Intratumor heterogeneity and clonal evolution in an aggressive papillary thyroid cancer and matched metastases

Soazig Le Pennec¹, Tomasz Konopka¹, David Gacquer¹, Danai Fimereli¹, Maxime Tarabichi¹, Gil Tomás¹, Frédérique Savagner^{3,4}, Myriam Decaussin-Petrucci⁵, Christophe Trésallet⁶, Guy Andry⁷, Denis Larsimont⁷, Vincent Detours^{1,*} and Carine Maenhaut^{1,2,*}

¹IRIBHM, ²WELBIO, Université libre de Bruxelles (ULB), Campus Erasme, 808 Route de Lennik, 1070 Brussels, Belgium

³CHU d'Angers, Bâtiment IRIS, 4 rue Larrey, Angers F-49033, France

⁴EA 3143, Université d'Angers, F-49033 Angers, France

⁵Service d'Anatomie et Cytologie Pathologiques, Centre de Biologie Sud – Bâtiment 3D, Centre Hospitalier Lyon Sud, 69495 Pierre Bénite Cedex, France

⁶Hôpital Pitié-Salpêtrière, Université Pierre et Marie Curie, 47 Boulevard de l'Hôpital, 75013 Paris, France

⁷Institut Jules Bordet, 121 Boulevard de Waterloo, 1000 Brussels, Belgium

*V Detours and C Maenhaut contributed equally to this work

Correspondence should be addressed to C Maenhaut or V Detours
Emails
cmaenhau@ulb.ac.be or vdetours@ulb.ac.be

Abstract

The contribution of intratumor heterogeneity to thyroid metastatic cancers is still unknown. The clonal relationships between the primary thyroid tumors and lymph nodes (LN) or distant metastases are also poorly understood. The objective of this study was to determine the phylogenetic relationships between matched primary thyroid tumors and metastases. We searched for non-synonymous single-nucleotide variants (nsSNVs), gene fusions, alternative transcripts, and loss of heterozygosity (LOH) by paired-end massively parallel sequencing of cDNA (RNA-Seq) in a patient diagnosed with an aggressive papillary thyroid cancer (PTC). Seven tumor samples from a stage IVc PTC patient were analyzed by RNA-Seq: two areas from the primary tumor, four areas from two LN metastases, and one area from a pleural metastasis (PLM). A large panel of other thyroid tumors was used for Sanger sequencing screening. We identified seven new nsSNVs. Some of these were early events clonally present in both the primary PTC and the three matched metastases. Other nsSNVs were private to the primary tumor, the LN metastases and/or the PLM. Three new gene fusions were identified. A novel cancer-specific KAZN alternative transcript was detected in this aggressive PTC and in dozens of additional thyroid tumors. The PLM harbored an exclusive whole-chromosome 19 LOH. We have presented the first, to our knowledge, deep sequencing study comparing the mutational spectra in a PTC and both LN and distant metastases. This study has yielded novel findings concerning intra-tumor heterogeneity, clonal evolution and metastases dissemination in thyroid cancer.

Key Words

- ▶ clonal evolution
- ▶ RNA-Seq
- ▶ tumor heterogeneity
- ▶ thyroid tumors

Endocrine-Related Cancer
(2015) 22, 205–216

Introduction

Follicular-cell-derived thyroid cancers include several histological types including papillary thyroid cancer (PTC), follicular thyroid cancer (FTC), poorly differentiated thyroid cancer (PDTC), and anaplastic thyroid cancer (ATC; Sipos & Mazzaferri 2010). PTC and FTC are differentiated thyroid carcinomas (DTCs). They generally exhibit an indolent clinical course but some develop local invasion and, less frequently, distant metastases. PDTC and ATC are associated with high mortality. Point mutations, genomic rearrangements, and chromosomal copy number alterations have been linked to thyroid carcinogenesis. Mutations in *BRAF* and *RAS* genes and *RET/PTC* and *PAX8/PPAR γ* rearrangements are often found in DTCs (Xing et al. 2013). These genetic defects are believed to be responsible for the initiation of thyroid carcinogenesis, but the mechanisms involved in thyroid cancer metastasis are not well defined. Characterization of tumor and metastases heterogeneity and evolution is very important for the treatment and management of metastatic disease. However, the contribution of intratumor heterogeneity to thyroid metastatic cancers is still unknown. The clonal relationships between the primary thyroid tumors and lymph nodes (LN) or distant metastases are also poorly understood. This could be partly explained by limited access to tissue samples from associated primary tumors and metastatic tissues, especially for distant metastatic tissues. Previous studies examining thyroid tumor heterogeneity presented discordant conclusions (Kim et al. 2006, Park et al. 2006, Abrosimov et al. 2007, Giannini et al. 2007, Guerra et al. 2012, de Biase et al. 2014, Walts et al. 2014). These divergent conclusions could be explained by differences in the cohorts studied and in addition by the use of technologies that allowed for detection of only one or a few pre-determined genetic aberrations associated with thyroid cancer, such as those in *BRAF*. The use of next generation sequencing provides unprecedented opportunities to analyze the mutational spectra in cancer samples.

In the current study, we performed paired-end massively parallel sequencing of cDNA (RNA-Seq) from a patient diagnosed with an aggressive PTC to determine the phylogenetic relationships between matched primary thyroid tumor and metastases. To our knowledge, this is the first deep sequencing study comparing a primary thyroid tumor with LNs and distant metastases of the same patient on the basis of multiple genetic alterations, i.e. non-synonymous single-nucleotide variants (nsSNVs),

gene fusions, alternative transcripts, and loss of heterozygosity (LOH).

Materials and methods

Case description

A 71-year-old male was diagnosed with a PTC in the right lobe of the thyroid with concurrent metastatic involvement of the right LNs and many pulmonary metastases. This corresponded to a stage IVc PTC (pT4aN1bM1). He was treated with a total thyroidectomy and dissection of the right LNs (six LN metastases in the recurrential and antero-posterior mediastinal area and two metastases in the jugulo-carotid area; Fig. 1). Iodine-131 radiotherapy was administered 1 and 7 months after surgery. Eight months after thyroidectomy, a thoracoscopy was performed for a recurrent pleural effusion of neoplastic origin and two pleural metastases were collected. Eleven months after thyroidectomy, multiple right cervical metastases with invasion of neck soft tissues were observed and dissection of LNs was performed; two malignant LN were collected. The patient died 12 months after thyroidectomy.

Histopathological examination revealed a PDTC originating from an area of follicular variant PTC. Results of immunohistochemical analysis confirmed the thyroid origin of the metastases with staining for thyroid transcription factor 1 and thyroglobulin. Sanger sequencing results revealed monoallelic Q61R *NRAS* somatic mutation in both the primary tumor and the three metastases analyzed.

Tissues samples

For RNA-Seq analysis, tumor and normal thyroid tissues were obtained from the Jules Bordet Institute tissue bank (Brussels, Belgium). The patient gave written informed consent. We studied different areas of the primary PTC (PRT1 and PRT2), one pleural metastasis (PLM), and two LN metastases (LN1.1, LN1.2 and LN2.1, LN2.2). In anatomicopathologists' estimations tumor cell fractions of 90% for the primary tumor and the two LN metastases and 95% for the PLM were reported. For Sanger sequencing analyses in additional samples (in search of SNVs, gene fusions, and alternative transcripts), tissues were obtained from the following tissue banks: Jules Bordet Institute, Pitié-Salpêtrière Hospital (Paris, France), Angers Hospital (Angers, France), and Lyon Sud Hospital (Lyon, France).

Figure 1

Case and samples description. The different surgeries and iodine-131 radiotherapies are indicated. Both mRNA and genomic DNA samples were obtained from the primary thyroid tumor (PRT1 and PRT2), the lymph node metastases 1 (LN1.1 and LN1.2), and 2 (LN2.1 and LN2.2) and the pleural

metastasis (PLM). Only genomic DNA was available for normal thyroid tissue (N). Histopathological images ($\times 10$) of the normal thyroid tissue, the primary tumor, and the three metastases are also shown.

In accordance with the French Public Health Code (articles L. 1243-4 and R. 1243-61), samples were issued from care and were reclassified for research. As available RNA quantity was low for some samples, the panel was not exactly the same for all the Sanger sequencing analyses. For the SNVs screening panel, 115 samples (primary tumors and associated metastases) obtained from 91 patients diagnosed with thyroid carcinoma (87 PTC, one PDTC, and three ATC) were analyzed. For the gene fusions panel, 94 samples (primary tumors and associated metastases) obtained from 62 patients diagnosed with thyroid carcinoma (59 PTC, one PDTC, and two ATC) were analyzed. For the alternative transcript panel, samples from 73 PTC patients (96 samples of primary PTC and associated metastases), 11 FTC, one PDTC, nine ATC, 16 follicular adenomas (FA), 12 hyperfunctioning autonomous adenomas (AA), 60 normal thyroid tissues adjacent to KAZN-alternative-transcript-positive and -negative (23 and

37 respectively) tumors, and a pool of 23 normal thyroid tissues were analyzed. Protocols were approved by the ethics committees of the institutions. The tissues were immediately dissected, snap-frozen in liquid nitrogen, and stored at -80°C until RNA and DNA processing were performed.

Extraction and quality assessment of nucleic acids

DNA from tumors and normal thyroid tissues was extracted using the DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's recommendations. Total RNA was extracted from tissues using TRIzol (Invitrogen), followed by purification on RNeasy columns (Qiagen). RNA and DNA concentrations were spectrophotometrically quantified, and the integrity of nucleic acids was checked using an automated gel electrophoresis system (Experion, Bio-Rad).

Whole-genome amplification and RT

In order to increase the amount of available genomic DNA from normal and tumor samples, whole-genome amplification was performed on 20 ng genomic DNA using the REPLI-g Kit (Qiagen) following the manufacturer's instructions. To obtain cDNA, after a DNase treatment using DNase I Amplification Grade (Invitrogen), RT of 1 µg RNA was performed using SuperScript II RNase H Reverse Transcriptase (Invitrogen) according to the manufacturer's recommendations.

Sanger sequencing of genomic DNA and cDNA samples

To look for mutations and gene fusions, PCRs were performed on 80 ng of amplified genomic DNA or 100 ng of cDNA using the Recombinant Taq DNA Polymerase Kit (Invitrogen) and appropriate primer pairs (primer sequences and PCR conditions are provided in [Supplementary Table 1](#), see section on [supplementary data](#) given at the end of this article). PCR products were purified with the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's instructions. Sequencing was performed with the BigDye Terminator V3.1 Cycle Sequencing Kit (Applied Biosystems) on an ABI PRISM 3130 sequencer (Applied Biosystems) using the genetic analysis program 3130-XI.

Transcriptome sequencing, variant calling, and allelic proportions analyses

RNA-Seq analysis was performed on eight RNA samples: a reference pool of 23 normal thyroid tissues (used as a technical control) and seven tumor samples from the same patient (PRT1, PRT2, LN1.1, LN1.2, LN2.1, LN2.2, and PLM). The normal thyroid tissue from this patient was excluded from the analysis because of the poor quality of its RNA sample. Paired-end reads of 51 bases were generated using HiSeq 2000 (Illumina, San Diego, CA, USA) on two lanes of the sequencer's flow cell.

Raw data were scanned using the TriageTools program ([Fimereli et al. 2013](#)) to search for exact duplicates and only unique reads were passed on to the alignment. Sequences were mapped to the human genome (*Homo sapiens* UCSC hg19) using TopHat v2 ([Trapnell et al. 2009](#)). Variant calling was performed using an in-house toolkit (<http://sourceforge.net/projects/bamformatics/>) that considers each covered locus to compute the allelic proportion P and an associated uncertainty δP (estimated using a Poisson error model). The formulae are:

$$P = \frac{a}{a + \bar{a}} \quad \text{and} \quad \delta P = \frac{\sqrt{(a + \xi)(\bar{a} + \xi)}}{(a + \bar{a} + 2\xi)^{3/2}}$$

Where a is the number of reads showing the alternative base and $\bar{a}$ is the number of reads showing other bases (the reference base and/or other bases). The parameter ξ is a dark count and is hard-coded to the value 1. The score z for a given variant, referred to as a variant quality, is computed using

$$z = 10 \frac{P}{\delta P}$$

SNVs were filtered according to the following conditions: i) the score z required to report a variant should be ≥ 18 ; ii) the allelic proportion at the position should be ≥ 0.05 for the SNV; iii) the number of reads showing a base different from the reference should be ≥ 3 ; iv) the distance from the 5'-end and from the 3'-end of a read (including soft-clipped and trimmed bases) for a variant to be recorded should be ≥ 5 ; and v) the read mapping quality should be ≥ 4 . After filtering, SNVs were manually inspected using Integrative Genomics Viewer (IGV) ([Robinson et al. 2011](#), [Thorvaldsdóttir et al. 2013](#)) to further screen for sequencing and alignments errors. Called variants were compared with the dbSNP135 database. The calls that matched the database exactly (same locus and exact genotype) were not considered for discovery of new somatic SNVs.

In order to detect regions of putative LOH, allelic proportions of variants in paired matched samples were measured and segmented using the Locsmoc algorithm ([Tarabichi et al. 2012](#)). Locsmoc allows for the detection of unusual allelic proportions in entire regions by considering groups of nearby variants even though the coverage of the individual variant loci may be insufficient to call them separately. Aberrations detected by the algorithm are highlighted on full-genome segmentation heatmaps.

Candidate fusion transcripts identification

RNA-Seq reads were passed to two software packages, deFuse (deFuse-0.4.3) ([McPherson et al. 2011](#)) and TopHat-Fusion (TopHat v2) ([Kim & Salzberg 2011](#)). For TopHat-Fusion, indexes were downloaded and parameters set according to authors' recommendations (<http://tophat-fusion.sourceforge.net/tutorial.html>). In order to detect both inter- and intra-chromosomal rearrangements and read-through transcripts, we set the parameters to consider alignments to be fusion candidates when the two 'sides' of the event either reside on different chromosomes or reside on the same chromosome and are separated by at least 100 bp. For deFuse, we also used the default parameters.

To explore the *KAZN* alternative transcript in more detail with RNA-Seq data, we attempted *de-novo* assembly of *KAZN*-related transcripts. We first used Triagetools to extract reads with sequence similarity to genomic region chr1: 14 219 052–15 444 544, which corresponds to *KAZN* and *C10RF196*. We then applied Trinity (trinityrna-seq_r20131110) (Grabherr *et al.* 2011) to assemble the extracted reads into contigs with lengths of 200 bp or more.

Results

Somatic point mutations revealed clonal heterogeneity between primary tumor and metastases and within metastases

Analysis of RNA-Seq data confirmed the NRASQ61R somatic mutation in all samples. We also identified 33 new nsSNVs by RNA-Seq and validated them using Sanger sequencing (Supplementary Table 2, see section on supplementary data given at the end of this article). Among these, 26 were also detected by Sanger sequencing in DNA from the patient's normal thyroid but not present in the dbSNP135 database, indicating that these SNVs were rare germline polymorphisms. The remaining seven were somatic nsSNVs (Table 1). The new nsSNVs occurred in six genes: the neurolysin (metallopeptidase M3 family) (*NLN*), actin-like 6A (*ACTL6A*), *DEAH* (Asp-Glu-Ala-His) box polypeptide 16 (*DHX16*), eukaryotic translation initiation factor 2- α kinase 1 (*EIF2AK1*), adenosine monophosphate deaminase 2 (*AMPD2*), and tRNA nucleotidyl transferase, CCA-adding, 1 (*TRNT1*).

The allelic proportions of the somatic mutations were consistent between RNA- and DNA-based Sanger sequencing and RNA-Seq (data not shown). *NRAS* p.Q61R, *NLN* p.S122R, and *NLN* p.A125V were clonal mutations detected in the primary tumor, the two LN metastases, and the PLM. This indicates that these SNVs were early events that arose before the development of metastases in a common ancestral clone. *ACTL6A* p.K379E was detected only in PRT1 and PRT2 in sub-clonal proportions. The other nsSNVs were metastases-specific. The *DHX16* p.D189Y mutation was present in the four areas of the two LN metastases, indicating that these two metastases originated from a common ancestor. In the LN1.2 area, an *EIF2AK1* p.I604L SNV was also detected in subclonal proportions, indicating greater intra-tumor heterogeneity in the LN1 than in the LN2 metastasis. The same *EIF2AK1* mutation was also detected in the PLM sample. Even if we cannot exclude the possibility that this SNV was already present in a subclone of the primary tumor not detected by RNA-Seq,

Table 1 Frequency of somatic nsSNVs detected by RNA-Seq in the primary thyroid tumor and the three metastases

Position	Gene	Ref	Alt	AA change	Primary tumor						Lymph node 1						Lymph node 2						Pleural		
					PRT1			PRT2			LN1.1			LN1.2			LN2.1			LN2.2			PLM		
					AP	cov	S	AP	cov	S	AP	cov	S	AP	cov	S	AP	cov	S	AP	cov	S	AP	cov	S
chr1: 115256529	<i>NRAS</i>	T	C	p.Q61R	0.48	21	Y	0.49	53	Y	0.38	71	Y	0.50	26	Y	0.42	55	Y	0.51	37	Y	0.28	40	Y
chr5: 65058849	<i>NLN</i>	A	C	p.S122R	0.48	21	Y	0.46	41	Y	0.33	33	Y	0.42	12	Y	0.25	20	Y	0.33	18	Y	0.41	17	Y
chr5: 65058859	<i>NLN</i>	C	T	p.A125V	0.50	18	Y	0.43	40	Y	0.24	33	Y	0.40	10	Y	0.40	20	Y	0.25	12	Y	0.40	15	Y
chr3: 179304346	<i>ACTL6A</i>	A	G	p.K379E	0.27	26	Y	0.15	27	Y	0	78	N	0	20	N	0	64	N	0	28	N	0	29	N
chr6: 30638611	<i>DHX16</i>	C	A	p.D189Y	0	14	N	0	33	N	0.40	52	Y	0.15	27	Y	0.08	39	Y	0.22	23	Y	0	23	N
chr7: 6064387	<i>EIF2AK1</i>	T	G	p.I604L	0	92	N	0	167	N	0	277	N	0.18	94	Y	0	199	N	0	96	N	0.27	66	Y
chr1: 110170817	<i>AMPD2</i>	G	T	p.R452L	0	8	N	0	24	N	0	39	N	0	13	N	0	31	N	0	16	N	0.40	15	Y
chr3: 3170855	<i>TRNT1</i>	G	C	p.G44A	0	11	N	0	10	N	0	36	N	0	6	N	0	22	N	0	10	N	0.40	15	Y

RNA-Seq analysis was done on samples from the primary thyroid tumor (PRT1 and PRT2), the lymph node metastases 1 (LN1.1 and LN1.2), and 2 (LN2.1 and LN2.2) and the pleural metastasis (PLM). The genome positions are shown as chr, coordinates; Ref, reference base; Alt, alternative base; AA change, amino acid change; AP, alternative allelic proportion; cov, coverage, i.e. number of reads at the locus; S, detection of the mutation by Sanger sequencing in both tumor DNA and RNA (Y) or in neither (N).

this rather indicates that the clone that seeded the PLM originated from LN1.2. *AMPD2* p.R452L and *TRNT1* p.G44A were private SNVs detected, in clonal proportions, only in the PLM. This may indicate that they were acquired *de novo* in the pleura or during dissemination of cancer cells.

None of the seven new mutations has been reported in other tumor types according to the catalogue of somatic mutations in cancer (COSMIC v67). *NLN* p.S122R and *NLN* p.A125V are located in the N-terminal neurolysin/thimet oligopeptidase domain of NLN protein, and *AMPD2* p.R452L is located in the AMP deaminase domain of AMPD2 enzyme (Supplementary Table 3, see section on supplementary data given at the end of this article).

To determine whether these SNVs are recurrent in thyroid cancer, we searched for the presence of *NLN* p.S122R, *NLN* p.A125V, *DHX16* p.D189Y, *EIF2AK1* p.I604L, *AMPD2* p.R452L, and *TRNT1* p.G44A by RT-PCR followed by Sanger sequencing in 115 supplementary thyroid tumor samples. None of the mutations was detected in these tumors.

The PLM harbored an exclusive whole-chromosome 19 LOH

In order to detect the regions of putative LOH in the tumor samples studied by RNA-Seq, allelic proportions analysis from variant calls was performed on paired matched

samples. Allelic aberrations were observed on whole-chromosome 19 in the PLM sample relative to PRT1 or PRT2 (Fig. 2), indicating the presence of LOH of the entire chromosome 19 in the PLM. Chromosome 19 LOH was also observed in the PLM when the PLM sample was compared with the LN1.1, LN1.2, LN2.1, or LN2.2 samples (data not shown).

To validate this LOH, RT-PCR followed by Sanger sequencing analysis was performed on genomic DNA samples from the primary thyroid tumor, the two LN metastases, and the PLM at different loci matching germline SNVs all along the chromosome 19. Sanger sequencing results were in accordance with the allelic proportions detected by the RNA-Seq analysis (Table 2 and Supplementary Fig. 1, see section on supplementary data given at the end of this article) and confirmed the presence of LOH of the entire chromosome 19 in the distant metastasis.

Identification of new gene fusions and of a new alternative transcript

We identified four novel gene fusions detected by both deFuse and TopHat-Fusion algorithms in at least one of the seven RNA-Seq tumor samples (Supplementary Table 4, see section on supplementary data given at the end of this article). These four fusions were validated using RT-PCR

Figure 2

Allelic proportions analysis. In order to detect the regions of putative loss of heterozygosity in the primary thyroid tumor (PRT1 and PRT2), the lymph node metastases 1 (LN1.1 and LN1.2), and 2 (LN2.1 and LN2.2) and the pleural metastasis (PLM), allelic proportions analysis from variant calls was

performed on paired matched samples data (each sample vs PRT1 and each sample vs PRT2). Allelic aberrations are highlighted in pink on full-genome segmentation heatmaps.

Table 2 Validation of loss of heterozygosity of chromosome 19 in the pleural metastasis by Sanger sequencing

Position	Gene	Ref	Alt	dbsNP135	Primary tumor				Lymph node 1				Lymph node 2				Pleural	
					PRT1		PRT2		LN1.1		LN1.2		LN2.1		LN2.2		PLM	
					AP	S DNA	AP	S DNA	AP	S DNA	AP	S DNA	AP	S DNA	AP	S DNA	AP	S DNA
chr19: 648972	RNF126	G	A		0.29	G + A	0.47	G + A	0.49	G + A	0.40	G + A	0.36	G + A	0.61	G + A	0.96	A (+G)
chr19: 1923155	SCAMP4	C	T		0.55	C + T	0.46	C + T	0.48	C + T	0.46	C + T	0.46	C + T	0.56	C + T	0.16	C (+T)
chr19: 2807014	THOP1	A	G		0.57	A + G	0.50	A + G	0.47	A + G	0.42	A + G	0.45	A + G	0.52	A + G	0.85	G (+A)
chr19: 4362349	SH3GL1	A	G	rs146037902	0.55	A + G	0.57	A + G	0.41	A + G	0.39	A + G	0.41	A + G	0.55	A + G	0.82	G (+A)
chr19: 40331117	FBL	G	A	rs139739777	0.48	G + A	0.41	G + A	0.43	G + A	0.44	G + A	0.41	G + A	0.42	G + A	0.16	G (+A)
chr19: 40357398	FCGBP	C	G	rs140253776	0.35	C + G	0.34	C + G	0.32	C + G	0.54	C + G	0.26	C + G	0.38	C + G	0.97	G
chr19: 41129562	LTBP4	G	A		0.70	G + A	0.54	G + A	0.62	G + A	0.54	G + A	0.63	G + A	0.53	G + A	0.01	G
chr19: 54652265	CNOT3	C	T		0.54	C + T	0.18	C + T	0.59	C + T	0.70	C + T	0.43	C + T	0.44	C + T	0	C (+T)
chr19: 56297164	NLRP11	T	G	rs116820715	0.55	T + G	0.42	T + G	0.64	T + G	0.67	T + G	0.70	T + G	0.61	T + G	1.00	G

To search for loss of heterozygosity, RT-PCR followed by Sanger sequencing analysis was done on genomic DNA samples from the primary thyroid tumor (PRT1 and PRT2), the lymph node metastases 1 (LN1.1 and LN1.2), and 2 (LN2.1 and LN2.2) and the pleural metastasis (PLM) at different loci of germline SNVs. The genome positions of the germline SNVs are shown as chr:coordinates. rsID of SNVs annotated in dbSNP135 database are indicated. Ref, reference base; Alt, alternative base; AP, alternative allelic proportion; DNA Sanger sequencing results (S DNA) are shown as Ref + Alt when the two peaks were detected in equivalent proportions and as Ref (+Alt) or Alt (+Ref) when one of the peaks, the base indicated in brackets, was detected in very low proportions. For each locus, two peaks of equivalent size, corresponding to the two alleles of the SNV, were observed in the primary tumor and the LN samples. In contrast, in the pleural metastasis the balance of the two alleles changed: only one peak, or one very weak peak and one very high peak, was observed.

followed by Sanger sequencing across the fusion point (Fig. 3). Of these fusions, two involve genes on separate chromosomes (*C11ORF58-SBSPON* and *ARNTL-PTPRD*), one is an intra-chromosomal fusion (*KIAA2026-MIR31HG*) and one (*KAZN-C1ORF196*) is predicted by deFuse to result from alternative splicing that, unlike the other three fusions, generates an in-frame protein. *C11ORF58-SBSPON*, *ARNTL-PTPRD*, and *KIAA2026-MIR31HG* fusions were detected by the algorithms in the PLM sample only and were more strongly detected by RT-PCR in the PLM sample than in the other tumor areas. Interestingly, in each tumor area the three fusions were detected in equivalent proportions, indicating that these fusions were together present in the same cells. They were detected at equivalent levels in PRT1 and PRT2, and in LN2.1 and LN2.2, but were detected at higher levels in the LN1.2 than in the LN1.1 sample. The *KAZN-C1ORF196* fusion was detected by the algorithms in PRT2, LN1.1, LN1.2, and LN2.2 samples and strongly detected by RT-PCR in all the RNA-Seq tumor areas. The gene *C1ORF196* has been removed from the most recent versions of the genome annotation, therefore we will refer to the *KAZN-C1ORF196* fusion as 'KAZN alternative transcript' hereafter. To explore the KAZN alternative transcript in more detail with RNA-Seq data, we attempted *de novo* assembly of KAZN-related transcripts. This analysis produced contigs that confirmed the presence of an alternative KAZN transcript with a breakpoint situated in the 5'UTR of the canonical KAZN A isoform, but failed to generate a full-length transcript sequence (Supplementary Figures 2 and 3, see section on supplementary data given at the end of this article). The sequence gaps could explain the presence of several different contigs instead of a full-length contig. The full form of the novel transcript (and protein) is therefore still unclear and additional experiments using an independent method will be necessary for a complete characterization of the KAZN alternative transcript sequence in the future.

To determine whether these fusions were recurrent in thyroid tumors, we explored a large panel of thyroid samples by RT-PCR followed by Sanger sequencing. The *C11ORF58-SBSPON* fusion was excluded from the screening analysis as the Sanger sequence corresponding to the *C11ORF58* part of the fusion was highly similar to nucleotide sequences located on two different chromosomes (chromosomes 11 and 14). We examined the *ARNTL-PTPRD* and *KIAA2026-MIR31HG* fusions in 94 additional thyroid tumor samples. None of the samples contained the fusions. In contrast, we were able to detect the KAZN alternative transcript in a large number of

Figure 3

Biological validation of gene fusions identified from the RNA-Seq analysis by both the deFuse and TopHat-Fusion software. To search for the presence of gene fusions, RT-PCR followed by Sanger sequencing analysis was performed on cDNA samples from the primary thyroid tumor (PRT1 and PRT2), the lymph node metastases 1 (LN1.1 and LN1.2) and 2 (LN2.1 and

LN2.2), and the pleural metastasis (PLM). RT-PCR-derived products confirming the *C11ORF58-SBSPON*, *ARNTL-PTPRD*, *KIAA2026-MIR31HG*, and *KAZN-C1ORF196* fusions were detected on agarose gels. *GAPDH* was used as a positive control. Representative chromatograms of Sanger sequences at the fusion junctions are shown.

thyroid tumors (146 tumor samples were interrogated). This transcript was present in 85%, 55%, and 11% of the PTC, FTC, and ATC patients respectively and in the PDTC sample. It was very weakly detected in one FA but absent in the AAs. Notably, this transcript was very weakly detected in only two normal tissues adjacent to *KAZN*-alternative-transcript-positive PTC (23 normal tissues analyzed). It was absent in the 37 normal thyroid tissues adjacent to *KAZN*-alternative-transcript-negative tumors and in a pool of 23 normal thyroid tissues.

Discussion

Our RNA-Seq analysis of an aggressive PTC and its associated metastases in a single patient identified several

types of genetic alterations, i.e. nsSNVs, LOH, gene fusions, and an alternative transcript. We cannot exclude the possibility that additional aberrations could be detected with deeper sequencing or with additional sequencing assays. Nonetheless, the data already paint a detailed picture of the patient's cancer history.

We searched for unusual allelic proportions in genomic areas. Our analyses revealed similar allelic proportions in the primary tumor and the LN metastases, but highlighted the presence of LOH of the entire chromosome 19 in the PLM. Interestingly, results from several comparative genomic hybridization (CGH) studies have indicated that a gain of chromosome 19 is positively associated with thyroid tumor aggressivity and recurrence, including lung metastases (Tallini *et al.* 1999, Wada *et al.* 2002).

Figure 4

Genetic evolution between the primary PTC and associated metastases. (A) Summary of genetic alterations detected by RNA-Seq and RT-PCR followed by Sanger sequencing analyses in the primary PTC and associated lymph nodes and pleural metastases. nsSNVs were detected in clonal or subclonal proportions by combined analysis of RNA-Seq allelic proportions and Sanger chromatograms (for RNA and DNA samples). WT indicates that only the WT nucleotide was detected at the SNV locus. LOH of the entire chromosome 19 was indicated by the results of allelic proportions analysis from variant calls performed on RNA-Seq data from paired matched samples and validated by Sanger sequencing analysis of genomic DNA from the tumor samples. The presence of the three fusions *C11ORF58-SBSPON*, *ARNTL-PTPRD*, *KIAA2026-MIR31HG*, and the *KAZN* alternative transcript

detected by the deFuse and TopHat-Fusion algorithms was validated by RT-PCR followed by Sanger sequencing analysis of RNA samples. After migration on agarose gels, the RT-PCR products were either undetected (-), very weakly detected (+/-), well detected (+), very well detected (++), or strongly detected (+++). (B) Clonal evolution between the primary PTC and associated metastases. Dissimilarities between samples were scored using Manhattan distances based on features in (A); SNVs were scored by their allelic proportions as detected by RNA-Seq and validated by Sanger sequencing analysis in genomic DNA, LOH and fusions were scored into five bins with values in (0,1). Dissimilarities were then visualized in a multi-dimensional scaling (MDS) plot.

We screened for the new nsSNVs and gene fusions in dozens of additional tumors by RT-PCR and Sanger sequencing. None of these alterations was detected in additional samples, indicating that they may be passenger, i.e. irrelevant to oncogenesis, or rare driver events for thyroid cancer. This is in accordance with the results of a deep sequencing study by Berger *et al.* (2010): each fusion product (initially detected by RNA-Seq) was detectable by RT-PCR only in the original melanoma sample from which it was discovered and not in the additional 90 interrogated. This result is also consistent with recent exome data of the Cancer Genome Atlas (TCGA) on 401 PTCs: very few recurrent mutations are present in PTCs beyond *BRAF*, *NRAS*, and *HRAS* (http://gdac.broadinstitute.org/runs/analyses__2014_01_15/reports/cancer/THCA/MutSigNozzleReportMerged/nozzle.html). Smalldridge *et al.* (2014) have recently identified new recurrent fusion transcripts in 20 PTC. However, as shown for ovarian cancer, the recurrence of such fusions is questionable and additional studies using a larger cohort will be necessary to determine the frequency and the recurrence of these fusions in thyroid cancer (Salzman *et al.* 2011, Micci *et al.* 2014).

The novel unannotated *KAZN* alternative transcript was recurrently detected by RT-PCR in 11% of the nine ATC,

55% of the 11 FTC, and 85% of the 73 PTC patients. Interestingly, except for one case, when both primary PTC and associated metastases were available, the presence or absence of this transcript was concordant between the different samples for a patient. This indicates that the acquisition of the *KAZN* alternative transcript is an early event in thyroid tumorigenesis. This transcript was also very weakly detected in one of the 16 FA, but absent in the 12 AA. Notably, among all the normal tissues, the *KAZN* transcript was detectable at very low levels in only two normal tissues adjacent to *KAZN*-alternative-transcript-positive PTC, possibly resulting from the presence of contaminating tumor cells. This indicates that this *KAZN* transcript is cancer-specific and that it might be used as a new diagnostic marker for distinguishing benign thyroid tumors from malignant thyroid tumors. Other *KAZN* transcripts have been shown to be involved in intercellular adhesion, cell differentiation, signal transduction, cytoskeletal organization, and apoptosis (Sevilla *et al.* 2008, Nachat *et al.* 2009, Wang *et al.* 2009). The disruption of these pathways can promote cancer. Functional analyses will be necessary to determine the biological significance of the *KAZN* alternative transcript we detected in our study, but these results indicate that this transcript might play a role in the development and/or progression of thyroid cancer.

Our study also established the phylogenetic relationships between the primary PTC and its associated LN and PLM. Some of the genetic alterations identified by RNA-Seq were ubiquitously detected in all the seven tumor samples from the patient and others were detected only in some tumor areas (Fig. 4). Analysis of the distribution of these alterations highlights several aspects of intra-tumor heterogeneity, clonal evolution, and metastases dissemination in thyroid cancer.

Firstly, the primary PTC, the two LN metastases, and the PLM were composed of several cellular subclones in this patient. Despite the presence of these subclones, the two areas of the primary tumor and the two areas of the LN metastasis 2 presented similar genetic profiles. In contrast, the two areas of the LN metastasis 1 presented more divergent mutations and fusions. Similarly, Anaka et al. (2013) have identified significant heterogeneity in chromosomal aberrations in different regions of a LN metastasis in a melanoma patient. As mentioned by others (Unger et al. 2008, Gerlinger et al. 2012), this molecular heterogeneity highlights the importance of sampling multiple areas of the same tumor to better ascertain the range of genomic alterations characterizing its progression. Indeed the heterogeneous presence of genomic alterations may impair patient genotyping and subsequent prognostic classification and targeted therapy.

Secondly, as has been described in similar studies of breast, pancreatic, and ovarian cancers (Ding et al. 2010, Yachida et al. 2010, Bashashati et al. 2013), the ancestral mutations present in the primary PTC were propagated in the metastases. However, additional genomic aberrations were detected in the metastases and there was a greater genetic divergence between the PLM and the primary PTC than between the LN metastases and the primary PTC. This indicates accumulation of genetic alterations after clonal divergence of the PLM from the ancestral clone. The metastases were collected after radiotherapy. We cannot exclude the possibility that radioiodine treatment may be involved in the occurrence of the metastases private SNVs. Nonetheless, the PLM, collected before the LN metastases, presented more private aberrations than the LN metastases, indicating that occurrence of these private mutations is more a consequence of cancer genomic instability than of radiotherapy. Furthermore, one private SNV was also detected in the primary PTC samples collected before radiotherapy. Therefore, as many LN and pulmonary metastases were already detected at the time of diagnosis and thyroidectomy, it is likely that these private mutations were acquired in the primary PTC and metastases after the dissemination of metastatic cells as a

consequence of tumor genomic instability. In accordance with our results, results from CGH studies have indicated that, compared with primary tumors, hematogenous metastases from colorectal carcinomas presented more chromosomal alterations than LN metastases, and lung metastases presented more alterations than liver metastases (Knösel et al. 2004, 2005). In addition, a DNA sequencing analysis of metastatic pancreatic cancer has revealed that lung lesions were further evolved than other metastases (Campbell et al. 2010). Additional studies will be necessary to determine whether this accumulation of genetic aberrations found in the PLM is a consequence of a lung-specific microenvironment or whether these alterations are linked to the ability of cancer cells to spread and/or implant to the lung. In the latter case, this may represent a way to find curative therapies targeting lung metastases.

Thirdly, whether cancer cells simultaneously spread from a primary tumor to regional LN and distant site(s) or first form metastases in the LN that subsequently disseminate themselves further is still a matter of debate (Sleeman et al. 2012). Genomic alterations observed in this PTC patient indicate that LN and distant metastases had a common origin. Moreover, it is likely that the PLM originated from a subclone of a LN metastasis that we have studied using RNA-Seq (LN1.2). A correlation between the number of LN metastases and lung metastases has been shown in PTC, the presence of more than 20 involved nodes being indicative of a high risk of lung metastasis (Machens & Dralle 2012). Our results indicate that in PTC, besides acting as an indicator of increased probability of the formation of distant metastases, the formation of LN metastases contributes, at least in some cases, to the seeding of these distant metastases.

In summary, during our study we identified a novel cancer-specific KAZN alternative transcript in thyroid tumors and revealed intra-tumor heterogeneity and tumor clonal evolution in an aggressive PTC on the basis of nsSNVs, gene fusions, and LOH. These new insights are potentially important for improving therapeutic targeting and diagnosis of patients with aggressive thyroid cancer.

Supplementary data

This is linked to the online version of the paper at <http://dx.doi.org/10.1530/ERC-14-0351>.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

This work was supported by grants from the WELBIO, CIMATH-2, Fonds de la Recherche Scientifique FNRS-FRSM.

Author contribution statement

C Maenhaut and V Detours were involved in study conception; S Le Pennec, C Maenhaut, and V Detours helped in study design; S Le Pennec, T Konopka, D Gacquer, and D Fimereli helped in executing the study; S Le Pennec, T Konopka, M Tarabichi, G Tomás, and F Savagner helped with data interpretation; F Savagner, M Decaussin-Petrucci, C Trésallet, G Andry, and D Larsimont provided tissues and carried out anatomopathological analyses; S Le Pennec, C Maenhaut, and V Detours helped in article preparation.

Acknowledgements

The authors thank Jacques Emile Dumont for his great help with manuscript revision and Chantal Degraef for excellent technical assistance.

References

- Abrosimov A, Saenko V, Rogounovitch T, Namba H, Lushnikov E, Mitsutake N & Yamashita S 2007 Different structural components of conventional papillary thyroid carcinoma display mostly identical *BRAF* status. *International Journal of Cancer. Journal International du Cancer* **120** 196–200. (doi:10.1002/ijc.22290)
- Anaka M, Hudson C, Lo P-H, Do H, Caballero OL, Davis ID, Dobrovic A, Cebon J & Behren A 2013 Intratumoral genetic heterogeneity in metastatic melanoma is accompanied by variation in malignant behaviors. *BMC Medical Genomics* **6** 40. (doi:10.1186/1755-8794-6-40)
- Bashashati A, Ha G, Tone A, Ding J, Prentice LM, Roth A, Rosner J, Shumansky K, Kalloger S, Senz J et al. 2013 Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling: evolutionary trajectories of ovarian cancers. *Journal of Pathology* **231** 21–34. (doi:10.1002/path.4230)
- Berger MF, Levin JZ, Vijayendran K, Sivachenko A, Adiconis X, Maguire J, Johnson LA, Robinson J, Verhaak RG, Sougnez C et al. 2010 Integrative analysis of the melanoma transcriptome. *Genome Research* **20** 413–427. (doi:10.1101/gr.103697.109)
- Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, Morsberger LA, Latimer C, McLaren S, Lin M-L et al. 2010 The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467** 1109–1113. (doi:10.1038/nature09460)
- de Biase D, Cesari V, Visani M, Casadei GP, Cremonini N, Gandolfi G, Sancisi V, Ragazzi M, Pession A, Ciarrocchi A et al. 2014 High sensitivity *BRAF* mutation analysis: *BRAF* V600E is acquired early during tumor development but is heterogeneously distributed in a subset of papillary thyroid carcinomas. *Journal of Clinical Endocrinology and Metabolism* **99** E1530–E1538. (doi:10.1210/jc.2013-4389)
- Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, Harris CC, McLellan MD, Fulton RS, Fulton LL et al. 2010 Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464** 999–1005. (doi:10.1038/nature08989)
- Fimereli D, Detours V & Konopka T 2013 TriageTools: tools for partitioning and prioritizing analysis of high-throughput sequencing data. *Nucleic Acids Research* **41** e86. (doi:10.1093/nar/gkt094)
- Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P et al. 2012 Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *New England Journal of Medicine* **366** 883–892. (doi:10.1056/NEJMoa1113205)
- Giannini R, Ugolini C, Lupi C, Proietti A, Elisei R, Salvatore G, Berti P, Materazzi G, Miccoli P, Santoro M et al. 2007 The heterogeneous distribution of *BRAF* mutation supports the independent clonal origin of distinct tumor foci in multifocal papillary thyroid carcinoma. *Journal of Clinical Endocrinology and Metabolism* **92** 3511–3516. (doi:10.1210/jc.2007-0594)
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q et al. 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29** 644–652. (doi:10.1038/nbt.1883)
- Guerra A, Fugazzola L, Marotta V, Cirillo M, Rossi S, Cirello V, Forno I, Moccia T, Budillon A & Vitale M 2012 A high percentage of *BRAF*^{V600E} alleles in papillary thyroid carcinoma predicts a poorer outcome. *Journal of Clinical Endocrinology and Metabolism* **97** 2333–2340. (doi:10.1210/jc.2011-3106)
- Kim D & Salzberg SL 2011 TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biology* **12** R72. (doi:10.1186/gb-2011-12-8-r72)
- Kim J, Giuliano AE, Turner RR, Gaffney RE, Umetani N, Kitago M, Elashoff D & Hoon DS 2006 Lymphatic mapping establishes the role of *BRAF* gene mutation in papillary thyroid carcinoma. *Annals of Surgery* **244** 799–804. (doi:10.1097/01.sla.0000224751.80858.13)
- Knösel T, Schlüns K, Stein U, Schwabe H, Schlag PM, Dietel M & Petersen I 2004 Chromosomal alterations during lymphatic and liver metastasis formation of colorectal cancer. *Neoplasia* **6** 23–28. (doi:10.1016/S1476-5586(04)80050-2)
- Knösel T, Schlüns K, Dietel M & Petersen I 2005 Chromosomal alterations in lung metastases of colorectal carcinomas: associations with tissue specific tumor dissemination. *Clinical & Experimental Metastasis* **22** 533–538. (doi:10.1007/s10585-005-5239-7)
- Machens A & Dralle H 2012 Correlation between the number of lymph node metastases and lung metastasis in papillary thyroid cancer. *Journal of Clinical Endocrinology and Metabolism* **97** 4375–4382. (doi:10.1210/jc.2012-1257)
- McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MG, Griffith M, Heravi Moussavi A, Senz J, Melnyk N et al. 2011 deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Computational Biology* **7** e1001138. (doi:10.1371/journal.pcbi.1001138)
- Micci F, Panagopoulos I, Thorsen J, Davidson B, Tropé CG & Heim S 2014 Low frequency of *ESRRB-C11orf20* fusion gene in ovarian carcinomas. *PLoS Biology* **12** e1001784. (doi:10.1371/journal.pbio.1001784)
- Nachat R, Cipolat S, Sevilla LM, Chhatriwala M, Groot KR & Watt FM 2009 KazrinE is a desmosome-associated liprin that colocalises with acetylated microtubules. *Journal of Cell Science* **122** 4035–4041. (doi:10.1242/jcs.047266)
- Park SY, Park YJ, Lee YJ, Lee HS, Choi SH, Choe G, Jang H-C, Park SH, Park DJ & Cho BY 2006 Analysis of differential *BRAF*^{V600E} mutational status in multifocal papillary thyroid carcinoma: evidence of independent clonal origin in distinct tumor foci. *Cancer* **107** 1831–1838. (doi:10.1002/cncr.22218)
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G & Mesirov JP 2011 Integrative genomics viewer. *Nature Biotechnology* **29** 24–26. (doi:10.1038/nbt.1754)
- Salzman J, Marinelli RJ, Wang PL, Green AE, Nielsen JS, Nelson BH, Drescher CW & Brown PO 2011 *ESRRB-C11orf20* is a recurrent gene fusion in serous ovarian carcinoma. *PLoS Biology* **9** e1001156. (doi:10.1371/journal.pbio.1001156)
- Sevilla LM, Nachat R, Groot KR & Watt FM 2008 Kazrin regulates keratinocyte cytoskeletal networks, intercellular junctions and differentiation. *Journal of Cell Science* **121** 3561–3569. (doi:10.1242/jcs.029538)
- Sipos JA & Mazzaferri EL 2010 Thyroid cancer epidemiology and prognostic variables. *Clinical Oncology* **22** 395–404. (doi:10.1016/j.clon.2010.05.004)
- Sleeman JP, Cady B & Pantel K 2012 The connectivity of lymphogenous and hematogenous tumor cell dissemination: biological insights and

- clinical implications. *Clinical & Experimental Metastasis* **29** 737–746. (doi:10.1007/s10585-012-9489-x)
- Smallridge RC, Chindris A-M, Asmann YW, Casler JD, Serie DJ, Reddi HV, Cradic KW, Rivera M, Grebe SK, Necela BM *et al.* 2014 RNA sequencing identifies multiple fusion transcripts, differentially expressed genes, and reduced expression of immune function genes in *BRAF* (V600E) mutant vs *BRAF* wild-type papillary thyroid carcinoma. *Journal of Clinical Endocrinology and Metabolism* **99** E338–E347. (doi:10.1210/jc.2013-2792)
- Tallini G, Hsueh A, Liu S, Garcia-Rostan G, Speicher MR & Ward DC 1999 Frequent chromosomal DNA unbalance in thyroid oncocyctic (Hürthle cell) neoplasms detected by comparative genomic hybridization. *Laboratory Investigation; A Journal of Technical Methods and Pathology* **79** 547–555.
- Tarabichi M, Detours V & Konopka T 2012 Piecewise polynomial representations of genomic tracks. *PLoS ONE* **7** e48941. (doi:10.1371/journal.pone.0048941)
- Thorvaldsdóttir H, Robinson JT & Mesirov JP 2013 Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14** 178–192. (doi:10.1093/bib/bbs017)
- Trapnell C, Pachter L & Salzberg SL 2009 TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25** 1105–1111. (doi:10.1093/bioinformatics/btp120)
- Unger K, Malisch E, Thomas G, Braselmann H, Walch A, Jackl G, Lewis P, Lengfelder E, Bogdanova T, Wienberg J *et al.* 2008 Array CGH demonstrates characteristic aberration signatures in human papillary thyroid carcinomas governed by RET/PTC. *Oncogene* **27** 4592–4602. (doi:10.1038/onc.2008.99)
- Wada N, Duh Q-Y, Miura D, Brunaud L, Wong MG & Clark OH 2002 Chromosomal aberrations by comparative genomic hybridization in Hürthle cell thyroid carcinomas are associated with tumor recurrence. *Journal of Clinical Endocrinology and Metabolism* **87** 4595–4601. (doi:10.1210/jc.2002-020339)
- Walts AE, Pao A, Sacks W & Bose S 2014 *BRAF* genetic heterogeneity in papillary thyroid carcinoma and its metastasis. *Human Pathology* **45** 935–941. (doi:10.1016/j.humpath.2013.12.005)
- Wang Q, Liu M, Li X, Chen L & Tang H 2009 Kazrin F is involved in apoptosis and interacts with BAX and ARC. *Acta Biochimica et Biophysica Sinica* **41** 763–772. (doi:10.1093/abbs/gmp065)
- Xing M, Haugen BR & Schlumberger M 2013 Progress in molecular-based management of differentiated thyroid cancer. *Lancet* **381** 1058–1069. (doi:10.1016/S0140-6736(13)60109-9)
- Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, Kamiyama M, Hruban RH, Eshleman JR, Nowak MA *et al.* 2010 Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467** 1114–1117. (doi:10.1038/nature09515)

Received in final form 9 February 2015

Accepted 12 February 2015

Made available online as an Accepted Preprint

17 February 2015