Kalfakakou
Individual data status	Public
Created by	Despoina Kalfakakou
Date created	2020-05-26 11:51:40 +03:00 (EEST)
Last edited by	N/A
Date last edited	N/A

Options

Phenotypes

Breast Cancer (Breast) Add phenotype for this disease

Phenotype ID	Age of onset	Histology	Behaviour	Grade	ER	PR	HER2	BRCA/FH
000001902 32y	N/A	Ductal	Invasive	III	Negative	Negative	Negative	Weak
000001904 32y	N/A	Ductal	Invasive	III	Negative	Negative	Negative	Weak

Screenings

Screening ID	Template	Technique	Type	Genes screened	Variants found
000005483	DNA	SEQ-HQ-1		1	1

Variants

1 entry on 1 page. Showing entry 1. [1/0 per page]

Chr	Allele	Effect	DNA change (genomic) (hg19)	Reference	DB-ID	Genetic origin	Gene	DNA change (cDNA)	Protein
17	Parent #2	+/+	g.4126641A>G	[Constantopoulos et al. 2013]	BRCA1_000201	Germline	BRCA1	NM_007294.3:c.2487>C NM_007297.3:c.1057>C NM_007298.3:c.2487>C NM_007299.3:c.2487>C NM_007300.3:c.2487>C	p.L428Pfs*1 p.L428Pfs*1 p.L428Pfs*1 p.L428Pfs*1