Characterization of the mutational landscape of anaplastic thyroid cancer via whole-exome sequencing

John W. Kunstman1,2,†, C. Christofer Juhlin1,6,†, Gerald Goh3,4,†, Taylor C. Brown1,2, Adam Stenman6, James M. Healy1,2, Jill C.Rubinstein1,2, Murim Choi3,4, Nimrod Kiss6, Carol Nelson-Williams3,4, Shrikant Mane3, David L. Rimm5, Manju L. Prasad5, Anders Höög6, Jan Zedenius7, Catharina Larsson6, Reju Korah1,2, Richard P. Lifton3,4 and Tobias Carling1,2,*

1Yale Endocrine Neoplasia Laboratory, 2Department of Surgery, 3Department of Genetics, 4Howard Hughes Medical Institute and 5Department of Pathology, Yale School of Medicine, New Haven, CT 06520, USA, 6Department of Oncology–Pathology and 7Department of Molecular Medicine and Surgery, Karolinska Institutet, Karolinska University Hospital CCK, SE-171 76 Stockholm, Sweden

*To whom correspondence should be addressed at: Yale School of Medicine, 333 Cedar Street, FMB130A, Box 208062, New Haven, CT 06520, USA. Tel: +1 2037372036; Fax: +1 2037852498; Email: tobiaskarling@yale.edu

Abstract
Anaplastic thyroid carcinoma (ATC) is a frequently lethal malignancy that is often unresponsive to available therapeutic strategies. The tumorigenesis of ATC and its relationship to the widely prevalent well-differentiated thyroid carcinomas are unclear. We have analyzed 22 cases of ATC as well as 4 established ATC cell lines using whole-exome sequencing. A total of 2674 somatic mutations (121/sample) were detected. Ontology analysis revealed that the majority of variants aggregated in the MAPK, ErbB and RAS signaling pathways. Mutations in genes related to malignancy not previously associated with thyroid tumorigenesis were observed, including mTOR, NF1, NF2, MLH1, MLH3, MSH5, MSH6, ERBB2, EIF1AX and USH2A; some of which were recurrent and were investigated in 24 additional ATC cases and 8 ATC cell lines.

Somatic mutations in established thyroid cancer genes were detected in 14 of 22 (64%) tumors and included recurrent mutations in BRAF, TP53 and RAS-family genes (6 cases each), as well as PIK3CA (2 cases) and single cases of CDKN1B, CDKN2C, CTNNB1 and RET mutations. BRAF V600E and RAS mutations were mutually exclusive; all ATC cell lines exhibited a combination of mutations in either BRAF and TP53 or NRAS and TP53. A hypermutator phenotype in two cases with >8 times higher mutational burden than the remaining mean was identified; both cases harbored unique somatic mutations in MLH mismatch-repair genes. This first comprehensive exome-wide analysis of the mutational landscape of ATC identifies novel genes potentially associated with ATC tumorigenesis, some of which may be targets for future therapeutic intervention.

Introduction
Anaplastic thyroid carcinoma (ATC) is a rare endocrine malignancy with therapeutic options of limited effectiveness. In contrast to the much more common malignancies of the thyroid follicular epithelium, such as papillary or follicular thyroid cancers (collectively known as ‘well-differentiated thyroid...
cancers’—WDTC—which boast excellent cure rates and long-term survival, prognosis in ATC patients is exceedingly poor, nearly uniformly fatal, and has changed little in the past half century. In the USA, ATC is responsible for 1–2% of all thyroid cancers, but accounts for over 50% of deaths attributable to thyroid malignancy (1,2). Over 80% of patients present with loco-regional invasion and approximately half have distant metastases at the time of diagnosis (3–5). Considering the explosive clinical course of ATC, the recommended therapy is generally aggressive multimodal treatment for most patients (4). Despite maximal intervention, the median survival time is <6 months in most series, and cancer-specific mortality at 1 year exceeds 80% (4,6,7). The dismal prognosis and high prevalence of distant disease at the time of diagnosis highlights the urgent need for novel, targeted systemic treatments for ATC.

Approximately 20–25% of patients with ATC have a history of a previous WDTC and an additional 20–30% have a coexisting WDTC found following surgery for ATC (3,4,8,9). Although ATC can also arise de novo or in the setting of chronic goiter in occasional patients, these data suggest that a significant fraction of ATC cases are the result of progression from WDTC. Furthermore, genetic analysis of ATC has characterized frequent somatic mutations in several genes such as BRAF and RAS that are commonly mutated in WDTC as well, implying a common initial tumorigenic pathway (10–12). Mutations unique to ATC, in genes such as TP53, CTNNB1 and others, are thought to be involved in dedifferentiation and progression to ATC. However, due to the rarity of ATC and correspondingly small sample size of most studies evaluating its genetic landscape, conclusively determining the incidence and impact of these mutations is difficult. For example, CTNNB1 mutations have been variably observed at a rate of 0–61.3% of ATC cases (13–15). Similarly, widespread copy number and structural genomic instability have been repeatedly demonstrated in ATC with copy number gain present in over 80% of tumors, but wide spectrum of variation reported coupled with small sample sizes limit the generalizability of these studies (2,16). While current knowledge of the molecular pathogenesis of ATC has led to several clinical trials of existing targeted pharmaceuticals, the results have thus far not shown dramatic improvements in outcomes and the precise molecular mechanisms of ATC dedifferentiation and tumorigenesis remain unknown (17–19), although individual successes have been described (20,21).

Whole-exome capture coupled with next-generation sequencing technology is a proven method for identifying functionally relevant genetic variants underlying both Mendelian and complex disease states, such as neoplasia (22–24). The ability of whole-exome sequencing (WES) to resolve single-nucleotide variants is particularly suited for characterizing previously unknown drivers of tumorigenesis and survey the landscape of somatic mutations present in a malignancy (25,26). Furthermore, WES has been successfully utilized in the clinical environment for both genetic diagnosis and tumor genotyping (23,27,28); a possibly useful application in ATC, where conventional cancer therapies are largely ineffective and targeted therapies have not been widely applied but have been shown to be beneficial in individual cases. Thus, we hypothesized that WES would be an ideal platform to (i) better characterize the mutational landscape of ATC, (ii) identify novel potential driver mutations for further study in ATC tumorigenesis and (iii) detect possible candidates for targeted therapy in patients with ATC. Accordingly, we applied WES to a cohort of 22 unique cases of ATC and four established ATC cell lines in an effort to further these goals.

Results

Exome sequencing cohort

Demographic and pathologic information describing the exome sequencing cohort are shown in Table 1. Thirteen patients (59%) were women, nine (41%) were men and the mean age at surgery was 73 years (median 74 years). Tumor tissue from 19 of the patients (86%) were derived from primary tumor tissue, whereas three patients (14%) had previously undergone total thyroidectomy for papillary thyroid cancer and ATC specimens were obtained during subsequent procedures for recurrent disease or palliative purposes. Thirteen patients (59%) presented with distant metastases at diagnosis. Eight patients had been treated with neoadjuvant external beam irradiation. Ten patients had demonstrable evidence of concurrent or previous WDTC in either the ipsi- or contralateral thyroid lobe, nine with papillary thyroid cancer and one with follicular thyroid cancer.

Whole-exome sequencing

In the 22 matched tumor/normal pairs that underwent WES, a total of 2674 somatic variants were identified, 1972 of which were non-synonymous (74%). In total, 68% constituted coding missense mutations, 5% non-sense mutations and 1% were exon-intron boundary mutations. The number of somatic variants per tumor ranged from 4 to 1805, with a mean number of total and non-synonymous variants of 121.5 and 89.6 per tumor, respectively. Two samples demonstrated a significantly higher mutation burden compared with all other samples (T2—253 mutations, T11—1805 mutations). When samples T2 and T11 were excluded, the mean number of total and non-synonymous variants was 30.8 and 23.7 per tumor, respectively. An overview of the WES results by sample is shown in Figure 1, and a detailed list of the sequencing data for all 22 cohort tumor-normal pairs is included in Supplementary Material, Tables S1 and S2. The percentage of reads on target was 66% and 65% for tumor and normal tissue, respectively, and the percentage of bases covered >20 times was 96% (tumor) and 92% (normal tissue). Tumor samples were deliberately sequenced to a greater depth than normal tissue (mean 264 and 138 reads/sample, respectively) in order to minimize the impact of the perceived lower quality of DNA libraries derived from fresh-frozen, paraffin-embedded (FFPE) samples (T1–T15) and also to maximize detection of heterozygous mutations in tumor samples admixed with adjacent normal tissue. Thirty mutations identified by exome sequencing throughout the cohort were selected for confirmatory Sanger sequencing. The mutations selected for confirmation were chosen based on perceived relevance in tumorigenesis, availability of sample for confirmatory studies and to ensure mutations from multiple FFPE-derived samples were confirmed. As expected based on confirmatory sequencing from prior whole exome studies, 29 of 30 (96.7%) mutations examined were confirmed when subjected to Sanger sequencing (Supplementary Material, Table S3).

Potential novel driver gene mutations

While a majority of cases harbored somatic mutations in known thyroid malignancy-related genes (discussed below), 16 different cancer-related genes denoted as ‘driver genes’ by Vogelstein et al. (29) not typically identified as drivers of thyroid malignancy were found to be present among the studied tumors. Several of these mutations were present in multiple samples and are strongly correlated with non-thyroidal malignancies, such as NF1, mTOR.
followed by currently mutated genes were described oncogenes and tumor suppressors. The most common re-

tensively, three cases with mutations in ATC-related gene mutations

<table>
<thead>
<tr>
<th>Case</th>
<th>Source</th>
<th>Age</th>
<th>Presentation</th>
<th>Primary tumor size</th>
<th>Nodal stage</th>
<th>Metastases at diagnosis</th>
<th>Alive?</th>
<th>Survival (months)</th>
<th>Additional thyroid disease?</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>FFPE</td>
<td>53</td>
<td>Neck mass, cord paralysis</td>
<td>5.0 cm</td>
<td>N0</td>
<td>None</td>
<td>Yes</td>
<td>50.9</td>
<td>No</td>
</tr>
<tr>
<td>T2</td>
<td>FFPE</td>
<td>70</td>
<td>Stridor, dyspnea</td>
<td>8.0 cm</td>
<td>N1b</td>
<td>Bone</td>
<td>No</td>
<td>2.8</td>
<td>No</td>
</tr>
<tr>
<td>T3</td>
<td>FFPE</td>
<td>72</td>
<td>Surveillance ultrasound</td>
<td>N/A</td>
<td>N1b</td>
<td>Lung, chest wall</td>
<td>No</td>
<td>24.1</td>
<td>WDTC</td>
</tr>
<tr>
<td>T4</td>
<td>FFPE</td>
<td>70</td>
<td>Dysphagia</td>
<td>5.8 cm</td>
<td>N1b</td>
<td>Lung, brain, retroperitoneum</td>
<td>No</td>
<td>0.9</td>
<td>No</td>
</tr>
<tr>
<td>T5</td>
<td>FFPE</td>
<td>61</td>
<td>Neck mass</td>
<td>8.0 cm</td>
<td>N/A</td>
<td>None</td>
<td>No</td>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>T6</td>
<td>FFPE</td>
<td>45</td>
<td>Neck mass</td>
<td>7.5 cm</td>
<td>N1a</td>
<td>None</td>
<td>No</td>
<td>11.3</td>
<td>No</td>
</tr>
<tr>
<td>T7</td>
<td>FFPE</td>
<td>89</td>
<td>Neck mass</td>
<td>8.0 cm</td>
<td>N1b</td>
<td>Lung, chest wall</td>
<td>No</td>
<td>0.8</td>
<td>WDTC</td>
</tr>
<tr>
<td>T8</td>
<td>FFPE</td>
<td>78</td>
<td>Neck mass</td>
<td>3.7 cm</td>
<td>N1b</td>
<td>None</td>
<td>No</td>
<td>8.6</td>
<td>WDTC</td>
</tr>
<tr>
<td>T9</td>
<td>FFPE</td>
<td>59</td>
<td>Dysphagia</td>
<td>7.5 cm</td>
<td>N1a</td>
<td>None</td>
<td>No</td>
<td>44.3</td>
<td>No</td>
</tr>
<tr>
<td>T10</td>
<td>FFPE</td>
<td>59</td>
<td>Dysphagia</td>
<td>5.0 cm</td>
<td>N1b</td>
<td>Lung, chest wall</td>
<td>No</td>
<td>0.8</td>
<td>WDTC</td>
</tr>
<tr>
<td>T11</td>
<td>FFPE</td>
<td>65</td>
<td>Neck mass</td>
<td>8.0 cm</td>
<td>N1b</td>
<td>Lung, chest wall</td>
<td>No</td>
<td>9.9</td>
<td>WDTC</td>
</tr>
<tr>
<td>T12</td>
<td>FFPE</td>
<td>93</td>
<td>Airway collapse</td>
<td>8.0 cm</td>
<td>N1b</td>
<td>None</td>
<td>No</td>
<td>0.8</td>
<td>WDTC</td>
</tr>
<tr>
<td>T13</td>
<td>FFPE</td>
<td>75</td>
<td>Neck mass</td>
<td>8.0 cm</td>
<td>N1b</td>
<td>None</td>
<td>No</td>
<td>0.8</td>
<td>WDTC</td>
</tr>
<tr>
<td>T14</td>
<td>FFPE</td>
<td>63</td>
<td>Hemoptysis</td>
<td>3.7 cm</td>
<td>N1b</td>
<td>None</td>
<td>No</td>
<td>0.8</td>
<td>WDTC</td>
</tr>
<tr>
<td>T15</td>
<td>FFPE</td>
<td>79</td>
<td>Stridor, dyspnea</td>
<td>7.0 cm</td>
<td>N1b</td>
<td>None</td>
<td>No</td>
<td>0.8</td>
<td>WDTC</td>
</tr>
<tr>
<td>T16</td>
<td>Fresh Frozen</td>
<td>77</td>
<td>Neck mass</td>
<td>7.0 cm</td>
<td>N1b</td>
<td>None</td>
<td>No</td>
<td>0.8</td>
<td>WDTC</td>
</tr>
<tr>
<td>T17</td>
<td>Fresh Frozen</td>
<td>83</td>
<td>Neck mass</td>
<td>7.0 cm</td>
<td>N1b</td>
<td>None</td>
<td>No</td>
<td>11</td>
<td>Goiter</td>
</tr>
<tr>
<td>T18</td>
<td>Fresh Frozen</td>
<td>82</td>
<td>Neck mass</td>
<td>4.0 cm</td>
<td>N1b</td>
<td>None</td>
<td>No</td>
<td>1</td>
<td>WDTC</td>
</tr>
<tr>
<td>T19</td>
<td>Fresh Frozen</td>
<td>84</td>
<td>Neck mass</td>
<td>5.0 cm</td>
<td>N1b</td>
<td>None</td>
<td>No</td>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>T20</td>
<td>Fresh Frozen</td>
<td>72</td>
<td>Neck mass, dyspnea</td>
<td>10.0 cm</td>
<td>N1b</td>
<td>None</td>
<td>No</td>
<td>1</td>
<td>Goiter</td>
</tr>
<tr>
<td>T21</td>
<td>Fresh Frozen</td>
<td>89</td>
<td>Neck mass</td>
<td>8.0 cm</td>
<td>N1b</td>
<td>None</td>
<td>No</td>
<td>0.5</td>
<td>WDTC</td>
</tr>
<tr>
<td>T22</td>
<td>Fresh Frozen</td>
<td>85</td>
<td>Neck mass</td>
<td>4.0 cm</td>
<td>N1b</td>
<td>None</td>
<td>No</td>
<td>4</td>
<td>WDTC</td>
</tr>
</tbody>
</table>

FFPE, fresh-frozen paraffin-embedded; N/A, not available (lost to follow-up); WDTC, well-differentiated thyroid cancer.

Novel recurrently mutated COSMIC genes

COSMIC genes found to be recurrently mutated that have not previously been described in ATC development included four cases of non-synonymous USH2A missense mutations (4/22; 18%). Additionally, three cases with mutations in EIF1AX and HECTD1 were found (3/22; 14%). These are summarized in Figure 1, Table 2 (USH2A, EIF1AX), and Supplementary Material, Table S2 (HECTD1).

ATC-related gene mutations

Multiple genes previously found to be involved in thyroid neoplasia demonstrated somatic mutations in the study group. These findings are summarized in Table 3 and include several well-described oncogenes and tumor suppressors. The most common recurrently mutated genes were TP53 and BRAF with six cases each, followed by NRAS (3 cases), KRAS (2 cases), PIK3CA (2 cases) as well as HRAS, CDKN1B, CDKN2C, CTNNB1, HRAS and RET (one case each, Fig. 1, Table 3). Previously recognized ATC-related recurrent mutations present in the cohort included BRAF V600E (six cases), NRAS Q61R (two cases) and PIK3CA H1047R/L (one case each). The mean number of single-nucleotide variants in tumors harboring BRAF V600E mutations versus RAS mutations was 47.8 (±31.5) versus 22.2 (±6.7). For tumors harboring neither BRAF nor RAS mutations, the mean number of single-nucleotide variants (SNVs) was 24.5 (±18.3) (comparison of three groups; P-value 0.09). Previously reported frequencies of commonly described ATC gene variants are compared with the frequencies described here in Supplementary Material, Table S4.

Five cases did not exhibit somatic mutations in either the ATC-related or novel ATC gene variant gene groups (Fig. 1), and the mutational status of these cases are summarized in Supplementary Material, Table S5. The mean number of mutations present among these five cases was 12.4 (range 4–22). A subset of the cases (T6, T8) did exhibit damaging mutations with concurrent loss of heterozygosity (LOH) in COSMIC genes with largely unknown functions. The recent discoveries of RASAL1 as a thyroid-specific tumor suppressor gene with a mutational frequency of 17% in ATC (30) and the role of ALK rearrangements in thyroid tumorigenesis (31,32) prompted investigation of whether RASAL1 and ALK mutations were present in the cohort. No somatic mutations in ALK were found; however, a missense mutation in RASAL1 (R774C) was identified in case T11 (Fig. 1).

Gene ontology and pathway analyses

Using the DAVID (Database for Annotation, Visualization and Integrated Discovery) Bioinformatics Resources 6.7 database for identification of enriched biological processes in the pool of somatically altered genes, wide and generic ontology terms were identified. These include metabolic processes, intracellular signaling, phosphorylation and apoptosis. Further analyses using
KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway annotations identified several pathways with significant representation in the ATC cohort, including general pathways in cancer, MAPK kinase and ErbB pathways as enriched among the mutated genes. The involvement of the MAPK kinase and ErbB pathways were further characterized and validated using an independent ontology analysis (Genomatix Pathway System), yielding P-values of $2.01 \times 10^{-6}$ and $6.88 \times 10^{-5}$, respectively. This analysis identified somatic mutations among the tumor cohort in both the MAPK/ERK pathway ($BRAF$, $HRAS$, $KRAS$, $NRAS$, $RAF1$, $AKT2$ and $PIK3CA$).

Figure 1. Overall mutational profile of the 22 ATCs studied by exome sequencing. All panels are aligned with vertical tracks representing 22 individuals with ATC (left) and 4 established ATC cell lines (right). The underlying heatmap shows the distribution of somatic coding mutations in thyroid malignancy-related genes, novel ATC gene variants [established driver genes as suggested by Vogelstein et al. (24) or recurrently mutated COSMIC genes found in $\geq 3$ samples], and finally mutated Mut homolog genes in which a correlation to the amount of mutational rate was observed.

Table 2. Summary of individual mutations and associated protein coding change in non-synonymous mut homolog genes, $USH2A$, and $EIF1AX$ among all ATC tissue samples assayed

<table>
<thead>
<tr>
<th>Case</th>
<th>Gene</th>
<th>Location</th>
<th>Mutation (AA)</th>
<th>Position (AA/total AA)</th>
<th>P-value</th>
<th>Nucleotide position</th>
<th>Nucleotide change</th>
<th>PolyPhen2 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2</td>
<td>MLH1</td>
<td>Chr3</td>
<td>I19M</td>
<td>19/756</td>
<td>1.16E−31</td>
<td>37 010 099</td>
<td>C&gt;G</td>
<td>1</td>
</tr>
<tr>
<td>T3</td>
<td>MLH1</td>
<td>Chr3</td>
<td>I68M</td>
<td>68/756</td>
<td>2.63E−21</td>
<td>37 013 201</td>
<td>C&gt;T</td>
<td>1</td>
</tr>
<tr>
<td>T4</td>
<td>MLH1</td>
<td>Chr3</td>
<td>Q60X</td>
<td>60/756</td>
<td>8.23E−18</td>
<td>37 013 175</td>
<td>Truncating</td>
<td>0.999</td>
</tr>
<tr>
<td>T11</td>
<td>MLH3</td>
<td>Chr14</td>
<td>L264V</td>
<td>264/1429</td>
<td>5.10E−08</td>
<td>74 585 322</td>
<td>G&gt;C</td>
<td>0.968</td>
</tr>
<tr>
<td>T12</td>
<td>MSH5</td>
<td>Chr6</td>
<td>A199V</td>
<td>199/822</td>
<td>4.43E−06</td>
<td>31 819 953</td>
<td>C&gt;T</td>
<td>0.999</td>
</tr>
<tr>
<td>T10</td>
<td>MSH6</td>
<td>Chr2</td>
<td>D736H</td>
<td>736/1360</td>
<td>9.08E−06</td>
<td>47 880 832</td>
<td>G&gt;C</td>
<td>1</td>
</tr>
<tr>
<td>T5</td>
<td>USH2A</td>
<td>Chr1</td>
<td>I2189V</td>
<td>2189/5202</td>
<td>2.18E−22</td>
<td>214 238 944</td>
<td>T&gt;C</td>
<td>0.921</td>
</tr>
<tr>
<td>T11</td>
<td>USH2A</td>
<td>Chr1</td>
<td>D798V</td>
<td>798/5202</td>
<td>2.79E−05</td>
<td>214 486 966</td>
<td>T&gt;A</td>
<td>0.549</td>
</tr>
<tr>
<td>T21</td>
<td>USH2A</td>
<td>Chr1</td>
<td>E4571K</td>
<td>4571/5202</td>
<td>6.33E−13</td>
<td>213 914 165</td>
<td>G&gt;T</td>
<td>0.961</td>
</tr>
<tr>
<td>T22</td>
<td>USH2A</td>
<td>Chr1</td>
<td>L1727F</td>
<td>1727/5202</td>
<td>2.08E−34</td>
<td>214 323 540</td>
<td>G&gt;A</td>
<td>0.961</td>
</tr>
<tr>
<td>T9</td>
<td>EIF1AX</td>
<td>ChrX</td>
<td>Ex-in</td>
<td>1 bp upstream of exon 6</td>
<td>1.14E−88</td>
<td>20 058 647</td>
<td>C&gt;G</td>
<td>Splice site</td>
</tr>
<tr>
<td>T13</td>
<td>EIF1AX</td>
<td>ChrX</td>
<td>G9R</td>
<td>9/144</td>
<td>1.40E−15</td>
<td>20 066 653</td>
<td>C&gt;G</td>
<td>0.966</td>
</tr>
<tr>
<td>T21</td>
<td>EIF1AX</td>
<td>ChrX</td>
<td>P2R</td>
<td>2/144</td>
<td>2.57E−12</td>
<td>20 069 675</td>
<td>G&gt;C</td>
<td>0.028</td>
</tr>
</tbody>
</table>

AA, amino acid.
and the ErbB pathway (MAPK10, ERBB2, ERBB3, RAF1, RAC1 and NF2) (Fig. 2). Additional analysis using DAPPLE (Disease Association Protein–Protein Link Evaluator) version 2.0 to examine the significance of protein–protein interactions among the gene products of those genes identified as mutated by exome sequencing was performed. Cases T2 and T11 were excluded from this analysis as well as the gene ontology analysis as they harbored a much greater mutational burden due to mutations in the mismatch repair process (described above). A pattern was observed in which proteins involved in the ErbB/RAS/MAPK signaling pathways separated from apoptosis-related pathways (Supplementary Material, Fig. S1).

Exome sequencing of ATC cell lines

The results from the exome sequencing of four established ATC cell lines are detailed in Figure 1 and Supplementary Material, Table S6. The number of mutations ranged from 358 (BHT101) to 565 (ACT1), averaging 457 mutations per cell line. All four cell lines exhibited mutations in TP53. Three cell lines were found to have the BRAF V600E mutation (B305C, 8505C, BHT101). The remaining cell line (ACT1) had mutations in NRAS and EIF1AX. Furthermore, ACT1 and BHT101 both exhibited USH2A mutations.

LOH and copy number variation analyses

A schematic overview of the LOH profile of the 22 ATCs studied is shown in Figure 3. In general, LOH events were most frequently identified in chromosomes 9p, 13p and 22q, averaging 40% of the cases. Eight of the ATCs had discernible copy number variations (CNVs), and there were no significant focal or arm-level CNVs from the GISTIC (Genomic Identification of Significant Targets in Cancer) analysis.

Correlations to clinical characteristics

The mean somatic mutation burden for the group that received neoadjuvant external radiotherapy (n = 8) was 51.6 mutations, compared with the 161.5 mutations seen in the group that did not receive neoadjuvant treatment (n = 14; P = NS). However, when excluding the hypermutator tumors T2 and T11 (did, and did not, receive neoadjuvant radiotherapy, respectively), almost comparable levels were seen (35.1 versus 22.9 mutations respectively; P = NS). Furthermore, no significant correlations were seen between number of somatic mutations and length of survival (Spearman’s correlation r = 0.169, P = 0.592), lymph node status (Kruskal–Wallis P = 0.909, ANOVA P = 0.800) or metastases at presentation (t-test P = 0.250). There was no statistically significant difference in the number of mutations observed in the FFPE- versus fresh-frozen-derived tissues subjected to exome sequencing. All described correlations remained insignificant when cases with a hypermutator phenotype were excluded.

Validation cohort analysis

To further examine the novel mutations discussed above in mTOR and ERBB2 that could have significant clinical impact, a validation cohort of 24 primary ATC cases and 8 ATC cell lines were subjected to targeted Sanger sequencing of the relevant mutation loci. In addition, EIF1AX, recently shown to be a potential driver in papillary thyroid carcinoma (33) and was one of the most frequent recurrently mutated novel genes among the exome sequencing cohort in the current study (three primary tumors and
one cell line), was also selected for further examination, as was the well-recognized BRAF V600E mutation.

A mutation in EIF1AX (Chr X:20058647; G>A) was found in ATC cell line c643, which corresponds to the identical chromosomal position of a heterozygous somatic mutation found in exome sample T9 (Chr X:20058647; G>C, Supplementary Material, Tables S2 and S7). This mutation, a single base pair upstream from the splice site of exon 6, is predicted as damaging by computational analysis (34). Furthermore, a recurrent SNP in the 5′ UTR of EIF1AX (rs201653081, C>T) four base pairs upstream of exon 1
(Chr X:20141644) were observed in two validation cohort samples, greater than the expected minor allele frequency of 0.24%. No additional mutations at the examined mTOR or ERBB2 loci were found in the validation cohort (R164Q/M232T and D387N, respectively). Five BRAF V600E mutations were found; four were within the ATC cell lines and served as positive controls for the Sanger sequencing as they have been previously described (35).

Discussion
The dismal prognosis of ATC is directly related to the ineffectiveness of mainstream treatment options and the incomplete understanding of the mechanisms of ATC tumorigenesis. This study sought to address these challenges using an exome-wide sequencing approach in 22 cases of ATC to survey the frequency of established genetic events in tumorigenesis, identify potential novel driver genetic events and characterize future candidates for targeted therapy. In summary, the results confirm that the genetic background of ATC is highly heterogeneous, but several novel genetic events not previously described in ATC were demonstrated. Some of these, such as mutations mTOR, NF1, or ERBB2, are particularly intriguing as these genes have been targeted previously by pharmaceuticals currently in clinical or experimental use. Additionally, two tumors were found to have defects in a mismatch repair genes and feature a hypermutator phenotype, a previously poorly characterized finding in ATC. Finally, the mutational landscape of the cohort was described with known thyroid malignancy genes present at varying frequencies in most, but not all, ATC cases.

The prototypic progression to ATC can be considered as a multi-step de-differentiation process in many cases, arising from a pre-existing WDTC such as papillary thyroid cancer or follicular thyroid cancer. This is best illustrated by the frequent observation of mutations in the BRAF and RAS oncogenes in ATC. The well-described BRAF V600E variant is the most common mutation in papillary thyroid cancer, while NRAS/KRAS/HRAS variants are commonly found in follicular thyroid cancer. In the current study, the observed frequencies for the BRAF V600E and RAS mutations (27% for both) mirrored the prevalence reported in previous manuscripts [Supplementary Material, Table S4 (12,15,36-38)]. As BRAF and RAS mutations are thought to be the drivers for development of WDTC prior to ATC development, the prevalence of these two mutations in a particular ATC cohort would be expected to be static. The cumulative prevalence of ~50-60% for these two mutations in this and other ATC cohorts may represent the fraction of ATCs that genuinely arise from pre-existing WDTCs. All BRAF mutations in the current study were found to be the V600E variant and were mutually exclusive with RAS mutations, presumably reflecting the underlying WDTC variant (papillary or follicular thyroid cancer, respectively). For instance, in six cases of ATC found to harbor the BRAF V600E mutation described here, four had either been treated for pre-existing PTC or were found to have concurrent PTC elsewhere in the thyroid following resection.

The mechanism by which de-differentiation results in progression from WDTC to ATC have been a topic of frequent study with accumulation of mutations in recognized malignancy-associated genes such as TP53 and the phosphatidylinositol 3-kinase/ Akt (PI3-K/Akt) and Ras/Raf/mitogen-activated protein kinase (Ras/Raf/MAPK) pathways being regularly mentioned (12,15,39). Indeed, half of the cases of ATC with BRAF V600E mutations in this cohort harbored coexisting mutations in TP53; mice with loss of TP53 function with BRAF V600E mutations are known to develop ATC (40). Among those BRAF V600E cases that did not have TP53 mutations, two-thirds demonstrated activating mutations in PIK3CA (H1047R in T15/T11). In a mouse model, PIK3CA H1047R is unable to drive thyroid tumorigenesis independently; however, in mice with both PIK3CA H1047R and BRAF V600E mutations, ATC will occur (41). Notably, in these ATC cases featuring BRAF V600E in coexistence with PIK3CA or TP53 mutations, few, if any, mutations in other known ATC-related or COSMIC genes are observed.

Among cases harboring NRAS/HRAS/KRAS mutations, the de-differentiation process is less clear. Only a single TP53 mutation (T9) and no PIK3CA mutations were observed. RAS mutations in thyroid cancer appear to be independently activating of the PI3-K-Akt pathway (38), yet many such cases of WDTC do not progress to ATC. An intriguing candidate not previously reported in thyroid malignancy identified by exome sequencing is EIF1AX. This gene encodes a protein with known roles in cell proliferation and apoptosis. In the current study, EIF1AX mutations were found in 22% of ATC cases harboring RAS mutations. It was not observed in any cases without a coexisting RAS mutation, nor was it observed in any cases with BRAF V600E mutations. Two of the three mutations were missense and the third was at a splice site; all were predicted to be damaging (Table 2). Furthermore, cell line ACT-1, which features wild-type BRAF but mutant NRAS, also demonstrated the EIF1AX mutation in the same residue mutated in sample T21. Moreover, in the validation cohort assessed here, cell line c643 (known to be BRAF wild-type) demonstrated the same splice site mutation as sample T9 from the whole exome cohort mutations in EIF1AX have recently been identified (via WES) in a significant fraction of uveal melanomas (42). These mutations are postulated to be late events in melanoma progression and have prognostic significance (43,44). Moreover, the recently released analysis of the genetic landscape of papillary thyroid cancer by the Cancer Genome Atlas Research Network has also identified EIF1AX as a possible novel driver of thyroid tumorigenesis (33). While the contribution of EIF1AX to tumor progression remains unclear, it has been shown that increased EIF1AX activity triggers cell proliferation in vitro (45). As a recurrent event coexisting with RAS-mutated ATC cases, EIF1AX is an attractive target for further investigation.

ATC also appears to arise de novo, without previous overt signs of differentiated cancer (39). Potential driver mutations in ErbB pathway genes (ERBB2 (D387N), NF2 (E103X) and mTOR (R164Q) were identified in two cases. These three genes have driver properties based on experimental findings in unrelated tumor types, and both ERBB2 (HER-2) and mTOR are overexpressed or mutated in various malignancies (46,47). The specific ERBB2 and mTOR mutations described in T5 have not previously been annotated as COSMIC variants. The ERBB2 D387N mutation is located in the extracellular domain; this region has unknown biological significance but is a commonly mutated domain in urothelial cancer (48). The truncating NF2 mutation in Case T17 was found to overlap a region of LOH, implying that the bi-allelic inactivation of NF2 in this case could represent a crucial event in progression of this tumor. Such alterations in the ErbB pathway alterations may potentially have an independent driver role in the development of ATC. Several additional recurrently mutated genes annotated in the COSMIC database previously not described in ATC were also identified. In addition to EIF1AX (discussed above), the most frequent were USH2A (four cases) and HECTD1 (three cases). Of these cases, USH2A is of particular interest. The gene encodes uscherin, which is ubiquitously expressed and involved in extracellular matrix binding via basement membrane-based interaction with collagen IV and fibronectin. USH2A is a COSMIC-annotated gene with various frequencies of somatic
mutations observed in malignant lesions, and the gene was recently postulated to be one of the top 10 mutated genes across various human tumor types (49). Germ line mutations in USH2A are associated with Usher syndrome type 2A (OMIM: 276901), a disorder characterized by hearing deficiencies and retinitis pigmentosa (50). Potential alterations in ECM or cellular adhesion functions merit close attention in highly invasive and metastatic tumors such as ATC.

Two ATC cases exhibited a ‘hypermutator’ phenotype that has been observed in other cancers but has not been described in ATC (51,52). The probable etiology of this phenotype is the presence of highly damaging Mutl homolog mutations in both cases (MLH1 in T2 and MLH3 in T11; T11 also having a MSH5 mutation). Damaging mutations in this family of genes results in hypermutability in both human cancers and experimental yeast models (53,54). A single case of ATC has been reported in an individual with hereditary nonpolyposis colon cancer (HNPPC—Lynch syndrome) (55). Both cases here demonstrated TP53 mutations as well, but as both were BRAF and RAS wild-type, the Mutl homolog mutations may be crucial contributors to tumor initiation or progression. A third case demonstrated a mutation in the MutS homolog MSH6 (T10); this tumor did exhibit BRAF V600E.

Five ATC cases (23%) did not demonstrate somatic mutations in either known thyroid cancer-related genes or genes with known driver properties. The mechanism of ATC tumorigenesis in these samples is limited to speculation, but might be attributable to somatic mutations in COSMIC genes of unknown function (Supplementary Material, Table S5), alternatively due to epigenetic or genetic alterations such as gene fusions; this is noteworthy as several rearrangements are recognized to drive WDTC tumorigenesis, such as RET, ALK or NTRK3 fusions in PTC and PAX8/PPARG in FTC (33,39). Interestingly, the lack of potential driver mutations in 22.7% (5/22) cases in the exome sequencing cohort roughly mirrors the 15.3% (74/484) of cases found to harbor chromosomal rearrangements and translocations in a recent analysis of the genetic landscape of PTC (33). Regarding the most common previously reported gene variants characterized in ATC, significant heterogeneity exists in the reported frequency of mutation. However, no significant difference in the rate of TP53, BRAF or RAS-family mutations were observed when historical reports were compared with the frequencies observed in the current study (12,15,36–38,56) (Supplementary Material, Table S4). Garcia-Rostan et al. (57) initially reported mutations in exon 3 of CTNNB1 occurred in 65.5% of ATCs; a statistically significant difference in prevalence compared with the single CTNNB1 mutation is observed here (also in exon 3). However, subsequent reports have found that the rate of CTNNB1 mutation seems to approximate the prevalence observed here (0–4.5% of ATC) (14,15). Other uncommon genetic events reported in thyroid malignancy were observed only rarely or not at all (such as mutations in RASAL1, PTEN, or Akt), likely due to the small sample size mandated by the rarity of ATC and the infrequency of these events (16,30,58).

Ontology analysis of all somatic mutations demonstrated during WES revealed clusters in the MAPK and RAS/Erk3 pathways (Fig. 2). Prior studies had surmised the importance of these pathways in ATC tumorigenesis. However, a significant number of somatic mutations in genes associated with these pathways but previously not shown to be altered in ATCs were confirmed here. An overall enrichment of mutations in apoptosis-related genes was also appreciated. This persisted even when limiting analysis to COSMIC-annotated gene variants; presumably, these variants are involved in dedifferentiation as large-cohort landscape analyses of WDTC have not demonstrated similar findings. DAPPLE analysis confirmed the above findings; also identifying clusters in Erb2/RAS/MAPK pathway and apoptosis-related proteins (Supplementary Material, Fig. S1). This may correlate with the global up-regulation of TGF-β genes described by Pita et al. in ATC (13); as RAS/MAPK signaling is known to integrate with the TGF-β pathway to regulate transcriptional activity. This synergy appears to have a significant effect on gene expression in ATC and should be a topic for ongoing study (59,60).

Examination of the mutational landscape of the study cohort revealed several trends. Three broad, mutually exclusive groups of ATC can be characterized based on their defining mutation: (1) tumors harboring BRAF V600E, (2) tumors with mutations within Ras-family genes, (3) tumors without either BRAF or Ras mutations (non-BRAF/non-Ras). Presumably, the majority of tumors in Groups (1) and (2) arose from pre-existing WDTC, a subset of tumors in Group (3) may also have arisen from WDTC via driver mechanisms not completely detectable by exome sequencing (i.e. gene rearrangements) or by novel mechanisms suggested here (such as microsatellite instability). Within Groups (1) and (2), dedifferentiation occurs via a small number well-recognized mechanisms including additional mutations in genes such as TP53, PIK3CA and others; EIF1AX may be a novel driver within this group identified here by exome sequencing in ATC. Within the non-BRAF/non-Ras group of tumors, potential drivers of de-differentiation including NF1, ERBB2 and MTOR and hypermutator MLH mutations.

Although the genetic and epigenetic events that result in thyroid malignancy continue to be defined with increasing clarity, some patients with certain subtypes of thyroid cancer, such as ATC, persistently have dismal outcomes. Of the recognized or novel driver genes identified here, many have pharmaceuticals targeting their action already in clinical or research usage. While several intriguing findings meriting further investigation are described here, perhaps the most noteworthy conclusion from this study is the potential role for next-generation sequencing of all ATC cases upon diagnosis. Such an effort may identify candidates for empiric targeted therapy in order to mitigate the highly lethal burden of this disease.

Materials and Methods
Sample acquisition
The exome sequencing cohort included matched tumor and normal samples from 22 ATC cases. Regardless of tissue source, ATC samples are histologically heterogeneous as different areas of the tumor contain a mix of viable and necrotic tumor cells, normal parenchyma and stromal tissue. In order to maximize sample accrual, we utilized formalin-fixed, paraffin-embedded (FFPE) archival tumor samples for tissue acquisition for 15 out of 22 cases (T1–15; 68%). Recently, FFPE-derived tissue has been demonstrated to be viable for effective WES (61,62). These patients underwent surgical resection at Yale-New Haven Hospital (New Haven, CT, USA). Briefly, all surgical pathology reports containing ‘anaplastic’ and ‘thyroid’ search terms from 1988 to 2012 were manually reviewed and tissue blocks were obtained for appropriate cases. Blocks with adequate remaining tissue were sectioned and re-reviewed by an experienced pathologist to confirm the presence of ATC and matched normal tissue for WES. The histopathological diagnosis was established according to WHO classification. Three 1 mm tissue cores were obtained and post-core sectioning was performed and reviewed to ensure tissue fidelity. Paraffin was then enzymatically removed and genomic DNA was
extracted and sheared via sonication. Seven additional cases of ATC (T16–22, 32%) were obtained from fresh-frozen tissue samples with matched normal tissue from fresh-frozen thyroid (n = 5) or leukocyte DNA (n = 2) from the Karolinska University Hospital (Stockholm, Sweden). Genomic DNA was extracted using the DNeasy Blood and Tissue DNA isolation kit (Qiagen, Hilden, Germany) in accordance with the manufacturer’s instruction. All fresh-frozen specimens were sectioned in parallel to the DNA extraction and subsequently re-analyzed by light microscopy to verify appropriate representation of tumor or normal tissue. The procurement of tissue and subsequent genomic analyses were approved by the Yale University Institutional Review Board, New Haven, CT, USA, and the local ethical review board at Karolinska Institutet, Stockholm, Sweden.

The validation cohort consisted of samples of fresh-frozen tissues from a total of 16 histologically confirmed ATCs collected at the Karolinska University Hospital, Stockholm, Sweden. Tissue samples were snap-frozen following resection and genomic DNA was extracted using the DNeasy Blood and Tissue DNA isolation kit (Qiagen, Hilden, Germany) in accordance with the manufacturer’s instruction. Fourteen of sixteen tissue samples were examined and confirmed by an experienced endocrine pathologist to demonstrate adequate representation of ATC tumor cells (Supplementary Material, Table S7).

ATC cell lines
Four established ATC cell lines (8305C, 8505C, ACT-1 and BHT101) were also exome-sequenced in addition to the 22 matched pairs of ATC and normal tissues described above using identical methodology for DNA isolation. 8305C and 8505C were purchased from Sigma, MO, USA. BHT101 was purchased from DSMZ Germany, while ACT-1 was generously provided by R.E. Schweppe, University of Colorado School of Medicine, Aurora, CO, USA (originator—Naoyoshi Onoda, Osaka City University, Japan). Cell lines used for the exome analysis (8505C, 8305C, BHT101 and ACT-1) were authenticated to be distinct and of human ATC origin by short tandem repeat and single-nucleotide polymorphism array analysis (83). Additional ATC cell lines used for the validation cohort (Supplementary Material, Table S7) were kindly provided by Dr Nils-Erik Heldin (Uppsala University, Sweden). These lines were reported previously to be of human origin and authenticated by genotyping of short tandem repeats (35,64).

Exome capture, massively parallel sequencing and analysis
Genomic DNA samples generating an adequate high-quality library were subjected to exome capture and sequencing as previously described (23). Briefly, adaptors of known sequence were ligated to genomic DNA fragments which were amplified by ligation-mediated PCR and were then subject to capture utilizing the NimbleGen 2.1M human exome array. Exome-specific DNA was eluted and then underwent 74 base paired-end sequencing on the Illumina HiSeq 2000 instrument according to the manufacturer’s instructions. SAMtools was utilized for base calling and removal of PCR duplicates. Reads were then mapped to the human reference genome GRCh37/hg18 using the ELAND program. Significance of somatic variant calls was assessed by comparing reference and non-reference reads utilizing Fisher’s exact test and were manually curated on a sample-by-sample basis to determine high-quality reads. Known variants in annotated databases were excluded and novel variants were evaluated for impact on transcriptional and/or translational processing as well as sequence conservation.

CNV prediction and LOH analysis
LOH loci were identified by comparing B allele frequencies (BAF) of a known array of SNPs in matched tumor and normal samples. Regions with apparent BAF shift were then manually curated to determine LOH. Comparative analysis of coverage depth between tumor and normal samples in 500 kb capture intervals was utilized to identify regions of somatic CNV. Significance was assessed by randomly distributing CNVs in >108 permutations and assessing the likelihood of observing the distribution by chance alone. Again, a false-discovery cut-off of <0.25 was considered significant. A GISTIC-like peel algorithm was utilized to assess the significance of individual CNV peaks.

Sequence validation
Thirty somatic mutations of interest were selected for validation via Sanger sequencing. Flanking primers were designed using an NCBI PrimerBlast (http://blast.ncbi.nlm.nih.gov) and genomic DNA was amplified via PCR and verified to be of correct size by agarose gel electrophoresis (Supplementary Material, Table S3). Putative mutations were sequenced using forward and reverse primers and confirmed by two independent reads. All cases in the validation cohort were analyzed for the demonstrated mTOR R164Q/M2327I, ErbB2 D387N, BRAF V600E mutations defined during exome sequencing and the seven coding exons of EIF1AX were sequenced in a similar fashion. Chromatograms were analyzed using CodonCode Aligner software (CodonCode Corporation; Centerville, MA, USA).

Ontology analysis
We applied the DAVID Bioinformatics Resources 6.7 database (http://david.abcc.ncifcrf.gov) to aid in the identification of significantly altered biological processes and pathways among the 22 ATC cases. By using a functional annotation tool and a user-customized gene background, gene-annotation enrichment analyses as well as KEGG pathway mapping were performed. Significantly enriched GO terms and pathways were selected based on Benjamini-corrected P-values, to allow a balance between discovery of statistically significant genes and restriction of false-positive occurrences. Pathway analyses were also carried out using the Genomatix Pathway System (www.genomatix.de), in which information is extracted from public databases to present >400 canonical pathways as well as extended networks based on literature data rather than GO terms. We furthermore employed the Disease Association Protein-Protein Link Evaluator (DAPPLE—www.broadinstitute.org/mpg/dapple/dapple.php) to evaluate important physical associations between proteins encoded by mutated genes in the ATC cohort. The analysis was made allowing 10,000 permutations and a common interactor binding degree cut-off of 2. Associations are reported with the Bonferroni-corrected P-values representing the statistical significance of a number of network connectivity parameters reported in the literature.

Statistical analyses
Statistical calculations were carried out using the IBM SPSS Statistics 19 software (IBM, Armonk, NY, USA). Tests for normality of continuous variables were performed using the Shapiro–Wilk test, and Spearman’s correlation, t-test, Kruskal–Wallis and
ANOVA tests were used to test for significance between various clinical parameters and overall mutational burden. P-values of <0.05 were considered as statistically significant.

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

The authors would like to thank Aruna Madan, MD, Lisa Ånfalk and John Overton, PhD for their invaluable assistance in sample collection, submission and assistance. Dr Robert Udelsman, MD, MBA, FACS, FACE, is thanked for his unwavering support of this project.

Conflict of Interest statement. The funding sources had no role in the design, conduct, or reporting of this study and the authors have no conflict of interest to report.

Funding

This work was supported by the Damon Runyon Cancer Research Foundation (T.C.); the Stockholm County Council (C.C.J.); the Swedish Cancer Society (C.C.J., J.Z.); the Swedish Research Council Foundation (T.C.); the Stockholm County Council (C.C.J.); the Swedish Research Council (C.L., A.S.); the Agency for Science, Technology and Research, Singapore (G.G.); and the Howard Hughes Medical Institute (R.P.L.).

References


