Heliyon 7 (2021) e06408

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

DNA repair pathway activation features in follicular and papillary thyroid tumors, interrogated using 95 experimental RNA sequencing profiles

Uliana Vladimirova [a,e], Pavel Rumiantsev [b,e], Marianna Zolotovskaia [e], Eugene Albert [c], Aleksander Abrosimov [b], Konstantin Slashchuk [b], Petr Nikiforovich [b], Olga Chukhacheva [b], Nurshat Gaifullin [f], Maria Suntsova [a], Galina Zakharova [b], Alexander Glusker [a], Daniil Nikitin [c,g], Andrew Garazha [c], Xinmin Li [h], Dmitriy Kamashev [a], Alexei Drobyshev [a], Irina Kochergina-Nikitskaya [g], Maxim Sorokin [a,c,d], Anton Buzdin [a,c,d,g,*]

[a] I.M. Sechenov First Moscow State Medical University, Moscow, 119991, Russia
[b] Endocrinology Research Centre, Moscow, 117312, Russia
[c] Omicsway Corp., Walnut, CA, 91789, USA
[d] Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, 141701, Russia
[e] Pirogov Russian National Research Medical University, Moscow, 117997, Russia
[f] Faculty of Fundamental Medicine, Lomonosov Moscow State University, Moscow, 119992, Russia
[g] Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, 117997, Russia
[h] Department of Pathology and Laboratory Medicine, University of California, Los Angeles, CA, 90095, USA

ARTICLE INFO

Keywords:
DNA repair
Follicular thyroid adenoma
RNA sequencing
Thyroid cancer
Personalized medicine
Molecular biomarkers
Molecular pathway analysis
Radiation iodine therapy resistance
Papillary thyroid cancer
Systems biology

ABSTRACT

DNA repair can prevent mutations and cancer development, but it can also restore damaged tumor cells after chemo and radiation therapy. We performed RNA sequencing on 95 human pathological thyroid biosamples including 17 follicular adenomas, 23 follicular cancers, 3 medullar cancers, 51 papillary cancers and 1 poorly differentiated cancer. The gene expression profiles are annotated here with the clinical and histological diagnoses and, for papillary cancers, with *BRAF* gene V600E mutation status. DNA repair molecular pathway analysis showed strongly upregulated pathway activation levels for most of the differential pathways in the papillary cancer and moderately upregulated pattern in the follicular cancer, when compared to the follicular adenomas. This was observed for the BRCA1, ATM, p53, excision repair, and mismatch repair pathways. This finding was validated using independent thyroid tumor expression dataset PRJEB11591. We also analyzed gene expression patterns linked with the radioiodine resistant thyroid tumors (n = 13) and identified 871 differential genes that according to Gene Ontology analysis formed two functional groups: (i) response to topologically incorrect protein and (ii) aldo-keto reductase (NADP) activity. We also found RNA sequencing reads for two hybrid transcripts: one in-frame fusion for well-known *NCOA4-RET* translocation, and another frameshift fusion of *ALK* oncogene with a new partner *ARHGAP12*. The latter could probably support increased expression of truncated *ALK* downstream from 4th exon out of 28. Both fusions were found in papillary thyroid cancers of follicular histologic subtype with node metastases, one of them (*NCOA4-RET*) for the radioactive iodine resistant tumor. The differences in DNA repair activation patterns may help to improve therapy of different thyroid cancer types under investigation and the data communicated may serve for finding additional markers of radioiodine resistance.

1. Introduction

About 300 000 new cases of thyroid cancer are recorded globally each year [1]. Thyroid cancer is also the fifth most common cancer in women [1, 2], and it can move to the fourth place during the next decade [3]. Thyroid cancer is a common term for several heterogenous malignancies of thyroid gland, and its clinical manifestation, treatment and prognosis are generally defined by specific tumor subtypes. There are three main types of cells from which thyroid cancers are derived: follicular thyroid cells, Hurthle cells, and parafollicular C cells [4]. The most common

* Corresponding author.
E-mail address: buzdin@oncobox.com (A. Buzdin).

https://doi.org/10.1016/j.heliyon.2021.e06408
Received 4 September 2020; Received in revised form 22 December 2020; Accepted 26 February 2021

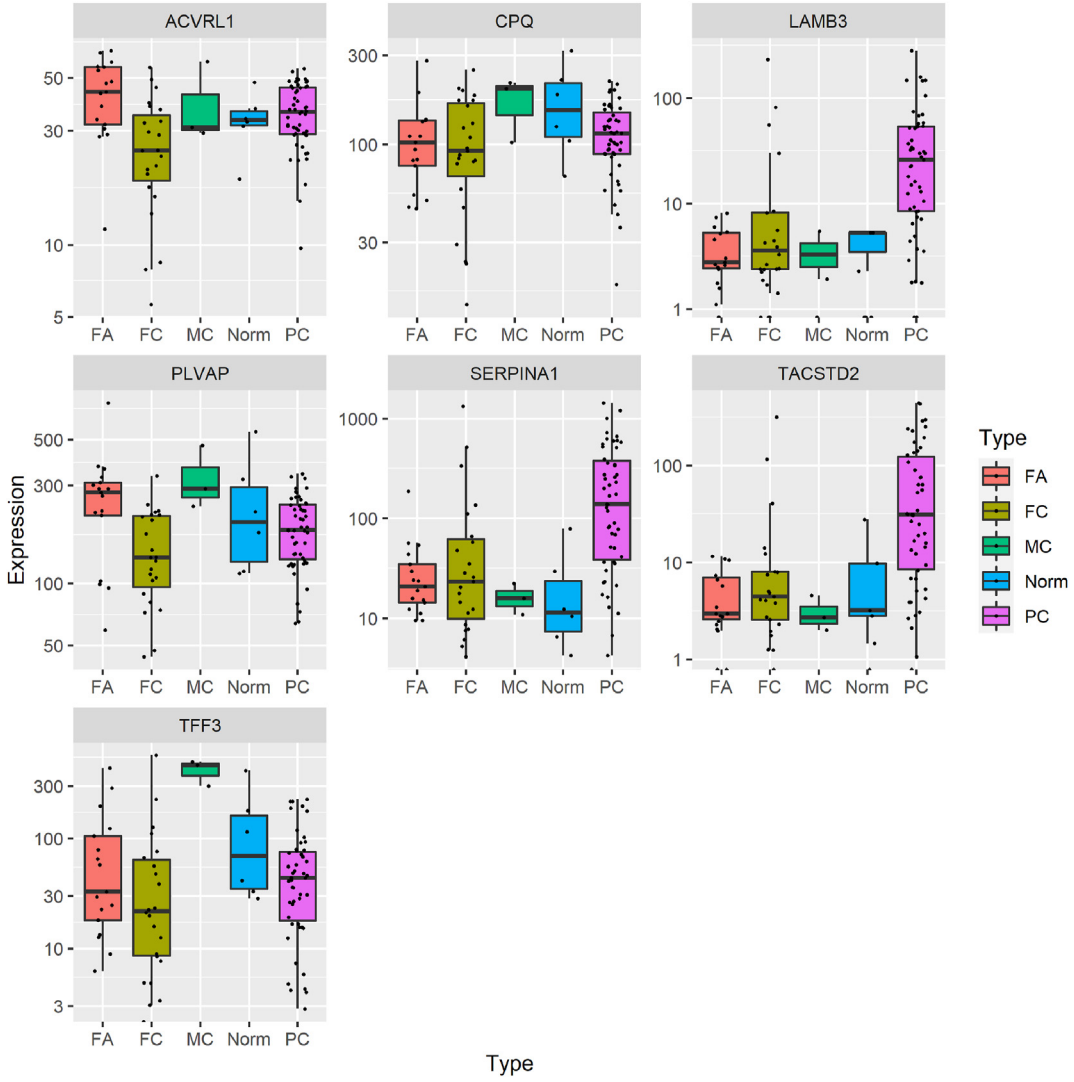


Figure 1. Differential expression of previously reported marker genes in different thyroid tumor types. Genes *SERPINA1, TACSTD2, LAMB3* were reported as the markers upregulated in PC; genes *ACVRL1, PLVAP, TFF3* and *CPQ* were reported as the markers upregulated in FA. Each box shows gene expression distribution in each group, the ends of the box are the upper and lower quartiles, median in shown by horizontal line inside the boxes, whiskers extend from the upper and lower hinges of the boxes towards the highest and lowest values, respectively, at most 1.5 * interquartile range of box. A black point corresponds to a tissue sample. Expression counts were processed as reads per million (RPM).

subtypes of thyroid cancer are follicular thyroid cell-derived tumors, including papillary thyroid cancer (PC), follicular thyroid cancer (FC), and poorly differentiated thyroid cancer [5]. In addition, a small proportion of thyroid malignancies belongs to parafollicular C cell-derived medullary thyroid cancer (MC) [6]. Molecular characterization of thyroid cancer subtypes can partially solve the problem of understanding specific carcinogenesis mechanisms and selecting appropriate treatments.

In addition, resistance of recurrent thyroid cancers to radioactive iodine treatment, confirmed by observations, poses another challenge for early diagnostics of tumors prone not to respond to the standard therapy regimens [7]. Such diagnostics could be used to individually tune treatments, e.g. by adjusting dosage of therapeutic agents and timeline of treatment [8]. Two decades ago, a connection between thyroid cancer radioactive iodine resistance and activity of sodium iodide symporter gene *SLC5A5* was discovered [9]. Also, *BRAF* and *TERT* mutations were found as the predictors of radio iodine refractory thyroid cancer, especially in duet [10, 11, 12]. However, it is clear that the problem has not been solved yet as there are no effective instruments in clinics to predict thyroid cancer resistance to radiation therapy [13].

DNA damage is one of the major causes that trigger the development of tumors [14, 15]. However, it also is an important focus in chemo and radiation therapy [16, 17, 18]. DNA repair mechanisms restore damaged DNA and may have a dual effect on the tumor: they prevent new somatic mutations and may restrict tumor development, but they also restore tumor cells with damaged DNA after chemo and radiation therapy [19]. Thus, when planning the chemo and/or radiation therapy, the possibility of DNA repair must be taken into account [20]. Low activities of DNA repair genes can make tumor cells more sensitive to the therapy, but at the same time not-repaired damaged tumor cells can form more new clones of treatment resistant cells [21, 22]. High activities of certain DNA repair genes, however, can counteract therapy of tumors [22, 23]. Therefore, the selection of relevant therapy should ideally be based on the state of DNA repair in an individual tumor. Recently, defects in DNA repair were shown to play an important role in the clonal evolution of thyroid cancer [24]. Currently, the thyroid cancer subtypes need more detailed characterization in terms of molecular aspects of DNA repair. This is probably due to the apparent shortage of publicly available gene expression data associated with histological diagnoses of thyroid tumors [25, 26, 27].

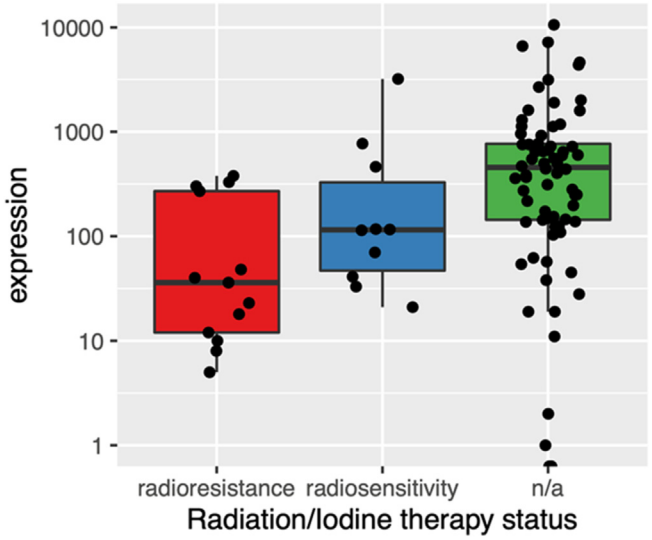


Figure 2. Differential expression of gene *SLC5A5* in radioactive iodine resistant, sensitive and all other thyroid tumors. Expression – DESeq2 normalized read counts. Gene expression values for radioactive iodine resistant tumors (*radioresistance*), sensitive (*radiosensitivity*) thyroid tumors and tumors with undefined radioactive iodine therapy status profiled in this study are boxed. Gene expression is plotted in the log10-trasformed scale. Each box shows gene expression distribution in each group, the ends of the box are the upper and lower quartiles, median in shown by horizontal line inside the boxes, whiskers extend from the upper and lower hinges of the boxes towards the highest and lowest values, respectively, at most 1.5 * interquartile range of box. A black point corresponds to a tissue sample. Drawn using ggplot package in R.

RNA sequencing is currently a method of choice for generating cancer gene expression data [28]. The Genomic Data Commons portal at National Cancer Institute (GDC) [29] provides limited access to 502 RNA sequencing profiles of thyroid cancer samples (accessed at June 30, 2020). Nevertheless, GDC dataset can be used for only a limited number of applications because it contains data for only papillary histotype of thyroid cancer (PC) and does not provide immediate open access to raw reads, thus complicating data analysis and limiting its usefulness.

Therefore, more open-access collections of clinically annotated profiles are needed for different types of malignant and benign thyroid tumors and their histotypes, that would be compatible with the reference data for healthy thyroid tissues.

Here we present the subtype-specific characteristic of DNA repair pathway activation in human thyroid tumors. We performed RNA sequencing on 95 human pathological thyroid biosamples including 17 follicular adenomas, 23 follicular cancers, 3 medullar cancers, 51 papillary cancers and 1 poorly differentiated cancer. The gene expression profiles were annotated with the clinical and histological diagnoses and, for papillary cancers, with *BRAF* gene V600E mutation status. Gene expression data were obtained using the same equipment, reagents, and protocols as for the previously published ANTE database, including six profiles for healthy thyroid tissues [30]. DNA repair molecular pathway analysis showed strongly upregulated pathway activation levels for most of the differential pathways in the papillary cancer and moderately upregulated pattern in the follicular cancer, when compared to the follicular adenomas. This was observed in the BRCA1, ATM, p53, excision repair, and mismatch repair pathways. However, the G2/M checkpoint pathway followed the opposite trend. We also analyzed gene expression patterns linked with the radioiodine resistant thyroid tumors (n = 13) and identified 871 differential genes that according to Gene Ontology analysis formed two functional groups: (i) response to topologically incorrect protein and (ii) aldo-keto reductase (NADP) activity. We also found RNA sequencing reads for two hybrid transcripts: one in-frame fusion for well-known *NCOA4-RET* translocation, and another frameshift fusion of *ALK* oncogene with a new partner *ARHGAP12*. The latter could probably support increased expression of truncated *ALK* downstream from 4th exon out of 28. The differences in DNA repair activation patterns may help to improve therapy of different thyroid cancer types under investigation and the data communicated may serve for finding additional markers of radioiodine resistance.

2. Results

2.1. Expression of known molecular mRNA markers of thyroid cancer

Overall, 51 papillary thyroid cancer (PC), 23 follicular thyroid cancer (FC), 3 medullary thyroid cancer (MC), 1 poorly differentiated thyroid cancer (PDC), and 17 follicular thyroid adenoma (FA) samples were investigated. Clinical annotations and quality control metrics of tumor specimens investigated are summarized in the Supplementary files 1 and 2.

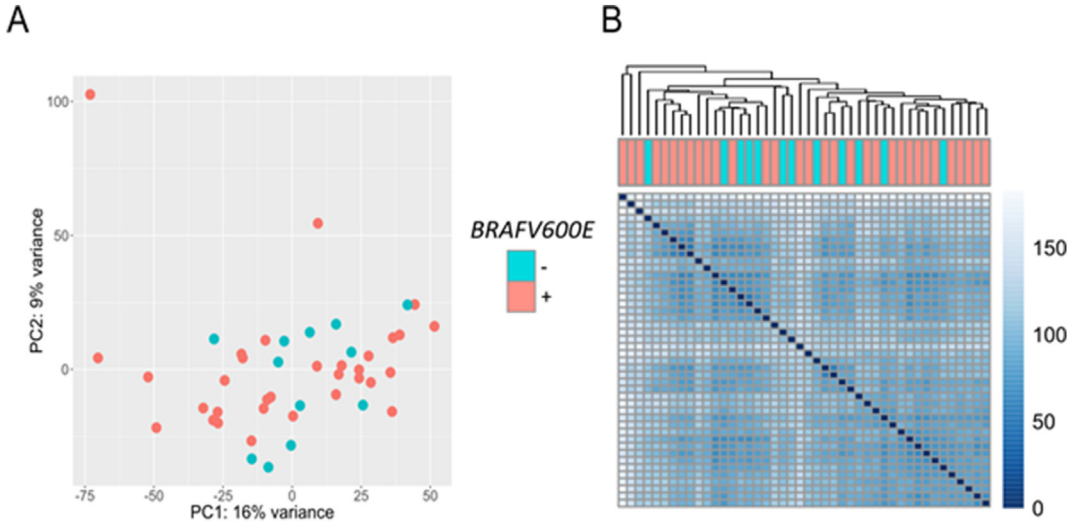


Figure 3. *BRAF* V600E mutation status in PC samples. A – plot for principle component analysis in normalized gene expression space for all PC profiles investigated. B – hierarchical clustering based on Euclidian distance in gene space in relation to *BRAF* V600E mutation status. Drawn using ggplot package in R, clustering algorithm "ward.d2". Expression counts were normalized with DESeq2.

Table 1. Alterations of pathway activation levels for 38 DNA repair pathways in differential comparisons of follicular adenoma (FA), follicular cancer (FC), and papillary cancer (PC) tissues.

Pathway ID	Pathway name	PC vs FC[1,2]	FC + PC vs FA[2,3]	PC vs FA[2,4]
p1	ATM Pathway	↑, 2.8e-02	↑[5], 4.1e-02	↑, 3.6e-03
p2	ATM Pathway Cell Survival	↑, 1.5e-02	↑[5], 6.8e-02	↑[5], 8.7e-03
p3	ATM Pathway G2-Mitosis progression	↑, 1.4e-02	↓, 9.8e-01	↑, 3.3e-01
p4	Biocarta atm signaling pathway	↑[5], 8.3e-01	↑, 4.0e-01	↑, 4.7e-01
p5	Biocarta cell cycle G2M checkpoint pathway	↓, 2.8e-02	↓[5], 4.1e-02	↓[5], 8.1e-04
p6	BRCA1 Pathway	↑, 1.0e+00	↑[5], 4.2e-02	↑[5], 2.8e-02
p7	BRCA1 Pathway Chromatin Remodeling	↑[5], 1.5e-01	↑, 1.3e-01	↑, 9.8e-02
p8	BRCA1 Pathway Homologous Recombination Repair	↓, 1.1e-01	↑, 8.5e-02	↑, 1.9e-01
p9	BRCA1 Pathway Mismatch Repair	↓, 9.8e-02	↑, 1.7e-01	↑, 4.1e-01
p10	DNA Repair Mechanisms Pathway	↑, 5.6e-01	↑, 8.5e-02	↑, 6.4e-02
p11	KEGG Base excision repair pathway	↑, 2.8e-02	↑, 8.5e-02	↑, 2.0e-02
p12	KEGG Fanconi anemia pathway	↓, 1.7e-02	↑, 8.7e-01	↓, 4.8e-01
p13	KEGG Homologous recombination pathway	↑, 2.5e-01	↑, 8.5e-02	↑, 3.4e-02
p14	KEGG Mismatch repair pathway	↓, 1.0e+00	↑, 1.4e-01	↑, 1.9e-01
p15	KEGG Non homologous end joining pathway	↑, 1.0e+00	↑, 1.4e-01	↑, 1.9e-01
p16	KEGG Nucleotide excision repair pathway	↓, 5.6e-01	↑, 1.0e-01	↑, 1.6e-01
p17	Mismatch Repair in Eukaryotes Pathway	↓, 7.7e-01	↑, 8.5e-02	↑, 9.8e-02
p18	NCI ATM pathway	↑, 2.5e-01	↑[5], 4.1e-02	↑[5], 1.5e-02
p19	NCI ATM Pathway (G1 S transition checkpoint)	↑, 6.5e-01	↑[5], 4.1e-02	↑[5], 1.6e-02
p20	NCI ATR signaling pathway	↓, 2.1e-01	↑, 1.4e-01	↑, 2.7e-01
p21	NCI ATR signaling Pathway (Pathway negative regulation of transcription during mitosis via CHEK1)	↓, 2.1e-01	↑, 1.7e-01	↑, 3.2e-01
p22	NCI ATR signaling Pathway (regulation of double strand break repair via homologous recombination)	↓[5], 1.8e-01	↑, 1.2e-01	↑, 2.2e-01
p23	NCI ATR signaling Pathway (response to G2 M transition DNA damage checkpoint signal)	↓, 2.1e-01	↑, 1.8e-01	↑, 3.4e-01
p24	NCI DNA PK pathway in nonhomologous end joining Pathway (double strand break repair via nonhomologous end joining)	↑, 5.6e-01	↑, 1.4e-01	↑, 1.4e-01
p25	NCI DNA PK pathway in nonhomologous end joining Pathway (V D J recombination)	↑, 1.0e+00	↑, 1.7e-01	↑, 2.2e-01
p26	NCI Fanconi anemia pathway	↓[5], 1.5e-01	↑, 1.1e-01	↑, 2.2e-01
p27	NCI Fanconi anemia Pathway (regulation of double strand break repair via homologous recombination)	↓[5], 1.5e-01	↑, 1.0e-01	↑, 2.0e-01
p28	NCI Fanconi anemia Pathway (Sister Chromatid Exchange Process)	↓, 2.1e-01	↑, 1.7e-01	↑, 3.1e-01
p29	NHEJ mechanisms of DSBs repair effect	↓[5], 2.8e-01	↑, 1.4e-01	↑, 2.6e-01
p30	Nucleotide excision repair effect	↑, 1.1e-01	↑, 5.0e-02	↑, 1.1e-02
p31	p53 Signaling Pathway	↑, 2.0e-03	↑, 4.1e-02	↑, 8.1e-04
p32	p53 Signaling Pathway DNA Repair	↑, 5.3e-01	↑, 1.4e-01	↑, 9.8e-02
p33	p53 Signaling Pathway Gene Expression DNA Replication and Repair via TP53	↑, 7.2e-02	↑, 9.1e-02	↑, 3.4e-02
p34	reactome Fanconi Anemia pathway	↑, 7.7e-01	↑, 1.1e-01	↑, 9.8e-02
p35	Reactome Formation of transcription coupled NER TC NER repair complex pathway	↑, 1.0e+00	↑, 1.7e-01	↑, 2.2e-01
p36	Reactome Mismatch repair MMR directed by MSH2, MSH3, MutSbeta pathway	↑, 1.3e-01	↑, 8.5e-02	↑, 3.4e-02
p37	Reactome Mismatch repair MMR directed by MSH2, MSH6, MutSalpha pathway	↑, 9.8e-02	↑, 6.1e-02	↑, 1.6e-02
p38	Reactome Repair synthesis for gap filling by DNA polymerase in TC NER pathway	↑, 5.0e-01	↑[5], 1.0e-01	↑[5], 7.4e-02

[1] Comparison of papillary cancer versus follicular cancer tissues.
[2] q-value (FDR-adjusted p-value) for Wilcoxon test is given for every molecular pathway PAL in every comparison. Sign "↑" means that the first comparative group has higher mean PAL than the second, sign "↓" denotes the opposite situation.
[3] Comparison of follicular and papillary cancer tissues versus follicular adenoma.
[4] Comparison of papillary cancer versus follicular adenoma tissues.
[5] The coefficient of quartile variation (CQV) is higher for compared groups together than for each separate group.

To initially characterize molecular data presented here we analyzed expression levels of several mRNA molecules previously reported to be differentially regulated in thyroid tumors. For example, *SERPINA1* [30], *LAMB3* [31], and *TACSTD2* [32] genes are known to be overexpressed in PC, which was also the case in our dataset (Figure 1, Supplementary file 3). On the other hand, in line with the overall mediocre concordance between FC expression biomarkers reported in the previous studies, we did not observe any differences between FA and FC samples for two out of four marker genes proposed in a recent meta-analysis paper [25]: *CPQ* and *TFF3*. However, the other two other marker genes *ACVRL1* and *PLVAP* demonstrated upregulated trends in FA versus other tumors (Figure 1, Supplementary file 3).

We also compared expression of sodium iodide symporter gene *SLC5A5* in the biosamples from tumors, which were resistant to radioactive iodine therapy, with all other thyroid tumors profiled in this study. In line with the previous data [33, 34, 35], in our dataset *SLC5A5* was statistically significantly downregulated ($p = 3.44*10^{-3}$) in radioactive iodine-resistant tumor samples (Figure 2).

Thyroid hormone biosynthesis requires active hydrogen peroxide production which may also be responsible for the high level of oxidative DNA damage. The NADPH oxidase/dual oxidase family members are the only known oxidoreductases whose primary function is production of reactive oxygen species, and their increased expression levels were previously shown in thyroid cancer [36]. We compared expressions of *NOX1, NOX3, NOX4, NOX5, DUOX1,* and *DUOX2* genes in our experimental FA, FC, and PC samples and found no significant differences between the above thyroid tumor types (data not shown).

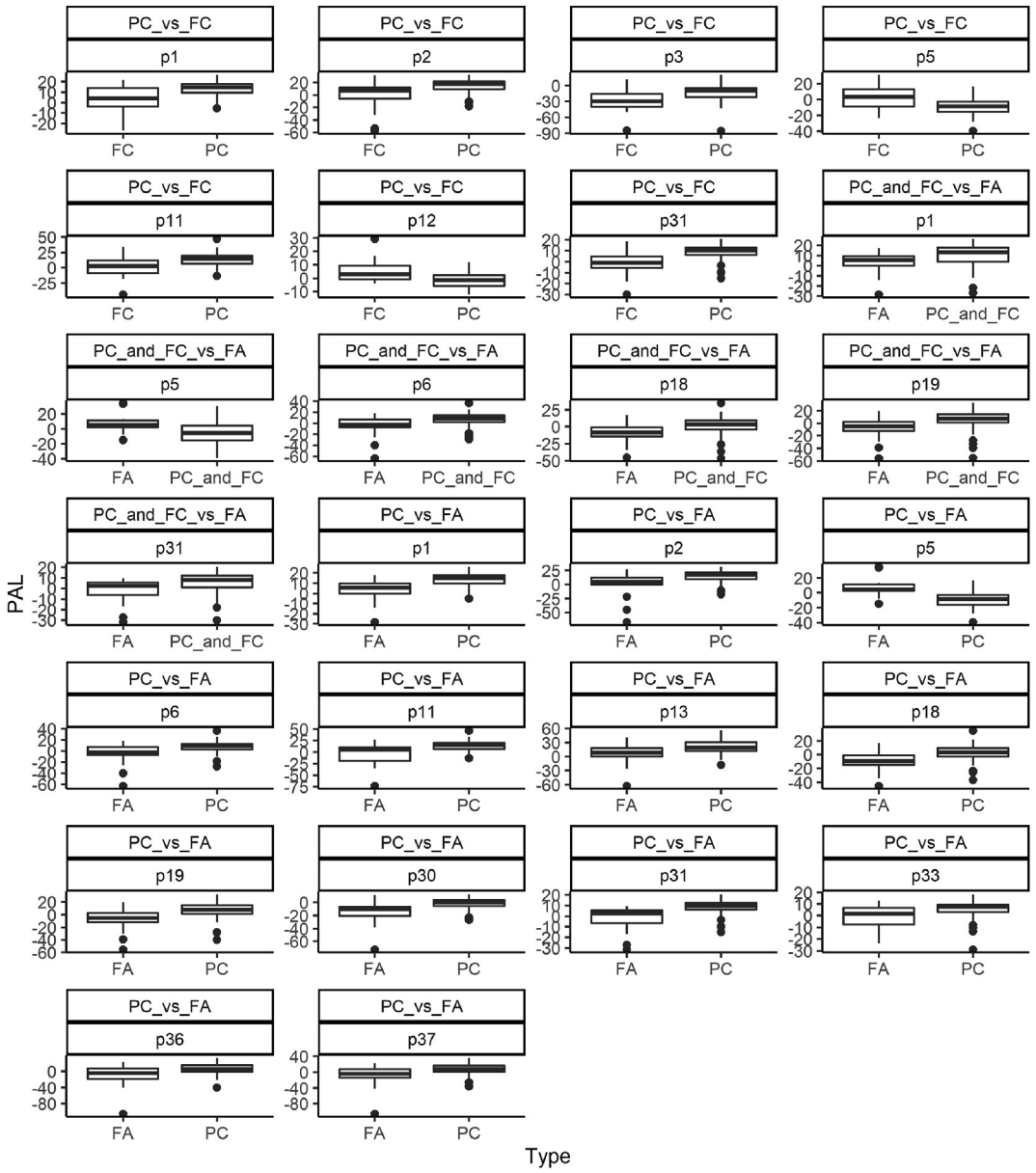


Figure 4. Boxplot representation of PAL values for the differential pathways corresponding to comparisons from Table 1. Pathway name abbreviations: p1 - ATM Pathway, p2 - ATM Pathway Cell Survival, p3 - ATM Pathway G2-Mitosis progression, p5 - Biocarta cell cycle G2M checkpoint pathway, p6 - BRCA1 Pathway, p11 - KEGG Base excision repair pathway, p12 - KEGG Mismatch repair pathway, p13 - KEGG Nucleotide excision repair pathway, p18 - NCI ATM pathway, p19 - NCI ATM Pathway (G1 S transition checkpoint), p30 - Nucleotide excision repair effect, p31 - p53 Signaling Pathway, p33 - p53 Signaling Pathway Gene Expression DNA Replication and Repair via TP53, p36 - Reactome Mismatch repair MMR directed by MSH2, MSH3, MutS beta pathway, p37 - Reactome Mismatch repair MMR directed by MSH2, MSH6, MutS alpha pathway.

2.2. *BRAF (V600E) mutation status*

In PC patients, *BRAF* (V600E) mutation status is an important diagnostic and prognostic biomarker for tumor aggressiveness and relapse [37]. We therefore screened all PC specimens for this mutation and identified 15 mutant and 27 wild-type cases (Supplementary file 1). Notably, neither PCA nor Euclidean distance-based hierarchical clustering identified any specific pattern associated with mutated or wild-type status *BRAF* V600E in the tumor samples (Figure 3). We found no differentially expressed genes with $q < 0.1$ by Wilcoxon test for the groups of *BRAF* (V600E) mutation positive and negative PCs.

2.3. *Activities of DNA repair pathways in different types of thyroid tumors*

We then attempted to evaluate the differences in the activities of DNA repair pathways between the normal thyroid samples, benign tumors, and thyroid cancer subtypes.

Pathway activation level (PAL) is an integral parameter that characterizes activation of a molecular pathway using high-throughput gene expression data [38, 39]. It utilizes aggregation of logarithmic fold-change expression levels (case compared with control samples) for the gene products involved in a certain molecular pathway, using functional flags indicating specific activator or repressor roles of a gene product in the molecular pathway [38]. Positive PAL value indicates

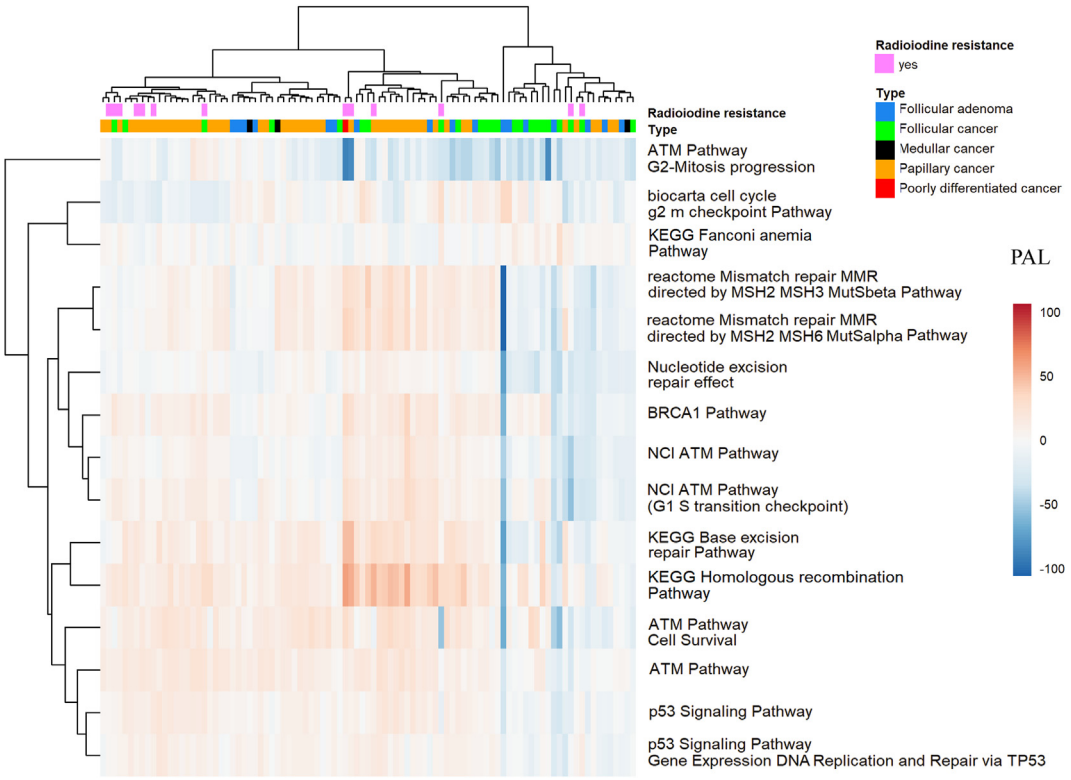


Figure 5. Heatmap of pathway activation levels for 15 differential DNA repair pathways in all thyroid tumor samples investigated.

upregulation of the molecular pathway, negative value – its down-regulation, whereas zero means unaffected level compared with the controls [38].

For our assay, we selected 38 DNA repair pathways, which were obtained by manual curation of Reactome [40], NCI Pathway Interaction [41], Kyoto Encyclopedia of Genes and Genomes [42], Biocarta [43], and Qiagen [44] databases, and which include at least ten gene products. For these pathways, we then calculated PAL values for the tumor samples while taking six healthy thyroid tissue gene expression profiles from ANTE database [45] as the controls. Then we compared PAL values for the selected DNA repair pathways between the normal samples, benign tumors, and the thyroid cancer subtypes. In total, we performed nine comparisons to assess differential PAL activation trends: malignant and benign (PC + FC + FA) were compared with the normal samples (*i*); follicular adenoma (FA) was compared with the normal samples (*ii*); FC and PC were compared with normal samples (*iii*); FC and PC were compared with FA (*iv*); FC was compared with PC (*v*), FA was compared with FC (*vi*); PC was compared with FA (*vii*), PC was compared with the normal samples (*viii*); and FC was compared with the normal samples (*ix*).

We found statistically significantly differentially regulated DNA repair pathways in only three out of these nine comparisons (Table 1, Supplementary File 4). In particular, six pathways were differential in the comparison (*iv*) FC + PC vs FA; four pathways were differential in the comparison (*v*) FC vs PC; twelve pathways were differential in the comparison (*vii*) PC vs FA (Table 1).

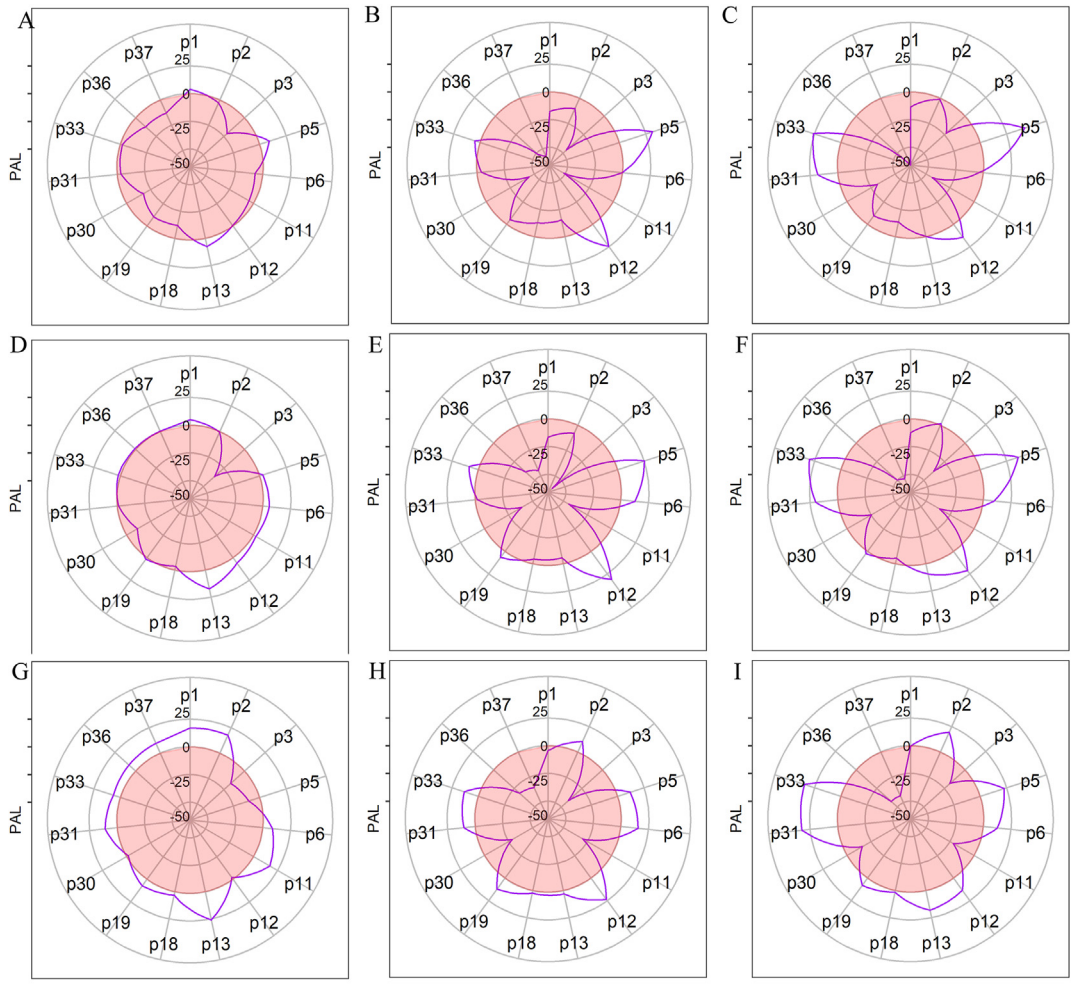
The boxplot presentations of PAL values for all differential pathway comparisons are shown in Figure 4. We also assessed variation of each group under comparison separately and together, both on pathway (Table 1, Supplementary file 4) and gene levels (Supplementary file 5). We calculated the following statistical characteristics: mean, standard deviation, interquartile range (IQR), Range (the difference between maximum and minimum), ratio of IQR and Range, the coefficient of quartile variation (CQV), confidence interval (CI) for CQV (Supplementary file 4). At the level of pathway activation, CQV was calculated for (PAL-minimal (PAL)) because PAL can take negative values [46]. We marked the pathways, for which CQV of two compared groups together was higher than CQV of each separate group (Table 1, Supplementary file 4), there were six such differential pathways (Table 1): *ATM Pathway, ATM Pathway Cell Survival, BRCA1 Pathway, NCI ATM pathway, NCI ATM Pathway (G1 S transition checkpoint), Biocarta cell cycle G2M checkpoint pathway*.

We then applied the same comparison parameters for individual DNA repair pathway member genes to identify genes which contribute to differential regulation of the pathways (Supplementary file 5). For example, Biocarta cell cycle G2M checkpoint pathway had seven differentially expressed genes for PC vs FA comparison (*ATR, CCNB1, CDKN2D, MDM2, PI4KA, RASGRF1, SFN*), Supplementary file 5. Most of differential genes were up-regulated in PC (Supplementary file 5) which resulted in up-regulation of most of the pathways. However, *Biocarta cell cycle G2M checkpoint pathway* was down-regulated in PC because most of the differential genes have negative activator/repressor role coefficients (ARR) for this pathway [42, 43, 47].

In total, 15 individual DNA repair pathways were differential in at least one comparison. Among them, five pathways were differential simultaneously in two comparisons, and three – in three comparisons (Table 1).

While PAL values were positive for certain pathways in PC and FC, the PAL differences between the tumors and normal thyroid samples were not significant. To test that we calculated PAL level for each normal sample (Supplementary file 6). PAL level for a normal sample can be interpreted as an activation level of a pathway in a normal sample when compared to the average of all normal samples. This is needed to test the significance of the differences when comparing the group of normal samples with other groups. Lack of significance between normal samples and tumors is linked with a high heterogeneity of PAL values in normal samples (Supplementary file 4).

Thus, no pathway activation changes distinctive for all tumor types versus normal samples have been found. There were also no differentially regulated patterns between all benign and malignant thyroid neoplasms.



p1 - ATM Pathway,
p2 - ATM Pathway Cell Survival,
p3 - ATM Pathway G2-Mitosis progression,
p5 - biocarta cell cycle g2 m checkpoint Main Pathway,
p6 - BRCA1 Pathway,
p11 - KEGG Base excision repair Main Pathway,
p12 - KEGG Fanconi anemia Main Pathway
p13 - KEGG Homologous recombination Main Pathway,
p18 - NCI ATM Main Pathway,
p19 - NCI ATM Pathway(G1 S transition checkpoint),
p30 - Nucleotide excision repair effect,
p31 - p53 Signaling Pathway,
p33 - p53 Signaling Pathway Gene Expression DNA Replication and Repair via TP53,
p36 - reactome Mismatch repair MMR directed by MSH2 MSH3 MutSbeta Main Pathway,
p37 - reactome Mismatch repair MMR directed by MSH2 MSH6 MutSalpha Main Pathway,

Figure 6. Radar-charts of 15 differential DNA repair pathway activation profile for (A) follicular adenoma, ANTE normal samples used for PAL calculation, (B) follicular adenoma, GTEx normal samples used for PAL calculation, (C) follicular adenoma, TCGA normal samples used for PAL calculation, (D) follicular cancer, ANTE normal samples used for PAL calculation, (E) follicular cancer, CTEx normal samples used for PAL calculation, (F) follicular cancer, TCGA normal samples used for PAL calculation, (G) papillary cancer, ANTE normal samples used for PAL calculation, (H) papillary cancer, GTEx normal samples used for PAL calculation, (I) papillary cancer, TCGA normal samples used for PAL calculation. Positive pathway activation values (PALs) are shown in the outer white area, negative values –in the inner red circle area.

We conclude, therefore, that we could not identify marker DNA repair pathway alterations for thyroid tumorigenesis in general, but several DNA repair pathways were characteristic for follicular and papillary neoplasms.

We then assessed clustering of thyroid tumors according to the PAL values of fifteen DNA repair pathways found to be significantly differential in the previous comparisons (Figure 5). We observed a trend with either PC-, or FC and FA-enriched clusters (Figure 5). Unlike the other cancer types, the PC samples had mostly positive PAL values for the DNA repair pathways, which suggested their overall stronger activation. However, there were also two pathways related to passing through G2 phase of the cell cycle, that were downregulated in most of the thyroid cancer specimens including PC: *ATM Pathway G2-Mitosis progression* and *Biocarta cell cycle g2 m checkpoint pathway*.

To further characterize tumor type-specific pathway activation profiles, we used average PAL values to build radar-charts for the 15 differential DNA repair pathways for the FA, FC, and PC tumor types (Figure 6 A,D,G). We did not consider medullar thyroid cancer and poorly differentiated thyroid cancer because these tumor types did not have enough samples to enable statistically robust investigation. For most of the pathways, this analysis revealed downregulated activation pattern in FA, strongly enhanced pattern in PC, and intermediate pattern in FC (Figure 6, A,D,G). However, the observed activation patterns did not relate to the two pathways associated with passing through the G2

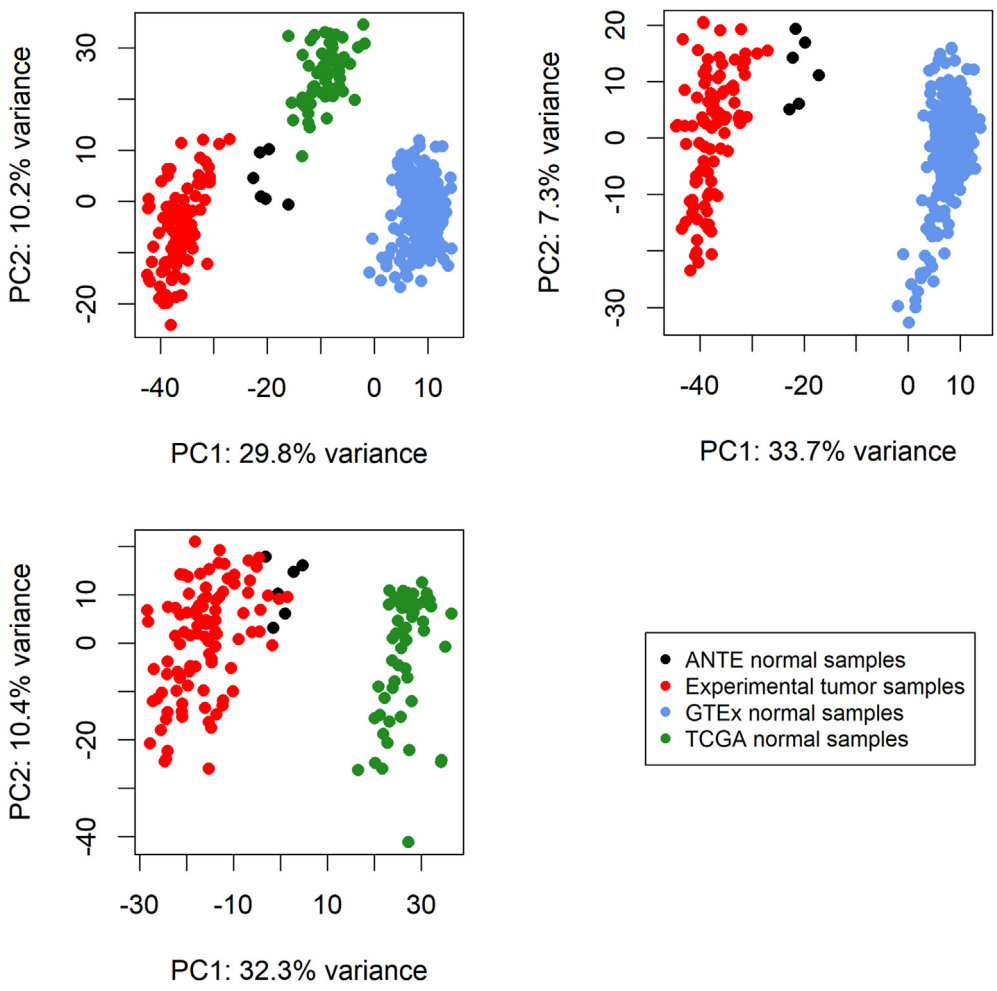


Figure 7. Principal component analysis of gene expression profiles for experimental thyroid tumor samples, and normal thyroid tissue samples from ANTE, GTEx and TCGA databases.

phase of the cell cycle. The *ATM Pathway G2-Mitosis progression* pathway was strongly downregulated in all the tumor types investigated. The *Biocarta cell cycle g2 m checkpoint pathway* showed another trend: it was upregulated in FA and to a lesser extent in FC, and downregulated in PC (Figure 6, A,D,G).

We repeated the above comparative analysis with greater amount of control samples by using normal thyroid RNA sequencing profiles from public databases GTEx (446 samples) and TCGA (58 samples) which showed, however, strong platform-specific bias (Figure 7). Unlike for the ANTE dataset, different equipment, reagents and protocols for RNA isolation, ribosomal RNA depletion, library preparation and sequencing were used for the GTEx and TCGA data. This led to significant changes of the absolute pathway activation levels when calculated using GTEx and TCGA norms (Supplementary File 4). However, the intratumoral comparisons resulted in the same conclusions as before, and we obtained the same 15 differential pathways using both types of new normal datasets (Table 1, Supplementary File 4). Although pathway activation level (PAL) values depended on the normal dataset used, the trends of PAL expression in thyroid tumor types was consistent regardless of the normal set used: minimal overall activation of DNA repair pathways in FA samples (Figure 6 A, B, C), medium in FC samples (Figure 6 D,E,F), and maximal in PC samples (Figure 6 G,H,I).

We then reconstructed the joint molecular networks by combining the differential DNA repair pathways for each of the above three comparisons (Table 1): PC vs FC, PC + FC vs FA, and PC vs FA (Figures 3, 4, and 5, Supplementary file 5). For example, the joined network for the comparison "FC vs PC" included *p53 signaling pathway, KEGG Base excision repair pathway, biocarta cell cycle g2 m checkpoint pathway*, *KEGG Fanconi anemia pathway* and *ATM pathway* with its branches *responsible for cell survival* (*i*) and *passing through G2 stage during cell cycle progression* (*ii*) (Figure 8).

Overall, in PC compared with FC, we detected upregulated *ATM, p53 and base excision repair pathways* and downregulated *KEGG Fanconi anemia and cell cycle g2-m checkpoint pathways* (Table 1). *P53 signaling pathway* is involved in many mechanisms of DNA repair [48, 49, 50]. ATM-related pathways participate in double strand DNA breaks repair [46]. Fanconi anemia pathway coordinate a complex mechanism that enlists elements of three classic DNA repair types: homologous recombination, nucleotide excision repair, and mutagenic translesion synthesis in response to genotoxic insults [16, 47]. Thus, we can highlight that PC compared with FC has more active double strand DNA breaks repair and base excision repair pathways, and also more intensively involves homologous recombination, nucleotide excision repair, and mutagenic translesion synthesis mechanisms.

Furthermore, when comparing PC + FC versus FA, in addition to the upregulated *ATM and p53 pathways* we also detected upregulated *BRCA1 pathway* and *NCI ATM pathway* with its branch *G1 S transition checkpoint.* In turn, the *Cell cycle G2M checkpoint pathway* was downregulated in PC + FC compared with FA (Figure 9, Table 1). *G1 S transition checkpoint* branch of *NCI ATM pathway* was upregulated, unlike the G2M checkpoint pathway. The *BRCA1 pathway* is responsible for the regulation of DNA damage-induced cell cycle checkpoints and controls multiple DNA repair processes, including homologous recombination, non-homologous end-joining, and single-strand annealing [51, 52]. Thus, we conclude that PC

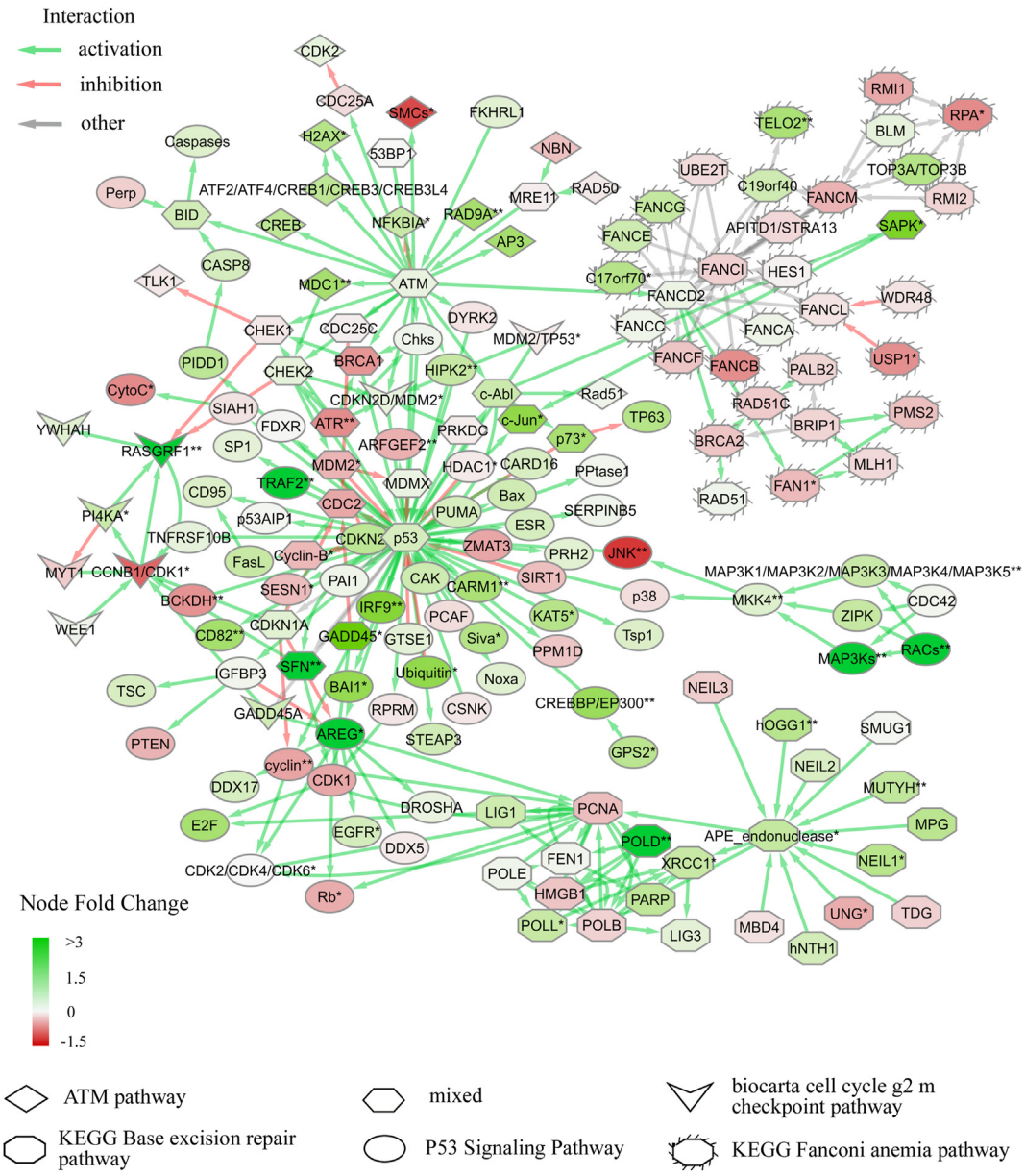


Figure 8. Reconstructed differential DNA repair molecular network for the comparison of Papillary cancer (PC) vs Follicular cancer (FC). Node Fold Change for each node was calculated as the sum of fold changes for the genes which were included in the node. Fold change for each gene was calculated by taking logarithm base 2 of ratio of geometric means of gene expression in compared groups. Green nodes represent genes that are up-regulated in PC (down-regulated in FC); red nodes - genes that are down-regulated in PC (up-regulated in FC). Arrow color indicates type of interaction. Shape of a node reflects source molecular pathway(s). Asterisks (*) indicate nodes which include differential genes (PC vs FC) with q-value <0.05; double asterisks (**) - nodes which include differential genes with $q < 0.01$. ATM pathway represented here also includes its differentially regulated branches that can be considered as separate pathways: *ATM Pathway Cell Survival* and *ATM Pathway G2-Mitosis progression*.

+ FC have higher activation levels of double strand DNA breaks repair, if compared with FA.

Finally, in the comparison of PC vs FA we observed the greatest number of differential pathways. The previously mentioned pathways (*ATM pathway* with its branch *Cell Survival, p53 pathway, BRCA1 pathway, Biocarta cell cycle G2M checkpoint pathway, KEGG Base excision repair pathway*) followed the same regulation trends, but there were also several new pathways. These were *KEGG Homologous recombination pathway*, *Nucleotide excision repair effect pathway, Reactome Mismatch repair MMR directed by MSH2, MSH6, MutSalpha pathway*, *Reactome Mismatch repair MMR directed by MSH2, MSH3, MutS beta pathway*, and specific branch of p53 pathway *Gene Expression DNA Replication and Repair via TP53* (Figure 10, Table 1). All the pathways were upregulated in PC compared with FA except for the cell cycle G2M checkpoint pathway, which was downregulated. We, therefore, conclude that PC, if compared with FA, had more active double strand DNA breaks repair, nucleotide excision and mismatch repair mechanisms, but less active G2M checkpoint transition pathway.

We then analyzed gene compositions of the reconstructed comparison-specific DNA repair networks and found statistically significantly differential genes (according to Wilcoxon test) that may represent the key regulated pathway nodes. We adjusted Wilcoxon *p*-value by Benjamini-Hochberg false discovery rate test and obtained the differential gene set with *q*-value less than 0.05 for each reconstructed pathway: totally 68, 73, and 91 genes for the three comparisons, respectively (Supplementary file 5). The gene compositions in these three lists were highly redundant and totally represented 96 individual genes. Top-10 of these differential genes sorted by the fold change value are shown on Table 2.

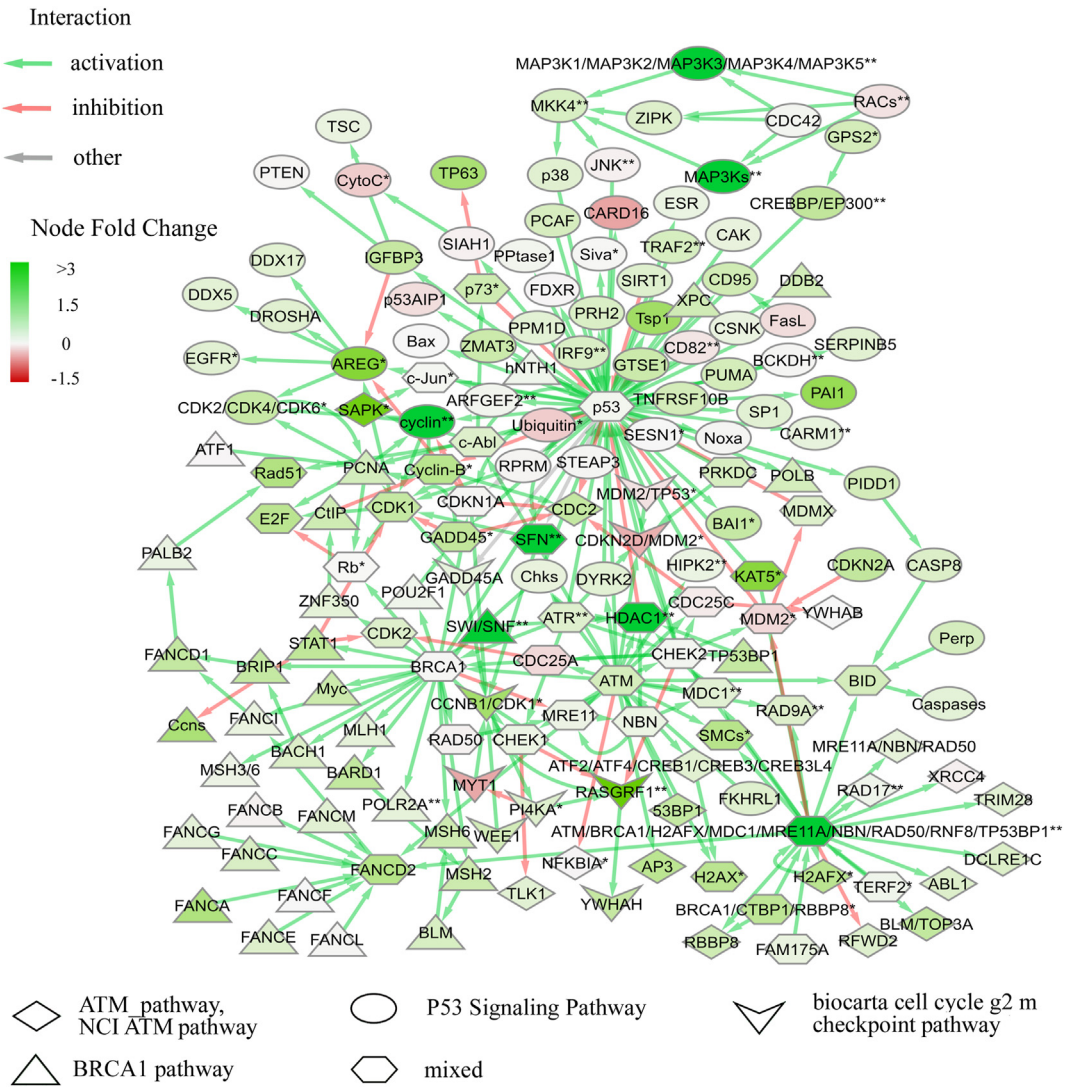


Figure 9. Reconstructed differential DNA repair molecular network for the comparison of Follicular and Papillary cancers (FC + PC) vs Follicular adenoma (FA). Node Fold Change for each node was calculated as the sum of fold changes of the genes which were included in the node. Fold change for each gene was calculated by taking logarithm base 2 of ratio of geometric means of gene expression in compared groups. Green nodes represent genes that are up-regulated in FC + PC (down-regulated in FA); red nodes – genes that are down-regulated in FC + PC (up-regulated in FA). Arrow color indicates type of interaction. Shape of a node reflects source molecular pathway(s). Asterisks (*) indicate nodes which include differential genes (FC + PC vs FA) with q-value <0.05; double asterisks (**) - nodes which include differential genes with q < 0.01. NCI ATM pathway represented here also includes its differentially regulated branch that can be considered as separate pathway - *G1 S transition checkpoint.*

In some cases, the transcriptional activity of genes can be unrelated to the corresponding protein levels due to specific translation regulation mechanisms. To investigate whether transcriptional changes reflect DNA repair effector protein concentrations in cancer tissues we used published datasets to assess correlation of expression at the transcriptomic and proteomic levels for four DNA repair effector genes: *ATM, BRCA1, H2AFX*, and *TP53* in different cancers. We found that transcription and protein concentration were statistically significantly positively correlated in most of the cases, which can support adequacy of the transcriptome-based approach to DNA repair pathway analyzes (Figure 11).

2.4. DNA repair pathway activities in thyroid tumors and in adjacent pathologically normal tissues

We explored previously published thyroid tumor gene expression dataset PRJEB11591 [53,54] to validate the activation trend revealed for the activities of DNA repair pathways in three types of thyroid tumors. The literature dataset contained RNA sequencing data for 25 FAs, 30 minimally invasive FCs, 48 follicular variants of PC (FVs), 77 classical papillary thyroid carcinomas (PCs) and 81 adjacent pathologically normal thyroid tissues. The adjacent normal samples corresponded to 18 FAs, 19 FCs, 11 FVs and 33 PCs. As before, we performed differential pathway analyses (for 38 DNA repair pathways) and obtained the statistically significant q-value in at least one comparison for 36 out of 38 DNA repair pathways investigated (Supplementary file 7). We used 81 adjacent normal samples as the controls to calculate PAL of DNA repair pathways.

Using these data we confirmed the previously revealed tendency that the majority of differential DNA repair pathways have minimal activation signature in FA, medium in FC and maximal in PC (Figure 12). FV (follicular variant of PC) had intermediate pathway activation values between FC and PC (Figure 12). DNA repair pathway activation levels in adjacent pathologically normal tissues were very similar to the levels observed in the corresponding tumors (R Pearson = 0.8492, p-value < 2.2e-16, Figure 12). This finding may reflect tumor influence on gene expression in adjacent pathologically normal tissue.

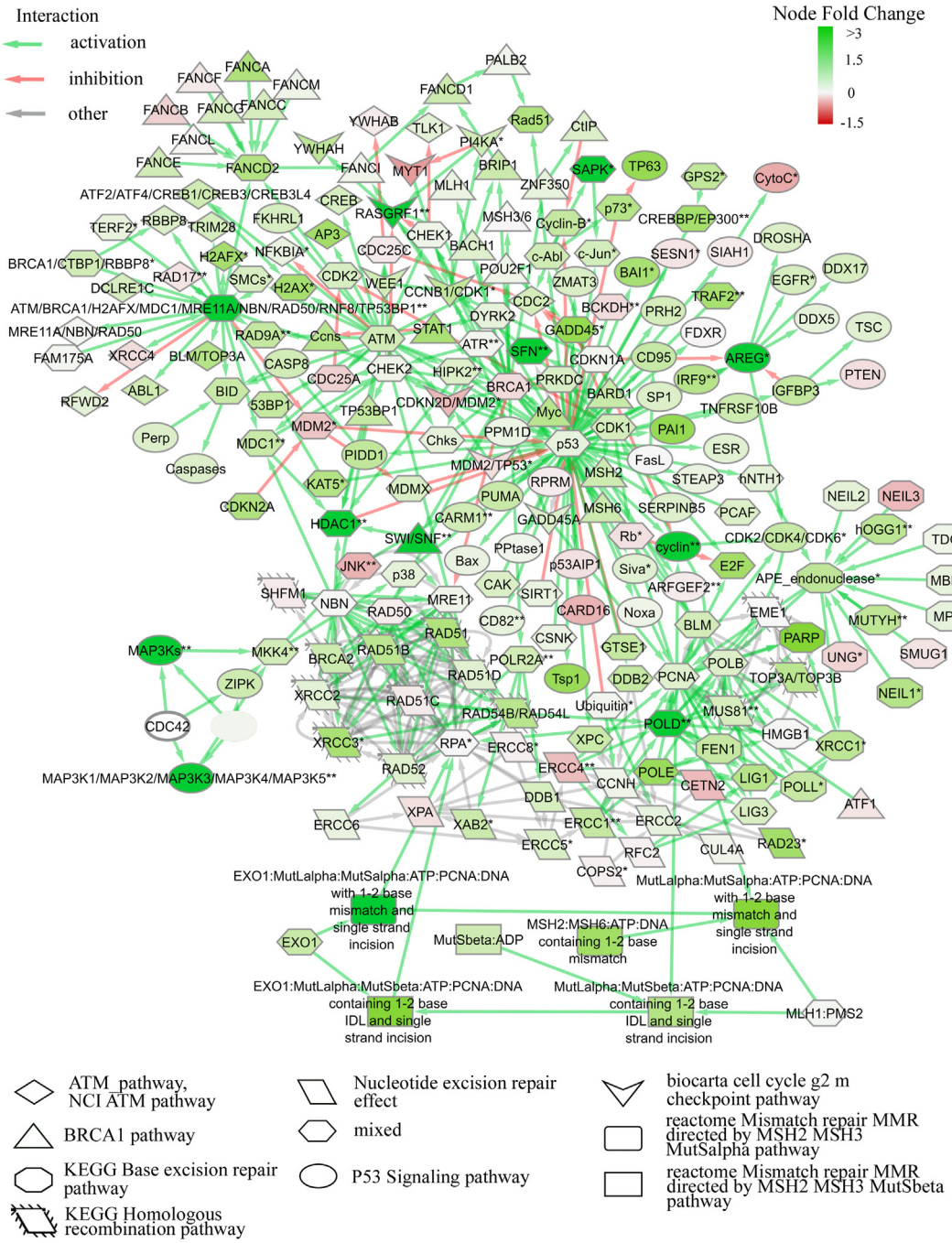


Figure 10. Reconstructed differential DNA repair molecular network for the comparison of Papillary cancer (PC) vs Follicular adenoma (FA). Node Fold Change for each node was calculated as the sum of fold changes of the genes which were included in the node. Fold change for each gene was calculated by taking logarithm base 2 of ratio of geometric means of gene expression in compared groups. Green nodes represent genes that are up-regulated in PC (down-regulated in FA); red nodes – genes that are down-regulated in PC (up-regulated in FA). Arrow color indicates type of interaction. Shape of a node reflects source molecular pathway(s). Asterisks (*) indicate nodes which include differential genes (PC vs FA) with q-value <0.05; double asterisks (**) - nodes which include differential genes with $q < 0.01$. ATM pathway, NCI ATM pathway, and p53 signaling pathway represented here also include their differentially regulated branches that can be considered as separate pathways: *Cell Survival, G1 S transition checkpoint*, and *Gene Expression DNA Replication and Repair via TP53*, accordingly.

2.5. *APOBEC activity in thyroid tumors*

Mutational signature of thyroid cancer specimens suggests involvement of DNA-modifying APOBEC family enzymes [37]. APOBEC family cytidine deaminases convert cytidine to uracil, coupled to activity of the base excision repair and DNA replication processes [37, 38, 39]. We explored expression of *APOBEC* family genes in experimental samples under investigation for thyroid tumor type-specific patterns (Figure 13). Four genes showed tumor type-specific differential expression, which was increased in PC compared to FA and FC: *APOBEC3C, APOBEC3F, APOBEC3G*, and *APOBEC4* (Figure 13A). Furthermore, we also explored *APOBEC* family gene signature which was the sum of logarithmic expression levels of all family members under study: *AICDA, APOBEC2, APOBEC3A, APOBEC3A_B, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4*. The *APOBEC* signature was also differential for the comparisons of FA vs PC, and FC vs PC (Figure 13B). The direction of expression changes of *APOBEC* family gene signature and separate genes *(APOBEC3C, APOBEC3F, APOBEC3G*, and

Table 2. Top 10 differential DNA repair genes for thyroid tumor comparisons.

Gene ID	PC vs FC [1,2]	FC + PC vs FA[2,3]	PC vs FA[2,4]	Gene product function according to GeneCards database [62]
SFN	1.42	1.62	2.06	Cell cycle checkpoint protein, regulates mitotic translation by binding with translation initiation factors
RAC2	1.40	0.37	0.80	Metabolizes small guanosine triphosphate (GTP) and thus regulates secretion, phagocytosis, cell polarization, motility, etc
RASGRF1	1.35	1.27	1.68	Stimulates dissociation of GDP from RAS proteins
IGF2	1.18	0.56	0.93	Insulin growth factor involved in growth and development
IRF9	1.00	-	-	Transcription factor mediating interferon signaling by IFN-alpha and IFN-beta
JUN	0.90	-	-	Transcription factor that recognizes and binds to the enhancer heptamer motif 5′-TGA[CG]TCA-3′
POLD4	0.89	-	-	Enhances the activity of DNA polymerase delta and takes part in replication fork repair and stabilization
ADGRB1	-0.87	-	-	Adhesion G Protein-Coupled Receptor B1 enhances the engulfment of apoptotic cells. Activates the Rho pathway in a G-protein-dependent manner
MAP3K11	0.86	0.42	0.69	Preferentially activates MAPK8/JNK kinase, and functions as a positive regulator of JNK signaling pathway
MAP3K14	0.80	-	-	Binds to TRAF2 and stimulates NF-kappaB activity
ATR	-	0.68	0.86	Serine/threonine kinase and DNA damage sensor, activating cell cycle checkpoint signaling upon DNA stress, phosphorylates and activates several proteins involved in the inhibition of DNA replication and mitosis; promotes DNA repair, recombination, and apoptosis.
CCNB1	-	0.53	0.71	Regulates mitosis by complexing with p34 to form the maturation-promoting factor necessary for proper control of the G2/M transition phase of the cell cycle.
CDKN2D	-	0.40	0.64	Cyclin-dependent kinase inhibitor which forms a stable complex with CDK4 or CDK6, and prevent their activation, thereby regulating cell cycle G1 phase progression.
MDM2	-	0.39	0.64	Nuclear-localized E3 ubiquitin ligase which can promote carcinogenesis by targeting for proteasomal degradation tumor suppressor proteins including P53.
PI4KA	-	0.37	-	Phosphatidylinositol 4-kinase which catalyzes the first step in the biosynthesis of phosphatidylinositol 4-bisphosphate.
H2AFX	-	-	0.69	Variant histone that is phosphorylated as a reaction on DNA double-strand breaks, and replaces conventional H2A histone in nucleosomes

[1] Comparison of PC versus FC tissues.
[2] Gene fold change (log base 2). Sign "-" means that the gene is not included in top-10 of the most up- or downregulated differential genes.
[3] Comparison of FC + PC versus FA tissues.
[4] Comparison of PC versus FA tissues.

APOBEC4) coincides with direction of expression changes in DNA repair pathways (expression is higher in PC than in FC and FA).

2.6. Gene expression pattern specific for radioactive iodine resistant tumors

We then attempted to identify the radioiodine resistance (RAIR)-specific gene expression patterns using our experimental RNA sequencing dataset. To this end, we compared 13 RAIR samples with 62 thyroid cancer samples: 18 follicular cancers and 44 papillary cancers. We did not consider 17 samples of follicular adenoma and 3 samples of medullar cancer in this comparison. We found no statistically significant differences for all 38 DNA repair pathways, thus, we investigated expression on gene level. We selected 217 differentially regulated genes using Wilcoxon statistical test with Benjamini-Hochberg FDR-adjusted p-value$<$0.05. However, well known RAIR biomarker gene *SLC5A5* was not included in the list because it did not pass the FDR criterion. We, therefore, modified the candidate gene search criterion so that it would be maximally strict and at the same time would allow *SLC5A5* ($p = 0.00344$) to be selected. Thus, we marked as differential all the genes with Wilcoxon test p-value less than 0.0035.

As a result, we identified 871 differential genes (Supplementary File 8) and then characterized partly their functional impacts using Gene Ontology (GO) analysis. Only 550 of these genes had functional GO annotation, and five statistically significantly enriched GO terms were identified in total (Figure 14). ClueGo [55] software analysis showed that these terms belong to two functional groups.

The first group deals with response to topologically incorrect (misfolded and unfolded) proteins. Unfolded protein response (UPR) is a cellular process related to the endoplasmic reticulum (ER) stress [56]. The information about the protein folding status is sensed in the ER lumen and transmitted to the cytoplasm and the nucleus. UPR signaling pathway includes transcriptional induction of UPR genes, global attenuation of translation, and ER-associated protein degradation. If the protein folding defect is not corrected, cells undergo apoptosis [57]. We found no previous references in the scientific literature to the association between the response to unfolded proteins and radioiodine resistance in thyroid cancer. However, induction of UPR was reported to promote radiosensibilization in radioresistant colorectal cancer cells [58].

The second functional group included only one GO term: aldo-keto reductase (NADP) activity. The aldo-keto reductase (AKR) superfamily comprises several enzymes that catalyze redox transformations involved in biosynthesis, intermediary metabolism, and detoxification [59]. We found no publications linking AKRs with radioiodine resistance of thyroid cancer. However, it was previously reported that small interfering RNA specific to *AKR1C3* gene could significantly enhance sensitivity to radiation therapy in non-small cell lung cancer cells [53]. In contrast, overexpression of *AKR1C3* also increased radioresistance of prostate and esophageal cancer cells [54, 60]. Meanwhile, increased aldehyde reductase expression mediates acquired radioresistance of laryngeal cancer via p53-related mechanism [61].

In addition, five differential genes out of 871 identified belonged to the DNA repair pathways (Table 3). One of them (*GTF2H2C)* is involved in nucleotide excision repair and RNA transcription [62]. For the other four genes, a link to radiation therapy was previously mentioned in the literature.

RFC5 gene product is part of the Rad17-RFC complex, which is one of the mechanisms of cellular recognition of a radiation insult [63]. The complex can dock on fractured DNA or near the site of damaged DNA and

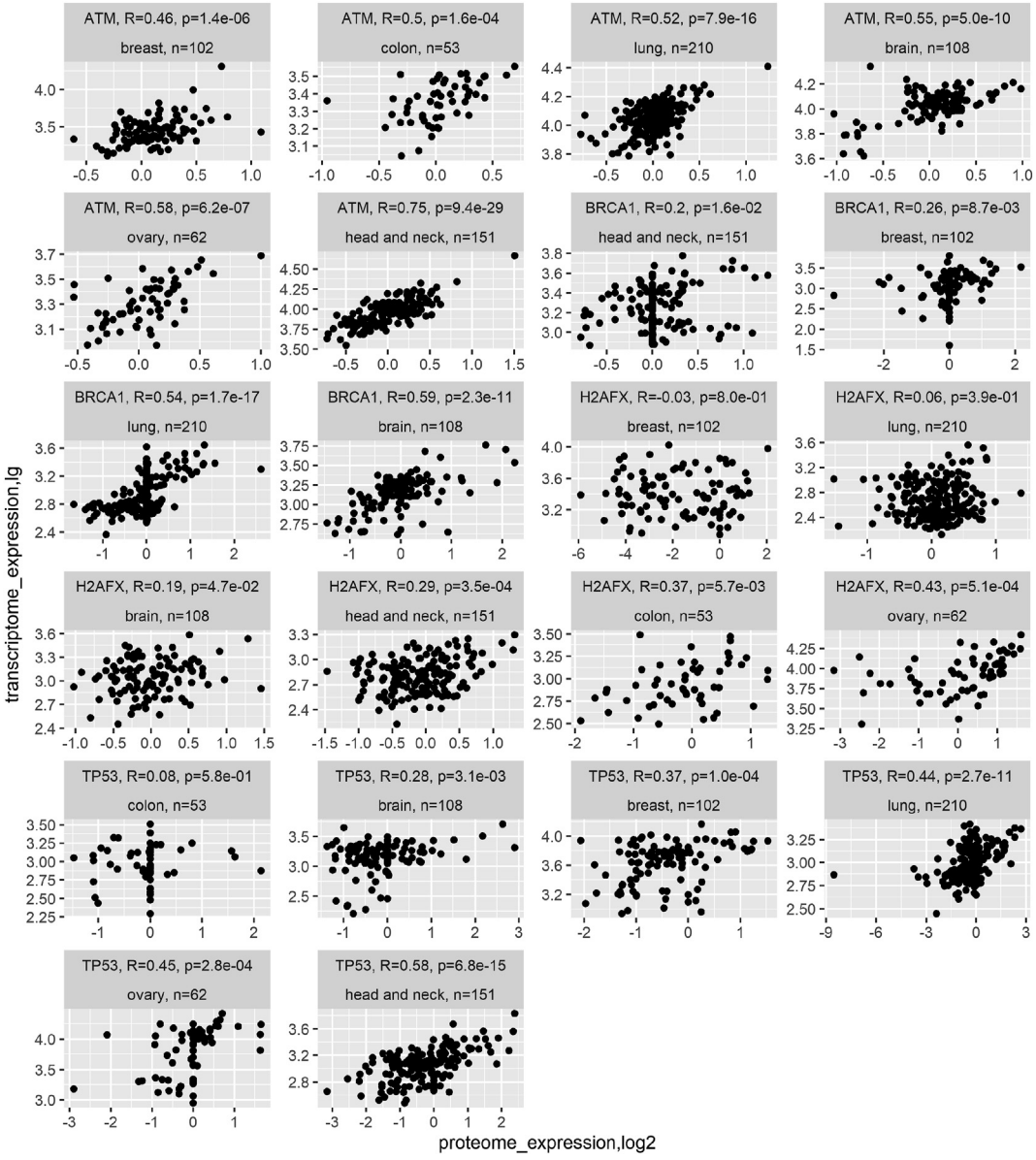


Figure 11. Correlation of expression at transcriptome and proteome levels for *ATM, BRCA1, H2AFX,* and *TP53* genes in six human cancer types.

initiate downstream signaling, including activation of ATM and ATR kinases [63]. Then the ATM and ATR signaling pathways activate downstream effectors and thus regulate cell cycle arrest, DNA repair, and apoptosis [63, 64, 65, 66, 67].

Cdc42 belongs to the Rho subfamily of GTPases, which plays a key role in the initiation and progression of cancer by transmitting oncogenic signals from the cell surface receptors and from activated guanine nucleotide exchange factors (GEF) signaling [62]. Alternatively, Cdc42 can be activated by the PI3K signaling through activation of PIP3-regulated GEFs, such as P-Rex, Vav, Sos, and SWAP70 [68]. Recently, Rac1 inhibition has been shown to sensitize pancreatic and breast cancer cells to radiation therapy, therefore, Rac and, similarly, Cdc42 may be regarded as the putative targets for chemosensitization of radiotherapy [69, 70, 71].

RBX1 contributes to the G1 phase-specific degradation of Exonuclease 1 (EXO1), which is responsible for the protruding 3′ single stranded DNA formation during processing of double-strand breaks. RBX1 knockdown decreases the degradation of EXO1 [72], and its depletion causes radiosensibilization of cervical cancer cells and deficiency in DNA double-strand breaks repair [72]. *RBX1* silencing significantly sensitizes radioresistant glioblastoma and lung cancer cells to the radiation with sensitivity enhancement ratios of 1.5 and 1.3, respectively [73].

Finally, the *RPS27A* gene was reported to be over-expressed in breast fibroadenomas, colorectal, renal cancers, chronic myeloid leukemia and acute leukemia [74, 75, 76]. Ubiquitin is encoded by four different genes that are highly homologous in eukaryotes. *UBA52* and *RPS27A* genes code for a single copy of ubiquitin fused to the ribosomal proteins L40 and S27a, respectively. Knock-down of ubiquitin increased the radiosensitivity in H1299 cells, thus *RPS27A* may contribute to radioresistance of tumor cells [77].

2.7. Fusion gene transcripts

We screened RNA sequencing reads for the clinically relevant hybrid transcripts with eleven oncogenes: *ALK, ROS1, RET, NTRK1, NTRK2, NTRK3, FGFR1, FGFR2, FGFR3, BCR,* and *ABL1*. We found reads directly supporting two fusion transcripts with genes *ALK* and *RET*. One was in-frame fusion for well-known *NCOA4-RET* translocation [107] (Figure 15A), and another one was frameshift fusion of *ALK* oncogene with a previously unknown upstream partner *ARHGAP12* (Figure 16A).

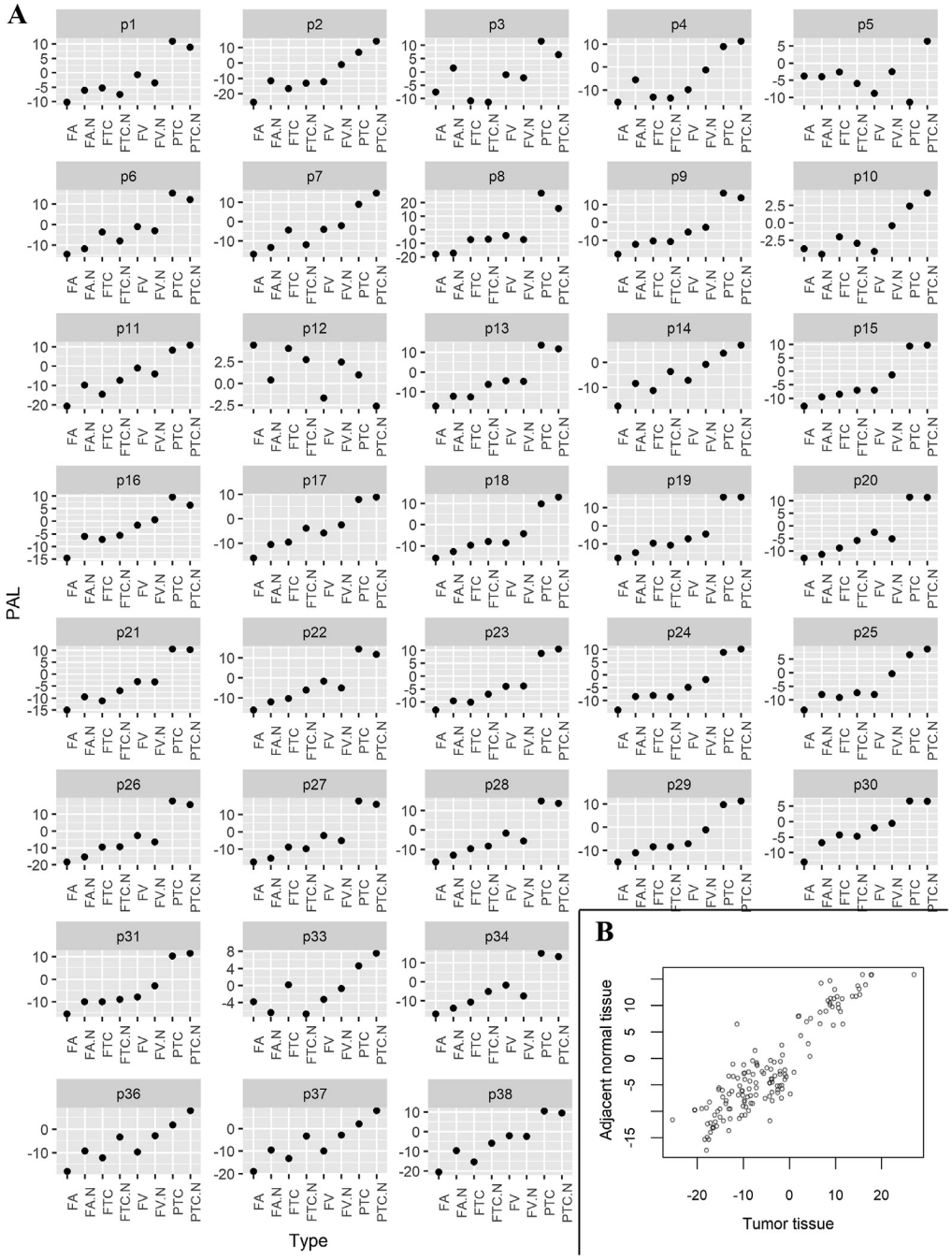


Figure 12. Average PAL values for 36 differential DNA repair pathways. (A) calculated using PRJEB11591 dataset for follicular adenoma (FA), follicular cancer (FC), follicular variant of papillary cancer (FV), classical variant of papillary cancer (PC) and their matched adjacent pathologically normal tissues: FA.N, FC.N, FV.N, PC.N, respectively. Pathway name abbreviations: p1 - ATM Pathway, p2 - ATM Pathway Cell Survival, p3 - ATM Pathway G2-Mitosis progression, p4 - Biocarta atm signaling pathway, p5 - Biocarta cell cycle G2M checkpoint pathway, p6 - BRCA1 Pathway, p7 - BRCA1 Pathway Chromatin Remodeling, p8 - BRCA1 Pathway Homologous Recombination Repair, p9 - BRCA1 Pathway Mismatch Repair, p10 - DNA Repair Mechanisms Pathway, p11 - KEGG Base excision repair pathway, p12 - KEGG Fanconi anemia pathway, p13 - KEGG Homologous recombination pathway, p14 - KEGG Mismatch repair pathway, p15 - KEGG Non homologous end joining pathway, p16 - KEGG Nucleotide excision repair pathway, p17 - Mismatch Repair in Eukaryotes Pathway, p18 - NCI ATM pathway, p19 - NCI ATM Pathway (G1 S transition checkpoint), p20 - NCI ATR signaling pathway, p21 - NCI ATR signaling Pathway (Pathway negative regulation of transcription during mitosis via CHEK1), p22 - NCI ATR signaling Pathway (regulation of double strand break repair via homologous recombination), p23 - NCI ATR signaling Pathway (response to G2 M transition DNA damage checkpoint signal), p24 - NCI DNA PK pathway in nonhomologous end joining Pathway (double strand break repair via nonhomologous end joining), p25 - NCI DNA PK pathway in nonhomologous end joining Pathway (V D J recombination), p26 - NCI Fanconi anemia pathway, p27 - NCI Fanconi anemia Pathway (regulation of double strand break repair via homologous recombination), p28 - NCI Fanconi anemia Pathway (Sister Chromatid Exchange Process), p29 - NHEJ mechanisms of DSBs repair effect, p30 - Nucleotide excision repair effect, p31 - p53 Signaling Pathway, p33 - p53 Signaling Pathway Gene Expression DNA Replication and Repair via TP53, p34 - reactome Fanconi Anemia pathway, p36 - Reactome Mismatch repair MMR directed by MSH2, MSH3, MutS beta pathway, p37 - Reactome Mismatch repair MMR directed by MSH2, MSH6, MutS alpha pathway, p38 - Reactome Repair synthesis for gap filling by DNA polymerase in TC NER pathway. (B) Average PAL for tumor tissues and the corresponding adjacent pathologically normal tissues. Each dot represents a "pathway – thyroid tumor type" case, totally 144 cases represented (36 pathways in four tumor types).

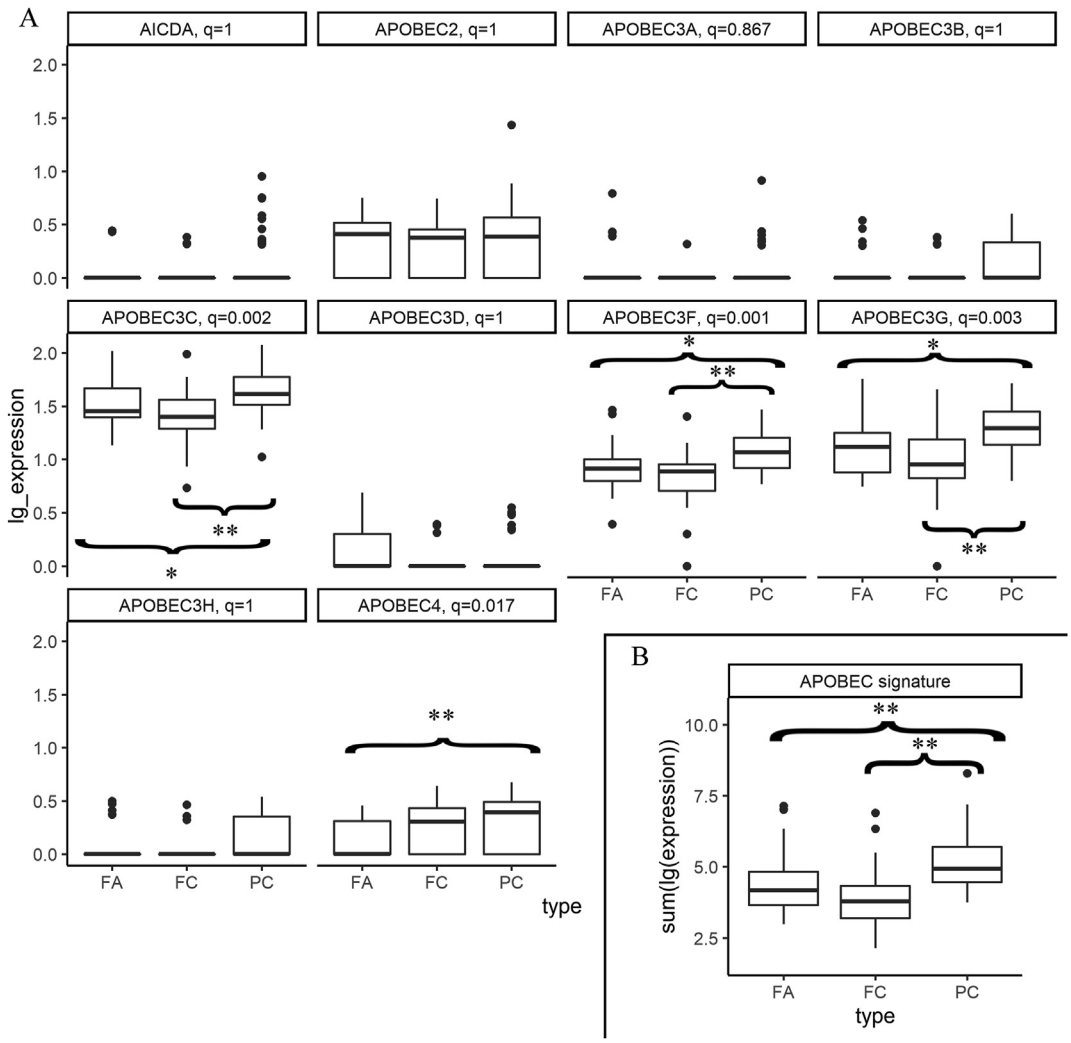


Figure 13. (A) Boxplot representation of logarithmic expression of *APOBEC* genes in thyroid tumor types: FA (n = 17), FC (n = 23), PC (n = 51). q-value of Kruskel-Wallis test is shown. Asterisks (*) indicate differential pairs of cancer types with Wilcoxon–test p-value <0.05; double asterisks (**) - with p-value <0.01. (B) *APOBEC* gene signature for three thyroid tumor types: FA (n = 17), FC (n = 23), PC (n = 51). Double asterisks (**) denote pairs with p-value <0.01 by Wilcoxon–test.

Seventh exon of *NCOA4* was fused with the eleventh exon of *RET*. In turn, third exon of *ARHGAP12* was fused with the fourth exon of *ALK*. In both cases, exon coverage also supported fusion status for the *ALK* (Figure 16B) and *RET* (Figure 15B) genes as the relative coverage downstream to the fusion site was higher than upstream (Figures 6B, 7B).

Both fusions were found in papillary thyroid cancers of follicular histologic subtype with node metastases, one of them (*NCOA4-RET*) in the radioactive iodine resistant tumor. We didn't detect any outstanding DNA repair pathway activation features in the samples with fusion transcripts (Figure 17).

3. Discussion

Here we present clinically annotated RNA sequencing data for 95 human thyroid tumor biosamples corresponding to 17 follicular adenomas, 23 follicular, 51 papillary, 1 poorly differentiated, and 3 medullar thyroid cancers cases. These molecular data are fully compatible with the previously published ANTE database, which includes RNA sequencing profiles for healthy human tissues, including thyroid tissue [45]. The gene expression dataset is assembled in a machine-readable format and annotated by the available patients' data, including type and histological subtype of thyroid tumor, sex, age, diagnosis, and *BRAF* (V600E) mutation status and gene fusion status. This dataset will hopefully be useful to those interested in thyroid tumor biology, especially in the context of relevant data accumulation from different sources. The use of this material in further analytic investigations can make it possible to identify and validate thyroid tumor type-specific expression biomarkers.

These data allowed us to analyze characteristic features of DNA repair pathway regulation in different thyroid tumor types. Follicular and papillary cancers demonstrated higher activation level of DNA repair pathways than the follicular adenomas. Our results suggest that papillary cancers have more activated DNA repair pathways than follicular neoplasms, with the remarkable exception of a G2M transition checkpoint pathway which showed an opposite trend. We found that most of DNA repair pathways were minimally activated (among tumors) in follicular adenomas. This trend was also confirmed for literature dataset PRJEB11591. We hypothesize that lower DNA repair pathway activation strengths in follicular adenomas compared to thyroid tumors can be linked with relatively lower level of DNA replication defects. Alternatively, this effect can be also connected with hypoxia which frequently accompanies benign tumors and inhibits DNA repair pathways [78].

Thus, our results demonstrate that DNA repair activity is increased in malignant compared to benign tumors. The same trend was also established for the tumor-adjacent pathologically normal tissues. These two evidences highlight the link between tumor progression and increased level of DNA repair and are consistent with the previous literature data on DNA repair in thyroid tumors [24, 79]. In agreement with that, the increased level of DNA damage marker 8-oxo-dG had been previously

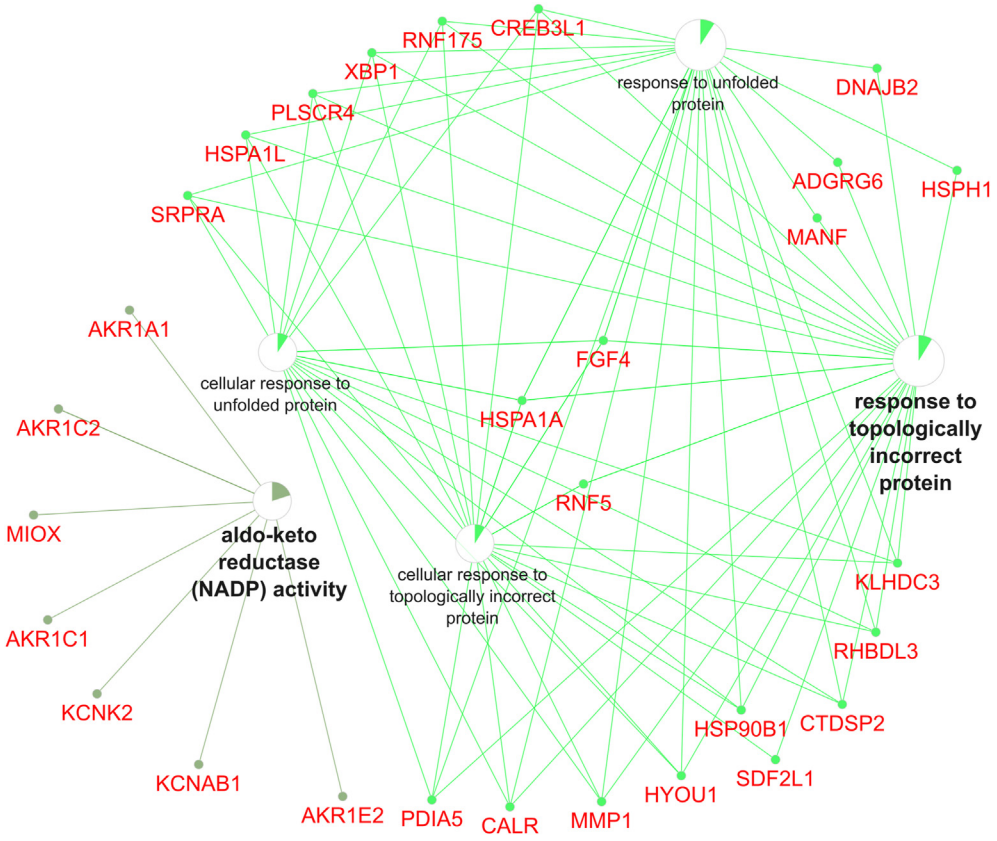


Figure 14. Statistically significantly enriched gene ontology (GO) molecular processes (GO-terms) identified for 871 differential genes between radioiodine resistant cancers and other thyroid cancer samples. Differential gene names are shown in red. Circle size inversely reflects FDR-adjusted p-value for a differential GO-term identified. Filled part of a circle represents percentage share of matched differential genes in gene set of the respective GO-term. Links between the terms are taken from the STRING database.

reported for thyroid tumors [80]. The same trend was also observed for another reactive oxygen marker, cytoplasmic 4-HNE, which characterizes lipid peroxidation level [80].

We also reconstructed DNA repair networks that were differential in the FC versus PC, FC + PC versus FA, and PC versus FA comparisons, and identified statistically significantly differential nodes. It is of note that the differential gene lists were very similar in these three comparisons. This fact may suggest that molecular differences between the follicular and papillary neoplasms are bigger than between the malignant and benign tumors.

Table 3. DNA repair genes linked with the radioiodine resistant thyroid cancer samples.

Gene ID	Gene function according to GeneCards database [62]	Relevant DNA repair pathways
CDC42	*CDC42* encodes a small GTPase of the Rho subfamily that regulates signaling pathways controlling cell morphology, migration, endocytosis, and cell cycle progression.	• p53 Signaling Pathway
GTF2H2C	Component of the core-TFIIH basal transcription factor involved in nucleotide excision repair of DNA and, when complexed to CDK activating kinase, in RNA transcription by RNA polymerase II.	• KEGG Nucleotide excision repair pathwayp2p • Reactome Formation of transcription coupled NER TC NER repair complex pathway
RFC5	*RFC5* encodes the smallest subunit of the factor C complex required for DNA replication. This subunit interacts with the C-terminus of PCNA protein and loads it onto DNA during S-phase of the cell cycle. *RFC5* is a member of the AAA+ (ATPases associated with various cellular activities) family and forms a core complex with the other two subunits that possesses DNA-dependent ATPase activity.	• Mismatch Repair in Eukaryotes Pathway, • KEGG Mismatch repair pathway, • KEGG Nucleotide excision repair pathway, • NCI ATR signaling Pathway regulation of double strand break repair via homologous recombination, • NCI ATR signaling Pathway negative regulation of transcription during mitosis via CHEK1, • NCI ATR signaling Pathway response to G2 M transition DNA damage checkpoint signal, • Reactome Repair synthesis for gap filling by DNA polymerase in TC NER pathway, • NCI ATR signaling pathway, • NCI Fanconi anemia pathway • NCI Fanconi anemia pathway (Sister Chromatid Exchange Process) • NCI Fanconi anemia pathway (regulation of double strand break repair via homologous recombination)
RBX1	*RBX1* encodes a RING finger-like domain-containing protein that interacts with cullin proteins and likely plays a role in ubiquitination processes necessary for cell cycle progression and DNA repair.	• KEGG Nucleotide excision repair pathway
RPS27A	Ubiquitin has a major role in targeting cellular proteins for proteasomal degradation, is synthesized as a precursor protein consisting of either polyubiquitin chains or a single ubiquitin fused to an unrelated protein. *RPS27A* encodes a fusion protein consisting of ubiquitin at the N terminus and ribosomal protein S27a at the C terminus.	• reactome Fanconi Anemia pathway

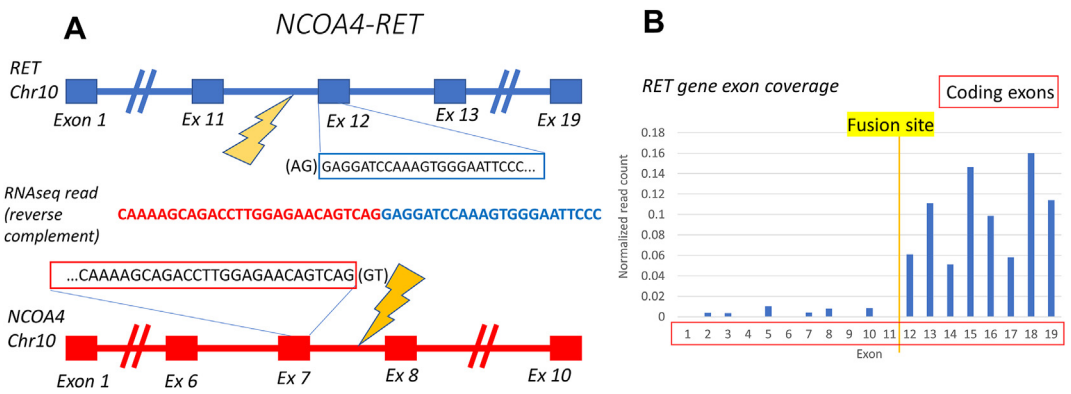


Figure 15. Schematic representation of *NCOA4-RET* fusion transcript identified. *A*, gene structures upstream and downstream of fusion site. *B*, *RET* gene exon coverage by normalized RNA sequencing reads in RAIR4 sample.

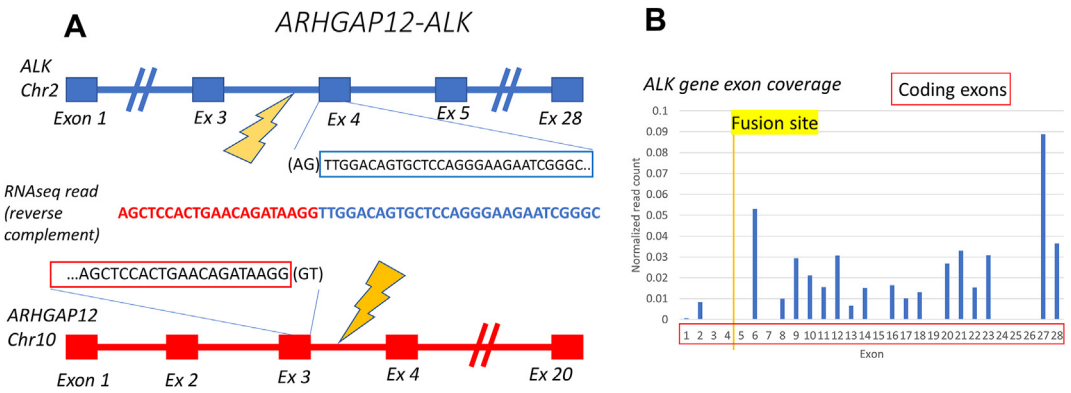


Figure 16. Schematic representation of *ARHGAP12*-ALK fusion transcript identified. (A) gene structures upstream and downstream of fusion site. (B) *ALK* gene exon coverage by normalized RNA sequencing reads in TC12 sample.

In this study we identified statistically differential features of DNA repair pathway activation between follicular and papillary thyroid cancers. Uncovering these differences may not only contribute to understanding of molecular mechanisms of thyroid tumors but also may help further developing new molecular-based therapeutic approaches for different types and subtypes of TCs. DNA repair has dual significance by simultaneously slowing down tumor evolution but also by protecting tumor cells from the treatment of replication targeting therapeutics. Thus, controlled inhibition of DNA repair activities can be considered personalized therapeutic option [81, 82]. On the other hand, we found no differential DNA repair pathways between radioiodine resistant and sensitive tumor samples. This may suggest that the molecular differences between thyroid tumor types are more significant than differences between radioiodine resistant and sensitive tumors.

Another key point of this study deals with the expression patterns of tumor-adjacent pathologically normal tissues learned from the literature dataset PRJEB11591. We showed that activation levels of DNA repair pathways strongly correlated between the tumor and adjacent normal tissues (R Pearson = 0.8492, p-value < 2.2e-16, Figure 12B). This observation is consistent with some previous studies [83, 84] and suggests that tumor-adjacent pathologically normal tissues can be strongly affected by the tumor and must be considered with care when comparing cancerous and healthy tissues.

Annotations for 23 molecular profiles published here include the response data on radioactive iodine therapy, of them 10 were for the treatment responders and 13 for the non-responders. This sampling is modest in terms of robust biomarker discovery. However, these datasets can be combined with the previously published or future data collections *e.g.* using harmonization of gene expression profiles data formats [85]. Alternatively, advanced methods of reducing data dimensionality can be applied in order to make use of machine learning and other artificial intelligence approaches to further analyze the data [86, 87, 88].

Finally, we identified fusion transcripts for *ALK* and *RET* oncogenes. One of them represents a well-known *NCOA4-RET* translocation, and another one is a new hybrid transcript of *ALK* oncogene with a previously unknown upstream fusion partner *ARHGAP12*. The latter can probably drive increased transcription of truncated *ALK* downstream from the fusion site because its normalized expression is one-two orders of magnitude higher in the normal tissues than for the *ALK* gene (UCSC Genome Browser, https://genome.ucsc.edu). This was also the case for our experimental dataset of thyroid tumor tissues, where mean normalized *ARHGAP12* expression was ~12-fold greater than the expression of *ALK*.

We hope that the data communicated here could be used for fundamental cancer research applications and for revising lists of tumor specific gene expression biomarkers, including those that distinguish between malignant and benign thyroid tumors and predict response to iodine radiation therapy.

4. Materials and methods

4.1. Biosamples

Biosamples were obtained from patients diagnosed with primary thyroid tumors who had undergone surgery at Endocrinology Research Centre, Moscow. For all the biosamples, informed written consents to participate in the study were collected from the patients or their legal representatives. The study was conducted in accordance with the Declaration of Helsinki ethical principles. The consent procedure and the design of the study were approved by the local ethical committee of the

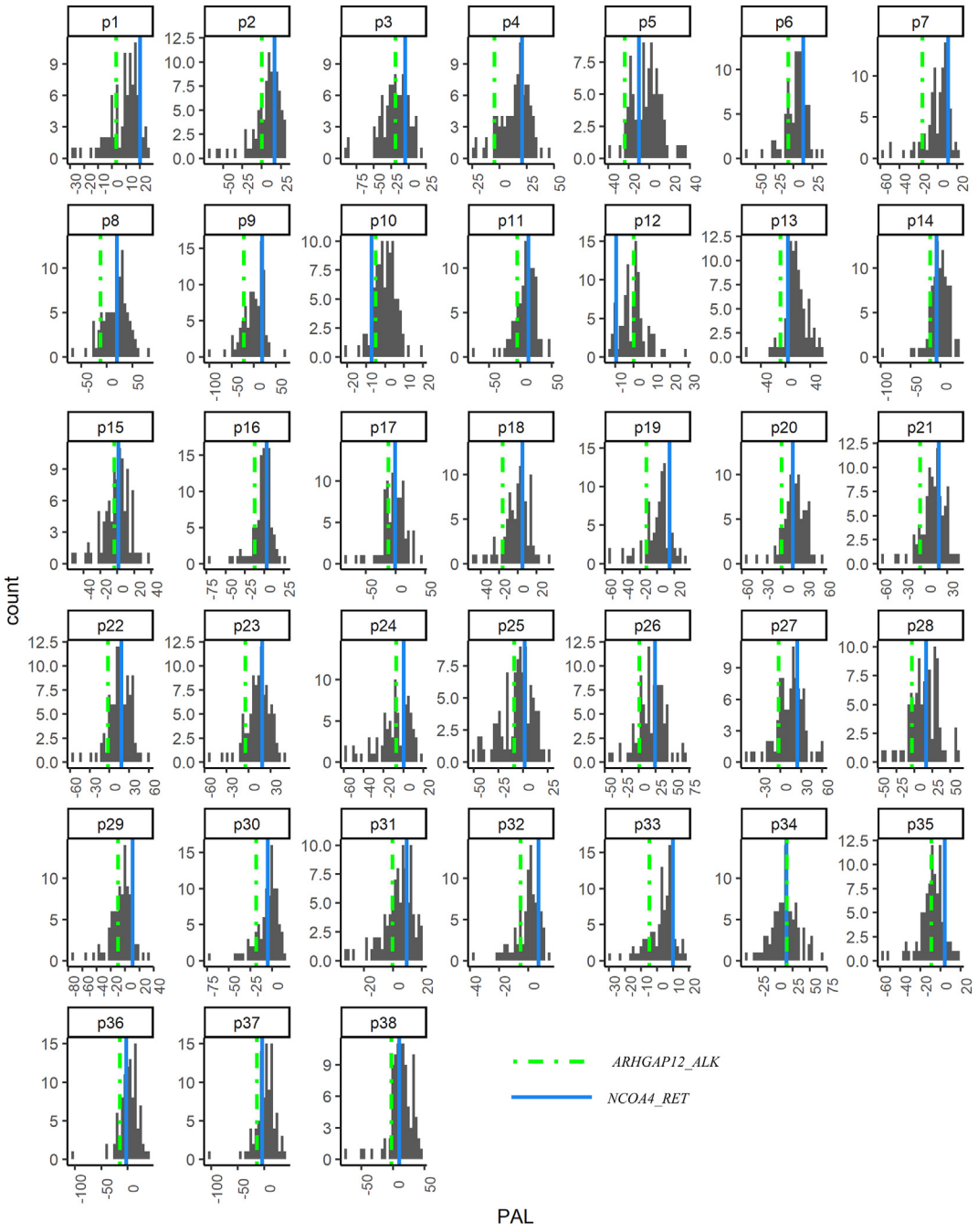


Figure 17. PAL distribution for 38 DNA repair pathways in experimental thyroid tumor samples. Color denotes samples with gene fusions. Pathway name abbreviations: p1 - ATM Pathway, p2 - ATM Pathway Cell Survival, p3 - ATM Pathway G2-Mitosis progression, p4 - Biocarta atm signaling pathway, p5 - Biocarta cell cycle G2M checkpoint pathway, p6 - BRCA1 Pathway, p7 - BRCA1 Pathway Chromatin Remodeling, p8 - BRCA1 Pathway Homologous Recombination Repair, p9 - BRCA1 Pathway Mismatch Repair, p10 - DNA Repair Mechanisms Pathway, p11 - KEGG Base excision repair pathway, p12 - KEGG Fanconi anemia pathway, p13 - KEGG Homologous recombination pathway, p14 - KEGG Mismatch repair pathway, p15 - KEGG Non homologous end joining pathway, p16 - KEGG Nucleotide excision repair pathway, p17 - Mismatch Repair in Eukaryotes Pathway, p18 - NCI ATM pathway, p19 - NCI ATM Pathway (G1 S transition checkpoint), p20 - NCI ATR signaling pathway, p21 - NCI ATR signaling Pathway (Pathway negative regulation of transcription during mitosis via CHEK1), p22 - NCI ATR signaling Pathway (regulation of double strand break repair via homologous recombination), p23 - NCI ATR signaling Pathway (response to G2 M transition DNA damage checkpoint signal), p24 - NCI DNA PK pathway in nonhomologous end joining Pathway (double strand break repair via nonhomologous end joining), p25 - NCI DNA PK pathway in nonhomologous end joining Pathway (V D J recombination), p26 - NCI Fanconi anemia pathway, p27 - NCI Fanconi anemia Pathway (regulation of double strand break repair via homologous recombination), p28 - NCI Fanconi anemia Pathway (Sister Chromatid Exchange Process), p29 - NHEJ mechanisms of DSBs repair effect, p30 - Nucleotide excision repair effect, p31 - p53 Signaling Pathway, p32 - p53 Signaling Pathway DNA Repair, p33 - p53 Signaling Pathway Gene Expression DNA Replication and Repair via TP53, p34 - reactome Fanconi Anemia pathway, p35 - Reactome Formation of transcription coupled NER TC NER repair complex pathway, p36 - Reactome Mismatch repair MMR directed by MSH2, MSH3, MutS beta pathway, p37 - Reactome Mismatch repair MMR directed by MSH2, MSH6, MutS alpha pathway, p38 - Reactome Repair synthesis for gap filling by DNA polymerase in TC NER pathway.

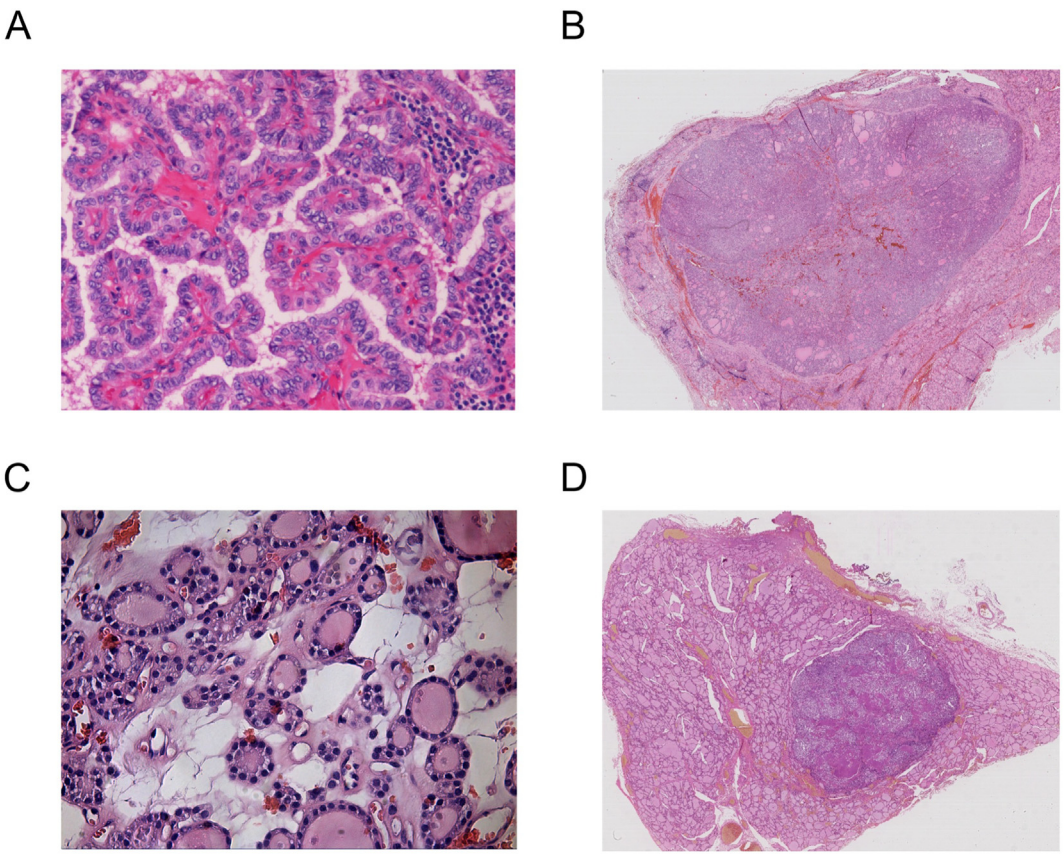


Figure 18. Representative microphotographs of eosin-hematoxylin stained histology preparations of different thyroid tumor types. A – papillary thyroid cancer (TC), magnification X200, sample ID: TC_65; B - Follicular thyroid cancer (FC), magnification X50, sample ID: TC_119; C – Follicular adenoma (FA), magnification X100, sample ID: TC_22; D – medullary thyroid cancer (MC), magnification X50, sample ID: TC_17.

Endocrinology Research Centre, Moscow, protocol number 19 (October 25, 2017, Chairman Dr. E.N. Andreeva). From the patients TC16, TC18, TC19, TC26 and TC127, the consent was obtained for disclosure of sex, histological tumor type, and molecular data including RNA sequencing data but excluding whole-genome and/or whole-exome sequencing data; from the patients TC11, TC12, TC17, TC37 and TC125 - for disclosure of sex, histological tumor type, diagnosis, and molecular data including RNA sequencing data but excluding whole-genome and/or whole-exome sequencing data; for the rest of the patients - for disclosure of sex, age, histological tumor type, diagnosis, and molecular data including RNA sequencing data but excluding whole-genome and/or whole-exome sequencing data.

Tissue specimens were formalin-fixed and embedded in paraffin. Tumors were evaluated by a pathologist to determine histological type and to estimate proportion of tumor cells. All the specimens used in this study contained at least 60% of tumor cells.

Overall, 51 papillary thyroid cancer (PC), 23 follicular thyroid cancer (FC), 3 medullary thyroid cancer (MC), 1 poorly differentiated thyroid cancer (PDC), and 17 follicular thyroid adenoma (FA) samples were investigated. There were 21 male and 74 female patients. The mean age of patients was 47 years (range 11–88 years). Additionally, 6 healthy thyroid tissue samples from ANTE database [45] were used to normalize gene expression levels in molecular pathway analysis. Clinical annotations of tumor specimens investigated are summarized in the Supplementary file 1.

4.2. RNA sequencing

RNA libraries were generated, sequenced, and processed as described in [45]. RNA extraction was performed with the RecoverAll™ Total Nucleic Acid Isolation Kit (Invitrogen), following the manufacturer's protocol. RNA Integrity Number (RIN) was measured using Agilent 2100 bioanalyzer. RNA concentration was measured with Agilent RNA 6000 Nano or Qubit RNA Assay Kits. Depletion of ribosomal RNA was performed using RNA Hyper with RiboErase (KAPA Biosystem) Kit. Library concentrations and quality were measured with Qubit ds DNA HS Assay kit (Life Technologies) and Agilent Tapestation (Agilent). The samples were sequenced using Illumina HiSeq 3000 equipment for single end sequencing, 50 bp read length, for approximately 30 million raw reads per sample. Data quality check was conducted using Illumina SAV. De-multiplexing was performed using Illumina Bcl2fastq2 v 2.17 software [45].

4.3. Processing of RNA sequencing data

RNA sequencing FASTQ files were processed with STAR aligner [89]. We used 'GeneCounts' mode with the Ensembl human transcriptome annotation (Build version GRCh38 and transcript annotation GRCh38.89). Complete HGNC dataset [90], version of July 13, 2017, was used to convert Ensembl gene IDs to HGNC gene symbols.

4.4. Evaluation of BRAF (V600E) mutation status

Detection of *BRAF* (V600E) mutation was performed using allele-specific PCR with dual-labeled probe as described in [91]. PCR cycling conditions were pre-denaturation step 95°C – 2 min, followed by 50 cycles of denaturation (94°C, 10 s), annealing, and elongation (60°C, 15 s). Mutation status was considered positive if the mutant allele content was 3% or more of the wild-type allele.

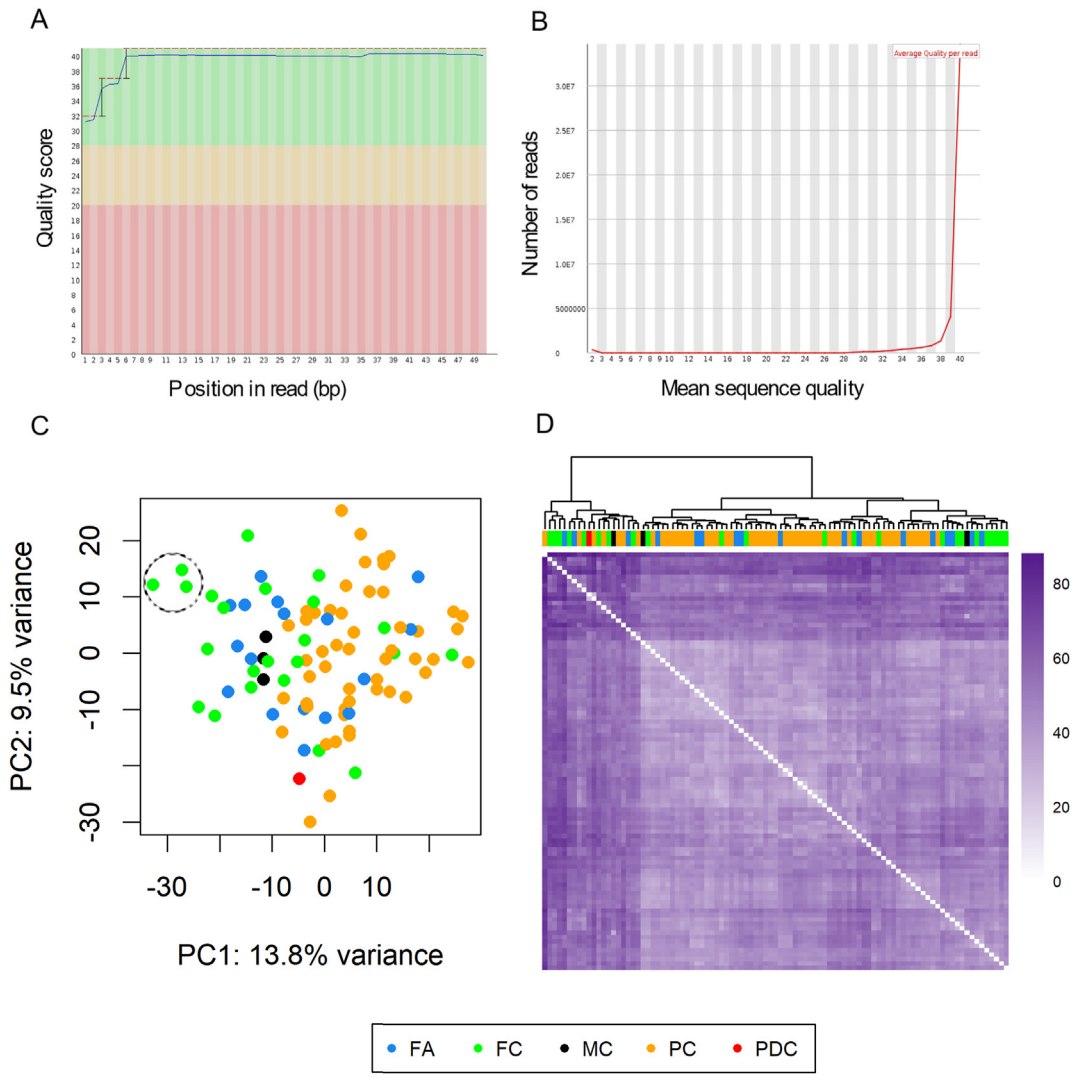


Figure 19. Quality control of RNA sequencing profiles. A - fastQC per base quality score for sample TC76. The plot was drawn using fastQC software. B - fastQC per sequence quality score for sample TC76. The plot was drawn using fastQC software. C - plot for principal component analysis in normalized gene expression space (lg(RPM)) for all RNA sequencing profiles investigated. The plot was drawn using graphics package in R language. D - hierarchical clustering of all RNA sequencing profiles based on Euclidian distance in normalized gene expression space (lg(RPM)). The plot was drawn using pheatmap package in R, clustering algorithm "ward.d2".

4.5. Histological characterization of samples

All samples were subjected to histological analysis, representative microphotographs of preparations of the thyroid tumor types investigated are shown in Figure 18.

Specific histology subtypes of the specimens are given in the Supplementary file 1. For further analyses, we grouped specimens by major tumor types: PC, FC, MC, and FA.

4.6. Quality control of mRNA sequencing profiles

Detailed annotation of RNA sequencing reads mapping statistic produced by STAR aligner [89] and quality control metrics produced by NCBI MAGIC software [79] are shown in the Supplementary file 2. Minimal number of reads uniquely mapped on mRNA sequences of known genes was ~3,4*10^6 with the mean of ~9,2*10^6 (Supplementary file 2). Average mapping rate, taking into account only uniquely mapped reads, was 78%. Typical per base and per sequence quality scores visualized by fastQC software [80] demonstrated high values for individual base pairs and read sequences (Figure 19 A, B).

Principal component analysis (PCA) and hierarchical clustering based on Euclidian distance in normalized counts (lg(RPM)) space did not reveal any outliers (Figure 19 C, D). In comparison with other FC and PC samples, three FC samples formed a distant cluster (TC2, TC4, TC124; Figure 19C, circled). These three cluster members had 150 differentially expressed genes with q-value<0.05 by Wilcoxon test (Supplementary File 9). The gene contribution coefficients for the first two principal components are given in Supplementary File 9.

We performed sequencing of two technical replicates of samples TC63 and TC116, Spearman's correlation coefficient between gene expression profiles was 0.95 for both cases. We also found that the replicates tightly clustered on the dendrogram (Figure 20) which confirmed data consistency of the RNA sequencing protocol used.

Overall, our data quality assay did not identify any significant internal problems or internal bias in the RNA sequencing data presented. We did not include the poorly differentiated thyroid cancer sample in our further analysis because it cannot be used for group comparisons. We used reads per million (RPM) normalization of expression data. We filtered low-expressed values (less than one) to reduce data noise.

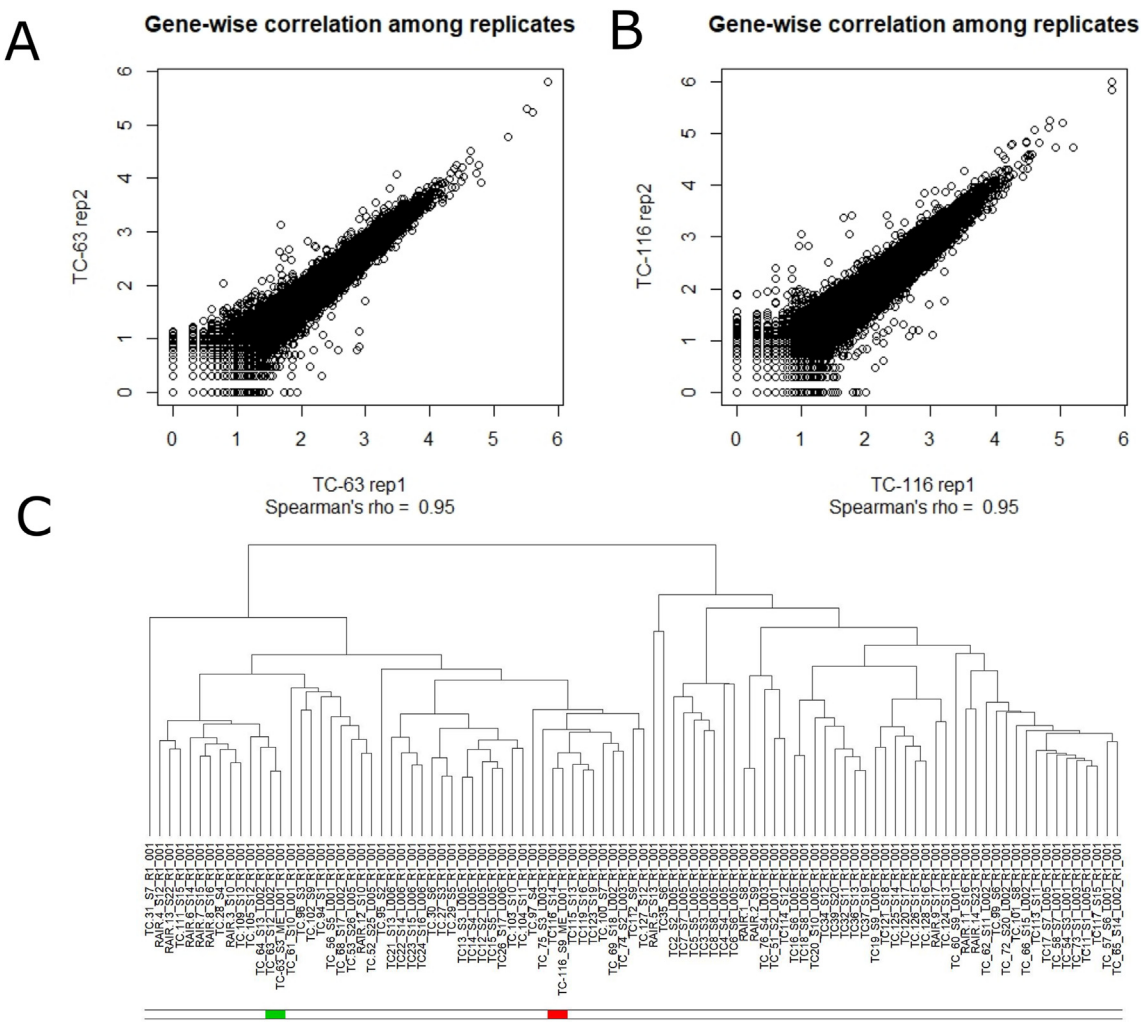


Figure 20. Comparison of RNA sequencing gene expression profiles of two technical replicates: for TC-63 sample *(A)*, and for TC-116 sample (*B*). Clustering dendrogram of experimental tumor samples expression profiles (*C*), technical replicates are shown in color.

4.7. Data records

Original gene expression data were deposited in Gene Expression Omnibus database (GEO) under accession number GSE138042. Raw fastq data are available in SRA archive under accession number PRJNA588725. Description of clinically relevant (age, sex, diagnosis, tumor histotype, *BRAF* (V600E) mutation status) and technical (RIN, date of sequencing) information is given in the Supplementary file 1. Results of RNA sequencing reads quality assay are given for mRNA in the Supplementary file 2.

4.8. Structures of DNA repair pathways

The gene structures and molecular architectures of 38 DNA repair pathways were obtained from the public databases Reactome [40], NCI Pathway Interaction Database [41], Kyoto Encyclopedia of Genes and Genomes [42], Biocarta [43], and Qiagen [44], and manually curated as described in [81]. Only molecular pathways which include ten or more genes [82] were considered in calculations of pathway activation level (PAL) in order to increase statistical accuracy.

4.9. Alternative thyroid tissue datasets

We used published RNA sequencing gene expression profiles of normal thyroid tissues from TCGA [83] and GTEx [84] project databases, amounting to 58 and 446 profiles, correspondingly. Samples barcodes are shown in Supplementary File 6. Raw gene counts were processed as reads per million (RPM).

Alternatively, thyroid tumor and normal thyroid RNA sequencing profiles from dataset PRJEB11591 [92,93] were used. The dataset includes data for 25 FA, 30 minimally invasive FC, 48 follicular variant of PC (FV), 77 classical papillary thyroid carcinomas (PC) and 81 adjacent normal thyroid tissues. FASTQ files were processed by STAR aligner [89] to get raw gene counts that were processed as RPM.

The proteome datasets from CPTAC [94] and corresponding transcriptome datasets from gdc-portal [29] were used to assess correlation between expression on transcriptome and proteome level for key genes in tumor and normal tissues. We totally used 102 paired RNA sequencing and proteome profiles of breast, 53 - colon, 210 – lung, 108 – brain, 62 – ovary, 151 – head and neck tissues.

4.10. Calculation of pathway activation level

Pathway activation level (PAL) characterizes cumulative changes in expression levels of genes belonging to a certain molecular pathway [38, 39, 95]. PAL is calculated as follows:

$$PAL_p = \sum_n ARR_{np} * \lg(CNR_n) \Big/ \sum_n |ARR_{np}|,$$

where PAL_p is PAL for pathway p, CNRn is case-to-normal ratio, the ratio of gene n expression level in a tumor sample under study to an average level in the control group; ARR (activator/repressor role) is a Boolean flag that depends on the function of gene *n* product in pathway *p*. ARR value is −1 if gene product n inhibits pathway p; 1 if n activates the pathway; 0 if n has ambiguous or unclear role in the pathway; 0.5 or −0.5, if n is more a pathway activator or its inhibitor, respectively. For convenience, in this article we use modified PAL = PAL*100. PAL values for all samples under investigation are given in Supplementary File 6.

4.11. Gene ontology analysis and visualization of molecular pathways

Gene Ontology (GO) analysis was performed using ClueGO software [55] with GO database from 28.04.2020. GO-analysis results and molecular pathways were visualized using Cytoscape 3.8.0 [96].

4.12. Identification of fusion transcripts

Fusion transcripts were initially screened using STAR-Fusion software. Preliminary files containing fusion candidates for genes *ALK, ROS1, RET, NTRK1, NTRK2, NTRK3, FGFR1, FGFR2, FGFR3, BCR, ABL1* were generated and the corresponding RNA sequencing reads were extracted. The output data were manually inspected using UCSC BLAT and UCSC Browser (https://genome.ucsc.edu/) to interrogate fusion candidates according to the following criteria: (i) does the read cover exon junction of two different transcripts, (ii) if the junction point exactly corresponds to exon termini of known genes with canonic splice sites, (iii) if both transcripts are in the same orientation. The inputs with three positive flags were considered as those supporting fusion transcripts.

Declarations

Author contribution statement

Anton Buzdin, Maxim Sorokin: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Uliana Vladimirova: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Pavel Rumiantsev, Dmitriy Kamashev: Conceived and designed the experiments; Analyzed and interpreted the data.

Marianna Arsenovna Zolotovskaia: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Eugene Albert: Analyzed and interpreted the data; Wrote the paper.

Aleksander Abrosimov: Performed the experiments.

Konstantin Slashchuk, Petr Nikiforovich, Olga Chukhacheva: Contributed reagents, materials, analysis tools or data.

Nurshat Gaifullin: Conceived and designed the experiments; Wrote the paper.

Maria Suntsova, Xinmin Li: Conceived and designed the experiments; Performed the experiments.

Galina Zakharova, Andrew Garazha: Analyzed and interpreted the data.

Alexander Glusker: Conceived and designed the experiments.

Daniil Nikitin: Analyzed and interpreted the data; Wrote the paper.

Alexei Drobyshev, Irina Kochergina-Nikitskaya: Performed the experiments; Analyzed and interpreted the data.

Funding statement

This work was supported by Omicsway Corp., USA, and by Russian Foundation for Basic Research grant (19-29-01108).

Data availability statement

Data associated with this study has been deposited at GEO and SRA repositories and published under GEO accession id GSE138042.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2021.e06408.

References

[1] C. La Vecchia, M. Malvezzi, C. Bosetti, W. Garavello, P. Bertuccio, F. Levi, E. Negri, Thyroid cancer mortality and incidence: a global overview, Int. J. Canc. 136 (2015) 2187–2195.
[2] M.E. Cabanillas, D.G. McFadden, C. Durante, Thyroid cancer, Lancet 388 (2016) 2783–2795.
[3] L. Rahib, B.D. Smith, R. Aizenberg, A.B. Rosenzweig, J.M. Fleshman, L.M. Matrisian, Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States, Canc. Res. 74 (2014) 2913–2921.
[4] S.L. Asa, The current histologic classification of thyroid cancer, Endocrinol Metab. Clin. N. Am. 48 (2019) 1–22.
[5] E. Younis, Oncogenesis of thyroid cancer, Asian Pac. J. Cancer Prev. APJCP 18 (2017) 1191–1199.
[6] F. Raue, K. Frank-Raue, Epidemiology and clinical presentation of medullary thyroid carcinoma, Recent Results Canc. Res. 204 (2015) 61–90.
[7] C. de la Fouchardière, M. Decaussin-Petrucci, J. Berthiller, F. Descotes, J. Lopez, J.C. Lifante, J.L. Peix, A.L. Giraudet, A. Delahaye, S. Masson, C. Bournaud-Salinas, F. Borson Chazot, Predictive factors of outcome in poorly differentiated thyroid carcinomas, Eur. J. Canc. 92 (2018) 40–47.
[8] E.Y. Ibrahim, N.L. Busaidy, Treatment and surveillance of advanced, metastatic iodine-resistant differentiated thyroid cancer, Curr. Opin. Oncol. 29 (2017) 151–158.
[9] H. Shimura, K. Haraguchi, A. Miyazaki, T. Endo, T. Onaya, Iodide uptake and experimental 131I therapy in transplanted undifferentiated thyroid cancer cells expressing the Na+/I- symporter gene, Endocrinology 138 (1997) 4493–4496.
[10] Y. Luo, H. Jiang, W. Xu, X. Wang, B. Ma, T. Liao, Y. Wang, Clinical, pathological, and molecular characteristics correlating to the occurrence of radioiodine refractory differentiated thyroid carcinoma: a systematic review and meta-analysis, Front. Oncol. 10 (2020) 549882.
[11] J. Liu, R. Liu, X. Shen, G. Zhu, B. Li, M. Xing, The genetic duet of BRAF V600E and TERT promoter mutations robustly predicts loss of radioiodine avidity in recurrent papillary thyroid cancer, J. Nucl. Med. 61 (2020) 177–182.
[12] X. Yang, J. Li, X. Li, Z. Liang, W. Gao, J. Liang, S. Cheng, Y. Lin, TERT promoter mutation predicts radioiodine-refractory character in distant metastatic differentiated thyroid cancer, J. Nucl. Med. 58 (2017) 258–265.
[13] M. Aashiq, D.A. Silverman, S. Na'ara, H. Takahashi, M. Amit, Radioiodine-refractory thyroid cancer: molecular basis of redifferentiation therapies, management, and novel therapies, Cancers 11 (2019) 1382.
[14] J. Sarhan, B.C. Liu, H.I. Muendlein, C.G. Weindel, I. Smirnova, A.Y. Tang, V. Ilyukha, M. Sorokin, A. Buzdin, K.A. Fitzgerald, A. Poltorak, Constitutive interferon signaling maintains critical threshold of MLKL expression to license necroptosis, Cell Death Differ. 26 (2019) 332–347.
[15] V. Zorin, A. Zorina, N. Smetanina, P. Kopnin, I.V. Ozerov, S. Leonov, A. Isaev, D. Klokov, A.N. Osipov, Diffuse colonies of human skin fibroblasts in relation to cellular senescence and proliferation, Aging (Albany. NY) 9 (2017) 1404–1413.
[16] A. Torgovnick, B. Schumacher, DNA repair mechanisms in cancer development and therapy, Front. Genet. 6 (2015).
[17] M. Sorokin, R. Kholodenko, A. Grekhova, M. Suntsova, M. Pustovalova, N. Vorobyeva, I. Kholodenko, G. Malakhova, A. Garazha, A. Nedoluzhko, R. Vasilov, E. Poddubskaya, O. Kovalchuk, L. Adamyan, V. Prassolov, D. Allina, D. Kuzmin, K. Ignatev, A. Osipov, A. Buzdin, Acquired resistance to tyrosine kinase inhibitors may be linked with the decreased sensitivity to X-ray irradiation, Oncotarget 9 (2018) 5111–5124.
[18] D. Kalasauskas, M. Sorokin, B. Sprang, A. Elmasri, S. Viehweg, G. Salinas, L. Opitz, M. Rave-Fraenk, W. Schulz-Schaeffer, S.R. Kantelhardt, A. Giese, A. Buzdin, E.L. Kim, Diversity of clinically relevant outcomes resulting from hypofractionated radiation in human glioma stem cells mirrors distinct patterns of transcriptomic changes, Cancers 12 (2020).
[19] K. Kiwerska, K. Szyfter, DNA repair in cancer initiation, progression, and therapy—a double-edged sword, J. Appl. Genet. 60 (2019) 329–334.
[20] E.L. Kim, M. Sorokin, S.R. Kantelhardt, D. Kalasauskas, B. Sprang, J. Fauss, F. Ringel, A. Garazha, E. Albert, N. Gaifullin, C. Hartmann, N. Naumann, S.-E.E. Bikar, A. Giese, A. Buzdin, Intratumoral heterogeneity and longitudinal changes in gene expression predict differential drug sensitivity in newly diagnosed and recurrent glioblastoma, Cancers 12 (2020) 520.

[21] A. Noda, Radiation-induced unrepairable DSBs: their role in the late effects of radiation and possible applications to biodosimetry, J. Radiat. Res. 59 (2018) 114–120.

[22] G.A.S. Nogueira, E.F.D. Costa, L. Lopes-Aguiar, T.R.P. Lima, M.B. Visacri, E.C. Pincinato, G.J. Lourenço, L. Calonga, F.V. Mariano, A.M. de A.M. Altemani, J.M.C. Altemani, P. Moriel, C.T. Chone, C.D. Ramos, C.S.P. Lima, Polymorphisms in DNA mismatch repair pathway genes predict toxicity and response to cisplatin chemoradiation in head and neck squamous cell carcinoma patients, Oncotarget 9 (2018) 29538–29547.

[23] L.P. Martin, T.C. Hamilton, R.J. Schilder, Platinum resistance: the role of DNA repair pathways, Clin. Canc. Res. 14 (2008) 1291–1295.

[24] J.O. Paulsson, S. Backman, N. Wang, A. Stenman, J. Crona, J. Thutkawkorapin, M. Ghaderi, E. Tham, P. Stålberg, J. Zedenius, C.C. Juhlin, Whole-genome sequencing of synchronous thyroid carcinomas identifies aberrant DNA repair in thyroid cancer dedifferentiation, J. Pathol. 250 (2020) 183–194.

[25] B. Wojtas, A. Pfeifer, M. Oczko-Wojciechowska, J. Krajewska, A. Czarniecka, A. Kukulska, M. Eszlinger, T. Musholt, T. Stokowy, M. Swierniak, E. Stobiecka, E. Chmielik, D. Rusinek, T. Tyszkiewicz, M. Halczok, S. Hauptmann, D. Lange, M. Jarzab, R. Paschke, B. Jarzab, Gene expression (mRNA) markers for differentiating between malignant and benign follicular thyroid tumours, Int. J. Mol. Sci. 18 (2017) 1184.

[26] Z. Pan, L. Li, Q. Fang, Y. Qian, Y. Zhang, J. Zhu, M. Ge, P. Huang, Integrated bioinformatics analysis of master regulators in anaplastic thyroid carcinoma, BioMed Res. Int. 2019 (2019).

[27] A. Prete, P. Borges de Souza, S. Censi, M. Muzza, N. Nucci, M. Sponziello, Update on fundamental mechanisms of thyroid cancer, Front. Endocrinol. 11 (2020) 102.

[28] A. Buzdin, M. Sorokin, A. Garazha, A. Glusker, A. Aleshin, E. Poddubskaya, M. Sekacheva, E. Kim, N. Gaifullin, A. Giese, A. Seryakov, P. Rumiantsev, S. Moshkovskii, A. Moiseev, RNA sequencing for research and diagnostics in clinical oncology, Semin. Canc. Biol. 60 (2019) 311–323.

[29] R.L. Grossman, A.P. Heath, V. Ferretti, H.E. Varmus, D.R. Lowy, W.A. Kibbe, L.M. Staudt, Toward a shared vision for cancer genomic data, N. Engl. J. Med. 375 (2016) 1109–1112.

[30] K. Vierlinger, M.H. Mansfeld, O. Koperek, C. Nöhammer, K. Kaserer, F. Leisch, Identification of SERPINA1 as single marker for papillary thyroid carcinoma through microarray meta analysis and quantification of its discriminatory power in independent validation, BMC Med. Genom. 4 (2011) 30.

[31] J.A. Kim, Cooperative instruction of signaling and metabolic pathways on the epigenetic landscape, Mol. Cell. 41 (2018) 264–270.

[32] A. Simms, R.P. Jacob, C. Cohen, M.T. Siddiqui, TROP-2 expression in papillary thyroid carcinoma: potential Diagnostic Utility, Diagn. Cytopathol. 44 (2016) 26–31.

[33] R.M. de Morais, A.B. Sobrinho, C.M. de S. Silva, J.R. de Oliveira, I.C.R. da Silva, O. de T. Nóbrega, The role of the NIS (SLC5A5) gene in papillary thyroid cancer: a systematic review, Internet J. Endocrinol. 2018 (2018).

[34] E.C. Morari, M.A. Marcello, A.C.T. Guilhen, L.L. Cunha, P. Latuff, F.A. Soares, J. Vassallo, L.S. Ward, Use of sodium iodide symporter expression in differentiated thyroid carcinomas, Clin. Endocrinol. 75 (2011) 247–254.

[35] C. Tavares, M.J. Coelho, C. Eloy, M. Melo, A.G. da Rocha, A. Pestana, R. Batista, L.B. Ferreira, E. Rios, S. Selmi-Ruby, B. Cavadas, L. Pereira, M.S. Simões, P. Soares, NIS expression in thyroid tumors, relation with prognosis clinicopathological and molecular features, Endocr. Connect. 7 (2018) 78–90.

[36] R. Ameziane-El-Hassani, M. Schlumberger, C. Dupuy, NADPH oxidases: new actors in thyroid cancer? Nat. Rev. Endocrinol. 12 (2016) 485–494.

[37] A. Czarniecka, M. Oczko-Wojciechowska, M. Barczyński, BRAF V600E mutation in prognostication of papillary thyroid cancer (PTC) recurrence, Gland Surg. 5 (2016) 495–505.

[38] N. Borisov, M. Sorokin, A. Garazha, A. Buzdin, Quantitation of molecular pathway activation using RNA sequencing data, Methods Mol. Biol. 2063 (2020) 189–206.

[39] A. Buzdin, M. Sorokin, A. Garazha, M. Sekacheva, E. Kim, N. Zhukov, Y. Wang, X. Li, S. Kar, C. Hartmann, A. Samii, A. Giese, N. Borisov, Molecular pathway activation - new type of biomarkers for tumor morphology and personalized selection of target drugs, Semin. Canc. Biol. 53 (2018) 110–124.

[40] D. Croft, A.F. Mundo, R. Haw, M. Milacic, J. Weiser, G. Wu, M. Caudy, P. Garapati, M. Gillespie, M.R. Kamdar, B. Jassal, S. Jupe, L. Matthews, B. May, S. Palatnik, K. Rothfels, V. Shamovsky, H. Song, M. Williams, E. Birney, H. Hermjakob, L. Stein, P. D'Eustachio, The Reactome pathway knowledgebase, Nucleic Acids Res. 42 (2014) D472–D477.

[41] C.F. Schaefer, K. Anthony, S. Krupa, J. Buchoff, M. Day, T. Hannay, K.H. Buetow, PID: the pathway interaction database, Nucleic Acids Res. 37 (2009) D674–D679.

[42] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (2000) 27–30. http://www.ncbi.nlm.nih.gov/pubmed/10592173. (Accessed 20 August 2018).

[43] D. Nishimura, BioCarta, Biotech Softw, Internet Rep. 2 (2001) 117–120.

[44] QIAGEN - Pathway-Central (n.d), https://www.qiagen.com/us/shop/genes-and-pathways/pathway-central/. (Accessed 19 September 2018).

[45] M. Suntsova, N. Gaifullin, D. Allina, A. Reshetun, X. Li, L. Mendeleeva, V. Surin, A. Sergeeva, P. Spirin, V. Prassolov, A. Morgan, A. Garazha, M. Sorokin, A. Buzdin, Atlas of RNA sequencing profiles for normal human tissues, Sci. Data. 6 (2019) 36.

[46] N. Chatterjee, G.C. Walker, Mechanisms of DNA damage, repair, and mutagenesis, Environ. Mol. Mutagen. 58 (2017) 235–263.

[47] G.L. Moldovan, A.D. D'Andrea, How the fanconi anemia pathway guards the genome, Annu. Rev. Genet. 43 (2009) 223–249.

[48] V. Menon, L. Povirk, Involvement of p53 in the repair of DNA double strand breaks: multifaceted roles of p53 in homologous recombination repair (HRR) and non-homologous end joining (NHEJ), Subcell. Biochem. 85 (2014) 321–336.

[49] A.B. Williams, B. Schumacher, p53 in the DNA-damage-repair process, Cold Spring Harb. Perspect. Med. 6 (2016).

[50] H.C. Reinhardt, B. Schumacher, The p53 network: cellular and systemic DNA damage responses in aging and cancer, Trends Genet. 28 (2012) 128–136.

[51] J. Wu, L.Y. Lu, X. Yu, The role of BRCA1 in DNA damage response, Protein Cell 1 (2010) 117–123.

[52] R. Roy, J. Chun, S.N. Powell, BRCA1 and BRCA2: different roles in a common pathway of genome protection, Nat. Rev. Canc. 12 (2012) 68–78.

[53] L. Xie, J. Yu, W. Guo, L. Wei, Y. Liu, X. Wang, X. Song, Aldo-keto reductase 1C3 may be a new radioresistance marker in non-small-cell lung cancer, Canc. Gene Ther. 20 (2013) 260–266.

[54] W. Xiong, J. Zhao, H. Yu, X. Li, S. Sun, Y. Li, Q. Xia, C. Zhang, Q. He, X. Gao, L. Zhang, D. Zhou, Elevated expression of AKR1C3 increases resistance of cancer cells to ionizing radiation via modulation of oxidative stress, PloS One 9 (2014), e111911.

[55] G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.-H. Fridman, F. Pagès, Z. Trajanoski, J. Galon, ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks, Bioinformatics 25 (2009) 1091–1093.

[56] R. Bravo, V. Parra, D. Gatica, A.E. Rodriguez, N. Torrealba, F. Paredes, Z.V. Wang, A. Zorzano, J.A. Hill, E. Jaimovich, A.F.G. Quest, S. Lavandero, Endoplasmic reticulum and the unfolded protein response. Dynamics and metabolic integration, in: Int. Rev. Cell Mol. Biol., Elsevier Inc., 2013, pp. 215–290.

[57] C.Y. Liu, R.J. Kaufman, The unfolded protein response, J. Cell Sci. 116 (2003) 1861–1862.

[58] T.M. Drake, J.E. Ritchie, C. Kanthou, J.J. Staves, R. Narramore, L. Wyld, Targeting the endoplasmic reticulum mediates radiation sensitivity in colorectal cancer, Exp. Mol. Pathol. 98 (2015) 532–539.

[59] O.A. Barski, S.M. Tipparaju, A. Bhatnagar, The aldo-keto reductase superfamily and its role in drug metabolism and detoxification, Drug Metab. Rev. 40 (2008) 553–624.

[60] S.Q. Sun, X. Gu, X.S. Gao, Y. Li, H. Yu, W. Xiong, H. Yu, W. Wang, Y. Li, Y. Teng, D. Zhou, Overexpression of AKR1C3 significantly enhances human prostate cancer cells resistance to radiation, Oncotarget 7 (2016) 48050–48058.

[61] J.-S. Kim, J. Wook Chang, J. Kuk Park, S.-G. Hwang, Increased aldehyde reductase expression mediates acquired radioresistance of laryngeal cancer cells via modulating p53 View supplementary material, Canc. Biol. Ther. 13 (2012) 638–646.

[62] M. Safran, I. Dalah, J. Alexander, N. Rosen, T.I. Stein, M. Shmoish, N. Nativ, I. Bahir, T. Doniger, H. Krug, A. Sirota-Madi, T. Olender, Y. Golan, G. Stelzer, A. Harel, D. Lancet, GeneCards Version 3: the Human Gene Integrator, Database. 2010, 2010, baq020.

[63] D.A. Palacios, M. Miyake, C.J. Rosser, Radiosensitization in prostate cancer: mechanisms and targets, BMC Urol. 13 (2013) 4.

[64] S.T. Kim, D.S. Lim, C.E. Canman, M.B. Kastan, Substrate specificities and identification of putative substrates of ATM kinase family members, J. Biol. Chem. 274 (1999) 37538–37543.

[65] J.A. Wright, K.S. Keegan, D.R. Herendeen, N.J. Bentley, A.M. Carr, M.F. Hoekstra, P. Concannon, Protein kinase mutants of human ATR increase sensitivity to UV and ionizing radiation and abrogate cell cycle checkpoint control, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 7445–7450.

[66] R.T. Abraham, Checkpoint signaling: epigenetic events sound the DNA strand-breaks alarm to the ATM protein kinase, Bioessays 25 (2003) 627–630.

[67] C.J. Bakkenist, M.B. Kastan, DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation, Nature 421 (2003) 499–506.

[68] E.C. Lien, C.C. Dibble, A. Toker, PI3K signaling in cancer: beyond AKT, Curr. Opin. Cell Biol. 45 (2017) 62–71.

[69] Y. Yan, P.M. Greer, P.T. Cao, R.H. Kolb, K.H. Cowan, RAC1 GTPase plays an important role in γ-irradiation induced G 2/M checkpoint activation, Breast Cancer Res. 14 (2012) R60.

[70] Y. Yan, A. Hein, A. Etekpo, K. Burchett, C.L. Oncotarget, undefined, Inhibition of RAC1 GTPase sensitizes pancreatic cancer cells to γ-irradiation, Ncbi.Nlm.Nih.GovPaperpile, 2014 (n.d.), https://www.ncbi.nlm.nih.gov/pmc/articles/pmc4279370/. (Accessed 23 June 2020).

[71] M. Del Mar Maldonado, S. Dharmawardhane, Targeting rac and Cdc42 GT pases in cancer, Canc. Res. 78 (2018) 3101–3111.

[72] Y. Xie, Y.K. Liu, Z.P. Guo, H. Guan, X.D. Liu, D.F. Xie, Y.G. Jiang, T. Ma, P.K. Zhou, RBX1 prompts degradation of EXO1 to limit the homologous recombination pathway of DNA double-strand break repair in G1 phase, Cell Death Differ. 27 (2020) 1383–1397.

[73] L. Jia, J.S. Bickel, J. Wu, M.A. Morgan, H. Li, J. Yang, X. Yu, R.C. Chan, Y. Sun, RBX1 (RING box protein 1) E3 ubiquitin ligase is required for genomic integrity by modulating DNA replication licensing proteins, J. Biol. Chem. 286 (2011) 3379–3386.

[74] S.M. Adams, M.G.F. Sharp, R.A. Walker, W.J. Brammar, J.M. Varley, Differential expression of translation-associated genes in benign and malignant human breast tumours, Br. J. Canc. 65 (1992) 65–71.

[75] J.M. Wong, K.-I. Manine, H. Yow, E.N. Rivers, T.S. Ravikumar, G.D. Steele, L.B. Chen, Ubiquitin-Ribosomal Protein S27a Gene Overexpressed in Human Colorectal Carcinoma Is an Early Growth Response Gene1, 1916. https://cancerres.aacrjournals.org/content/53/8/1916.short. (Accessed 30 November 2020).

[76] H. Wang, J. Yu, L. Zhang, Y. Xiong, S. Chen, H. Xing, Z. Tian, K. Tang, H. Wei, Q. Rao, M. Wang, J. Wang, RPS27a promotes proliferation, regulates cell cycle progression and inhibits apoptosis of leukemia cells, Biochem. Biophys. Res. Commun. 446 (2014) 1204–1210.

[77] Y. Tang, Y. Geng, J. Luo, W. Shen, W. Zhu, C. Meng, M. Li, X. Zhou, S. Zhang, J. Cao, Downregulation of ubiquitin inhibits the proliferation and radioresistance of non-small cell lung cancer cells in vitro and in vivo, Sci. Rep. 5 (2015) 1–12.
[78] S.E. Scanlon, P.M. Glazer, Multifaceted control of DNA repair pathways by the hypoxic tumor microenvironment, DNA Repair 32 (2015) 180–189.
[79] D. Thierry-Mieg, J. Thierry-Mieg, AceView: a comprehensive cDNA-supported gene and transcripts annotation, Genome Biol. 7 (Suppl 1) (2006) S12, 1-14.
[80] S. Andrews, others, FastQC: a Quality Control Tool for High Throughput Sequence Data, 2010.
[81] M.A. Zolotovskaia, V.S. Tkachev, A.P. Seryakov, D.V. Kuzmin, D.E. Kamashev, M.I. Sorokin, S.A. Roumiantsev, A.A. Buzdin, Mutation enrichment and transcriptomic activation signatures of 419 molecular pathways in cancer, Cancers 12 (2020).
[82] N. Borisov, M. Suntsova, M. Sorokin, A. Garazha, O. Kovalchuk, A. Aliper, E. Ilnitskaya, K. Lezhnina, M. Korzinkin, V. Tkachev, V. Saenko, Y. Saenko, D.G. Sokov, N.M. Gaifullin, K. Kashintsev, V. Shirokorad, I. Shabalina, A. Zhavoronkov, B. Mishra, C.R. Cantor, A. Buzdin, Data aggregation at the level of molecular pathways improves stability of experimental transcriptomic and proteomic data, Cell Cycle 16 (2017) 1810–1823.
[83] S. Zheng, A.D. Cherniack, N. Dewal, R.A. Moffitt, L. Danilova, B.A. Murray, A.M. Lerario, T. Else, T.A. Knijnenburg, G. Ciriello, S. Kim, G. Assie, O. Morozova, R. Akbani, J. Shih, K.A. Hoadley, T.K. Choueiri, J. Waldmann, O. Mete, A.G. Robertson, H.-T. Wu, B.J. Raphael, L. Shao, M. Meyerson, M.J. Demeure, F. Beuschlein, A.J. Gill, S.B. Sidhu, M.Q. Almeida, M.C.B. V Fragoso, L.M. Cope, E. Kebebew, M.A. Habra, T.G. Whitsett, K.J. Bussey, W.E. Rainey, S.L. Asa, J. Bertherat, M. Fassnacht, D.A. Wheeler, S. Cancer Genome Atlas Research Network, G.D. Hammer, T.J. Giordano, R.G.W. Verhaak, A.D. Cherniack, N. Dewal, R.A. Moffitt, L. Danilova, B.A. Murray, A.M. Lerario, T. Else, T.A. Knijnenburg, G. Ciriello, S. Kim, G. Assié, O. Morozova, R. Akbani, J. Shih, K.A. Hoadley, T.K. Choueiri, J. Waldmann, O. Mete, A.G. Robertson, H.-T. Wu, B.J. Raphael, M. Meyerson, F. Beuschlein, A.J. Gill, S.B. Sidhu, M. Almeida, M.C.B. Fragoso, L.M. Cope, E. Kebebew, M.A. Habra, T.G. Whitsett, K.J. Bussey, W.E. Rainey, S.L. Asa, J. Bertherat, M. Fassnacht, D.A. Wheeler, C. Benz, A. Ally, M. Balasundaram, R. Bowlby, D. Brooks, Y.S.N. Butterfield, R. Carlsen, N. Dhalla, R. Guin, R.A. Holt, S.J.M. Jones, K. Kasaian, D. Lee, H.I. Li, L. Lim, Y. Ma, M.A. Marra, M. Mayo, R.A. Moore, A.J. Mungall, K. Mungall, S. Sadeghi, J.E. Schein, P. Sipahimalani, A. Tam, N. Thiessen, P.J. Park, M. Kroiss, J. Gao, C. Sander, N. Schultz, C.D. Jones, R. Kucherlapati, P.A. Mieczkowski, J.S. Parker, C.M. Perou, D. Tan, U. Veluvolu, M.D. Wilkerson, D.N. Hayes, M. Ladanyi, M. Quinkler, J.T. Auman, A.C. Latronico, B.B. Mendonca, M. Sibony, Z. Sanborn, M. Bellair, C. Buhay, K. Covington, M. Dahdouli, H. Dinh, H. Doddapaneni, B. Downs, J. Drummond, R. Gibbs, W. Hale, Y. Han, A. Hawes, J. Hu, N. Kakkar, D. Kalra, Z. Khan, C. Kovar, S. Lee, L. Lewis, M. Morgan, D. Morton, D. Muzny, J. Santibanez, L. Xi, B. Dousset, L. Groussin, R. Libé, L. Chin, S. Reynolds, I. Shmulevich, S. Chudamani, J. Liu, L. Lolla, Y. Wu, J.J. Yeh, S. Balu, T. Bodenheimer, A.P. Hoyle, S.R. Jefferys, S. Meng, L.E. Mose, Y. Shi, J.V. Simons, M.G. Soloway, J. . Wu, W. Zhang, K.R.M. Shaw, J.A. Demchok, I. Felau, M. Sheth, R.Z. Wang, L. Yang, J.C. Zenklusen, J. Julia, Zhang, T. Davidsen, C. Crawford, C.M. Hutter, H.J. Sofia, J. Roach, W. Bshara, C. Gaudioso, C. Morrison, P. Soon, S. Alonso, J. Baboud, T. Pihl, R. Raman, Q. Sun, Y. Wan, R. Naresh, H. Arachchi, R. Beroukhim, S.L. Carter, J. Cho, S. Frazer, S.B. Gabriel, G. Getz, D.I. . Heiman, J. Kim, M.S. . Lawrence, P. Lin, M.S. Noble, G. Saksena, S.E. Schumacher, C. Sougnez, D. Voet, H. Zhang, J. Bowen, S. Coppens, J.M. Gastier-Foster, M. Gerken, C. Helsel, K.M. Leraas, T.M. Lichtenberg, N.C. Ramirez, L. Wise, E. Zmuda, S. Baylin, J.G. Herman, J. LoBello, A. Watanabe, D. Haussler, A. Radenbaugh, A. Rao, J. Zhu, D.K. Bartsch, S. Sbiera, B. Allolio, T. Deutschbein, C. Ronchi, V.M. Raymond, M. Vinco, L. Shao, L. Amble, M.S. Bootwalla, P.H. Lai, D.J. Van Den Berg, D.J. Weisenberger, B. Robinson, Z. Ju, H. Kim, S. Ling, W. Liu, Y. Lu, G.B. Mills, K. Sircar, Q. Wang, K. Yoshihara, P.W. Laird, Y. Fan, W. Wang, E. Shinbrot, M. Reincke, J.N. Weinstein, S. Meier, T. Defreitas, G.D. Hammer, T.J. Giordano, R.G.W. Verhaak, Comprehensive pan-genomic characterization of adrenocortical carcinoma, Canc. Cell 29 (2016) 723–736.
[84] L.J. Carithers, H.M. Moore, The genotype-tissue expression (GTEx) project, Biopreserv. Biobanking 13 (2015) 307–308.
[85] N. Borisov, I. Shabalina, V. Tkachev, M. Sorokin, A. Garazha, A. Pulin, I.I. Eremin, A. Buzdin, Shambhala: a platform-agnostic data harmonizer for gene expression data, BMC Bioinf. 20 (2019) 66.
[86] N. Borisov, A. Buzdin, New paradigm of machine learning (ML) in personalized oncology: data trimming for squeezing more biomarkers from clinical datasets, Front. Oncol. 9 (2019) 658.
[87] N. Borisov, V. Tkachev, M. Suntsova, O. Kovalchuk, A. Zhavoronkov, I. Muchnik, A. Buzdin, A method of gene expression data transfer from cell lines to cancer patients for machine-learning prediction of drug efficiency, Cell Cycle 17 (2018) 486–491.
[88] V. Tkachev, M. Sorokin, A. Mescheryakov, A. Simonov, A. Garazha, A. Buzdin, I. Muchnik, N. Borisov, FLOating-window projective separator (FloWPS): a data trimming tool for support vector machines (svm) to improve robustness of the classifier, Front. Genet. 9 (2018) 717.
[89] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (2013) 15–21.
[90] B. Yates, B. Braschi, K.A. Gray, R.L. Seal, S. Tweedie, E.A. Bruford, Genenames.org: the HGNC and VGNC resources in 2017, Nucleic Acids Res. 45 (2017) D619–D625.
[91] S.E. Titov, M.K. Ivanov, E. V Karpinskaya, E. V Tsivlikova, S.P. Shevchenko, Y.A. Veryaskina, L.G. Akhmerova, T.L. Poloz, O.A. Klimova, L.F. Gulyaeva, I.F. Zhimulev, N.N. Kolesnikov, miRNA profiling, detection of BRAF V600E mutation and RET-PTC1 translocation in patients from Novosibirsk oblast (Russia) with different types of thyroid tumors, BMC Canc. 16 (2016) 201.
[92] S.-K. Yoo, S. Lee, S. Kim, H.-G. Jee, B.-A. Kim, H. Cho, Y.S. Song, S.W. Cho, J.-K. Won, J.-Y. Shin, D.J. Park, J.-I. Kim, K.E. Lee, Y.J. Park, J.-S. Seo, Comprehensive analysis of the transcriptional and mutational landscape of follicular and papillary thyroid cancers, PLoS Genet. 12 (2016), e1006239.
[93] S.-K. Yoo, Y.S. Song, E.K. Lee, J. Hwang, H.H. Kim, G. Jung, Y.A. Kim, S. Kim, S.W. Cho, J.-K. Won, E.-J. Chung, J.-Y. Shin, K.E. Lee, J.-I. Kim, Y.J. Park, J.-S. Seo, Integrative analysis of genomic and transcriptomic characteristics associated with progression of aggressive thyroid cancer, Nat. Commun. 10 (2019) 2764.
[94] N.J. Edwards, M. Oberti, R.R. Thangudu, S. Cai, P.B. McGarvey, S. Jacob, S. Madhavan, K.A. Ketchum, The CPTAC data portal: a resource for cancer proteomics research, J. Proteome Res. 14 (2015) 2707–2713.
[95] A.M. Aliper, M.B. Korzinkin, N.B. Kuzmina, A.A. Zenin, L.S. Venkova, P.Y. Smirnov, A.A. Zhavoronkov, A.A. Buzdin, N.M. Borisov, Mathematical justification of expression-based pathway activation scoring (PAS), Methods Mol. Biol. 1613 (2017) 31–51.
[96] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software Environment for integrated models of biomolecular interaction networks, Genome Res. 13 (2003) 2498–2504.