Elucidation of the molecular basis of genetic susceptibility to breast cancer has been one of the most successful areas of biomedical research in the last two decades [Oláh, HAS-Inaugural Lecture, 2013]. Breast cancer is the most common cancer in women. The majority of breast cancers are sporadic, but about 10 to 15% of all breast cancers are familial. The genetic architecture of hereditary breast cancer is complex and involves germline pathogenic variants in *high* and *moderate risk genes* and *polygenetic factors*. Approximately 20% of this familial risk is explained by pathogenic variants in the high risk genes **BRCA1** and **BRCA2**, 2-5% by variants in other breast cancer genes, and 18% by the known *common low risk variants*, mostly single nucleotide polymorphisms (SNPs). It has been estimated that all common low risk variants together, known and predicted to exist, will jointly explain 41% of the familial relative risk in breast cancer [Michailidou et al, 2017].
Genetic testing for inherited mutations in BRCA1 and BRCA2 and some very rare high risk genes has become integral to the care of women with a severe family history of breast or ovarian cancer, but an unknown number of patients receive negative (i.e., wild-type) results when they actually carry a pathogenic mutation of cancer predisposing gene(s). In other words, the predisposing event responsible for a large number of syndromic families is unknown (**'missing heritability'**).

The Goals of our ongoing studies (supported by this and a previous OTKA grant) are/were as follows:

- Identification and analysis of *BRCA1/2-related malignancies in Hungary*;
- Implementing the new *NGS technology* into research and generating clinical utility (genetic testing and managing high risk patients);
- Elucidation of *novel genetic factors of germline susceptibility* in Hungarian patients/families with tumors of cancer syndromes and other rare malignancies;
- Identification of the *missing heritability* in Hungarian breast/ovarian cancer families tested negative for mutations of cancer syndrome genes (e.g. BRCA1 and BRCA2);
- *Polygenic susceptibility*: identification of modifiers of cancer risk in BRCA1/2 mutation carriers (in our international collaborations);
- Clinical aspects: identification of *genotype-phenotype associations* of germline variants for *disease subtypes* (in collaboration with Departments of Pathology and Surgery at the Institute, and with large consortia);
- Genomics and epidemiology research to gain new insight into the *genetic architecture* and mechanisms underlying *breast and ovarian cancers* (in collaboration with multiple, prominent *disease-based consortia* listed below).
- Our ULTIMATE GOAL is to incorporate this new knowledge on disease-associated variants into risk-assessment and *improved cancer risk predictions* and *assist management of high-risk individuals*.

The AIMs of this project were:

(1) to explore *novel factors of genetic predisposition (germline susceptibility) to hereditary/familial cancers* with a focus on female and male breast cancer (BC) and ovarian cancer (OC) and

(2) to *gain* new insight into the *polygenic architecture of cancer susceptibility* and *cancer risk* by using new genotyping technologies, such as next generation sequencing (NGS), Genome-Wide Association Study (**GWAS**) and, most recently, transcriptome-wide association study (**TWAS**).

These approaches rely on availability of large collections of genomes/transcriptomes (DNA and RNA samples) and clinical data to undertake *large-scale genotyping* studies of genomics and epidemiology research on the *"big data"* generated – and the bigger, the better.

Our project has utilized (1) *cutting edge hereditary cancer research* work of the Oláh lab, including those in the emerging *"big data era"*, (2) our *results of >11,000 genetic tests* on cancer susceptibility genes that identified over 1000 deleterious germline BRCA1/2 mutations and hundreds of non-BRCA1/2-related BC-families *in the last 25 years*, (3) our existing *large biobank* sample- and data set from genetic testing, (4) implementing up-to-date *NGS* methods by installing the first Illumina MiSeq platform in Hungary in 2012, and (5) institutional and domestic collaborations and active participation of the HUNBOCS (Hungarian Breast- and Ovarian Cancer Study Group, Principal Investigator: Edit Oláh) in very successful, *ongoing large international collaborative projects on genomics and epidemiology research of breast-, and ovarian cancer*, contributing to multiple disease-based worldwide consortia, including, among others, the CIMBA (The Consortium of Investigators of Modifiers of BRCA1/2, http://cimba.ccge.medschl.cam.ac.uk/) and the *OncoArray Network*, a consortium of consortia established to discover germline genetic variants predisposing to different human cancers (e.g., breast, colon, lung, ovary, endometrium and prostate cancers).

The ***specific aims*** of this project were:

Objective A: **Search for germline variants in genes responsible for the maintenance of genome integrity** (Gene Set I).
Objective B: **Search for germline variants in genes linked to breast cancer (BC) by GWAS studies** (Gene Set II).
Objective C: **Testing germline variants as potential risk modifiers in known BRCA mutation carriers.**
Objective D: **Functional analysis of the most promising candidate alleles.**
Objective E: **Development of a method for genome-wide testing for a potential novel class of mutations**

The "high risk" breast cancer families and cases were previously selected for BRCA1/2 genetic testing and for further research analysis. (See *Figure 1.*)

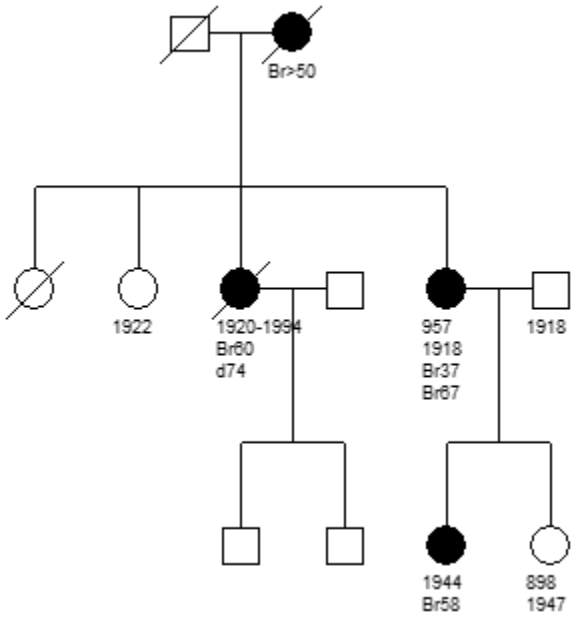

Figure 1.: Pedigree of a "high-risk" breast cancer (BC) family

RESULTS

– **Generation of Study Material by Identification and Analysis of BRCA1/2-related Malignancies in Hungary**

In the last 25 years (since the discovery of BRCA1/2 genes), the Oláh Lab analysed above 11,000 patients with female- or male breast cancer, ovarian cancer and identified above 1,000 distinct deleterious germline variants in BRCA1 and BRCA2 genes, the main susceptibility genes of breast- and ovarian cancer. Above 400 pathogenic variants were identified in the last five years, using NGS technology (Illumina, MiSeq platform and Reporter software) and Sanger-sequencing. *Pathogenic BRCA1/2 variants* included nonsense mutations and small frameshift deletions/insertions (indels), canonical splice variants and copy number variations (CNVs). *Large genomic rearrangements* – identified by MLPA – occurred in both genes but were more prevalent in BRCA1 than in BRCA2 genes due to large number of Alu repeats in the genomic region containing the BRCA1 gene. The mutation pattern and potential *founder* mutations (common mutations in a population arising from a small number of individuals) demonstrated in **female and male breast cancer, in ovarian cancer and in other rare malignancies**, is well in line with those reported in our first publications (Ramus-Köte et al, Am J Hum Genet 1997; Ramus et al, Nature Genet 1997; Csókay et al Cancer Res 1999; van der Looij et al Int J Cancer 2000] [*manuscript in preparation*].

– In a recent *worldwide study*, we reported the prevalence and spectrum of deleterious germline mutations in BRCA1 and BRCA2 genes of 29,700 families with BRCA1 and BRCA2 mutations ascertained from 69 centres in 49 countries on 6 continents, including several hundreds of *Hungarian breast- and ovarian cancer families*. Our study comprehensively describes the characteristics of the 1,650 unique BRCA1 and 1,731 unique BRCA2 *deleterious (disease-associated) mutations* identified in the CIMBA database. We observed substantial variation in mutation type and frequency by geographical region and race/ethnicity.
[*Rebbeck et al. – incl. Oláh E, Mátrai Z. and Papp J, Hum Mut, 2018*].

– We published the first *transheterozygote* mutation carrier case (i.e. patient who has inherited deleterious mutations both in BRCA1 and BRCA2 genes) in a Hungarian patient with breast and ovarian cancer [*Ramus Nat. Genet. 1977*]. Recently, we contributed to a large study characterizing the breast tumor markers and phenotypes of BRCA1-BRCA2 transheterozygotes. We concluded that the clinical phenotypes of these patients closely resemble those with only BRCA1 mutations, while their breast tumour marker characteristics are phenotypically intermediate to BRCA1 and BRCA2 carriers. [*Rebbeck et al – incl. Oláh Breast Cancer Res 2016*].

– **Implementing the new NGS technology** (Illumina MiSeq platform) **into research and clinical genetic testing**
In 2011, to elucidate germline variants of BRCA1/2, we conducted the first targeted next-generation sequencing (NGS) using a Roche FLX Titanium platform. To bring the NGS methods applied by us up-to-date, in 2012 we implemented NGS by installing the first Illumina MiSeq platform in Hungary.

Our department also developed an in-house bioinformatics workflow to help the secondary evaluation of our sequencing runs. That workflow turned out to be superior to MiSeq Reporter, the software tool installed on the sequencer. By using our in-house bioinformatics procedure, we became capable of preventing all types of miscalls that were typical of the earlier versions of MiSeq Reporter. [*Vaszkó et al – manuscript was submitted for publication but rejected 2016*]

***Objective A*:**
Search for germline variants in genes responsible for the maintenance of genome integrity (Gene Set I)

The major high-penetrance breast cancer risk genes include BRCA1 and BRCA2 involved in DNA double-strand break repair through homologous recombination (HR) and in the inter-strand crosslink (ICL) repair as a part of the Fanconi Anemia (FA) pathway. We tested the hypothesis that some other known/still unidentified genes of FA pathway and beyond that are associated with the maintenance of genome integrity could explain the "THE MISSING HERITABILITY". This study included 350 **genetically unexplained Hungarian BC/OC cases** with a suspicion of susceptibility to breast cancer but ***negative for germline BRCA1/2 mutations.*** The families were selected between 1992 and 2014 through Clinical Genetic Service provided by our Department.

(1) The gene set we analysed included established breast cancer susceptibility genes (CHEK2, PALB2, FAM175A/Abraxas and others) and genes playing a role in the maintenance of genome integrity (RAD50, RAD51C, ERCC4, FANCM, among others). Library construction, enrichment for the selected genes, and massively parallel next-generation sequencing was done using Agilent SureSelect XT2 kit and, for a smaller cohort, the Illumina TruSight Cancer Panel. For the initial (primary and secondary) analysis of sequencing data, we used the software tools specific to our Illumina MiSeq sequencing instrument as well as freely available data analysis packages (BWA, GATK, SAMTools and others).

Analyses of NGS data led us to the identification of 12 predisposing variants in 8 genes (ATM, CHEK2, ERCC2, FANCL, FANCM, NBN, PALB2, TP53) which could explain the disease of 14 from all 139 patients studied (10%). Further, we used several in silico tools to predict potential pathogenicity of missense and splice site variants, revealing the presence of another 21 likely pathogenic variants in our cohorts. In summary, our data indicated that rare truncating variants in different DNA repair genes, may be responsible for breast and ovarian cancer susceptibility or could act as risk-modifying variants. From the 21 probably pathogenic variants, 15 were located in DNA repair genes, which role in breast cancer susceptibility has been already established. The pathogenicity of these variants as well as the extent of their associated cancer risk undoubtedly requires further investigations.

The above results are now summarized in a manuscript draft.

(Pocza T, Papp J, Bozsik A, Vaszko T, Gyuris T, Balint LB, Olah E: Repair Gene Variants are Potential Risk Alleles for Cancer Susceptibility in Herediray Breast and Ovarian Cancer Patients)

(2) To extend the number of genes analyzed, RNA enrichment probes for 24 DNA polymerases (all human DNA polymerase genes playing a role in non-homologous end joining and other DNA repair processes) and also for 32 repair-associated enzymes were designed and ordered, altogether covering ~160 kb genomic regions. Library construction, enrichment for the selected genes, and massively parallel next-generation sequencing was done using Agilent SureSelect QXT kit and a HiSeq 2000 Sequencing System (Illumina). The bioinformatics analysis of sequencing data were done as above.

Variants (including copy number variations) of the targeted loci were assessed using deep sequencing of 90 BRCA-negative *male breast cancer cases* already selected for uniform histology (invasive ductal carcinoma cases).
The analysis and validation of the NGS results is still in progress, but our initial investigations already revealed the presence of several pathogenic/likely pathogenic variants in genes involved in the process of homologous recombination and/or in homologous recombination-associated repair (FANCM (3 cases), POLQ (3 cases), POLN (2 cases), and 1 case in XRCC3, MRE11, HUS1B, RAD1, RAD17, RAD51B, and RAD51D each). Moreover, we identified one case with the XXY syndrome.
These results will be published after the completion of the data analysis.

To extend our knowledge on the 'missing heritability' in breast cancer risk and on the 'polygenic cancer susceptibility' of BC and OC *(Objectives A, B, C)*, we participated in several LARGE-SCALE PROJECTS as parts of consortial efforts with world-leading genotyping centers via CIMBA to identify

- new susceptibility regions/SNP variants/candidate genes
- new clues about disease mechanisms (genetic architecture of BC susceptibility)
- genotype-phenotype associations in tumor subgroups (with Departments of Pathology& Surgery of our Institute)
- cross-cancer analysis to discover shared cancer risk factors
- Gene_x_Environment and Gene_x_Gene interactions
- risk profiling/polygenic risk score (PRS)

Analyses of inherited/germline variants in several chromosomal regions were undertaken in female and male breast cancer and ovarian cancers and their subtypes by genome-wide association studies (**GWAS**) and most recently, by TWAS and the RESULTS were published:

(1) Conducting a GWAS, we successfully identified significant independent associations with ten variants at nine novel loci in *estrogen-receptor- (ER)-negative breast cancer*. These new variants with previously reported variants explain ~16% of the familial risk of this breast cancer subtype.
[*Milne et al – incl. Bozsik and Olah, Nat Genet 2017*];

(2) Analysis of pooled data from multiple genome-wide genotyping projects identified twelve new susceptibility loci for *different epithelial ovarian cancer (EOC) histotypes* (six for serous EOC (3q28, 4q32.3, 8q21.11, 10q24.33, 18q11.2 and 22q12.1), two for mucinous EOC (3q22.3 and 9q31.1) and one for endometrioid EOC (5q12.3). Integrated analyses of genes and regulatory biofeatures at each locus predicted candidate susceptibility genes, including OBFC1, a new candidate susceptibility gene for low-grade and borderline serous EOC
[*Phelan et al – incl. Olah and Pocza, Nat Genet 2017*];

(3) Through the characterization of the *pathological features of BRCA1/2 mutation carrier* ***male breast*** *cancer* samples it was shown that they display markedly distinct pathologic characteristics compared with BRCA1/2-positive female breast cancers
[*Silvestri et al. – incl. Ivády G, Papp J, Olah E, Breast Cancer Res. 2016*];

(4) We investigated associations of common genetic variants with *breast and prostate cancer risks for male carriers of BRCA1/2 mutations*. Large differences in absolute cancer risks were observed at the extremes of the polygenic risk score distribution, meaning that these scores may provide informative cancer risk stratification for male carriers of BRCA1/2 mutations
[*Lecarpentier et al. – incl. Vaszko T, Kasler M, Olah E, J Clin Oncol. 2017*];

(5) A most recent TWAS among 97,000 women identified candidate susceptibility genes for epithelial ovarian cancer risk
[*Lu et al 2019*];

(6) **Cross-cancer analysis** undertaken by several consortia resulted in the discovery of ***shared cancer risk***. Using GWAS summary statistics across six cancer types based on a total of 296,215 cases and 301,319 controls of European ancestry, we observed statistically significant genetic correlations between lung and head/neck cancer, breast and ovarian cancer, breast and lung cancer and breast and colorectal cancer. It was also found that multiple cancers are genetically correlated with *non-cancer traits* including smoking, psychiatric diseases and metabolic characteristics. Functional enrichment analysis revealed a significant excess contribution of conserved and regulatory regions to cancer heritability. This comprehensive analysis of cross-cancer heritability suggests that solid tumors arising across tissues share in part a common germline genetic basis.
[*Jiang et al – incl. Olah, Nat Comm 2019*];

(7) Analysis of **genetic susceptibility to colorectal cancer** (to join 'Cross-cancer studies'):
In a study on 87 unrelated probands from familiar adenomatous polyposis families, 24 different pathogenic mutations in APC were identified in 65 patients, including 12 novel variants. APC-negative samples were also tested for MUTYH mutations and we were able to identify biallelic pathogenic mutations in 23 % of these cases. Our data represent the first comprehensive study delineating the mutation spectra of both APC and MUTYH in Hungarian FAP families, including clinical characterisation of APC- and MUTYH-associated phenotypes
[*Papp et al. Fam Cancer 2016*].

These results above, together with those we previously reported on other genes related to hereditary colorectal cancer syndromes can be utilized in large-scale cross-cancer studies. These would include e.g. the MLH1, MSH2, and STK11 genes (Papp et al. 2007; Papp et al., 2010) and our discovery of a novel genetic factor of cancer susceptibility, the terminal deletion of EPCAM/TACSTD1 in Lynch syndrome (Kovács et al. 2009). EPCAM deletion is now part of the routine screening of Lynch syndrome.

(8) We identified, fine-scale mapped and characterized several *additional* ***novel BC susceptibility*** *(modifier)* ***loci*** *in BRCA mutation carriers*, and delineated a possible ***functional mechanism*** behind multiple SNPs associated with breast and ovarian cancer risk:
(i) identified multiple SNPs at chromosome 19p13 that regulate ABHD8 and perhaps ANKLE1 expression and indicate common mechanisms underlying breast and ovarian cancer risk (*Lawrenson et al Nat Comm 2016*);
(ii) our fine-scale mapping of the chromosome 9p22.2 region identified candidate causal variants that modify cancer risk in BRCA1 and BRCA2 mutation carriers (*Vigorito et al PLoS ONE 2016*);
(iii) through fine-scale mapping of the 12p11 locus we were able to identify independent association signals and putative functional variants for breast cancer risk (*Zeng et al Breast Cancer Res 2016*)
(iv) we identified four previously unidentified loci that display genome-wide significant associations with estrogen-receptor (ER)-negative breast cancer (*Couch et al Nat Comm 2016*);
(v) Analysis of several thousand common genetic variants at 6q25 identified breast cancer risk variants that display different phenotype associations and regulate ESR1, RMND1 and CCDC170 (Dunning et al Nat Comm 2016);

(vi) Genome-wide association studies have identified breast cancer risk variants in over 150 genomic regions, but the mechanisms underlying risk remained largely unknown. These regions were explored by combining association analysis with in silico genomic feature annotations (using dense genotype data on > 217K subjects participating in the BCAC and CIMBA. All samples were genotyped using the OncoArrayTM or the iCOGS chip). We defined 205 independent risk-associated signals with the set of credible causal variants in each one. Prioritizing of genes as targets showed that known cancer drivers, transcription factors and genes in the developmental, apoptosis, immune system and **DNA integrity checkpoint** gene ontology pathways were over-represented among the highest confidence target genes. **Fine-mapping of all 150 breast cancer risk regions resulted in identification of 191 high confidence target genes.**
[Fachal L *et al – incl. Olah and Papp Nat Genetics 2019 (accepted for publication)*];

(9) NOVEL METHOD for genetic analysis of **Chromosome Y** was developed:
Chromosome Y's AZFc is one of the most unstable regions of the human genome, and numerous structures deducible from the reference sequence through inversion, deletion and duplication, or combinations thereof, have been reported. Among those rearrangements, gr/gr deletion seems to be a risk factor for testicular germ cell tumors (TGCTs) according to a study carried out by the International Testicular Cancer Linkage Consortium (ITCLC) and, according to our inital findings, it may also be associated with male breast cancer susceptibility. Different subtypes of gr/gr deletions remove different members of the eight gene families located in AZFc, therefore their effects may be not equivalent. It is the same with other AZFc deletions and duplications. However, subtyping of the AZFc rearrangements has been unreliable partly due to methodological shortcomings.
We developed a novel method called variant ratio analysis which is able to differentiate the partial deletion and partial duplication subtypes of the Deleted in Azoospermia (DAZ) gene family, on Chromosome Y. Besides

DAZ1/DAZ2 and DAZ3/DAZ4 deletions, not yet described rearrangements such as DAZ2/DAZ4 deletion and three duplication subtypes were also found by the utilization of the novel approach. We showed that it is possible to make distinction among the DAZ deletion and duplication subtypes with very high sensitivity by using a single, integrated approach. Based on the described principles, it is possible to develop a next generation sequencing-based method that would be easily applicable for rapid routine screening in clinical settings
[*Vaszkó et al, PLoS ONE 2016*].

***Objective B*:**
Search for germline variants in genes linked to breast cancer by GWAS studies (Gene Set II)

We tested the hypothesis that the regions designated by genome-wide association studies really contain susceptibility genes, whose role in tumor predisposition may not even be suspected on the basis of their already known functions.To complete this Objective, we selected 11 breast cancer-associated chromosome regions implicated by at least two independent genome-wide association studies performed between 2005-2012 (A Catalog of Published Genome-Wide Association Studies, http://www.ebi.ac.uk/gwas) and performed deep resequencing of 53 targeted genes positioned in a 1 Mb region surrounding the most significant (focal) SNPs of the selected GWAS loci in 120 early-onset female and 60 male primary invasive breast cancer patients. For control, we sequenced 180 healthy Hungarian samples in 6 pooled groups.

Library preparation was done with Agilent Sure Select XT2 enrichment method, next-generation sequencing was performed on an Illumina HiScanSQ Sequencing System. The yielded variant calls were appropriately quality-controlled and subjected to extensive in silico annotations and predictions (incl. missense, splice, expression regulation computations) to assess their possible pathogenicity.

The exhaustive sequencing led to the identification of 2845 high-confidence variants, 724 of them were unique, not listed in variant databases so far. Annotation of the variants resulted in the detection of 16 clear-cut mutations (11 nonsense, 4 canonical splice, one frameshift) as well as 26 predicted pathogenic variants with MAF≤0.05 in altogether 20 genes. Eight of them (SYNE1, ATE1, RTKN2, ZFYVE26, DAP, ROPN1L, CHD9, ZNF365) were of especially high interest, with several putatively causative variants genotyped within.
Interestingly, three of the suspected genes (ZFYVE26, ZNF365 and TACC2) are involved in centrosome pathway, a cellular procedure, which has an emerging role in predisposition to breast cancer. SYNE1 also organizes meiotic chromosome movements via binding telomeres to telocentrosomes and mediate the centrosome-nucleus coupling during neuronal migration in the cerebral cortex. Centrosomal aberrations have been detected in premalignant lesions and in situ tumors in the breast and in over 70% of invasive breast tumours, making the genes of this pathway compelling candidates in the etiology of breast cancer.

Follow-up studies as well as functional evaluations are required to confirm the inferred pathogenic nature of these variants.

Objective C:
Testing germline variants as potential RISK modifiers in known BRCA mutation carriers

(1) **BRCA1/2-related risk for BC, OC, bilateral BC and male cancers:**
We re-evaluated and published cancer risk in BRCA1/2 mutation carriers:
(i) Age-specific *risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers* were re-estimated, also risk modification by family cancer history and also by mutation location were evaluated, using data from ~10,000 mutation carriers. The cumulative breast cancer risk to age 80 years was 72% for BRCA1 and 69% for BRCA2 carriers. Breast cancer risk was higher if mutations were located outside *vs* within the regions bounded by positions c.2282-c.4071 in BRCA1 and c.2831-c.6401 in BRCA2. These findings demonstrate the potential importance of family history and mutation location in risk assessment
[*Kuchenbaecker et al – incl. Olah, JAMA 2017*];

(ii) Results of our international collaborative study on BRCA1/2-related risk, based on 31,000 carriers' samples, concluded that breast and ovarian *cancer risks varied by type and location of BRCA1/2 mutations.* With appropriate validation, these data may have implications for risk assessment and cancer prevention decision making for carriers of BRCA1 and BRCA2 mutations
[*Rebbeck et al – incl. Olah and Papp, JAMA 2016*].

(2) **Genomics and epidemiology research** to gain new insight into the **genetic architecture** and mechanisms underlying hereditary breast and ovarian cancers.
In collaboration with multiple, prominent disease-based consortia, we In order to identify **common variants** (mostly SNPs)**, new modifier chromosomal loci and candidate genes as potential risk modifiers** large-scale genotyping was undertaken using carriers of BRCA1/2 mutations from the Consortium of Investigators of Modifiers of BRCA1/2 (**CIMBA**) that includes the Hungarian Breast- and Ovarian Cancer Study Group (**HUNBOCS** – PI: Edit Olah).

The results are published in prestigious journals:
(i) Conducting a GWAS, we successfully identified significant independent associations with ten variants at nine novel loci (as we reported for Objective A). These variants with 10 previously reported in *estrogen-receptor- (ER)-negative disease* explain ~16% of the familial risk of this breast cancer subtype. Our findings may lead to improved risk prediction and inform further fine-mapping and functional work to better understand the biological basis of ER-negative breast cancer
[*Milne et al – incl. Bozsik and Olah, Nat Genet 2017*];

(ii) After identifying nine novel risk alleles for EOC subtypes, we performed meta-analysis on the results for high-grade serous ovarian cancer with the results from analysis of >31000 BRCA1 and BRCA2 mutation carriers. This approach identified 3 additional susceptibility loci at 2q13, 8q24.1 and 12q24.31
[*Phelan et al – incl. Olah and Pocza, Nat Genet 2017*];

(iii) Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers were re-estimated to better understand breast/ovarian cancer disease mechanisms. To this end age-specific risks of breast, ovarian, and contralateral breast cancer for BRCA mutation carriers were estimated and risk modification by family cancer history and also by mutation location were evaluated, using data from ~10000 mutation carriers. The cumulative breast cancer risk to age 80 yrs was 72% for BRCA1 and 69% for BRCA2 carriers. Breast cancer risk was higher if mutations were located outside vs within the regions bounded by positions c.2282-c.4071 in BRCA1 and c.2831-c.6401 in BRCA2. These findings demonstrate the potential importance of family history and mutation location in risk assessment
[*Kuchenbaecker et al – incl. Olah, JAMA 2017*];

(iv) we assessed the role of the K3326X (BRCA2) variant and found it associated with risk of developing breast and ovarian cancers independent of other pathogenic variants in BRCA2 gene
[*Meeks et al – incl. Olah, J Natl Cancer Inst 2016*];

(v) identified mitochondrial haplogroup T1a1 as a modifier of BRCA2-associated breast cancer risk
[*Blein et al – incl. Olah, Breast Cancer Res 2015*];

(vi) identified the chromosomal locus 11q22.3 as a new modifier in BRCA1 mutation carriers, revealing the presence of genetic variants showing differential allelic expression
[*Hamdi et al – incl. Olah, Breast Cancer Treatment 2017*];

(3) **Polygenic Susceptibility/ Polygenic Risk Scores (PRS)**
It is now widely accepted view that genetic susceptibility to cancer – particularly in genetically unexplained families – is linked to variant patterns rather than single alternative variants. This hypothesis was tested by genotyping both BRCA1/2 mutation negative and positive cancer cases with known family history and healthy controls.

(i) Using a 161-SNP polygenic risk score (PRS) to risk prediction of 101 **non-BRCA1/2 high-risk BC families** (including BC-affected and unaffected family members) within a long-term Dutch-Hungarian collaborative study, we demonstrated significant association between PRS and breast cancer. The currently known breast cancer (BC)-associated single nucleotide polymorphisms (SNPs) are presently not used to guide clinical management. Our results, however, support the application of the PRS in risk prediction and clinical management of women from **genetically unexplained breast cancer families.**
[*Lakeman et al – incl. Olah J Med Genet 2019*].

(ii). We investigated associations of common genetic variants with breast and prostate cancer risks for **male carriers of BRCA1/ 2 mutations** and implications for cancer risk prediction using **polygenic risk scores (PRS)** – for the first time to our knowledge – in >1800 male carriers of BRCA1/2 mutations from the CIMBA consortium.

Large differences in absolute cancer risks were observed at the extremes of the polygenic risk score distribution, meaning that these scores may provide informative cancer risk stratification for male carriers of BRCA1/2 mutations. We concluded that PRSs may provide informative cancer risk stratification for male carriers of BRCA1/2 mutations that might enable these men and their physicians to make informed decisions on the type and timing of breast and prostate cancer risk management.
[*Lecarpentier et al – incl. Vaszko, Kasler and Olah, J Clin Oncol 2017*].

***Objective D*:**
Functional analysis of the most promising candidate alleles

(1) As a continuation from the previous years, several novel candidate variants were selected for further studies within the framework of the ENIGMA consortium (Evidence-based Network for the Interpretation of Germline Mutant Alleles, http://enigmaconsortium.org)
FANCM is a member of the BRCA/FA molecular pathway which encodes for a translocase. Germline mutations were relatively frequent in our BRCA-negative families (see results for Objective A). Within an ENIGMA project (FANCM family study) we conduct modified segregation analysis to derive breast cancer risk estimates in carriers of truncating FANCM variants.

(2) To study the effect of FANCM on breast cancer risk further, we tested three recurrent truncating FANCM variants within the OncoArray Consortium. Three variants were tested for association with breast cancer risk in 67,112 breast cancer cases, 53,766 controls and 26,662 carriers of deleterious variants in BRCA1 or BRCA2 (CIMBA). We also studied the functional effect of these three variants after their lentiviral transduction into a FANCM–/– patient-derived cell line. The *functional results* indicated that all three variants were deleterious affecting cell survival and chromosome stability.
Our results also suggest that the effect of truncating variants on breast cancer risk may depend on their position in the gene. Furthermore, cell sensitivity to olaparib exposure, identifies a possible therapeutic option to treat FANCM-associated tumors.
[*Figlioli et al – incl. Olah and Pocza, np J Breast Cancer* (preliminarily approved) *2019*].

(3) Functional analysis of known BRCA variants with uncertain clinical significance (denoted as variants of unknown significance, VUS) was planned in order to determine their effects on homology-dependent recombination-driven double-strand break repair mechanism. As a first step, a limited number of candidate variants were selected for further studies (variants of unknown significance in BRCA1/2) within the framework of the ENIGMA.

(4) We have made haplotype and phenotype analysis of *high-risk families* (including BC cases and healthy relatives) with BRCA1 or BRCA2 *missense mutations* as we previously described for BC families with recurrent deleterious mutation in BRCA1/2 genes (Neuhausen et al, Am J Hum Genet 1996 & 1998). Missense mutations where haplotypes are co-segregating with malignant phenotype will be further evaluated for re-classification.

(5) It is a great challenge to translate the rapid expansion of sequencing capacity into useful knowledge and, in particular, interpret variant data of hereditary cancer genes to generate clinical utility. Therefore, genotype and phenotype data on pathogenic variants and on VUSs are submitted to global BRCA1/2 databases.

Objective E:
Development of a method for genome-wide testing for a potential novel class of mutations
The main goal was to establish a novel genetic approach to find germ-line genetic variants potentially predisposing to breast cancer using RNA extracted from peripheral blood-derived short-term cell cultures treated by a nonsense-mediated decay (NMD)-blocking agent. The conditions of the short-term culture were adapted in our laboratory, giving sufficient amount of high-quality RNA. In parallel, to determine the efficiency of the caffeine treatment on NMD inhibition, we selected several mutations resulting in premature stop codons and thus predicted to be NMD-sensitive. By comparing allele-specific expression patterns for these variants, we successfully determined the efficiency of the NMD mechanism and found it to be only 66-68%, leaving a considerable proportion of the mutant allele unaffected. Thus, our initially planned RNA-seq based approach would not have been useable to acceptably enrich the NMD-affected alleles, requiring an unaffordably high sequencing coverage (read depth) to comprehensively analyze patients' samples.

Consequently, we re-planned this part of the project without seriously diverging from our primary aims and, instead of a genome-wide search, we concentrated on retrotransposon-mediated genomic changes influencing the established breast cancer susceptibility genes BRCA1 and BRCA2.

For this we designed, tested an applied 31 PCR primer pairs to amplify the whole coding (and a large part of the non-coding) sequences of the BRCA genes, allowing us to amplify and characterize retrotransposon-mediated sequence changes (both deletion and duplications) up to several kilobases. Using this approach in combination with the gold standard MLPA (multiplex ligation-mediated probe amplification) we could already successfully identify several such variants.
These include 4 cases with partial duplication of the first two exons of the BRCA1 gene; 9 cases with deletions starting from exon 1 of BRCA1 (carriers of the del(ex1-2), del(ex1-3) and del(ex1-20) mutations); 7 cases with the BRCA1:del(ex8) variant;
9 cases with BRCA1 exon 13 duplication; 4 cases with deletion of exon 17 of BRCA1, and 12 cases with BRCA1: del(ex21-22). Altogether, this mutation type seems to be extremely infrequent in BRCA2, but relatively frequent in BRCA1, giving ~15% of all pathogenic mutations in BRCA1 gene.

Characterization of the breakpoints of these duplications/deletions showed a frequent (>90% of the cases) involvement of retrotransposable elements (Alu and LINE1).
Based on these initial results this modified approach seems to be an effective way to identify and characterize retrotransposon-mediated mutation mechanism and is also more easily adaptable to patient-oriented clinical genetics. Also, it is ready to be extended to test other susceptibility genes, where such sequence changes can be assessed and quantified for their risk-modifier effects.

The NEW KNOWLEDGE obtained have **immediate utilization** in clinical risk prediction, risk diagnosis (genetic testing) and clinical management of high-risk individuals. Our achievements were also used in constructing the European Consensus Statement and Expert Recommendations on Counseling and Testing for Breast Cancer Susceptibility Genes
[Singer *et al – incl. Olah, Eur J Cancer. 2019*].

ACKNOWLEDGEMENT: to OTKA for the funding of this project that permitted the continuation of our long-established collaborations with the prominent disease-based research consortia to elucidate the molecular basis of heritable susceptibility to cancer (mainly breast, ovarian and colorectal cancer).
We could confirm the motto of the American Association for Cancer Research (AACR): *"Cancer Research Saves Lives".*

Published papers with acknowledgment of OTKA-K112228 funding:

Figlioli G, […], **Olah E, […], Pócza T**, […], Peterlongo P. *Breast cancer risk magnitude of FANCM truncating mutations depends on their gene position*. np J Breast Cancer 2019 (preliminarily approved)

Fachal L, […], **Olah E, […], Papp J**, […],Dunning AM. *Fine-mapping of 150 breast cancer risk regions identifies 191 high confidence target genes.* Nat Genet. 2019; (accepted for publication)

Jiang X, […], Olah E, […], Lindström S. *Shared heritability and functional enrichment across six solid cancers.* Nat Commun. 2019;10(1):431. doi: 10.1038/s41467-018-08054-4.

Qian F, […], Olah E, […], Huo D. *Height and Body Mass Index as Modifiers of Breast Cancer Risk in BRCA1/2 Mutation Carriers: A Mendelian Randomization Study*. J Natl Cancer Inst. 2018. doi: 10.1093/jnci/djy132

Rebbeck TR, Friebel TM, Friedman E, Hamann U, Huo D, Kwong A, **Olah E, […], Matrai Z, Papp J**, […],Nathanson KL. *Mutational spectrum in a worldwide study of 29,700 families with BRCA1 or BRCA2 mutations*. Hum Mutat. 2018;39(5):593-620. doi: 10.1002/humu.23406

Milne RL, […], **Bozsik A**, **Olah E,** […],Simard J. *Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer.* Nat Genet. 2017;49(12):1767-1778. doi: 10.1038/ng.3785

Kuchenbaecker KB, […], Olah E, […], Olsson H. *Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers.* JAMA. 2017;317(23):2402-2416. doi: 10.1001/jama.2017.7112

Lecarpentier J, […], **Vaszko T, Kasler M, Olah E,** […], Ottini L. *Prediction of Breast and Prostate Cancer Risks in Male BRCA1 and BRCA2 Mutation Carriers Using Polygenic Risk Scores.* J Clin Oncol. 2017;35(20):2240-2250. doi: 10.1200/JCO.2016.69.4935

Phelan CM, […], Olah E, […], Pharoah PDP. *Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer.* Nat Genet. 2017;49(5):680-691. doi: 10.1038/ng.3826

Rebbeck TR, […], Olah E, […], Ramus SJ. *Inheritance of deleterious mutations at both BRCA1 and BRCA2 in an international sample of 32,295 women.* Breast Cancer Res. 2016;18(1):112

Hamdi Y, […], Olah E, […], Simard J. *Association of breast cancer risk in BRCA1 and BRCA2 mutation carriers with genetic variants showing differential allelic expression: identification of a modifier of breast cancer risk at locus 11q22.3.* Breast Cancer Res Treat. 2017;161(1):117-134. doi: 10.1007/s10549-016-4018-2

Lawrenson K, […], Olah E, […], Gayther SA. *Functional mechanisms underlying pleiotropic risk alleles at the 19p13.1 breast-ovarian cancer susceptibility locus.* Nat Commun. 2016;7:12675. doi: 10.1038/ncomms12675

Zeng C, […], Olah E, […], Zheng W. *Identification of independent association signals and putative functional variants for breast cancer risk through fine-scale mapping of the 12p11 locus.* Breast Cancer Res. 2016;18(1):64. doi: 10.1186/s13058-016-0718-0

Couch FJ, […], Olah E, […], Antoniou AC. *Identification of four novel susceptibility loci for oestrogen receptor negative breast cancer.* Nat Commun. 2016;7:11375. doi: 10.1038/ncomms11375

Papp J, Kovacs ME, Matrai Z, Orosz E, Kásler M, Børresen-Dale AL, **Olah E,** *Contribution of APC and MUTYH mutations to familial adenomatous polyposis susceptibility in Hungary.*
Fam Cancer. 2016 Jan;15(1):85-97. doi: 10.1007/s10689-015-9845-5.

Meeks HD, […], Olah E, […], Goldgar DE. *BRCA2 Polymorphic Stop Codon K3326X and the Risk of Breast, Prostate, and Ovarian Cancers*. J Natl Cancer Inst. 2015;108(2). pii: djv315. doi: 10.1093/jnci/djv315

Vaszkó T, Papp J, Krausz C3, Casamonti E, **Géczi L**, **Olah E**. *Discrimination of Deletion and Duplication Subtypes of the Deleted in Azoospermia Gene Family in the Context of Frequent Interloci Gene Conversion.* PLoS One. 2016 Oct 10;11(10):e0163936. doi: 10.1371/journal.pone.0163936. eCollection 2016.

Silvestri V, […], **Ivady G, Papp J, Olah E,** […], Ottini L. *Male breast cancer in BRCA1 and BRCA2 mutation carriers: pathology data from the Consortium of Investigators of Modifiers of BRCA1/2.* Breast Cancer Res. 2016;18(1):15. doi: 10.1186/s13058-016-0671-

Other publication :
Singer CF, Balmaña J, Bürki N, Delaloge S, Filieri ME, Gerdes AM, Grindedal EM, Han S, Johansson O, Kaufman B, Krajc M, Loman N, Olah E, Paluch-Shimon S, Plavetic ND, Pohlodek K, Rhiem K, Teixeira M, Evans DG: *Counseling and Testing for Breast Cancer Susceptibility Genes in Metastatic Breast Cancer – European Consensus Statement and Expert Recommendations*
Eur J Cancer. 2019 Jan;106:54-60. doi: 10.1016/j.ejca.2018.10.007. Epub 2018 Nov 22.

Budapest, 2019.

Edit Oláh