npg

This paper has been modified since advance online publication, an acknowledgement has been added

A gene expression signature distinguishes normal tissues of sporadic and radiation-induced papillary thyroid carcinomas

G Dom*,[1], M Tarabichi[1], K Unger[2], G Thomas[2], M Oczko-Wojciechowska[3], T Bogdanova[4], B Jarzab[3], JE Dumont[1], V Detours[1] and C Maenhaut[1,5]

[1]IRIBHM, Université Libre de Bruxelles, Campus Erasme, School of Medicine, Route de Lennik 808, Brussels B-1070, Belgium; [2]Human Cancer Studies Group, Department of Surgery and Cancer, Imperial College London, London W12 OHS, UK; [3]Department of Nuclear Medicine and Endocrine Oncology, Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Wybrzeze AK 15, Gliwice 44-101, Poland; [4]Laboratory of Morphology of Endocrine System, V.P. Komisarenko Institute of Endocrinology and Metabolism of National Academy of Medical Sciences of Ukraine, Vyshgorodska Street 69, Kyiv 04114, Ukraine; [5]Welbio, Université Libre de Bruxelles, Campus Erasme, CP602, Brussels B-1070, Belgium

BACKGROUND: Papillary thyroid cancer (PTC) incidence increased dramatically in children after the Chernobyl accident, providing a unique opportunity to investigate the molecular features of radiation-induced thyroid cancer. In contrast to the previous studies that included age-related confounding factors, we investigated mRNA expression in PTC and in the normal contralateral tissues of patients exposed and non-exposed to the Chernobyl fallout, using age- and ethnicity-matched non-irradiated cohorts.

METHODS: Forty-five patients were analysed by full-genome mRNA microarrays. Twenty-two patients have been exposed to the Chernobyl fallout; 23 others were age-matched and resident in the same regions of Ukraine, but were born after 1 March 1987, that is, were not exposed to ^{131}I.

RESULTS: A gene expression signature of 793 probes corresponding to 403 genes that permitted differentiation between normal tissues from patients exposed and from those who were not exposed to radiation was identified. The differences were confirmed by quantitative RT-PCR. Many deregulated pathways in the exposed normal tissues are related to cell proliferation.

CONCLUSION: Our results suggest that a higher proliferation rate in normal thyroid could be related to radiation-induced cancer either as a predisposition or as a consequence of radiation. The signature allows the identification of radiation-induced thyroid cancers.

Genetics and Genomics

Thyroid cancer is the most common form of solid neoplasm associated with radiation exposure. There has been a considerable increase in occurrence of papillary thyroid carcinomas (PTCs) after the Chernobyl power plant explosion, particularly in children and adolescents (Baverstock *et al*, 1992). This increase in incidence (up to a 100-fold) is present only in the areas of Belarus, Ukraine and Russia that lie closest to the site of the Chernobyl nuclear power plant. The incidence of thyroid cancer in these age groups is very low in unexposed populations, which provides some evidence that the majority of thyroid cancers occurring in this population is a direct result of exposure to radiation (Malone *et al*, 1991). In radiation-induced PTC, the histology and disease stage are related to the young age of patients rather than to the triggering event (Williams *et al*, 2004; Jarzab *et al*, 2005a). Spontaneous and post-Chernobyl PTC are characterised by the constitutive activation of effectors along the RAS-RAF-MAP kinase signalling pathway: in adult PTC, *BRAF* somatic mutations (frequency: 36–69%) and *RET/PTC* rearrangements (frequency: <30%) represent the most

common genetic alterations (Cohen *et al*, 2003; Kimura *et al*, 2003; Soares *et al*, 2003). In paediatric PTC (spontaneous and radio-induced), *RET/PTC* rearrangements are the most prevalent alteration (60–80%), while *BRAF* point mutation is only observed in about 4% of the cases (Nikiforov, 2002; Xing, 2005).

A number of different studies have been undertaken that set out to identify a radiation signature by comparing sporadic PTC, whose ethiology is unknown, and radiation-induced PTC. So far, four transcriptomic studies comparing radiation-induced and spontaneous thyroid cancer have been reported. We have shown that post-Chernobyl PTC had the same global molecular pheno-type as spontaneous PTC (Detours *et al*, 2005; Detours *et al*, 2007). However, they were distinguishable with molecular signatures of responses to γ-radiation and H_2O_2, and with genes involved in homologous recombination (Detours *et al*, 2007). In another study, Port *et al* (2007) reported seven genes that discriminated post-Chernobyl from German spontaneous PTC. Recently, by investigating copy number and gene expression alterations in post-Chernobyl PTC, Stein *et al* (2010) identified 141 gene expression changes presented as potential biomarkers of radiation exposure to the thyroid. As mentioned by the authors themselves, these studies harbour potential confounding factors, namely the age and the ethnicity of the patients, because young post-Chernobyl patients

were compared with adult Western European patients. Hence, besides age, differences in iodine supply, heterogeneity of stage and pathological variant-related factors may explain the reported differences in gene expression. Moreover, the overlap between those studies in term of radiation-specific signatures is quite low.

The prospective collection of thyroid tumours from patients who were born after the Chernobyl accident by the Chernobyl Tissue Bank (CTB) (www.chernobyltissuebank.com) provides a unique opportunity to compare exposed and non-exposed cases, but this time with age- and ethnicity-matched cohorts. This approach, trying to minimise variability linked to age and ethnicity, has resulted in the identification of a gain of chromosome band 7q11 associated with radiation exposure (Hess *et al*, 2011). In the study reported here, we compared the gene expression profiles of the normal contralateral tissues of PTC patients exposed and not exposed to radioiodine in the fallout from Chernobyl. This analysis provides the opportunity to assess the existence of a susceptibility to radiation that could be responsible for tumour development. We report the identification of a gene expression signature that permits discrimination between exposed and non-exposed normal thyroid tissues.

MATERIALS AND METHODS

Tissue samples

Paired RNA samples of tumoural and non-tumoural thyroid tissues were obtained from Ukraine via the CTB ($n = \sim 150$ www.chernobyltissuebank.com). Diagnoses were confirmed by the members of the International Pathology Panel of the CTB. The CTB is an established research tissue bank and is approved by both the Institutional ethics committees of the contributing organisations (in the case of this study, the Institute of Endocrinology and Metabolism, Kiev and Imperial College London), and by the Institutional Review Board of the National Cancer Institute of the United States. The available patient information, clinical and gene alteration data relative to these samples are presented in Supplementary Table 1. RNA quality was assessed using an automated gel electrophoresis system (Experion, Bio-Rad Laboratories, Nazareth Eke, Belgium). The presence of *RET/PTC* rearrangement or *BRAF* mutation in tumours was based on real-time quantitative RT-PCR (qRT-PCR) (Taqman) analyses, and genomic DNA sequencing after PCR amplification of exon 15, respectively (Powell *et al*, 2005).

Microarray experiments

The quality of RNA was assessed using an automated electrophoresis system (Experion, Bio-Rad). Only samples with RNA Quality Indicator (RQI) >7.5 were kept for the microarray analyses (for most samples: RQI >8.5/9).

RNA amplification, and cDNA synthesis and labelling were performed following Affymetrix (Santa Clara, CA, USA) protocol. Two micrograms of RNA from 22 paired RNA samples from exposed thyroid tissues (tumour and adjacent tissue) and from 23 paired RNA samples from non-exposed thyroid tissues, together with five additional non-exposed tumour samples were hybridised on Affymetrix Human Genome U133 Plus 2.0 Arrays.

Analysis of expression data

CEL file data were subjected to normalisation by GCRMA. Hierarchical clustering and principal component analysis (PCA) were conducted with GenePattern (http://www.broad.mit.edu/cancer/software/genepattern/) (Reich *et al*, 2006). Significance Analysis Of Microarray (SAM) (Tusher *et al*, 2001) was used to search for single gene expression differences (1000 permutations), and GSEA (GenePattern, MsigDB) to search for multigene

signatures allowing to distinguish classes (Subramanian *et al*, 2005). Class prediction based on leave-one-out cross-validation was performed with the *k*-nearest neighbours algorithm (KNNXValidation, GenePattern), and two supervised classification algorithms were also used to search for the best classifiers, in R version 2.11.1: Support Vector Machine (SVM, packages e1071 1.5–24) (Meyer, 2011) and Random Forest (RF, package randomForest 4.6-2) (Liaw, 2011). They were used in an inner/outer cross-validation as implemented in the MCRestimate (2.4.0) package (Ruschhaupt, 2004) with parameters of partition ci = 5 and co = 10, and repeats cr = 10. Different ranges of parameters for each algorithm were tuned in the inner cross-validation loop: VAR numbers in $\{2^3, 2^5, 2^7\}$, SVM cost equal to 0.01 or 0.1), RF node size equal to 5 or 7. As a negative control the entire inner/outer cross-validation loop was repeated with 100 permutations of the sample labels, which gave an approximation of the *P*-value for the correct classification rate. As a positive control, the entire cross-validation loop was used to classify the samples regarding the sex of the patients, a classification task for which there should exist a perfect linear separation in the normal.

Covariate adjustment

The expression of each gene was decorrelated with respect to the age at operation by taking the residuals of a robust linear fitting model with respect to the age at operation for each gene (package MASS: function *lqs* method *lqs*).

Real-time qRT-PCR

Validation of microarray results was performed by real-time qRT-PCR (SYBR green method) (Eurogentec, Liege, Belgium). The primers were designed with the Primer-3 software (http://frodo.wi.mit.edu/primer3/) and are listed in Supplementary Table 2. All PCR efficiencies, obtained with four or five serial dilutions points (ranging from 20 ng to 20 or 200 pg), were above 90% and real-time qRT-PCR was performed in duplicate for each gene. NEDD8 and TTC1 expressions were used to normalise the data, as described previously (Delys *et al*, 2007).

RESULTS

Exposed and non-exposed tumours and normal adjacent tissues have similar global expression profiles

About 150 thyroid tissues samples were received from the CTB. Samples showing RQI below 7.5 were excluded from the study and 95 samples were kept for further analysis: 45 tumour/normal paired tissues (22 exposed, 23 non-exposed) and 5 tumours from non-exposed tissues for which the normal counterpart was not available. The samples were hybridised onto Affymetrix Human Genome U133 Plus 2.0 Arrays.

We first searched for global expression differences between exposed and non-exposed normal and tumour tissues, that is, extensive differences detectable when all the genes present on our arrays were considered. To search for biologically relevant subgroups among the samples, unsupervised analyses, including hierarchical clustering and PCA, were conducted. Both analyses showed a perfect separation between normal and tumour tissues (Figure 1). To look for consistent upregulated or downregulated genes across tumour and normal tissues, we used supervised methods such as SAM, which revealed 22 289 probes that significantly differentiated tumour and normal tissues. Thus, a large fraction of the transcriptome was significantly differently regulated in PTCs compared with normal thyroid tissues (FDR <5%).

To validate our microarray data, the modulation of the following eight genes, four upregulated and four downregulated in tumours

Genetics and Genomics

compared with normal tissues, was investigated by qRT-PCR: carbonic anhydrase 12, BH3-interacting domain death agonist, clusterin, cyclin D2, trefoil factor 3, low-density lipoprotein receptor-related protein 1B, dual specificity phosphatase 1 (DUSP1) and thrombospondin, type I, domain-containing 7A. These genes were selected because they were already identified as being important in carcinogenesis. Expressions were normalised with *TTC1* and *NEDD8*, which were identified in a previous work as being the best normalisation genes for PTC, resulting from their very stable, non-regulated, expression across the samples (Delys *et al*, 2007). Similar modulation patterns were found for the expression of the eight genes comparing microarray analyses with qRT-PCR (Supplementary Figure 1).

When considering all probes, hierarchical clustering and PCA did not separate exposed and non-exposed samples (Figure 1).
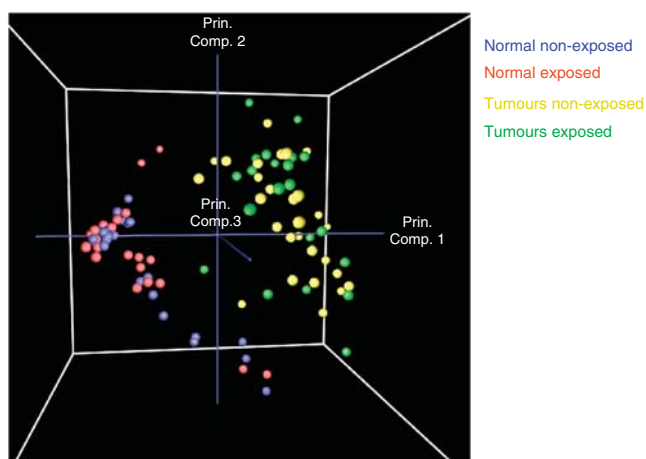

Figure 1 Global gene expression profiles of exposed and non-exposed normal and tumour tissues: PCA of the microarray data plotted with respect to first, second and third principal components. All probes were considered for the analysis. Tumour samples are shown in green (exposed) and in yellow (non-exposed), and normal samples are shown in red (exposed) and in blue (non-exposed). Abbreviation: Prin. Comp. = principal component.

Exposed and non-exposed samples did not separate either when the analysis was performed with only the normal samples or only the tumour samples. Similarly, pairing the tumour and normal samples from the same patient and considering the tumour/normal gene expression ratios led to the same result (data not shown). However, this does not exclude that a subset of genes might distinguish them. We investigated this hypothesis by conducting supervised analyses.

SAM analyses revealed differences between exposed and non-exposed normal tissues

Before the supervised analyses, we looked for the presence of potential confounding factors that may bias the results if they were unequally distributed within the two considered groups, that is, give a gene expression signature unrelated to the exposed/non-exposed conditions. We performed a systematic study of the following data (Supplementary Table 1): sex (25% males for exposed and 20% males for non-exposed), age at operation (median age at operation: 17 for exposed and 16.5 for non-exposed), date of operation, geographical origin (oblast) of the patients, PTC morphological subtype, TNM classification, presence of *BRAF* mutation or *RET/PTC* rearrangement, tumour size, percentage of epithelial cells in the samples, percentage of lymphocytic infiltration, localisation of the surgical pieces in the thyroid gland, RNA quality (small differences in RQI between exposed and non-exposed samples), hybridisation series (five different batches) and freezing time of the frozen tissue samples before RNA extraction. Only two factors were significantly associated with the radiation exposure status: the length of storage of frozen tissue samples before RNA extraction and the age of the patients at operation. The freezing time was for obvious reasons longer for the exposed samples, but there was no significant correlation between the storage length of the frozen tissue samples before RNA extraction and their quality (data not shown). Regarding age at operation, there was a small but significant difference (median: 6 months, $P = 0.006$) between the groups of exposed and non-exposed samples (Supplementary Figure 2). Significance Analysis of Microarray analysis identified genes with expression significantly associated with age. Data were adjusted in
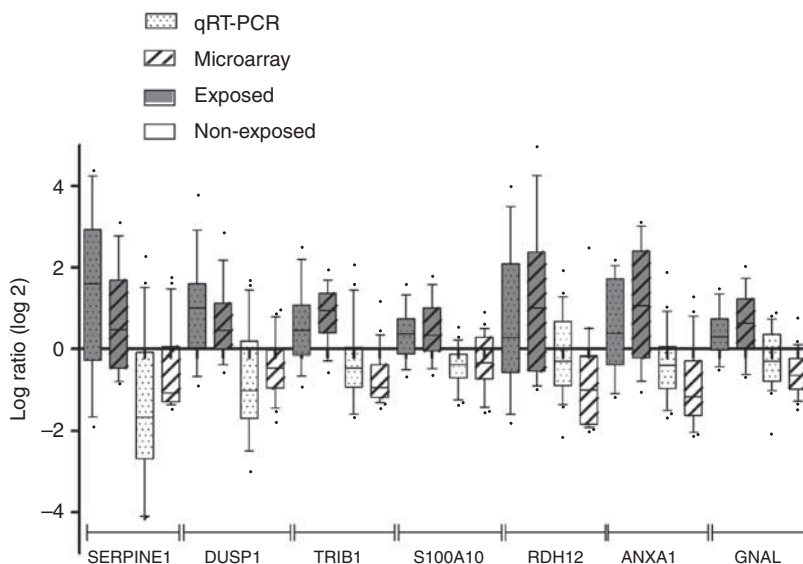
Figure 2 Comparison of differential gene expression data obtained by microarrays and qRT-PCR on exposed and non-exposed normal tissues. The upper and lower limits of each box stand for the upper and the lower quartiles, respectively; bold lines represent medians; and whiskers represent the 10–90 percentiles. Regulation of serpine peptidase inhibitor clade E (SERPINE1), DUSP1, tribbles homologue 1 (TRIB1), calcium-binding protein A10 (S100A10), retinol dehydrogenase 12 (RDH12), annexin A1 (ANXA1) and guanine nucleotide-binding protein G(olf) subunit alpha (GNAL) was confirmed on 13 exposed normal contralateral tissues and 20 non-exposed normal contralateral tissues.

order to remove age-related signals from the expression data (Materials and Methods). A hierarchical clustering on the age-adjusted data showed a perfect distinction between normal and tumour tissues, but still no distinction between exposed and non-exposed tissues (Supplementary Figure 3).

Significance Analysis of Microarray was used to compare exposed *vs* non-exposed tissues, on the age-adjusted data and identified differentially expressed genes between the normal tissues. Indeed, 793 probes, representing 403 genes, for the age-adjusted data (500 probes for the non-age-adjusted data) were found to be significantly upregulated in the exposed normal tissues ($q < 0.05$, q is a multiple-testing-adjusted confidence measure) (Supplementary Table 3, shows the 50 most regulated genes). Twenty-eight of these genes had a fold change higher than 2 (overall mean fold change: 1.53). No probe was found to be downregulated in the exposed normal tissues.

Quantitative RT-PCR analysis confirmed the expression differences for seven genes, that is, serpine peptidase inhibitor clade E, DUSP1, tribbles homologue 1, S100 calcium-binding protein A10, annexin A1, guanine nucleotide-binding protein G(olf) subunit alpha and retinol dehydrogenase 12, with similar modulation patterns for mRNA expression comparing microarray with qRT-PCR analyses (Figure 2).

When SAM was performed to compare age-adjusted expression values of exposed and non-exposed tumour samples, no significantly upregulated or downregulated probes were detected.

Validation with an external data set

The reliability of our signature was supported by similar results obtained with an external data set. Data sets for validation are very limited; however, in the context of a European Union-coordinated consortium, GENRISK-T, gene expression analyses on exposed and non-exposed samples were also carried out in the laboratory of B Jarzab (Poland). Owing to technical study/lab differences, microarrays profiles could not be meaningfully compared at the level of individual genes (Tamayo *et al*, 2007). Consequently, we used our 793 probes as a gene set and evaluated their collective expression with GSEA in the Polish data set. Collectively, these genes were regulated in the same direction in a significant manner ($P = 0.004$, NES = -1.773) (Supplementary Figure 4). These results showed that our signature was not restricted to our data set.

Biological meaning of the 793 probe signature

Investigations of the biological meaning of these 793 probes differentially expressed between normal non-exposed and exposed tissues were conducted with the DAVID (Database for Annotation, Visualisation and Integrated Discover) software (Dennis *et al*, 2003), which finds the most represented pathways or functions according to gene annotation databases such as KEGG and Gene Ontology. The most significantly altered KEGG pathways were related to cancer or proliferation, and included MAPK, insulin and mTOR signalling pathways, as well as cell adhesion, suggesting the presence of a proliferation signal in the transcriptome of the exposed normal tissues (Table 1).

The main global molecular functions that were significantly enriched in exposed normal samples were linked to nucleic acid processing, also suggesting a proliferative activity (Supplementary Table 4).

Supervised machine-learning classifiers distinguished exposed and non-exposed normal tissues with 30% error

Supervised machine-learning algorithms were used to search for a gene expression signature that predicts class membership for exposed and non-exposed normal tissues. K-nearest neighbours classification with leave-one-out cross-validation was chosen in a

Table 1 KEGG pathways enriched in exposed normal tissues and statistical significance following the analysis of the 793 probes signature with DAVID software

Term	P_{Val}	FDR
hsa05220: Chronic myeloid leukaemia	8.89^{E-06}	0.010354846
hsa04722: Neutrophin signalling pathway	2.28^{E-05}	0.02649372
hsa04010: MAPK signalling pathway	1.32^{E-04}	0.153409419
hsa 04910: Insulin signalling pathway	2.32^{E-04}	0.270422672
hsa05211: Renal cell carcinoma	6.64^{E-04}	0.770791067
hsa05212: Pancreatic cancer	8.19^{E-04}	0.949136812
hsa04810: Regulation of actin cytoskeleton	0.00126321	1.461223451
hsa03040: Spliceosome	0.00139603	1.613718043
hsa04150: mTOR signalling pathway	0.00201986	2.327103458
hsa04210: Apoptosis	0.00314594	3.602869581
hsa04510: Focal adhesion	0.00424634	4.834818104

Abbreviations: DAVID = Database for Annotation, Visualisaion and Integrated Discover; FDR = false discovery rate; MAPK = mitogen-activated protein kinase; mTOR = mammalian target of rapamycin.

Table 2 Error rates for supervised classification (based on all genes)

Classification algorithm	Exposed error	Non-exposed error	Global error
KNNXValidation	27	39	33
SVM	27	35	31
RF	31	26	29

Abbreviations: KNNXValidation = K-nearest neighbours classification with leave-one-out cross-validation; SVM = Support Vector Machine; RF = Random Forest.

first approach. The classification was run with the whole set of probes. Sixty-seven percentage of our samples were correctly classified. Furthermore, accuracies of 69% and 71% were obtained using two other linear classification algorithms, respectively, SVM and RF, and an inner/outer cross-validation protocol designed to prevent parameter and feature selection biases (Table 2). To control whether chance alone could explain these accuracies, the entire SVM and RF cross-validation loops were repeated with 100 random permutations of the sample labels. Equal or better accuracies were obtained for zero and three permutations, respectively. As a positive control, we used the exact same procedure to classify patients according to sex and obtained a 100% accuracy bringing perspective on the limits of the radiation-related transcriptional signal present in normal tissues.

DISCUSSION

The aim of this study was to investigate gene expression profiles in thyroid tumours that have arisen in the population exposed to the radioactive fallout from the Chernobyl accident (i.e., born before 26 April 1986), and to compare them with profiles of tumours of similar pathology, arising in an age-matched population, residing in the same geographical area, and born after 1 March 1987. Thus, contrary to previous studies, this work included a carefully matched control group and investigated a larger number of patients. Both tumours and their contralateral normal tissues were analysed in order to reveal a radiation signature. Although we may not exclude that our exposed cohort might contain some spontaneous PTC, they have been estimated to be <15% of the cases (Hess *et al*, 2011).

The microarray expression data confirmed previous results showing that a very large fraction of the transcriptome was dysregulated in the tumours (Huang *et al*, 2001; Jarzab *et al*, 2005b; Delys *et al*, 2007; Maenhaut *et al*, 2011). On a global scale, whereas unsupervised analyses clearly distinguished normal and tumour

<div style="writing-mode: vertical">**Genetics and Genomics**</div>

tissues, no distinction between the transcriptomes of exposed and non-exposed samples was observed. However, when using a supervised approach, SAM, differentially expressed genes between exposed and non-exposed normal tissues were detected, that is, a gene expression signature that permits discrimination between both types of samples. Such a difference was not observed among the tumours, probably because the latter have evolved into very diversified phenotypes, depending on the initial mutation, the local environment and other factors, and accordingly, are more heterogeneous (Figure 1). Thus, we may not exclude that differentially expressed genes might be revealed if a larger set of tumour samples was investigated.

Although age-matched patients were used in this study, contrary to the previous transcriptomic studies on radiation-induced cancer (Detours et al, 2005; Detours et al, 2007; Port et al, 2007; Stein et al, 2010), age was still a potential confounder with a median age difference of 6 months between the exposed and non-exposed patients. Consequently, the data were age-adjusted, that is, corrected for the correlation between age and gene expression. Age matching and age adjustment are important for several reasons. First, it has been observed that the incidence of thyroid cancer varies with age and is uncommon among children in normal conditions. Second, the risk of developing thyroid cancer after radiation exposure is higher during childhood, that is, the effects of radiation exposure on thyroid cancer development are age-dependent (Cardis et al, 2005). This is consistent with the decrease of thyroid cell proliferation with age (Coclet et al, 1989; Saad et al, 2006). Third, genetic alterations present in PTC vary with age, RET/PTC rearrangements being the most common abnormalities described in paediatric sporadic and radiation-induced PTC.

Seven hundred and ninety-three probes, corresponding to 403 genes, were shown to be differentially expressed between normal exposed and non-exposed samples in the age-adjusted data set. Although the overall differences in gene expression between the two groups were rather small, they were statistically significant ($q < 0.05$) and were confirmed by qRT-PCR for seven genes. In the field of carcinogenesis, Bozic et al (2010) showed that tumour development could result from the accumulation of multiple driver and passenger mutations, while each mutation on its own only has a little contribution to the process of cancer development. Similarly, multiple small gene expression differences could be the basis of susceptibility in our study of radiation-related PTC, and this signature may allow the identification of radiation-sensitive individuals.

There are so far no clear arguments that demonstrate that radiation affects everyone equally, and the propensity to develop cancer following exposure is likely to be variable. A genetic analysis of radiation-induced gene expression changes in immortalised human lymphocytes showed an extensive individual variation for several genes (Smirnov et al, 2009). Differences in genetic background underlie variation in the susceptibility to the effects of radiation in normal tissues (Chuang et al, 2006; Barnett et al, 2009).

A recent genome-wide association study (Takahashi et al, 2010) compared 500 000 polymorphisms in patients with PTC and in healthy Belarusian and Ukrainian subjects, all exposed to radiation from the Chernobyl fallout. An association between PTC and a polymorphism near the FOXE1 gene (TTF2) was found, but this polymorphism had also been reported previously in a non-irradiated Icelandic population (Gudmundsson et al, 2009). This study, however, had no demographically and ethnically matched control group of non-irradiated PTC patients. Thus, while it pointed out that radiation-induced and spontaneous PTC share the FOXE1 suceptibility loci, this study design could not unambiguously conclude on radiation-specific cancer predisposition loci.

Investigations about the biological meaning of our 793 probes signature highlighted significantly altered proliferation pathways, suggesting that the exposed normal tissues exhibit a proliferation signal in their transcriptome. This suggests that a higher proliferation rate would predispose to cancer after irradiation. Evidence of an association of the proliferative activity in thyroid cells with a risk of cancer after radiation exposure has been reported and may explain the higher risks of radiation-related thyroid cancer in children compared with adults (Saad et al, 2006). The levels of radiation observed after the Chernobyl accident were low to moderate, but no precise and individual radiation doses are available.

This signature might reflect radiation susceptibility, but various alternative interpretations should be considered. First, this signature might be a consequence of radiation. Radiation has potential DNA damaging and carcinogenic effects, and causes single- and double-strand breaks. Double-strand breaks are thought to be particularly important for cancer development, and represent the major effect of β-radiation, for example, ^{131}I, although the radiation effects are complex and numerous (Bourguignon et al, 2005; Harper and Elledge, 2007; Riley et al, 2008). However, DNA damage is repaired within the few hours or days after radiation exposure, and our signature would then reflect the long-lasting consequences of incorrectly repaired DNA damage. This damage would still be present in the normal tissues, but not severe enough to have induced tumourigenesis without (an) additional mutation(s) that generated the initial cancer cells.

To investigate whether radiation-related signatures could be detected in our samples, we constructed gene sets with radiation signatures published previously following analyses of post-Chernobyl PTC (Detours et al, 2007; Port et al, 2007; Stein et al, 2010; Ugolin et al, 2011) and used them in a GSEA-type analysis. None of them was found to be enriched in exposed or non-exposed normal tissues (Supplementary Table 5).

Second, this signal could be owing to the presence of microcarcinomas in the exposed normal tissues, as a result of radiation (Hayashi et al, 2010). This hypothesis is, however, unlikely, as analysis of many cases of the irradiated cohort by a panel of internationally recognised pathologists showed no evidence of an increase in microcarcinomas in this group. Moreover, to be detectable in whole-tissue gene expression analyses, such a presence should involve a significant part of the cell population.

Third, differences in iodine intake between the two cohorts might explain the signature. This proliferation signal could indeed be related to iodine deficiency or differential gland stimulation. It was proposed by some authors (Malone et al, 1991; Williams et al, 2008) that the morphological characteristics of Chernobyl-related childhood PTC were related to iodine intake and independent of radiation exposure. However, the existence of a difference in iodine dietary between the two studied Ukrainian cohorts is unlikely: Ukraine was iodine deficient before the Chernobyl disaster and is still deficient today, according to ICCIDD (International Council for Control of Iodine Deficiency Disorders) (www.iodinenetwork.net/documents/scorecard-2010.pdf) and to UNICEF. In addition, the majority of our exposed cases were between 0 and 2 years old at exposure, that is, born between 1984 and 1986, while most non-exposed cases were born between 1987 and 1990. The median date at operation was end 2001 for the exposed group and mid-2006 for the non-exposed group. Thus, the two groups had widely overlapping lifespan before surgery, and were therefore raised mostly in comparable historical context regarding iodine availability. Of course, these are global observations, and we cannot exclude individual differences in iodine intake.

Furthermore, as iodine deficiency results in an increase in TSH levels and in thyrocyte proliferation, we compared our signature with reported transcriptional signatures characterising stimulated thyroid tissues such as autonomous adenomas and familial non-autoimmune hyperthyroidism (Hebrant et al, 2009), or thyroid disorders linked to iodine deficiency such as follicular thyroid cancers. None of them was enriched in our exposed versus non-exposed normal tissues, again suggesting that our signature does not reflect differences in iodine dietary (data not shown).

In conclusion, by comparing the transcriptomes of normal contralateral tissues of PTC occurring in children exposed and non-exposed to the Chernobyl fallout, we have identified a gene expression signature that permits discrimination between both cohorts. This signature suggests the existence of a higher proliferation rate in the exposed normal thyroid tissues, which might predispose to cancer after radiation. Whether the signature reflects a susceptibility to radiation or a late effect of radiation, it gives, for a given tissue, an indication that the carcinoma was sporadic or caused by irradiation. It also suggests that decreasing the already slow renewal rate of thyroid cells (Coclet et al, 1989) by a preventive thyroxine treatment, suppressing the major trophic stimulus TSH, could prevent radiation-induced thyroid cancer. Of course, this hypothesis deserves to be tested.

REFERENCES

Barnett GC, West CM, Dunning AM, Elliott RM, Coles CE, Pharoah PD, Burnet NG (2009) Normal tissue reactions to radiotherapy: towards tailoring treatment dose by genotype. *Nat Rev Cancer* **9**(2): 134–142

Baverstock K, Egloff B, Pinchera A, Ruchti C, Williams D (1992) Thyroid-cancer after Chernobyl. *Nature* **359**(6390): 21–22

Bourguignon MH, Gisone PA, Perez MR, Michelin S, Dubner D, Giorgio MD, Carosella ED (2005) Genetic and epigenetic features in radiation sensitivity Part I: cell signalling in radiation response. *Eur J Nucl Med Mol Imaging* **32**(2): 229–246

Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, Karchin R, Kinzler KW, Vogelstein B, Nowak MA (2010) Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci USA* **107**(43): 18545–18550

Cardis E, Kesminiene A, Ivanov V, Malakhova I, Shibata Y, Khrouch V, Drozdovitch V, Maceika E, Zvonova I, Vlassov O, Bouville A, Goulko G, Hoshi M, Abrosimov A, Anoshko J, Astakhova L, Chekin S, Demidchik E, Galanti R, Ito M, Korobova E, Lushnikov E, Maksioutov M, Masyakin V, Nerovnia A, Parshin V, Parshkov E, Piliptsevich N, Pinchera A, Polyakov S, Shabeka N, Suonio E, Tenet V, Tsyb A, Yamashita S, Williams D (2005) Risk of thyroid cancer after exposure to 131I in childhood. *J Natl Cancer Inst* **97**(10): 724–732

Chuang EY, Chen X, Tsai MH, Yan H, Li CY, Mitchell JB, Nagasawa H, Wilson PF, Peng Y, Fitzek MM, Bedford JS, Little JB (2006) Abnormal gene expression profiles in unaffected parents of patients with hereditary-type retinoblastoma. *Cancer Res* **66**(7): 3428–3433

Coclet J, Foureau F, Ketelbant P, Galand P, Dumont JE (1989) Cell population kinetics in dog and human adult thyroid. *Clin Endocrinol* **31**(6): 655–665

Cohen J, Xing MZ, Mambo E, Guo ZM, Wu GG, Trink B, Beller U, Westra WH, Ladenson PW, Sidransky D (2003) BRAF mutation in papillary thyroid carcinoma. *J. Natl Cancer Inst* **95**(8): 625–627

Delys L, Detours V, Franc B, Thomas G, Bogdanova T, Tronko M, Libert F, Dumont JE, Maenhaut C (2007) Gene expression and the biological phenotype of papillary thyroid carcinomas. *Oncogene* **26**(57): 7894–7903

Dennis Jr. G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* **4**(5): 3

Detours V, Delys L, Libert F, Weiss SD, Bogdanova T, Dumont JE, Franc B, Thomas G, Maenhaut C (2007) Genome-wide gene expression profiling suggests distinct radiation susceptibilities in sporadic and post-Chernobyl papillary thyroid cancers. *Br J Cancer* **97**(6): 818–825

Detours V, Wattel S, Venet D, Hutsebaut N, Bogdanova T, Tronko MD, Dumont JE, Franc B, Thomas G, Maenhaut C (2005) Absence of a specific radiation signature in post-Chernobyl thyroid cancers. *Br J Cancer* **92**(8): 1545–1552

Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Sigurdsson A, Bergthorsson JT, He H, Blondal T, Geller F, Jakobsdottir M, Magnusdottir DN, Matthiasdottir S, Stacey SN, Skarphedinsson OB, Helgadottir H, Li W, Nagy R, Aguillo E, Faure E, Prats E, Saez B, Martinez M, Eyjolfsson GI, Bjornsdottir US, Holm H, Kristjansson K, Frigge ML, Kristvinsson H, Gulcher JR, Jonsson T, Rafnar T, Hjartarsson H, Mayordomo JI, de la CA, Hrafnkelsson J, Thorsteinsdottir U, Kong A, Stefansson K (2009) Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nat Genet* **41**(4): 460–464

Harper JW, Elledge SJ (2007) The DNA damage response: ten years after. *Mol Cell* **28**(5): 739–745

Hayashi Y, Lagarde F, Tsuda N, Funamoto S, Preston DL, Koyama K, Mabuchi K, Ron E, Kodama K, Tokuoka S (2010) Papillary micro-carcinoma of the thyroid among atomic bomb survivors: tumor characteristics and radiation risk. *Cancer* **116**(7): 1646–1655

Hebrant A, Van Sande J, Roger PP, Patey M, Klein M, Bournaud C, Savagner F, Leclere J, Dumont JE, van Staveren WC, Maenhaut C (2009) Thyroid gene expression in familial nonautoimmune hyperthyroidism shows common characteristics with hyperfunctioning autonomous adenomas. *J Clin Endocrinol Metab* **94**(7): 2602–2609

Hess J, Thomas G, Braselmann H, Bauer V, Bogdanova T, Wienberg J, Zitzelsberger H, Unger K (2011) Gain of chromosome band 7q11 in papillary thyroid carcinomas of young patients is associated with exposure to low-dose irradiation. *Proc Natl Acad Sci USA* **108**(23): 9595–9600

Huang Y, Prasad M, Lemon WJ, Hampel H, Wright FA, Kornacker K, LiVolsi V, Frankel W, Kloos RT, Eng C, Pellegata NS, De La Chapelle A (2001) Gene expression in papillary thyroid carcinoma reveals highly consistent profiles. *Proc Natl Acad Sci USA* **98**(26): 15044–15049

Jarzab B, Handkiewicz-Junak D, Wloch J (2005a) Juvenile differentiated thyroid carcinoma and the role of radioiodine in its treatment: a qualitative review. *Endocr Relat Cancer* **12**(4): 773–803

Jarzab B, Wiench M, Fujarewicz K, Simek K, Jarzab M, Oczko-Wojciechowska M, Wloch J, Czarniecka A, Chmielik E, Lange D, Pawlaczek A, Szpak S, Gubala E, Swierniak A (2005b) Gene expression profile of papillary thyroid cancer: sources of variability and diagnostic implications. *Cancer Res* **65**(4): 1587–1597

Kimura ET, Nikiforova MN, Zhu ZW, Knauf JA, Nikiforov YE, Fagin JA (2003) High prevalence of BRAF mutations in thyroid cancer: genetic evidence for constitutive activation of the RET/PTC-RAS-BRAF signaling pathway in papillary thyroid carcinoma. *Cancer Res* **63**(7): 1454–1457

Liaw A. (2011) Classification and Regression by Random Forest. Wiener M, R News 2/3, 18-22, 12-1-2011

Maenhaut C, Detours V, Dom G, Handkiewicz-Junak D, Oczko-Wojciechowska M, Jarzab B (2011) Gene expression profiles for radiation-induced thyroid cancer. *Clin Oncol* **23**(4): 282–288

Malone J, Unger J, Delange F, Lagasse R, Dumont JE (1991) Thyroid consequences of Chernobyl accident in the countries of the European community. *JEndocrinol Invest* **14**(8): 701–717

Meyer D. (2011) Support Vector Machines. The interface to libsvm in package e1070. 1–7, 5-22-2011

Nikiforov YE (2002) RET/PTC rearrangement in thyroid tumors. *Endocr Pathol* **13**(1): 3–16

Port M, Boltze C, Wang Y, Roper B, Meineke V, Abend M (2007) A radiation-induced gene signature distinguishes post-Chernobyl from sporadic papillary thyroid cancers. *Radiat Res* **168**(6): 639–649

Powell N, Jeremiah S, Morishita M, Dudley E, Bethel J, Bogdanova T, Tronko M, Thomas G (2005) Frequency of BRAF T1796A mutation in papillary thyroid carcinoma relates to age of patient at diagnosis and not to radiation exposure. *J Pathol* **205**(5): 558–564

Genetics and Genomics

1000

Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP (2006) GenePattern 2.0. *Nat Genet* **38**(5): 500–501

Riley T, Sontag E, Chen P, Levine A (2008) Transcriptional control of human p53-regulated genes. *Nat Rev Mol Cell Biol* **9**(5): 402–412

Ruschhaupt M. (2004) A compendium to ensure computational reproducibility in high-dimensional classification tasks. In *Statistical Applications in Genetics and Molecular Biology,* Huber W, Poutska A, Mansmann U (eds). Vol. 3, Issue 1, Article 37

Saad AG, Kumar S, Ron E, Lubin JH, Stanek J, Bove KE, Nikiforov YE (2006) Proliferative activity of human thyroid cells in various age groups and its correlation with the risk of thyroid cancer after radiation exposure. *J Clin Endocrinol Metab* **91**(7): 2672–2677

Smirnov DA, Morley M, Shin E, Spielman RS, Cheung VG (2009) Genetic analysis of radiation-induced changes in human gene expression. *Nature* **459**(7246): 587–591

Soares P, Trovisco V, Rocha AS, Lima J, Castro P, Preto A, Maximo V, Botelho T, Seruca R, Sobrinho-Simoes M (2003) BRAF mutations and RET/PTC rearrangements are alternative events in the etiopathogenesis of PTC. *Oncogene* **22**(29): 4578–4580

Stein L, Rothschild J, Luce J, Cowell JK, Thomas G, Bogdanova TI, Tronko MD, Hawthorn L (2010) Copy number and gene expression alterations in radiation-induced papillary thyroid carcinoma from Chernobyl pediatric patients. *Thyroid* **20**(5): 475–487

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**(43): 15545–15550

Takahashi M, Saenko VA, Rogounovitch TI, Kawaguchi T, Drozd VM, Takigawa-Imamura H, Akulevich NM, Ratanajaraya C, Mitsutake N, Takamura N, Danilova LI, Lushchik ML, Demidchik YE, Heath S, Yamada R, Lathrop M, Matsuda F, Yamashita S (2010) The FOXE1 locus is a major genetic determinant for radiation-related thyroid carcinoma in Chernobyl. *Hum Mol Genet* **19**(12): 2516–2523

Tamayo P, Scanfeld D, Ebert BL, Gillette MA, Roberts CW, Mesirov JP (2007) Metagene projection for cross-platform, cross-species characterization of global transcriptional states. *Proc Natl Acad Sci USA* **104**(14): 5959–5964

Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* **98**(9): 5116–5121

Ugolin N, Ory C, Lefevre E, Benhabiles N, Hofman P, Schlumberger M, Chevillard S (2011) Strategy to find molecular signatures in a small series of rare cancers: validation for radiation-induced breast and thyroid tumors. *PLoS One* **6**(8): e23581

Williams ED, Abrosimov A, Bogdanova T, Demidchik EP, Ito M, LiVolsi V, Lushnikov E, Rosai J, Sidorov Y, Tronko MD, Tsyb AF, Vowler SL, Thomas GA (2004) Thyroid carcinoma after Chernobyl latent period, morphology and aggressiveness. *Br J Cancer* **90**(11): 2219–2224

Williams ED, Abrosimov A, Bogdanova T, Demidchik EP, Ito M, LiVolsi V, Lushnikov E, Rosai J, Tronko MD, Tsyb AF, Vowler SL, Thomas GA (2008) Morphologic characteristics of Chernobyl-related childhood papillary thyroid carcinomas are independent of radiation exposure but vary with iodine intake. *Thyroid* **18**(8): 847–852

Xing M (2005) BRAF mutation in thyroid cancer. *Endocr Relat Cancer* **12**(2): 245–262

Genetics and Genomics