Identification and analysis of *BRCA1* and *BRCA2* gene mutations that play a key role in tumour evolution

Main points of the Ph.D. thesis

Márton Zsolt Enyedi

Supervisor: Dr. Lajos Haracska, D.Sc,

Scientific advisor

Hungarian Academy of Sciences, Biological Research Centre, Szeged

Institute of Genetics

University of Szeged, Faculty of Science and Informatics, Doctoral School of Biology

2017 Szeged

Introduction and background:

Cancer is a genetic disease in which changes in the genetic material are responsible for the transformation and clonal expansion of the cells. This process is often compared to the Darwinian evolutionary model, where changes in the genetic material occur gradually, offering selective advantages for the affected cells compared to the others. Inactivation of the tumour suppressor genes *BRCA1* and *BRCA2* is considered to be such a gradual evolutionary step during the development of breast and ovarian cancer. This is due to the crucial role of the BRCA proteins in the maintenance of genome integrity achieved by multiple mechanisms. Most importantly, they play a key role in homologous recombination, which is one of the most important repair mechanisms of DNA double-strand breaks.

We can distinguish between a hereditary and a sporadic form of *BRCA*-related tumorigenesis. In the case of inherited *BRCA* tumorigenesis, one allele of the gene has a germline-inactivating mutation. The remaining allele will be inactivated locally in the tumour cells by somatic alterations in a later phase of the tumour evolution. In the case of somatic *BRCA* tumorigenesis, both alleles are inactivated through somatic events. The presence of germline *BRCA* mutations can be observed in 10-20% of all breast and ovarian cancer patients, while the ratio of somatic *BRCA* mutations is between 5-10%. Due to the fact that germline mutations can increase the risk of developing breast and/or ovarian cancer by 90%, and both in case of somatic or germline *BRCA* mutant ovarian cancer there is a possibility to apply effective, targeted cancer therapy, the identification of *BRCA* mutations is a crucial healthcare issue.

Regarding all the above mentioned, it is comprehensible that the need for a reliable genetic test for the detection of mutations in the *BRCA1* and *BRCA2* genes emerged soon after their discovery in the early 90s. Since then, different methods have been used for mutation

diagnostics, among which the Sanger sequencing method proved to be the best. However, due to the lack of mutational hotspots in the *BRCA* genes and their relatively large size, the traditional capillary sequencing-based diagnostic process represents an expensive and time-consuming solution. The emergence and spread of benchtop next-generation sequencing systems offered the possibility to relocate the *BRCA1* and *BRCA2* mutation detection workflow onto these high-throughput platforms. This technology allows us to increase the throughput and detection sensitivity, enabling the identification of mutations from tumour samples even if they contain only trace amounts of tumour cells, which is a priority in understanding the process of tumour evolution and influencing its outcome. Various sources of DNA samples, different types of mutations, and the emergence of targeted cancer therapies all are require the development of a new diagnostic method for *BRCA1-2* that integrates the multiple needs and provides a reliable method for detecting hereditary and sporadic mutations in the *BRCA* genes.

Aims:

I aimed at investigating *BRCA*-related tumour evolution during my PhD work by developing a comprehensive *BRCA* diagnostic method that is based on multiplex PCR amplification, next-generation sequencing, and computational variant identification, offering a reliable tool for the detection of both germline and sporadic *BRCA*-inactivating mutations. I planned to implement this research and development process through the following steps:

1. Development of a method for the enrichment of *BRCA1* and *BRCA2* genes from genomic DNA samples and DNA-fragment library preparation for next-generation sequencing.

2. Contribution to the development of a bioinformatics software package for next-generation sequencing data analysis. Optimization and testing the performance of data analysis and mutation detection on *BRCA* mutant control DNA samples.

3. Validation of the developed diagnostic process on Ion Torrent PGM and Illumina MiSeq next-generation sequencers. Comparison of the mutation detection performance between the two sequencing platforms.

4. Testing the mutation detection workflow on tumour samples and studying the *BRCA*-related tumour evolution in ovarian cancers.

Materials and methods:

The target enrichment of the *BRCA1* and *BRCA2* genes from genomic DNA was achieved by multiplex PCR reactions. Primer sequences used in multiplex reactions were designed using the Primer3 and CloneManager software.

Multiplex PCR products were digested with a combination of restriction enzymes, and next-generation sequencing platform-specific adaptor sequences were ligated to the generated fragments. The resulting fragment libraries were quality checked by fragment analysis and quantified by qPCR. Next-generation sequencing was carried out on Illumina MiSeq or Ion Torrent PGM.

During the optimization of the sample preparation process, I was also involved in the development of a software for next-generation sequencing data analysis. The NGSeXplorer, in addition to the general modules of data analysis (de-multiplexing, mapping, variant calling), has some special algorithms that aim to reduce the high frequency of homopolymer sequencing errors specific to Ion Torrent PGM.

The *BRCA* mutation detection workflow was tested on a DNA sample set with previously characterized *BRCA1* or *BRCA2* germline mutations. This was followed by a validation process carried out on an independent sample set comparing the sequencing results and mutation detection of Illumina MiSeq and Ion Torrent PGM.

For the detection of partial or complete gene deletions or duplications, I calculated the dosage quotient from the coverage values of Illumina MiSeq sequencing. The results of this analysis were validated using MPLA test.

I tested the method on formalin-fixed, paraffin-embedded (FFPE) tumour samples as well. First, I checked the *BRCA* mutational status of ten ovarian cancer patients using paired

normal and tumour DNA. Following this, I also investigated a group of 25 ovarian cancer patients, this time sequencing only the tumour DNA. The identified mutations were validated with independent PCR and Sanger sequencing in all cases.

Results and discussion:

1. The *BRCA* diagnostic process was developed using two different sample sets: a test sample set and a validation sample set. Both sample sets contained DNA samples with germline mutations in the *BRCA1* or *BRCA2* genes. The test sample set was used for the optimization of the fragment library preparation, sequencing, and data analysis with special attention to the algorithms for Ion Torrent sequencing error reduction. By the end of the extensive optimization process, we achieved proper coverage data for the entire target region of *BRCA1* and *BRCA2*, and variant detection successfully identified the germline mutations for each sample.

2. After the initial optimization process, we blindly validated the diagnostic method on 20 samples each containing a unique pathogenic mutation. At the same time, we compared the mutation detection performance of the two most popular NGS platforms: Illumina MiSeq and Ion Torrent PGM. Illumina MiSeq correctly identified the known variants in all 20 retrospective samples with no false-positive or false-negative variant predictions. In the case of Ion Torrent PGM, the application of the variant filter set developed within the frame of NGSeXplorer led to a drastic reduction of false-positive results, illustrated by the specificity of 97.3%. Though, in one of the 20 samples a true pathogenic mutation was also discarded by one of the above mentioned filters resulting in an experimental sensitivity of 95.0%.

3. A comprehensive diagnostic workflow for *BRCA1* and *BRCA2* must include the ability of specific and accurate detection of large genomic rearrangements in these genes. For this reason, we tested the suitability of our diagnostic method for the detection of exon deletions and duplications. We had two test samples with known large deletions verified previously with MLPA analysis. The *BRCA* diagnostic method correctly identified these large deletions and duplications proving its suitability for the detection of large rearrangements.

4. We tested the performance of our method on DNA samples isolated from FFPE tissue blocks of ovarian cancer patients. From the first ten samples analysed, four harboured *BRCA1* mutations; three of them were germline and one was a somatic mutation. This latter was a novel *BRCA1* mutation (c.628C>T; p.Q210*) not reported previously in any database or reference. Finally, in this very same tumour sample we also managed to identify some steps of tumour evolution: one of the *BRCA1* alleles was inactivated by the stop-mutation mentioned above, while the remaining wild-type allele was lost during large genomic rearrangements (LOH - loss of heterozygosity).

5. As a next step, we analysed the *BRCA* mutational status of 25 additional ovarian cancer samples. We identified pathogenic mutations in nearly half of the samples, and the majority affected the germline. Two of the 25 analysed mutations were considered somatic alterations present only in the tumour cells. The frequency of the mutant allele was shifted towards homozygous direction in most of the cases, indicating that the remaining wild-type *BRCA* allele was inactivated through LOH. These results draw the attention to how important it is in a comprehensive *BRCA* diagnostic process to analyse the mutational status of the tumour sample as well as that of the germline.

Summary:

In summary, we developed a diagnostic method for the detection of *BRCA* mutations in germline and tumour. The method was successfully tested on both germline and tumour samples, and it was validated on a wide spectrum of pathogenic mutations using Illumina and Ion Torrent sequencers as well. Besides substitutions and small insertional and deletional events, we managed to identify deletions and duplications affecting entire exons. The *BRCA1* or the *BRCA2* gene was implicated in tumour evolution in a considerable percentage of the analysed ovarian cancer samples. In most of these cases, one allele of *BRCA1* or *BRCA2* was inactivated by a germline mutation, while the remaining wild-type allele was lost locally at the site of tumour formation during a later phase of tumour evolution (LOH - loss of heterozygosity). There were also some exceptions, when both *BRCA* alleles were inactivated by somatic events (mutations and large genomic rearrangements). The method is suitable for the detection of certain steps of *BRCA* tumour evolution using FFPE tumour samples, and, in combination with Illumina MiSeq sequencing, meets the sensitivity and specificity requirements for clinical diagnostics.

List of publications:

1. The two publication that form the basis of the doctoral procedure:

Enyedi MZ, Jaksa G, Pintér L, Sükösd F, Gyuris Z, Hajdu A, Határvölgyi E, Priskin K, Haracska L. Simultaneous detection of *BRCA* mutations and large genomic rearrangements in germline DNA and FFPE tumor samples. Oncotarget. 2016. doi: 10.18632/oncotarget, IF (2014): 6,35

E. Ruprecht, **M. Z. Enyedi,** A. Szabó, A. Fenesi.: Biomass removal by clipping and raking vs burning for the restoration of abandoned Stipa-dominated European steppe-like grassland. Applied Vegetation Science. 2015. DOI: 10.1111/avsc.12199. IF (2015): 2,30

MTMT number: 10051399

Total IF: 23,442

2. Publication in peer reviewed journals:

2.1. Publications supporting the dissertation:

Enyedi MZ, Jaksa G, Pintér L, Sükösd F, Gyuris Z, Hajdu A, Határvölgyi E, Priskin K, Haracska L. Simultaneous detection of *BRCA* mutations and large genomic rearrangements in germline DNA and FFPE tumor samples. Oncotarget. 2016. doi: 10.18632/oncotarget, IF (2014): 6,35

2.2. Additional publications:

Mórocz M, Zsigmond E, Tóth R, **Enyedi MZ**, Pintér L, Haracska L. DNA-dependent protease activity of human Spartan facilitates replication of DNA–protein crosslink-containing DNA. Nucleic Acids Res. 2017; gkw1315. doi: 10.1093/nar/gkw1315. IF (2015-2016): 9,202

E. Ruprecht, **M. Z. Enyedi,** A. Szabó, A. Fenesi.: Biomass removal by clipping and raking vs burning for the restoration of abandoned Stipa-dominated European steppe-like grassland. Applied Vegetation Science. 2015. DOI: 10.1111/avsc.12199. IF (2015): 2,30

E. Ruprecht, **M. Z. Enyedi,** R. L. Eckstein, T. W. Donath: Restorative removal of plant litter and vegetation 40 years after abandonment enhances re-emergence of steppe grassland vegetation. *Biological Conservation.* 2010. 143: 449-456. IF (2010): 4,29

E. Ruprecht, A. Szabó, **M. Z. Enyedi** & J. Dengler: Steppe-like grasslands in Transylvania (Romania): characterization and influence of management on species diversity and composition. Tuexenia. 2009. 29: 353-368.

Enyedi, M.-Zs., Ruprecht, E. & Deák, M: Long-term effects of the abandonment of grazing on steppe-like grasslands. Applied Vegetation Science. 2008. 11: 53-60. IF (2008): 1,30

3. Other scientific achievements:

Manuscripts:

Cs. Brasko, K. Smith, Cs. Molnar, L. Hegedus, A. Balind, N. Farago, L. Paavolainen, T. Balassa, A. Szkalisity, F. Sukosd, B. Balint, **M. Z. Enyedi**, I. Nagy, L. G. Puskas, L. Haracska, G. Tamas, & P. Horvath. Intelligent image-based in situ single-cell isolation. *Submitted.*

Conference lectures:

Sükösd F, **Enyedi M,** Hajdu A, Pankotai-Bodó G, Gyuris Z, Jaksa G, Pintér L, Haracska L. Ovárium karcinómák és örökletes emlőrákok *BRCA* mutáció vizsgálata. EME, Orvos-és Gyógyszerésztudományi Szakosztály, XXVI. Tudományos Ülésszak, Marosvásárhely (RO), 2016.

Enyedi M., Jaksa G., Pintér L., Gyuris Z., Keresztné Határvölgyi E., Priskin K., Haracska L.: A *BRCA1* és *BRCA2* gének diagnosztikája új generációs szekvenálással. EME, Orvos-és Gyógyszerésztudományi Szakosztály, XXV. Tudományos Ülésszak, Kolozsvár (RO), 2015.

P. Sarkozy**, Enyedi M**, P. Antal: Flow Index Based Characterization of Next Generation Sequencing Errors, 5th International Conference on Bioinformatics Models, Methods and Algorithms (Biostec), 2014.

Enyedi M-Zs. Szegedi Biológus Doktorandusz Konferencia: A *BRCA1* és *BRCA2* tumor szuppresszorok mutációinak és interakcióinak jellemzése a DNS-hibatolerancia útvonalban. Szekció előadói díj. 2013.

Enyedi M., Gyuris Z., Keresztné Határvölgyi E., Haracska L: Az újgenerációs szekvenálás alkalmazási lehetőségei a molekuláris tumordiagnosztikában. EME, Orvos-és Gyógyszerésztudományi Szakosztály, XXIII. Tudományos Ülésszak, Székelyudvarhely (RO), 2013.

M. Enyedi, Z. Gyuris, E. Határvölgyi, A. Hajdu, Sz. Bakó, F. Sükösd, L. Haracska: Mutational analysis of oncogenes along the EGFR signal transduction pathway. 3rd International Congress and 29th Annual Scientific Session of Romanian Society for Cell Biology, Arad (RO). 2011.

Enyedi, M.-Zs., Ruprecht, E., Józsa, J: Kísérleti beavatkozások hatása felhagyott szárazgyepek fajainak csírázási dinamikájára. VIII. Kolozsvári Biológus Napok. 2007.

Enyedi, M.-Zs.: Árvalányhajas sztyepprétek átalakulása felhagyásuk következtében az Erdélyi Mezőségen. Aktuális flóra és vegetációkutatás a Kárpát -medencében VII. Debrecen. 2006.

Enyedi, M.-Zs: Az Erdélyi Mezőség árvalányhajas sztyepprétjeinek előfordulását meghatározó tényezők. VI. Kolozsvári Biológus Napok.2005.