**Analysis of germ line variants in cancer-related genes of black female South African breast cancer patients**

Eygelaar, D.1,2,3, Jansen van Rensburg, E.1,3 and Joubert, F.1,2,3

1Department of Genetics, Biochemistry and Microbiology, University of Pretoria, Pretoria 0001, South Africa.

2Centre for Bioinformatics and Computational Biology, University of Pretoria, Pretoria 0001, South Africa.

3Genomics Research Institute, University of Pretoria, Pretoria 0001, South Africa.

**Abstract**

TBD.

**Introduction**

Breast cancer has been reported numerous times as the most prominent cancer in women 1,2 and is 100 times more likely to be identified in females compared to males 3,4. Globally, breast cancer has been ranked as having the second highest incidence when compared to other cancers 1. Until recently, breast cancer was rated the second most common cancer in North America 5 but new estimates suggest that breast cancer will overtake lung cancer as the most prominent cancer 6. In South Africa, previous studies estimated a lifetime breast cancer risk for women of 1 in 28, with 0.7% of all deaths accounted for by breast cancer 7.

Regarding early-onset breast cancer (under the age of 40), research has shown that the African American population has a higher disproportionate burden of aggressive early-onset breast cancer compared to other ethnical groups 5,8,9 and that 10-20% of these patients carried a BRCA1/BRCA2 mutation 10. While many studies on breast cancer susceptibility have focussed on variants in these high penetrance genes, this report additionally focusses on the genes related to breast cancer susceptibility and include TP53 (high penetrance), ATM, CHEK2, BRIP1, PALB2, RAD50, NBN, RB1 (medium penetrance) and PTEN, RAD51C, BARD1, STK11, CDH1 (low penetrance) 11. It has become clear that distinct population groups may be influenced by differing variants 12. The major available resources represent data primarily from European populations, together with some Asian populations. Some studies have included African American populations, but this data can be difficult to interpret in an African context due to extensively varying degrees of admixture in the groups involved. In general, cancer-related variant information is very scarce for African populations 13. While there are some papers available on founder populations in white Afrikaners in South Africa 14, most previous South African studies on black female cohorts have focussed primarily on BRCA1 and BRCA2 variants, with some studies on other single genes 15-17. Recently, other researchers have also embarked on studying unique variants in populations with under-represented genomic data 18-23.

The availability of whole genome, whole exome and directed panel sequencing is rapidly enhancing the possibility of performing population-specific studies for diseases with a strong genetic basis, such as cancer. While whole genome is still relatively expensive, whole exome and directed panel sequencing are becoming more affordable while directed panel sequencing is particularly attractive for diseases where commercial predesigned panels are available for the disease of interest. Currently, several companies are producing cancer sequencing panels available for both germ line and somatic studies.

For this study on breast cancer susceptibility genes in black South African females, the Illumina TruSight Cancel Panel was used, which is a capture-based panel targeting 94 cancer-related genes and 284 SNPs that were previously associated with a predisposition towards cancer. A total of 166 black African females were analysed, with ages ranging between 18 and 54 years.

**Materials and Methods**

*Sample collection and DNA extraction*

A total of 166 blood samples were previously collected from black South African females with breast cancer, who attended the Oncology Clinic at Steve Biko Hospital. These samples were collected with the patients consent, with ethics approvals KCT 265 and 260/2018. DNA was previously extracted from peripheral blood samples using the method described by Johns and Paulus-Thomas 24.

Samples were tested using conventional methods [???] for the presence of BRCA mutations, all the samples in this set tested negative for BRCA mutations. They were also tested for ….

Sequencing

DNA samples were sent to Omega Biotech in Georgia, USA, for sequencing using the Illumina TruSight Cancer Panel. The panel contains 94 cancer-related genes and 284 SNPs previously identified to be associated with a predisposition to cancer.

*Pre-processing*

Quality analysis was performed using FastQC 25. TheFastX toolkit was used to trim 5 and 95 nucleotides on the 5’ and 3’ ends of the 100bp paired-end reads respectively 26.

*Germ line variant calling*

Samples were analysed using the GATK best practices 27 approach by means of the BBCBIO pipeline 28. This includes mapping against h19 with BWA MEM 29, marking duplicates with Picard, base quality score recalibration, variant calling using the HaplotypeCaller in gVCF mode and specified cutoff-based filtering of variants with VariantFiltration using the BCBIO default filtering cut-offs.

*Variant annotation*

Functional variant annotation was done using Variant Effect Predictor (VEP), the default parameters were used in concordance with documentation 30. The filtered variants were uploaded to the VEP web interface, and additional output fields were activated as needed. Filtering of common variants was not performed in VEP.

*Variant selection*

In-house Python code was developed for the selection of variants with predicted functional effects. Variants related to splice site variants, missense variants, inframe deletions, inframe insertions, frameshift deletions and stop gained variants are included, other non-coding variants are not reported in detail here. Before selection, variants with an occurrence of ≥ 1% in the ExAcAfr database were removed 31.

For this study, the results of five functional effect predictors were considered, being LRT\_pred 32, MutationTaster 33, Provean 34, CADD 35 and FATHMM 36 with variants being selected if at least 4/6 methods predicted a variant to be deleterious. A threshold of 2.0 for GERP\_RS and 10.0 for CADD was used. For the other methods, a prediction of ‘D’ was selected.

As VEP provides results for multiple transcripts per gene, canonical transcripts are reported on here, as determined by mapping of REFSEQ identifiers to Ensembl canonical transcripts via UCSC tables 37. The mapping of genes to Reactome pathways was done using MSigDB tables 38. In the case of BRCA1, transcript 2 naming is also shown for historical reasons.

**Results**

*Variant analysis*

The study was performed using germ line DNA from 166 black female South African breast cancer patients, varying in age between 18 and 54 years (Supplementary Table 2). Sequencing was performed for 94 cancer-related genes and 284 SNPs that had previously been shown to be associated with an increased susceptibility to cancer. Variants of interest were selected when at least 3/5 methods predicted a deleterious effect.

Initially, a total of 1,560 variants were identified following GATK specified cutoff-based filtration of variants (Supplementary Table 1). VEP identified 135 (8.4%) of these as novel variants. In the final dataset following removal of synonymous coding variants and filtering by concordant (at least 3/5 methods) deleterious effect prediction, a total of 914 variants remained (Supplementary Table 2). In this data set the variant type categories were dominated by missense variants (751), followed by splice region variants (164), inframe deletions (45), frameshift variants (7), splice donor variants (6), inframe insertions (5), stop gained variants (3), splice acceptor variants (3) and a stop gained splice region variant (1) (Figure 1).

A total of 69 distinct variants were predicted to be deleterious by 5/5 functional effect predictors, 54 variants by 4/5 predictors and 54 variants by 3/5 predictors. At a threshold of 3/5 concordant predictors of a deleterious effect, 11 variants that have been annotated as benign or likely benign were included, vs. 0 variants that have been annotated as pathogenic or likely pathogenic. At a threshold of 4/5 concordant predictors of a deleterious effect, 11 variants that have been annotated as benign or likely benign were included, vs. 0 variants that have been annotated as pathogenic or likely pathogenic, and at a threshold of 5/5 concordant predictors of a deleterious effect, 26 variants that have been annotated as benign or likely benign were included, vs. 7 variants that have been annotated as pathogenic or likely pathogenic (Table 1).

A total of 66 genes showed variants, with PTCH1 (48/166), FANCI (39/166), SLX4 (39/166) and MLH1 (38/166) affected in the highest number of patients (Supplementary Figure 2). The most highly affected Reactome category was the DNA Repair Pathway, followed by the Regulation of Fanconi Anemia Pathway and the Fanconi Anemia Pathway.

*BRCA variants*

In the group of 166 patients a total of six BRCA1 and eight BRCA2 variants were present. The NM\_007294.3:c.5096G>A variant in BRB264 has been annotated in ClinVar as ‘Conflicting interpretations’, and the NM\_007294.3:c.4524G>A variant in BRB130 has been annotated as ‘Pathogenic’. Other BRCA variants have been annotated as ‘Uncertain significance’ or also ‘Conflicting interpretations’. There is one BRCA2 variant, being the frameshift variant NM\_000059.3:c.5771\_5774del in BRB290 which was also annotated as pathogenic (Table 2).

*Non-BRCA variants in genes associated with an increased susceptibility to breast cancer*

While TP53 is also regarded as a high-penetrance gene in breast cancer, only one TP53 variant with at least 3/5 deleterious predictions was identified.

In the medium penetrance gene set, only significantly disruptive variant types (eg. stop gained or frameshift variants) or variants with interesting ClinVar significance annotation effects were reported. Eleven of these variants were present, of which three were novel, but most likely significant, being an ATM stop gained variant in BRB14. Another, in ATM as well as a splice donor variant, NM\_000321.2:c.1127+1G>A, in the RB1 gene. There was one known CHEK2 pathogenic stop gained variant in BRB121. Additionally, there was a PALB2 likely pathogenic variant in BRB241 (Table 2).

No significantly disruptive variant types, or variants with interesting ClinVar significance annotation effects were found in the low penetrance gene set.

*Pathogenic or likely pathogenic variants in other genes*

Additionally, some known or likely pathogenic variants were found outside the set of 15 known breast cancer susceptibility genes were reported (Table 3). In BRB225 and BRB98, a FANCG known pathogenic frameshift variant was shown. In BRB114 and BRB161, an XPC known pathogenic splice acceptor variant was shown.

Only a single variant that has previously been annotated as pathogenic or likely pathogenic was not predicted to be deleterious by at least 3/5 functional effect prediction packages (thus filtered out of our selected variant set), being a HNF1A likely pathogenic missense variant in BRB18 (Table 4).

**Discussion**

As far as we are aware, this is the first breast cancer variant study on black South African females using a multigene panel. This set of samples, being BRCA deficient, gives new insight into cancer causing mutations, even though a total of 14 BRCA variants were still observed including a pathogenic frameshift and stop gained variant. A low number of known pathogenic variants was shown in the BRCA genes, which makes sense, given the circumstances.

The NM\_000059.3:c.5771\_5774del frameshift variant in BRCA2 for BRB290 is of particular interest because it has been previously identified as a possible non-Afrikaner South African founder mutation. Further studies will be needed to identify the impact of this variant.

A total of 11 variants with severely disruptive effects or demonstrated effects were identified in medium penetrance genes. This included one known stop gained pathogenic variant in CHEK2, one novel stop gained variant in ATM and RB1 and one PALB2 variant annotated as ‘with likely pathogenic allele’. The NM\_000051.3:c.162T>A stop gained variant in ATM and the NM\_000321.2:c.1127+1G>A splice donor variant are particularly interesting. In genes other than the 15 selected breast cancer susceptibility genes, a total four known pathogenic-type variants were found in FANCG and XPC.

Interestingly, only one TP53 mutation was identified in this work. A previous study also presented similar results 39, where a prominent breast cancer-associated gene (TP53) did not carry any known mutations. Another study on a subset of African American women (289), all with breast cancer, only one patient was found to carry a TP53 mutation 8. The p53 response is regulated by the methylation of specific arginines and multiple variants of the gene (p.R156H, p.R158C, p.R290C, and p.R333G) have been identified to nullify its gene specificity 40. More than 250 unique TP53 mutations have been linked to Li Fraumeni syndrome 41 and other cancers 42. Furthermore, none of the patients from this cohort were diagnosed with Li-Fraumeni syndrome. The International Agency for Research on Cancer (IARC) TP53 Germline Mutation Database (April 2016) for TP53 is grossly underrepresented in the African population, where only seven entries have been made 43. With this in mind, an underrepresentation of TP53 mutations in our study may be understandable.

The Reactome DNA Repair Pathway showed possibly deleterious variants in nearly all patients, followed by the Regulation of the Fanconi Anemia Pathway and the Fanconi Anemia. Enrichment analysis was not performed, as genes and pathways had effectively been pre-selected for cancer involvement using the Illumina Trusight Cancer Panel.

Much of the data in cancer susceptibility variant databases originates from European and some Asian studies. In the small number of African American studies performed, highly varying levels of admixture complicate extrapolation to African populations. The lack of known pathogenic BRCA and TP53 variants in this cohort, as well as low number of known pathogenic medium penetrance variants suggests the need for further in-depth study of South African and other African patients on a much larger scale, towards the elucidation of high and medium penetrance genes in these populations. This is clearly the case in many populations worldwide that have been underrepresented in genomic health studies.

While precision medicine is currently still mostly out of reach in African countries due to socio-economic reasons, the rapidly declining costs of genomic technologies will in future necessitate population-specific variant information, particularly in diseases such as cancer. A major shortfall to this study included: a relatively small and underrepresentation cohort of a black South African population as well as the absence of family history information regarding breast cancer.

**Acknowledgements**

This study was funded by the University of Pretoria Genomics Research Institute, the Cancer Association of South Africa, the South African Medical Research Council and the South African National Research Foundation (Grants IFR160218158695, BFG14081491267, CPRR13091237847 and the Rated Researcher Incentive Grant).

**Contributions**

FJ and EJvR conceived the project. Funding was obtained by FJ and EJvR. Samples were provided by EJvR. Analyses were performed by DE, EJvR and FJ. The manuscript was written by DE, EJvR and FJ.

**Data availability**

The fastq reads and the VCF file have been deposited in the in the European Genome-

Phenome Archive (EGA https://www.ebi.ac.uk/ega/home) with study number xxxxx. Access to data is determined by a Data Access Committee (DAC: xxxxx). Data access decisions can be passed to the EGA by emailing ega-helpdesk@ebi.ac.uk with the email address of each applicant and confirmation of the dataset(s) to provide access. The EGA will then create an EGA account with the relevant access permissions.

**Conflict of interest**

There is no conflict of interest.

**References**

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015;65(2):87-108.

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin.* 2016;66(1):7-30.

3. Borgen PI, Wong GY, Vlamis V, et al. Current management of male breast cancer. A review of 104 cases. *Annals of surgery.* 1992;215(5):451-457; discussion 457-459.

4. Fentiman IS, Fourquet A, Hortobagyi GN. Male breast cancer. *Lancet.* 2006;367(9510):595-604.

5. DeSantis C, Ma J, Bryan L, Jemal A. Breast cancer statistics, 2013. *CA Cancer J Clin.* 2014;64(1):52-62.

6. Society AC. Cancer Facts & Figures 2016. *American Cancer Society.* 2016:1-72.

7. Vorobiof DA, Sitas F, Vorobiof G. Breast cancer incidence in South Africa. *Journal of clinical oncology.* 2001;19(suppl 1):125s-127s.

8. Churpek JE, Walsh T, Zheng Y, et al. Inherited predisposition to breast cancer among African American women. *Breast Cancer Res Treat.* 2015;149(1):31-39.

9. Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *Jama.* 2006;295(21):2492-2502.

10. Sundquist M, Thorstenson S, Brudin L, Wingren S, Nordenskjold B. Incidence and prognosis in early onset breast cancer. *Breast (Edinburgh, Scotland).* 2002;11(1):30-35.

11. Wiesmüller L. Breast Cancer Susceptibility Genes. In: Schwab M, ed. *Encyclopedia of Cancer*. Berlin & Heidelberg: SpringerLink; 2011.

12. Neuhausen SL. Founder populations and their uses for breast cancer genetics. *Breast Cancer Res.* 2000;2(1):77-81.

13. Abbad A, Baba H, Dehbi H, et al. Genetics of breast cancer in African populations: A literature review. *Global Health, Epidemiology and Genomics.* 2018;3:E8.

14. Reeves MD, Yawitch TM, van der Merwe NC, van den Berg HJ, Dreyer G, van Rensburg EJ. BRCA1 mutations in South African breast and/or ovarian cancer families: evidence of a novel founder mutation in Afrikaner families. *Int J Cancer.* 2004;110(5):677-682.

15. Claes K, Machackova E, De Vos M, Poppe B, De Paepe A, Messiaen L. Mutation analysis of the BRCA1 and BRCA2 genes in the Belgian patient population and identification of a Belgian founder mutation BRCA1 IVS5 + 3A > G. *Dis Markers.* 1999;15(1-3):69-73.

16. van der Merwe NC, Hamel N, Schneider SR, Apffelstaedt JP, Wijnen JT, Foulkes WD. A founder BRCA2 mutation in non-Afrikaner breast cancer patients of the Western Cape of South Africa. *Clin Genet.* 2012;81(2):179-184.

17. Sluiter M, Mew S, van Rensburg EJ. PALB2 sequence variants in young South African breast cancer patients. *Fam Cancer.* 2009;8(4):347-353.

18. Ng PS, Wen WX, Fadlullah MZ, et al. Identification of germline alterations in breast cancer predisposition genes among Malaysian breast cancer patients using panel testing. *Clin Genet.* 2016.

19. Toh GT, Kang P, Lee SS, et al. BRCA1 and BRCA2 germline mutations in Malaysian women with early-onset breast cancer without a family history. *PloS one.* 2008;3(4):e2024.

20. Lee DS, Yoon SY, Looi LM, et al. Comparable frequency of BRCA1, BRCA2 and TP53 germline mutations in a multi-ethnic Asian cohort suggests TP53 screening should be offered together with BRCA1/2 screening to early-onset breast cancer patients. *Breast Cancer Res.* 2012;14(2):R66.

21. Tung N, Lin NU, Kidd J, et al. Frequency of Germline Mutations in 25 Cancer Susceptibility Genes in a Sequential Series of Patients With Breast Cancer. *J Clin Oncol.* 2016;34(13):1460-1468.

22. Tung N, Battelli C, Allen B, et al. Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel. *Cancer.* 2015;121(1):25-33.

23. Walker LC, Pearson JF, Wiggins GA, Giles GG, Hopper JL, Southey MC. Increased genomic burden of germline copy number variants is associated with early onset breast cancer: Australian breast cancer family registry. *Breast Cancer Res.* 2017;19(1):30.

24. Johns MB, Jr., Paulus-Thomas JE. Purification of human genomic DNA from whole blood using sodium perchlorate in place of phenol. *Analytical biochemistry.* 1989;180(2):276-278.

25. Andrews S. FastQC. 2018; <https://www.bioinformatics.babraham.ac.uk/projects/fastqc>.

26. Hannon G. The FastX Toolkit. 2018; <http://hannonlab.cshl.edu/fastx_toolkit/>.

27. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43(5):491-498.

28. Chapman B. bcbio: Blue Collar Bioinformatics. 2018; <https://github.com/bcbio/bcbio-nextgen>.

29. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760.

30. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17(1):122.

31. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536(7616):285-291.

32. Bumgarner R. DNA microarrays: Types, Applications and their future. *Current protocols in molecular biology / edited by Frederick M Ausubel [et al].* 2013;0 22:Unit-22.21.

33. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature methods.* 2014;11(4):361-362.

34. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PloS one.* 2012;7(10):e46688.

35. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46(3):310-315.

36. Shihab HA, Gough J, Cooper DN, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat.* 2013;34(1):57-65.

37. Karolchik D, Hinrichs AS, Furey TS, et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 2004;32(Database issue):D493-496.

38. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America.* 2005;102(43):15545-15550.

39. Tung N, Lin NU, Kidd J, et al. Frequency of Germline Mutations in 25 Cancer Susceptibility Genes in a Sequential Series of Patients With Breast Cancer. *Journal of Clinical Oncology.* 2016;34(13):1460-1468.

40. Jansson M, Durant ST, Cho EC, et al. Arginine methylation regulates the p53 response. *Nature cell biology.* 2008;10(12):1431-1439.

41. Varley JM. Germline TP53 mutations and Li-Fraumeni syndrome. *Hum Mutat.* 2003;21(3):313-320.

42. Kawaguchi T, Kato S, Otsuka K, et al. The relationship among p53 oligomer formation, structure and transcriptional activity using a comprehensive missense mutation library. *Oncogene.* 2005;24(46):6976-6981.

43. Bouaoun L, Sonkin D, Ardin M, et al. TP53 Variations in Human Cancers: New Lessons from the IARC TP53 Database and Genomics Data. *Hum Mutat.* 2016;37(9):865-876.

Table 1: Total numbers of distinct variants predicted as deleterious vs. broadly-grouped ClinVar significance annotations.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Deleterious Predictor Concordance | Total number of variants | ClinVar Significance Annotation | | | | | |
| **Benign or likely benign** | **Conflicting** | **Uncertain** | **Unknown** | **Novel** | **Pathogenic or likely pathogenic** |
| 3 / 5 | 54 | 25 | 10 | 8 | 11 | 0 | 0 |
| 4 / 5 | 54 | 11 | 13 | 19 | 11 | 0 | 0 |
| 5 / 5 | 69 | 26 | 6 | 15 | 7 | 8 | 7 |

Table 2: Variants in high and medium penetrance genes associated with an increased susceptibility to breast cancer. Only significantly disruptive variant types (eg. stop gained or frameshift variants) or variants with interesting ClinVar significance annotation effects are reported here (transcript 2 is provided in brackets for BRCA1 for historical reasons).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Patient | Age | Position | Gene Symbol | HGVSc | HGVSp | Variant  Category | dbSNP ID | ClinVar Significance |
| BRB264 | 42:3 | chr17:41215947 | BRCA1 | NM\_007300.3:c.5159G>A  (NM\_007294.3:c.5096G>A) | NP\_009231.2:p.Arg1720Gln  (NP\_009225.1:p. Arg1699Gln) | Missense | rs41293459 | Conflicting interpretations |
| BRB143 | 42:0 | chr17:41223249 | BRCA1 | NM\_007300.3:c.4745C>T  (NM\_007294.3:c.4682C>T) | NP\_009231.2:p.Thr1582Ile  (NP\_009225.1:p.Thr1561Ile) | Missense | rs56158747 | Conflicting interpretations |
| BRB146 | 52:2 | chr17:41223249 | BRCA1 | NM\_007300.3:c.4745C>T  (NM\_007294.3:c.4682C>T) | NP\_009231.2:p.Thr1582Ile  (NP\_009225.1:p.Thr1561Ile) | Missense | rs56158747 | Conflicting interpretations |
| BRB28 | 29:4 | chr17:41223249 | BRCA1 | NM\_007300.3:c.4745C>T  (NM\_007294.3:c.4682C>T) | NP\_009231.2:p.Thr1582Ile  (NP\_009225.1:p.Thr1561Ile) | Missense | rs56158747 | Conflicting interpretations |
| BRB42 | 49:11 | chr17:41223249 | BRCA1 | NM\_007300.3:c.4745C>T  (NM\_007294.3:c.4682C>T) | NP\_009231.2:p.Thr1582Ile  (NP\_009225.1:p.Thr1561Ile) | Missense | rs56158747 | Conflicting interpretations |
| BRB130 | 45:8 | chr17:41226499 | BRCA1 | NM\_007300.3:c.4587G>A  (NM\_007294.3:c.4524G>A) | NP\_009231.2:p.Trp1529Ter  (NP\_009225.1:p. Trp1508Ter) | Stop gained | rs80356885 | Pathogenic |
| BRB290 | 26:6 | chr13:32914260 | BRCA2 | NM\_000059.3:c.5771\_5774del | NP\_000050.2:p.Ile1924ArgfsTer38 | Frameshift variant | rs80359535 | Pathogenic |
| BRB160 | 40:3 | chr13:32972525 | BRCA2 | NM\_000059.3:c.9875C>T | NP\_000050.2:p.Pro3292Leu | Missense | rs56121817 | Conflicting interpretations |
| BRB99 | 45.3 | chr13:32972525 | BRCA2 | NM\_000059.3:c.9875C>T | NP\_000050.2:p.Pro3292Leu | Missense | rs56121817 | Conflicting interpretations |
| BRB59 | 39:0 | chr13:32912345 | BRCA2 | NM\_000059.3:c.3858\_3860del | NP\_000050.2:p.Lys1286del | Inframe deletion | rs772622136/rs80359406 | Conflicting interpretations |
| BRB9 | 36:5 | chr13:32912345 | BRCA2 | NM\_000059.3:c.3858\_3860del | NP\_000050.2:p.Lys1286del | Inframe deletion | rs772622136/rs80359406 | Conflicting interpretations |
| BRB193 | 43:5 | chr13:32913286 | BRCA2 | NM\_000059.3:c.4798\_4800del | NP\_000050.2:p.Asn1600del | Inframe deletion | rs276174851 | Uncertain significance |
| BRB268 | 46:2 | chr13:32913286 | BRCA2 | NM\_000059.3:c.4798\_4800del | NP\_000050.2:p.Asn1600del | Inframe deletion | rs276174851 | Uncertain significance |
| BRB98 | 43:3 | chr13:32913286 | BRCA2 | NM\_000059.3:c.4798\_4800del | NP\_000050.2:p.Asn1600del | Inframe deletion | rs276174851 | Uncertain significance |
| BRB234 | 39:10 | chr17:7578534 | TP53 | NM\_001276760.1:c.276\_278del | NP\_001263689.1:p.Asn92del | Inframe deletion | rs879254214 | Uncertain significance |
| BRB14 | 47:8 | chr11:108098592 | ATM | NM\_000051.3:c.162T>A | NP\_000042.3:p.Tyr54Ter | Stop gained | - | - |
| BRB146 | 52:2 | chr11:108143259 | ATM | NM\_000051.3:c.3078G>C | NP\_000042.3:p.Trp1026Cys | Missense variant, splice region variant | - | - |
| BRB121 | 54:0 | chr22:29130427 | CHEK2 | NM\_001005735.1:c.283C>T | NP\_001005735.1:p.Arg95Ter | Stop gained | rs587781269 | Pathogenic |
| BRB241 | 40:1 | chr16:23634452 | PALB2 | NM\_024675.3:c.2835-1G>C | - | Splice acceptor variant | rs515726099 | Likely pathogenic |
| BRB73 | 29:11 | chr13:48942741 | RB1 | NM\_000321.2:c.1127+1G>A | - | Splice donor variant | - | - |

Table 3: Variants in other genes associated with an increased susceptibility to breast cancer with pathogenic or likely pathogenic variants (variants with conflicting pathogenicity excluded here).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Patient | Age | Position | Gene Symbol | HGVSc | HGVSp | Variant  Category | dbSNP ID | ClinVar Significance |
| BRB98 | 43:3 | chr9:35077263 | FANCG | NM\_004629.1:c.637\_643del | NP\_004620.1:p.Tyr213LysfsTer6 | Frameshift variant | rs587776640 | Pathogenic |
| BRB225 | 34.4 | chr9:35077263 | FANCG | NM\_004629.1:c.637\_643del | NP\_004620.1:p.Tyr213LysfsTer6 | Frameshift variant | rs587776640 | Pathogenic |
| BRB114 | 47:1 | chr3:14190232 | XPC | NM\_004628.4:c.2251-1G>C | - | Splice acceptor variant | rs754673606 | Pathogenic |
| BRB161 | 29:6 | chr3:14190232 | XPC | NM\_004628.4:c.2251-1G>C | - | Splice acceptor variant | rs754673606 | Pathogenic |

Table 4: Variants not selected by 4/6 functional effect prediction software packages, with known or likely pathogenicity.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Patient | Age | Position | Gene Symbol | HGVSc | HGVSp | Variant  Category | dbSNP ID | ClinVar Significance |
| BRB18 | 31:9 | chr12:121432125 | HNF1A | NM\_000545.6:c.872C>G | NP\_000536.5:p.Pro291Arg | Missense variant | rs193922606 | With likely pathogenic allele |

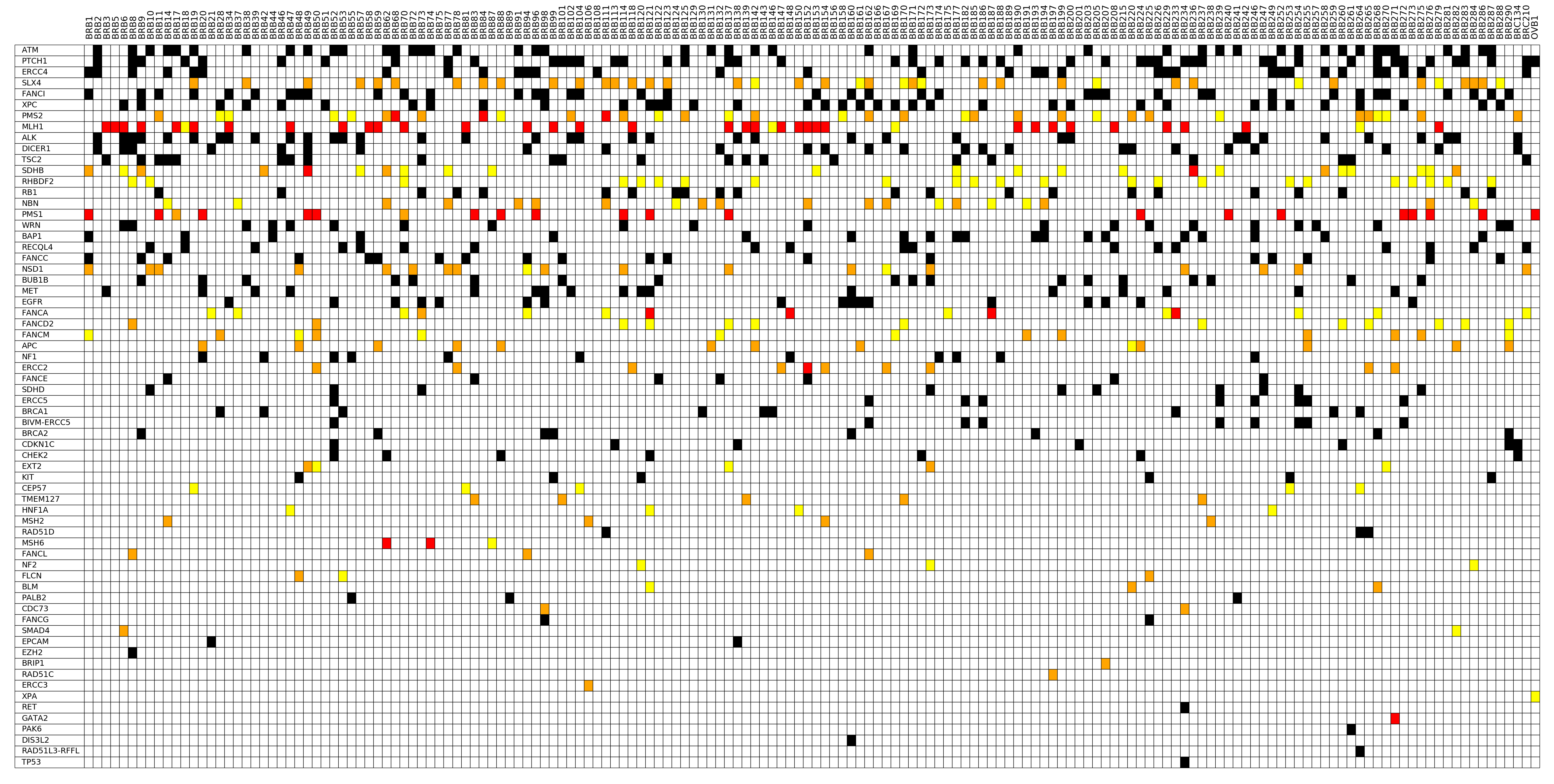


Figure 1: A matrix of patients vs. genes with variants (multiple variants may be present per gene, coloring is by the highest effect prediction score). Yellow indicates a deleterious effect predicted by 3/5 methods, orange by 4/5 methods, red by 5/5 methods and black by variants other than missense mutations.

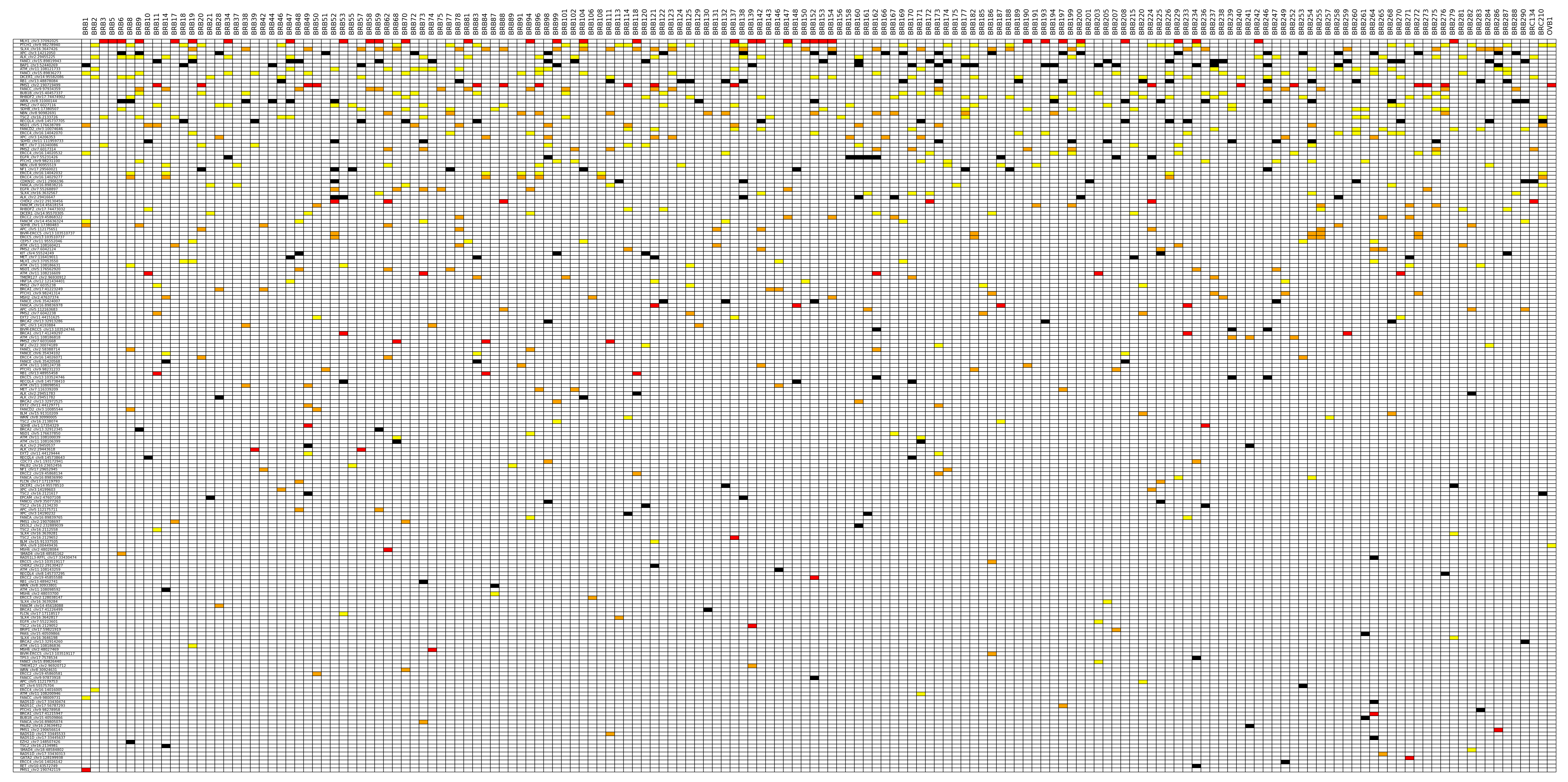


Figure 2: A matrix of variants vs. patients (coloring is by the highest effect prediction score). Yellow indicates a deleterious effect predicted by 3/5 methods, orange by 4/5 methods, red by 5/5 methods and black by variants other than missense mutations.

**Supplementary information**

Supplementary Table 1: All variants following GATK variant filtration (**online**).

Supplementary Table 2: All variants (excluding synonymous variants) predicted to be deleterious by three of the five functional prediction methods (**online**).