# Development of a new NGS-based method to detect malignant transformation of thyroid nodules

Thesis

**Barbara Kocsis-Deák**

Semmelweis University
Doctoral School of Clinical Medicine

Supervisor: Dr. Péter Lakatos, Dsc, academic lecturer

Offical reviewers: Dr. Henriett Pikó, Ph.D., research fellow
Dr. Bálint Ferenc Péterfia, Ph.D., senior research fellow

President of Comprehensive Examination Board:
Dr. Edit Buzás, DSc, academic lecturer

Members of the Board:
Dr. Attila Patócs Ph.D., associate professor
Dr. Gábor László Kovács, Ph.D., head of department

Budapest
2019

## 1. INTRODUCTION

Thyroid gland influences every mechanism in the human body. It regulates basic metabolism; therefore, its functioning is essential. If the thyroid gland cannot perform its tasks properly, the normal rythm of the human body may go wrong.

In Hungary, thyroid diseases are widespread, as they affect a huge ratio of the population. Besides the huge incidence of the most known thyroid disorders (hypo- and hyperthyroidism, or different kinds of inflammations), the incidence of thyroid tumors is continuously increasing worldwide. In Hungary, thyroid cancer is the most common endocrine tumor which usually develops from so-called 'cold nodules' (the prevalence of these in the thyroid gland is 4–7%); however, only some of these (5—10%) become malignant. Most of malignant thyroid tumors (90%) belong to the group of differentiated thyroid cancers (DTCs) which originate from the follicular epithelium. 70—80% of DTCs are papillary thyroid cancer (PTC).

Procedures in current clinical practice often detect malignant lesions late. Nowadays, having a modern method that is capable of making an early diagnosis; the specification of clinical diagnostics; and deciding if there is a risk of malignization is becoming more important. Due to the sudden improvements in genetics and informatics, a new doorway opened for a newer generation of molecular gene diagnostics which gives space to new generation sequencing (NGS)

techniques and the development of innovative tumor diagnostic procedures.

## 2. OBJECTIVE

In 2010, our team developed a panel based on the examination of 8 oncogenes. This panel helps diagnosing a thyroid lesion more accurately and foreshadow the potential malignancies. After this, our goal was to develop an extensive, more accurate and cost-efficient method.

1. Our primary objective was to develop a procedure for genetic testing based on new generation sequencing, which would be able to forecast the risk of malignization of currently cytologically benign cold nodules, and to help form a definitive diagnosis in ambiguous cytological cases.
2. To accomplish this goal, we first set out to develop a novel gene panel.
3. The new gene panel was designed to be validated on histologically confirmed tumor and non-tumor specimens.
4. We also wanted to make the new gene panel suitable for the evaluation of fine needle biopsy samples.
5. Our final objective was to make our method the most widespread and most cost-effective genetic diagnostic panel for the early detection of thyroid tumors; to predict the malignization of cold nodules; and to amend the results of ambiguous cytological test, with potential therapeutic consequences; internationally and in Hungary as well.

## 3. METHODS

### 3.1. Development of the genetic diagnostic panel

Our NGS method is based on the Ion Torrent PGM™ system by Thermo Fisher Scientific. We designed a unique AmpliSeq hot spot panel that contains 23 genes (*AKT1*, *APC*, *AXIN1*, *BRAF*, *C16orf3*, *CTNNB1*, *DICER1*, *EIF1AX*, *GNAS*, *HRAS*, *IDH1*, *KRAS*, *LPAR4*, *MET*, *NRAS*, *PIK3CA*, *PTEN*, *RET*, *SMAD4*, *TERT*, *TP53*, *TSHR* and *VHL*) and 568 known oncogene mutations. The genes and mutations included in our test were previously published in international literature in connection with thyroid tumors. The genetic diagnostic panel based on NGS can analyze several genes responsible for thyroid tumors from several thyroid tissue samples, simultaneously.

### 3.2. Collection of thyroid samples

We used two kinds of thyroid samples:

1. To optimize and validate our method, we analyzed thyroid tissue samples obtained during the operation of patients who were histologically confirmed to have papillary carcinomes. We collected fresh tissue samples intraoperatively from 40 patients with PTC (31 women and 9 men). Samples were collected from both the tumorous parts and from the normal, histologically healthy thyroid tissue as well. Overall, 67 samples have been examined, 39 of which were from papillary

thyroid carcinomes, and 28 of which were from histologically healthy thyroid glands. For 27 patients, the tissue samples from the PTC and normal thyroid parts were eligible for the study; from 12 patients only the malignant, while for 1 patient, only the healthy tissue was enrolled in the study.

2. Our procedure was tested on samples from possibly malignant nodules, obtained with ultrasound-guided fine needle aspiration biopsy (FNAB). Overall, 51 FNAB samples (obtained from nodules that were diagnosed as malignant via ultrasound) were analyzed.

### 3.3. Isolation of the genomic DNA

Genomic DNA was isolated from the cells of the tissue samples and the FNAB samples with the help of Roche High Pure PCR Template Preparation Kit (Roche Indianapolis, IN, USA). During isolation, the original protocol was followed. The quality of the isolated DNA was measured with NanoDrop spectrophotometer (NanoDrop Technologies, Montchanin, DE, USA) on 260/280 nm. Only the DNA with appropriate quality (where the 260/280 ratio was at least 1.8) was used in the study. The concentration of DNA was measured with the Qubit dsDNA HS assay Kit (Thermo Fisher, Waltham, MA, USA). Samples having too small concentration (<10 ng/ μl) were excluded from further

examination (the minimum DNA concentration for sequencing is 10 ng/µl).

### 3.4. New generation sequencing

Sequencing in a laboratory has 3 main steps: creating a library, preparing the emulsion PCR and sequencing.

The creation of the DNA library consisted of the amplification of specific regions of the isolated DNA.

During emulsion PCR, specific DNA regions are bound in an adequate concentration and ratio to the ion sequencing particles (ISP) and are amplified. From this step on, libraries included in the reaction are used in groups (5 samples/group), not individually, until the end of sequencing. If certain particles didn't bind any sequencable DNA segments to their surface, those pearls were removed from the template solution with an amplification process.

The last step of NGS in a laboratory is the sequencing process itself. The prepared ISP were taken into a sequencing chip; afterwords, the nucleotid sequence of the DNA fragments bound to the surface of the ISP was determined with an Ion Torrent PGM™ sequencing machine.

### 3.5. Analysis of genetic data

To evaluate the acquired data (at the end of the sequencing procedure its amount is several gigabytes) we used a platform

specific Torrent Suite v5.0.4 software. The genetic anomalies detected with the software analysis underwent a canonical variate analysis (CVA). With the help of these, we wanted to discover new, more accurate relations, by which we could create a score system that could help us separate malignant thyroid samples from benign ones.

Data from validating samples was statistically analyzed besides software analysis; whereas, results from the FNAB samples were only analyzed with software.

# 4. RESULTS

## 4.1. Mutation profile of thyroid tissue samples

From 67 samples 177 detected were non-synonymous mutations (that result in an amino acid sequence change); 145 of these were SNVs (142 missense and 3 nonsense); 31 INDELs (27 insertion, 3 deletion and 1 frameshift deletion) and 1 missense MNV. Fifty of these genetic differences were previously described in the COSMIC database; we identified further 127 new variants around the targeted sequences.

### *4.1.1. Mutations found in malignant tissue*

102 somatic mutations were identified in 20 genes from the PTC samples. The most polymorphic gene was found to be the *AXIN1* gene (with 16 different variations). The *BRAF* c.1799T>A (V600E) mutation occurred the most (found in 14 papillary thyroid carcinomes). On average, 5.74 mutations were detected in one PTC sample. The highest deviation number detected in one sample was 19 nucleotide change. Also, there were 2 PTC samples that did not have any detectable somatic mutation in the examined genes. Neither GNAS (the indicator of benign nodules) nor KRAS (which only causes benign thyroid gland neoplasy) were detected in any of the PTC samples.

#### *4.1.2. Mutations found in malignant tissue*

52 somatic mutations were detected in 15 genes altogether. *AXIN1* was the most polymorph gene, although its mutation pattern differed from those seen in PTC samples. *EIF1AX* c.71G>A mutation occured the most (in 2 samples).

### 4.2. Statistical analysis of genetic data

During CVA analysis, data were analyzed in two levels. One data matrix contained the mutations/variations that occured at least three times in the studied samples, while the other matrix summarized the changes on the level of genes.

#### *4.2.1. Statistical analysis via variants*

16 of the mutations proved to be the most informative (these variants occured in at least three samples), so the final CVA analysis was done on the level of these mutations. In the end of the analysis, all the samples got a certain value (F-value) by which the risk of developing a tumor could be estimated:

- positive F-values showed big correlation with having a tumor
- negative F-values showed tumor-free states.

The two patient groups are well separated along an axis; the results from the CVA analysis correlate well with the presence of a tumor.

The *BRAF* c.1799T>A mutation had the highest tumor marker potential, since it was detected in 14 tumor samples and was not present in any of the normal, tumor-free samples. We separated 5 other variants which had a huge role in the separation of the two groups (*TSHR* c.1373T>C, *APC* c.636_637insA, *LPAR4* c.137A>G, *TSHR* c.1963A>G and *SMAD4* c.964T>A). Therefore, genetic analysis can be useful even if we only examine these 6 point-mutations.

### *4.2.2. Statistical analysis via genes*

We performed a similar CVA analysis on 20 from the studied genes which showed at least one mutation. We gave every sample an F-value here as well; this was very helpful in estimating the risk of developing a tumor. On the level of genes, the line between the two groups could be pulled at 0.24 F-value. The two groups are well separated by the line.

- Tumor samples having an F-value greater than 0.24 are correlating with having a tumor present,
- while samples with an F-value smaller than 0.24 show a tumor-free state.

The presence of the BRAF gene mutations correlated most with the presence of the tumor. After the BRAF gene mutations, the mutations of the following 6 genes may be useful in estimating the risks: *TSHR*, *TP53*, *PIK3CA*, *VHL*, *AKT1* and *APC*. The first gene (by the F-value) that showed

mutations in a benign thyroid tumor was *EIF1AX*, followed by *RET*, *NRAS*, *DICER1* and *PTEN*.

The current health status of the patients/their histological results and the classification based on the statistical analysis correlated well in most cases. Most samples that were acquired from tumor-free patients were classified as healthy, while the samples from PTCs were classified as malignant.

Statistical analysis from mutations showed 79% sensitivity, 86% specificity, 89% PPV, and 75% NPV. Statistical analysis from genes showed 87% sensitivity, 68% specificity, 79% PPV, and 79% NPV.

### 4.3. The mutation profile of fine-needle aspiration biopsy samples

From the 51 FNAB analyzed samples, 35 showed a minimal clinically relevant genetic difference in one of the 23 examined genes, while 16 samples showed no such of difference.

Altogether 36 amino acid sequence change mutations were identified in the FNAB samples, in 15 genes: 24 SNVs (24 missense), 7 INDEL (1 insertion, 6 deletions), 2 splice site variants, 1 intragenic region mutation, and 2 MNVs were identified. From the differences we found, 27 were identified as new mutations.

From the 8 positives, i.e. cytologically malignant results, we found genetical abnormalities in 6 cases.

From 28 cytologically negative, i.e. benign abnormalities, some kind of genetic abnormality was identified in 20 cases, which could forecast the development of a malignant tumor later. We identified several mutations in these samples having a COSMIC identifier, the most known from these being the *HRAS* c.181C>A. The other 8 cytologically negative cases showed no change in the studied genes.

From the 14 samples classified as ambiguous (having the risk to develop malignity), no abnormality was found in 6 cases; however, some kind of mutation (such as *BRAF* c.1799T>A, *NRAS* c.181C>A) were identified in 8 samples.

With its 6 different variants, the AXIN1 gene was the most polymorphic gene, while the *LPAR4* c.137A>G mutation occured the most (in 11 samples): Citology showed 6 negative and 2 positive results, while 2 nodules had ambiguous results. The sample could not be evaluated (the smear was cell-free) in 1 case. The role of this gene is not yet clear, based on scientific literature.

The well-known *BRAF* c.1799T>A mutation was identified in 5 biopsy samples. Contrasted with the citology results, 4 samples were correctly classified as malignant, while 1 sample was classified as ambiguous and a new biopsy was advised.

We had 30 FNAB samples in which we discovered genetic abnormalities that were not in the BRAF gene. These mutations were identified in mostly cytologically negative (i.e. benign) samples.

## 5. CONCLUSIONS

With the help of our studies, we succeeded in developing a new diagnostic procedure based on new generation sequencing that is specific to examining thyroid genes. This procedure is the first in its category, and the most extensive gene panel in Europe. The results of sequencing were analyzed with software and a new statistical method, respectively. Because of this, our method can forecast the risk of malignization of a thyroid gland nodule with a big positive predictive value and sensitivity.

The importance of the BRAF driver gene (its importance in thyroid gland tumor genesis had been previously recognized) have been supported by our results. However, other genes (previously thought to be less important), were included in our study and came out useful as well. Several new, previously unknown mutations were identified in the genes included in our panel. These contribute to the discriminative power of our method.

Methods used routinely in everyday clinical practice could provide appropriate diagnosis in most cases. However, samples in the ambiguous category require further examination, additional sampling, and finally, surgical intervention (based on the current standards). If the study methods were paired with molecular genetic diagnostic techniques, it could provide additional information to the

citology results; it could decrease the number of ambiguous cases; furthermore, with the recognition of the accurate genetic background of cytologically benign abnormalities we could estimate their risk of malignization.

Our diagnostic gene panel was optimized on samples stored in our thyroid biobank with known histological backgrounds. Within the course of the study on these histological samples, we managed to make an accurately adaptive method based on NGS. The gene panel used on FNAB samples proved to be adequetly extensive. The results show that the method provides an adequate extension to decide in ambiguous cases; moreover, it can provide additional information about citologically benign samples.

## 6. LIST OF MY OWN PUBLICATIONS (published in Hungarian)

### 6.1 Publications related to the dissertation

Kocsis-Deák Barbara, dr. Balla Bernadett, dr. Tóbiás Bálint, Árvai Kristóf, dr. Putz Zsuzsanna, dr. Kósa János, dr. Lakatos Péter: *Molekuláris genetikai vizsgálatok a pajzsmirigy daganatainak diagnosztikájában.* Orvostovábbképző szemle, XXV. évf. 9. szám, 2018/09
IF: --

Balla Bernadett, Kocsis-Deák Barbara, Kósa János, Árvai Kristóf, Tobiás Bálint, Takács István, Lakatos Péter: *Az újgenerációs molekuláris diagnosztikai módszerek lehetőségei az endokrinológiában*. Magyar Tudomány, 2019/05
IF: --

Kocsis-Deák Barbara, dr. Balla Bernadett, Árvai Kristóf, dr. Tobiás Bálint, dr. Győri Gabriella, dr. Járay Balázs, dr. Székely Eszter, Podani János, dr. Kósa János, dr. Lakatos Péter: *A pajzsmirigygöbök genetikai vizsgálata újgenerációs szekvenáláson alapuló platformon kifejlesztett génpanel segítségével.* Orvosi Hetilap, 2019/36
IF: 0.564

Barbara Kocsis-Deák, Kristóf Árvai, Bernadett Balla, Bálint Tóbiás, Andrea Kohánka, Balázs Járay, János Horányi, János Podani, István Takács, Zsuzsanna Putz, János Kósa, Péter

Lakatos: *Targeted mutational profiling and a powerful risk score as additional tools for the diagnosis of papillary thyroid cancer*. Pathology & Oncology Research, 2019 Nov 22. Doi: 10.1007/s12253-019-00772-4.
IF: 2.433

### 6.2 Publications not related to the dissertation

Edina Nemes, Katalin Farkas, Barbara Kocsis-Deák, Andrea Drubi, Adrienn Sulák, Kornélia Tripolszki, Piroska Dósa, Ferenc Lakatos, Nikoletta Nagy, Márta Széll*: Phenotypic diversity of patients with LEOPARD syndrome carrying the worldwide recurrent p.Tyr279Cys PTPN11 mutation.* Archives of Dermatological Research, 2015/12
IF: 2.146

Katalin Farkas, Barbara Kocsis-Deák, Laura Cubells Sánchez, Ana Mercedes Victoria Martínez, János Varga, Alfredo Montoro Botella: *The CYLD p.R758X worldwide recurrent nonsense mutation detected in patients with multiple familial trichoepithelioma type 1, Brooke-Spiegler syndrome and familial cylindromatosis represents a mutational hotspot in the gene.* International Journal of Dermatology, 2016/02
IF: 2.666

Katalin Karászi, Szilvia Szabó, Kata Juhász, Péter Király, Barbara Kocsis-Deák, Beáta Hargitai, Tibor Krenács,

Petronella Hupuczi, Offer Erez, Zoltán Papp, Ilona Kovalszky, Nándor Gábor Than: *Increased placental expression of Placental Protein 5 (PP5) / Tissue Factor Pathway Inhibitor-2 (TFPI-2) in women with preeclampsia and HELLP syndrome: Relevance to impaired trophoblast invasion?* Placenta, 2019/01
IF: 2.733

Bernadett Balla, Sárvári M, János Kósa, Barbara Kocsis-Deák, Bálint Tobiás, Kristóf Árvai, István Takács, János Podani, Liposits Z, Péter Lakatos: *Long-term selective estrogen receptor-beta agonist treatment modulates gene expression in bone and bone marrow of ovariectomized rats.* J Steroid Biochem Mol Biol, 2019/04
IF: 3.785

Kósa János, Balla Bernadett, Kocsis-Deák Barbara, Árvai Kristóf, Tobiás Bálint, Takács István, Lakatos Péter: *Tumordiagnosztika vérből – A folyadékbiopszia*. Magyar Tudomány, 2019/05
IF: --