GENOMIC ANALYSIS OF HEAD AND NECK ENDOCRINE GLANDS

by

Katayoon Kasaian

B.Sc., The University of British Columbia, 2005
B.CS., The University of British Columbia, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Bioinformatics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2015

Abstract

Discovering biomarkers and molecular drivers of head and neck endocrine tumors was the inspiration for this thesis. Here, I describe the molecular evaluation of tumors of the thyroid and parathyroid endocrine glands for the purpose of identifying somatic driver alterations in these cancers. While molecular interplay of the germline genomic background of an individual and the somatic genome that emerges throughout the lifetime plays significant roles in increasing the susceptibility to cancer and in driving the malignant phenotype, the major known contributors to cancer remain the acquired somatic mutations. Analysis of a sporadic and recurring parathyroid carcinoma, with incidence of 1 per million population, revealed mutations in *mTOR*, *MLL2*, *CDKN2C* and *PIK3CA* and comparison of patient-matched primary and recurrent malignant tumors uncovered loss of *PIK3CA* activating mutation during the evolution of the tumor. Loss of the short arm of chromosome 1 along with somatic missense and truncating mutations in *CDKN2C* and *THRAP3* provided new evidence for the potential role of these as tumor suppressors. Hürthle cell thyroid carcinoma accounts for a small proportion of all thyroid cancers; however, this malignancy often presents at an advanced stage and poses unique challenges. Genomic analysis revealed large regions of copy number variation encompassing nearly the entire genomes accompanied also by near haploidization. Moreover, I identified loss-of-function mutations of the tumor suppressor gene *MEN1* in 4% of patients. Repeated alterations of the epigenetic machinery in anaplastic thyroid carcinoma, one of the most fatal of all adult solid malignancies, and novel gene fusions including *MKRN1-BRAF*, *FGFR2-OGDH* and

SS18-SLC5A11 are reported here. The transcriptomic analysis suggested known drug targets such as *FGFRs*, *VEGFRs*, *KIT* and *RET* to have low expressions in this cancer; however, through integrative data analysis, I identified the mTOR signaling pathway as a potential therapeutic target for anaplastic thyroid cancer. Molecular analysis of papillary thyroid carcinoma and benign thyroid nodules revealed very low mutation rates in these tumors with *CYP1B1*, *PTPRE, CTSH* and *RUNX1* emerging as promising diagnostic markers. The key somatic mutations identified in these studies can serve as novel diagnostic markers as well as therapeutic targets.

Preface

Portions of Chapter 1 have been published in a review paper and a book chapter: 1. Katayoon Kasaian, Steven JM Jones. (2011). A new frontier in personalized cancer therapy: mapping molecular changes. Future Oncology. 2011 Jul;7(7):873-94. doi: 10.2217/fon.11.63. Copyright by Future Medicine Ltd. I was the author of this review paper; this work was done under the supervision of SJMJ. 2. Katayoon Kasaian, Yvonne Y Li and Steven JM Jones. (2014). Chapter 9 - Bioinformatics for Cancer Genomics, In Cancer Genomics, edited by Graham Dellaire, Jason N Berman, Robert J Arceci, Academic Press, Boston, 2014, Pages 133-152, ISBN 9780123969675. Copyright by Elsevier. I was the lead author on this book chapter. I wrote the majority of the text with input from YYL in "Data Interpretation and Integration" section. I produced Table 1.1 and Figure 1.2, YYL Figures 1.1 and 1.3. This work was done under the supervision of SJMJ.

A version of Chapter 2 has been published as Katayoon Kasaian, Sam M Wiseman, Nina Thiessen, Karen L Mungall, Richard D Corbett, Jenny Q Qian, Ka Ming Nip, Ann He, Kane Tse, Eric Chuah, Richard J Varhol, Pawan Pandoh, Helen McDonald, Thomas Zeng, Angela Tam, Jacquie Schein, Inanc Birol, Andrew J Mungall, Richard A Moore, Yongjun Zhao, Martin Hirst, Marco A Marra, Blair A Walker, and Steven JM Jones. (2013). Complete genomic landscape of a recurring sporadic parathyroid carcinoma. Journal of Pathology, 230: 249–260. doi: 10.1002/path.4203. Copyright by Wiley. I was the lead researcher and author of this publication. I analyzed and interpreted data, designed validation experiments, performed

literature search, generated figures and wrote the manuscript. SJMJ, SMW, MAM conceived and designed study; SMW performed surgery; BAW provided pathology review; PP, AT, JS collected specimens and data; NT, AH processed transcriptome datasets; RDC, EC, RJV processed whole genome datasets; IB, KLM, KMN, JQQ conceived and performed *de novo* assembly of sequence data; TZ, KT, RAM performed validation experiments; HM, AJM, RAM, YZ, MH carried out library construction and sequencing experiments. The work described in this chapter and the associated methods were approved by the University of British Columbia's Research Ethics Board [Human ethics certificate # H08-01989 and biosafety certificate # B14-0044].

A version of Chapter 3 has been published as Katayoon Kasaian, Ana-Maria Chindris, Sam M Wiseman, Karen L Mungall, Thomas Zeng, Kane Tse, Jacqueline E Schein, Michael Rivera, Brian M Necela, Jennifer M Kachergus, John D Casler, Andrew J Mungall, Richard A Moore, Marco A Marra, John A Copland, E Aubrey Thompson, Robert C Smallridge, and Steven JM Jones. (2015). *MEN1* Mutations in Hürthle Cell (Oncocytic) Thyroid Carcinoma. Journal of Clinical Endocrinology and Metabolism. 2015 Apr;100(4):E611-5. doi: 10.1210/jc.2014-3622. Copyright by Endocrine Society. I was the lead researcher and author of this publication. I performed data analysis, designed validation experiment, generated figures, performed literature search and wrote the manuscript. SMW, MAM, JAC, RCS, SJMJ conceived and designed study; AMC, JES, BMN, JMK, EAT acquired samples and provided technical assistance; MR provided histopathological interpretation; JDC provided pathology specimens; KLM assembled two WGS

datasets; KT designed primers; TZ ran validation experiment; AJM constructed WGS libraries; RAM ran sequencing experiments. The work described in this chapter and the associated methods were approved by the University of British Columbia's Research Ethics Board [Human ethics certificate # H08-01989 and biosafety certificate # B14-0044].

A version of chapter 4 has been submitted for publication: Katayoon Kasaian, Sam M Wiseman, Blair A Walker, Jacqueline E Schein, Yongjun Zhao, Martin Hirst, Richard A Moore, Andrew J Mungall, Marco A Marra, and Steven JM Jones. (2015). The Genomic and Transcriptomic Landscape of Anaplastic Thyroid Cancer: Implications for Therapy. I was the lead researcher and author of this manuscript. I performed data analysis, generated figures, performed literature search and wrote manuscript. SMW, MAM, SJMJ conceived and designed study; BAW provided pathology review; JES acquired samples and provided technical assistance; YZ, MH, RAM and AJM constructed libraries and ran sequencing experiments. The work described in this chapter and the associated methods were approved by the University of British Columbia's Research Ethics Board [Human ethics certificate # H08-01989 and biosafety certificate # B14-0044].

Portions of Chapter 5 have either been published or are in preparation for submission. Section 5.3.1, "Cav1 and Gal3 Immunohistochemical Analysis", was published as Jay Shankar, Sam M Wiseman, Fanrui Meng, Katayoon Kasaian, Scott Strugnell, Alireza Mofid, Allen Gown, Steven JM Jones, and Ivan R Nabi. (2012). Coordinated expression of galectin-3 and caveolin-1 in thyroid cancer. Journal of Pathology, 228: 56–66. doi: 10.1002/path.4041. Copyright by Wiley. I

performed the statistical analysis for the tissue microarray experiment and contributed to the writing of Methods and Results sections describing this analysis. JS performed the western blots, RhoA GTP assay, immunoflouresence for Gal3 and Cav1, migration assay and siRNA-mediated knockdown, with AM assisting with the western blots FAK/actin immunoflouresence; SMW, SS, AG performed tissue microarray preparation, staining and scoring; FM performed quantification of the FRAP experiments; SMW, SJMJ and IRN conceived and designed study. The work described in this section and the above-mentioned publication and the associated methods were approved by the University of British Columbia's Research Ethics Board [certificate # H06-00217]. Section 5.2.1, "Prognostic Significance of Papillary Thyroid Carcinoma Presentation Mode", was published as Heywood Choi, Katayoon Kasaian, Adrienne Melck, Kaye Ong, Steven JM Jones, Adam White, Sam M Wiseman. (2015). Papillary Thyroid Carcinoma: Prognostic Significance of Cancer Presentation. American Journal of Surgery. pii: S0002-9610(15)00180-4. doi: 10.1016/j.amjsurg.2014.12.047. Copyright by Elsevier. I performed the statistical analysis and wrote the related portions of the Methods and Results sections. HC, AM, KO, AW performed retrospective review of patient data, constructed patient database and prepared data for analysis; SJMJ, SMW conceived and designed study. The work described in this section and the above-mentioned publication and the associated methods were approved by the University of British Columbia's Research Ethics Board [certificate # H13-00118]. Section 5.2.2, "Prognostic Significance of Tumor Laterality in Papillary Thyroid Cancer", is in preparation for submission: Sarah E Moore, Katayoon Kasaian, Steven JM Jones, Adrienne Melck, and Sam M Wiseman. (2015). Papillary Thyroid Cancer: Epidemiology of Bilateral Disease. I performed

the statistical analysis and wrote the related portions of the Methods and Results sections. SEM, AM performed retrospective review of patient data, constructed patient database and prepared data for analysis; SJMJ, SMW conceived and designed study. The work described in this section and the above-mentioned manuscript and the associated methods were approved by the University of British Columbia's Research Ethics Board [certificate # H12-03669]. Whole genome and transcriptome studies of benign thyroid nodules and papillary thyroid carcinoma described in sections 5.4 and 5.5 are based on unpublished work. I have conducted all data analysis for these sections. SMW and SJMJ conceived and designed study and the biospecimen, library construction and sequencing cores at the British Columbia Cancer Agency Genome Sciences Centre generated the data; the sequencing validation core designed primers for and ran *SLC34A2* validation experiment. The work described in these sections and the associated methods were approved by the University of British Columbia's Research Ethics Board [Human ethics certificate # H08-01989 and biosafety certificate # B14-0044].

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ATC	Anaplastic Thyroid Carcinoma
BH	Benjamini-Hochberg
BWA	Burrows-Wheeler Alignment
CaSR	Calcium-Sensing Receptor
CHIP-seq	Chromatin Immunoprecipitation followed by sequencing
CNV	Copy Number Variation
COSMIC	Catalogue of Somatic Mutations in Cancer
CT	Computerized Tomography
DAVID	Database for Annotation, Visualization and Integrated Discovery
DTC	Differentiated Thyroid Cancer
EGA	European Genome-phenome Archive
FFPE	Formalin Fixed Paraffin Embedded
FIHP	Familial Isolated Hyper-Parathyroidism
FNA	Fine Needle Aspiration
FTC	Follicular Thyroid Cancer
HPT-JT	Hyper-Parathyroidism Jaw Tumor
IGV	Integrative Genomics Viewer
KEGG	Kyoto Encyclopedia of Genes and Genomes
LOH	Loss Of Heterozygosity

MACIS	Metastasis, Age, Completeness of resection, Invasion, and Size
MEN	Multiple Endocrine Neoplasia
MMR	Mismatch Repair
MOJO	Minimum Overlap Junction Optimizer
MRI	Magnetic Resonance Imaging
MTC	Medullary Thyroid Cancer
NGS	Next Generation Sequencing
NHEJ	Non-Homologous End Joining
OMIM	Online Mendelian Inheritance in Man
PC	Parathyroid Carcinoma
PCR	Polymerase Chain Reaction
PHPT	Primary Hyperparathyroidism
PTC	Papillary Thyroid Carcinoma
PTH	Parathyroid Hormone
RPKM	Reads Per Kilobase of exon model per Million mapped reads
SEER	Surveillance, Epidemiology, and End Results
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variation
SV	Structural Variation
TCGA	The Cancer Genome Atlas
TMA	Tissue Microarray

UCSC	University of California Santa Cruz

Acknowledgements

I would like to thank Drs Steven Jones and Sam Wiseman for giving me the opportunity to be involved in this very exciting project. I would also like to thank Drs Rob Holt and Sohrab Shah for the guidance they provided me and for their careful examination of this thesis.

My most sincere gratitude and appreciation go to the Genome Sciences Centre family housed at the British Columbia Cancer Agency. Not only I have learned from and received tremendous support from everyone during my graduate work, but I also had the opportunity to know and spend time with some of the most amazing individuals I have ever met. It has been an absolute honor and a privilege to be a member of the GSC. Special thank yous go to Drs Obi Griffith, Gordon Robertson and Marco Marra for giving me invaluable opportunities at pivotal points during my studies.

This thesis in all its apparent weight and importance is only a very small part of who I have become; yet all I am and all I have, including this work, would not exist or otherwise would have been completely meaningless if it was not for my family.

Chapter 1: Introduction[1]

1.1 Head and Neck Endocrine Tumors

1.1.1 Thyroid Cancer

Thyroid cancer is a relatively rare disease, accounting for 1-5% of all cancers in females and less than 2% in males [1]. Although rare, thyroid carcinoma is the most common endocrine malignancy and its incidence rate has increased in most parts of the world [1]. According to statistics released by the Canadian Cancer Society in 2014, thyroid cancer is the most rapidly increasing of all major cancers in Canada. The incidence rate has also more than doubled in the United States [2], France [3], and Australia [4], with the average worldwide increase of 48.0% among males and 66.7% among females from 1973-1977 to 1998-2002 [1]. The increased use of techniques such as fine-needle aspiration biopsy and thyroid ultrasound, in addition to, physical examination which was the sole primary method of detection before the late 1990's, has led to the discovery of smaller thyroid nodules and has contributed to the increase in the incidence rate [2]. Although the stable mortality rate from these malignancies supports the hypothesis of over-diagnosis [2], some suggest that the increase might be real [1]. For instance, although the

[1] Portions of this chapter have been published, and the author contributions are provided in the Preface as per the University of British Columbia PhD thesis guidelines: Katayoon Kasaian, Steven JM Jones. (2011). A new frontier in personalized cancer therapy: mapping molecular changes. Future Oncology. 2011 Jul;7(7):873-94. doi: 10.2217/fon.11.63. Copyright by Future Medicine Ltd. Katayoon Kasaian, Yvonne Y Li and Steven JM Jones. (2014). Chapter 9 - Bioinformatics for Cancer Genomics, In Cancer Genomics, edited by Graham Dellaire, Jason N Berman, Robert J Arceci, Academic Press, Boston, 2014, Pages 133-152, ISBN 9780123969675. Copyright by Elsevier.

incidence of cancer diagnosis after the late 1990's increased in individuals from high socioeconomic backgrounds with better access to health care and diagnostic technologies, there was a steady increase in incidence in individuals from low socioeconomic backgrounds and the increasing diagnostic trend for larger malignant tumors remained the same in both groups [5].

With the discovery of smaller malignant nodules and the knowledge that these cancers are found to have an indolent course in a large subset of the population, the clinicians are faced with the challenge of stratifying patients based on their risk of recurrence and metastasis, and treating only those individuals at higher risk [5,6]. The current standard of care usually requires the removal of the thyroid gland from the patient presenting with a malignancy and hence most patients diagnosed with thyroid cancer undergo total thyroidectomy [2]. Although complications of total thyroidectomy including permanent hypoparathyroidism, recurrent laryngeal nerve damage and vocal cord paresis are uncommon, they are not negligible [7]. These concerns, and the need for hormone replacement therapy for life, suggest a need for more sensitive diagnostic tools for routine clinical use.

1.1.1.1 Papillary Thyroid Carcinoma

Papillary subtype of thyroid malignancies is the most common form of the disease, accounting for 80-85% of all thyroid cancers [8]. Papillary thyroid carcinoma (PTC) along with follicular and

Hürthle cell subtypes compromise the well-differentiated thyroid cancers (DTC) [9]. These are the cancers of the follicular cells of the gland, which produce the thyroid hormones T3 and T4 [10]. Papillary cancers generally have a favorable prognosis with 25-year survival rate at > 95% [8]; however, some patients are at high risk for developing recurrence and death [11]. It has been suggested that the step-wise de-differentiation of these malignancies could lead to the more aggressive forms such as the poorly differentiated and anaplastic (undifferentiated) carcinomas [9]. Several recurrent mutations have been described in PTCs, the most frequently observed mutation is the BRAF p.V600E activating mutation, present in 40-45% of cases [12]. Other observed genetic alterations include rearrangements of *RET/PTC*, particularly in individuals exposed to ionizing radiation [11] as well as other rare mutations such as *TRK* rearrangements [11] and *RAS* point mutations [13].

1.1.1.2 Anaplastic Thyroid Carcinoma

Anaplastic thyroid carcinoma (ATC), the undifferentiated subtype of thyroid malignancies lacking any evidence of follicular differentiation or even that of epithelial origin, is the least common form of the disease; it accounts for only 1-2% of all thyroid cancers [9]. Although the incident rate is low, the mortality and morbidity rates are extremely high [14]. ATC is the most aggressive type of thyroid cancer and one of the deadliest forms of all human malignancies [15]. Patients usually present with advanced disease demonstrating local invasion and distant metastases; 90% of patients die within 6 months of diagnosis with the median survival rate at

just 4 months [15]. There is an immediate need to characterize these malignancies on the molecular level to identify the driver mutations, which could subsequently assist in administering targeted therapeutics. Known alterations in ATCs include *TP53* point mutations which are otherwise rare in all other subtypes [16] as well as *BRAF* [14] and beta-catenin point mutations [17]. Compared with DTCs, anaplastic carcinomas are more likely to be aneuploid [14] and harbor large regions of gene copy loss and gain [18].

1.1.1.3 Hürthle Cell (Oncocytic) Thyroid Carcinoma

Hürthle cell or oncocytic thyroid carcinoma accounts for about 2-3% of thyroid cancers. A cell is referred to as an oncocytic cell if it demonstrates an abnormal accumulation of mitochondria in the cytoplasm. These cells have a granular appearance under the microscope and can be found in several, often metabolically active, organs such as kidney, thyroid, parathyroid, salivary and adrenal glands [19,20]. Pathological review of a thyroid tumor will indicate it to be an oncocytic nodule if 75% or more of its constituent cells are Hürthle cells. Those nodules with signs of invasion to local tissue or distant metastasis are diagnosed as malignant tumors. No molecular connections are established between benign and malignant oncocytic tumors and no comprehensive molecular profiling of these tumors has been performed before. To date, two studies have reported *NRAS* [21] and *GRIM19* [22] mutations in a small subset of oncocytic thyroid carcinomas. This malignancy has a poorer prognosis than the more common DTCs and yet no information regarding its molecular alterations is available. Currently, the treatment

options are surgery, radioactive iodine treatment, chemotherapy and in some cases external beam radiation therapy.

1.1.2 Benign Thyroid Nodules

Despite the rarity of thyroid cancer, benign thyroid lesions are found at a high frequency in the population and high-resolution ultrasound can detect thyroid nodules in 19-67% of randomly selected individuals [23]. The tumors can be detected by physical examination or incidentally through imaging done for other indications and are more common in women and the older population [24]. Histologically, an encapsulated tumor is referred to as an adenoma whereas one lacking a defined fibrous capsule and with poor boundaries with the normal tissue is referred to as a hyperplastic adenomatous nodule [25]. Benign tumors can also be classified as "cold", "normal", or "hot" indicative of decreased, normal or increased uptake of iodine, respectively [26]. "Cold" nodules are generally more likely to be malignant and thus should be examined more carefully [24]. Nodular goiter, another form of benign thyroid disease, refers to the enlargement of the gland with either single or multiple nodules affecting the normal structure of the organ [26]. Neither the size of the nodule or number of nodules in the gland is indicative of malignancy [24]. Factors such as iodine intake level, exposure to radiation, smoking, age and gender are known to influence the development of benign thyroid tumors; however, the interplay of these mediators and the individual's genetic makeup is the ultimate determining factor for the occurrence of these nodules [26]. Surgery, especially when tumors

are causing compressive symptoms, and radioiodine therapy, are the main modes of therapeutic intervention for these tumors [24]. To date, no comprehensive molecular characterization of benign thyroid tumors has been performed and hence the available knowledge on the molecular alterations leading to these lesions is very limited.

1.1.3 Parathyroid Carcinoma

Parathyroid carcinoma is an extremely rare cancer type. Uncontrolled cell division in the body's smallest organ, the parathyroid gland, causes an extensive and deadly imbalance in the blood calcium level given that the main function of this endocrine gland is controlling the level of calcium. A large percentage of patients with parathyroid carcinoma have multiple recurrences of the disease throughout their lifetime and the extreme calcium level, not the disease burden, is often the cause of death in these patients. Although some parathyroid tumors, usually benign nodules, are observed in individuals with multiple endocrine neoplasia 1 and 2A syndromes, parathyroid carcinomas tend to be sporadic [27]. High expression of *CCND1* has been observed in parathyroid cancer; however, an inversion involving CCND1 ultimately leading to its high expression was first reported in benign parathyroid adenomas [28,29]. Hence, the deregulation of this cell cycle activator is unlikely to be the sole driver of malignancy. The rare nature of this disease has hindered a detailed and comprehensive study of this cancer and no recurrent mutations has been identified in association with sporadic parathyroid carcinomas.

1.2 Cancer Genomics

In the late 1800s and early 1900s, David von Hansemann and Theodor Boveri were the first to propose the genetic basis of cancer explaining that aberrant mitosis can lead to the unequal distribution of chromosomes which can in turn produce malignant cells with the ability to grow without control [30,31]. The knowledge that cancer is a genetic disease has driven the scientific community for over 100 years in search of molecular mutations that are associated with various cancer types. Prior to the advent of high-throughput sequencing technologies, lower resolution methodologies were utilized in deciphering the biology of cancer. The majority of these efforts involved single-gene experiments or the examination of a gene family in a small cohort of patients using the then-novel Sanger sequencing. Such initiatives led to the discovery of activating point mutations in oncogenes such as *BRAF*, *KRAS*, *NRAS* and *HRAS* and loss of function mutations of tumor suppressors such as *TP53* with varying frequencies in different head and neck endocrine tumors. This suggested a potential utility for high-resolution sequencing techniques to unveil a more comprehensive profile of these tumors. Since the completion of the Human Genome Project, there has been a revolution in genomic technologies particularly the DNA sequencing methodologies. Advances in massively parallel and high-throughput next generation sequencing (NGS) have enabled cost-effective sequencing of a single human genome at an unprecedented rate, facilitating scientific endeavors never imagined possible before. These improvements have transformed the field of cancer genomics, allowing the complete molecular characterization of large cancer cohorts in hopes of identifying

common tumorigenic pathways and of individual cancer genomes allowing for the delivery of precision cancer medicine in the clinic.

Unraveling the genomic abnormalities that lead to cancer, potential therapeutic targets and the mechanisms behind tumor response or resistance to a particular treatment modality is integral to the advancement of cancer medicine. Therefore, the ultimate goal of the cancer genomics field is to fully explore the potential of the NGS technologies in characterizing different types of cancer on the molecular level, understanding the mechanism of the disease, identifying diagnostic, prognostic and predictive markers and finally translating this knowledge into patient-based therapies. Computational biology and bioinformatic techniques provide solutions for examining complete genetic material of a cancer sample for every type of mutation including single nucleotide variations (SNVs), small insertions and deletions (indels), copy number variations (CNVs), regions of loss of heterozygosity (LOH) as well as structural variations (SVs). These analysis tools and algorithms help to understand complex biological systems by systematically analyzing large data sets and by providing the necessary techniques for integrating various data types. This enables us to derive a global view of the healthy state of a cell and to identify how these are altered in the disease state. The utility of the vastly parallel sequencing machines is not limited to analyzing the genome; the epigenome, the transcriptome and the proteome of a cell can all be investigated through the use of these high throughput technologies (Figure 1.1).

The rest of this chapter reviews the strengths and limitations of different data modalities in cancer genomics and outlines some of the current bioinformatic algorithms and software for data analysis with an emphasis on whole genome and transcriptome analysis tools that were used in the described work of chapters 2, 3, 4 and 5. The specific aim of this thesis was to utilize the power of NGS technologies and the recent advances in bioinformatic software in the study of endocrine tumors of the head and neck particularly those of the thyroid and parathyroid glands. The goal was to identify novel and recurrent mutations associated with these tumors, describe the potential route of oncogenesis for these malignancies and devise diagnostic and therapeutic markers.

1.3 Data Types in Cancer Genomics

1.3.1 Whole Genome and Exome Sequence Data

Cancers arise due to mutations that provide the cell with a growth advantage. In sporadic or non-familial cases of cancer, these somatic events can be identified through the comparison of cancer and normal genomes of a patient. Whole genome shotgun sequencing provides the complete genetic landscape of a tumor specimen and this sequence data can be examined for the presence of various somatic alterations. Although the majority of mutations in any cancer sample fall outside the protein-coding regions, the scientific community for the most part has focused on examining the protein-coding changes not due to a lack of interest but perhaps to a

lack of comprehensive knowledge about the regulatory elements of the genome. Additionally, whole genome sequencing is still not affordable enough to be carried out for individual patients in clinical settings or even in every research laboratory and thus whole exome sequencing has become an appealing alternative. Sequencing only the exomes provides the information encoded in the complete coding region of the genome at a high depth of coverage and for a lower cost than whole genome sequencing. Currently, the NGS technologies provide such high sequence coverage that multiple exome libraries can be indexed, pooled and sequenced in a single experiment without losing any information while decreasing the cost even further. Whole exome sequence data can still unveil small mutations such as SNVs and indels. A few tools have been developed that promise the identification of regions of copy number loss and gain from the exome capture data [32,33]. In addition, examples of SVs detected from whole exome sequencing are evident from the literature [34]. However, most of the progress to date in finding somatic CNVs and SVs has been the result of whole genome sequencing experiments.

Examining the cell's DNA provides a static view of the mutations that could potentially be disrupting protein function. Cells are however dynamic entities, transcribing and translating the genetic information into protein products in accordance to their needs. Studying the dynamic profile of the cell through transcriptome sequencing or characterizing the protein collection of the cell can serve as a powerful tool for identifying disrupted pathways in a disease state.

1.3.2 Whole Transcriptome Sequence Data

It has long been known that there is a global change in the expression of genes in cancer cells compared with their normal counterparts. Some of these alterations, such as changes in the expression of oncogenes and tumor suppressors, will be drivers of the disease while others are the result of the malfunctioning cell and the fragile cancer genome. Using NGS technologies, the complete transcriptome of a cell can now be sequenced, providing a digital count of the expression of all genes. Through whole transcriptome sequencing, also referred to as RNA-seq, expressed mutations can be identified. *De novo* assembly of transcriptome data can also serve as a powerful tool for identifying events such as novel transcripts, skipped exons, retained introns or novel splicing events. Differential expression analysis between malignant and adjacent normal tissues can shed light on the altered pathways in the disease and help in developing diagnostic and prognostic panels; however, such analysis in cancer genomics is hindered due to the typically limited access to neighboring matched normal tissue. Patient's blood usually serves as the normal specimen and though it serves as a good reference for the tumor genome, the expression profile of the blood cells will be entirely different from that of a solid tumor, for instance. Different data types in cancer genomics have distinct strengths and limitations, generating as many datasets as possible using different modalities and their integration is the most promising solution in deciphering cancer signatures (Table 1.1).

1.3.3 Proteomic Data

High throughput techniques such as protein microarrays and mass spectrometry have been developed for studying the complete collection of a cell's proteins, often referred to as the proteome. Proteomic analysis of a biological sample can unveil all the proteins present, their amount, specific post-translation modifications and all protein-protein interactions. Through such analyses of cancer and matched normal tissues or various cancer subtypes, one can identify diagnostic and prognostic biomarkers as well as novel drug targets. Our knowledge of the human proteome however has lagged behind the efforts such as the Human Genome Project, which decoded the sequence of almost the entire genome, mostly due to lack of high-throughput technologies. Understanding the structure and function of proteins is an important step in cancer genomics, leading to conclusions about the function of mutated proteins, whether they contribute to disease initiation and progression and how they can be targeted. The Human Proteome Project launched in 2011 aims to identify the structure and function of at least one protein product of each protein-coding gene [35]. Such efforts combined with improvements in analytical tools and algorithms will lead to better understanding of the functional consequences of DNA mutations.

1.3.4 Epigenomic Data

Next generation sequencing technologies have also enabled the study of the epigenome, the transcriptional control of the cell. Mutations of several epigenetic enzymes are found in various cancers and thus there is increasing evidence that changes in the epigenome and the resultant alterations in the expressional profile of the cell could be the cause of many diseases including cancers. Examining the pattern of epigenetic marks associated with both the DNA and histone proteins throughout the genomes of the cancer and matched normal tissue can provide profound understanding of the changes leading to the disease state. Chromatin immunoprecipitation followed by sequencing (CHIP-seq) [36], with higher throughput and better sensitivity than CHIP-on-chip [37], provides a genome-wide view of specific DNA-protein interactions including histone modification marks. Profiling the methylation state of the genome is also now possible through techniques coupled with high throughput sequencing [38]. These methods are divided into those which enrich for methyl-DNA [39-41], those which utilize methylation-dependent restriction enzymes [42,43] and the third category which is based on direct bisulfite conversion [44-50].

Data generation has arguably become the easiest and the most efficient step in studying a cancer genome. The challenge now is to analyze the sheer volume of generated data and to integrate different types of mutational datasets such as SNVs, indels, CNVs, SVs, expression

profiles and the epigenetic alterations in order to draw a biologically correct and meaningful conclusion about the underlying cause of the disease and how to best treat it.

1.4 Data Analysis

1.4.1 Sequence Data Alignment and Assembly

High-throughput sequencing technologies produce large number of short reads in a relatively short period of time. Application of these technologies in cancer genomics depends on the ability to re-construct the complete genome from these short reads with great accuracy in a time- and memory-efficient manner. Generally, two options exist; one is to align the reads to the reference genome and the other is to perform a *de novo* assembly.

Simply put, alignment refers to the task of finding the location in the complete genome where a sequence read was generated from. This is in essence a string matching problem. Although the standard Smith-Waterman algorithm [51] widely used for the alignment of longer reads provides the most optimal solution, it becomes computationally intractable when working with a large number of short sequence reads. As a result, a growing number of algorithms for the alignment of NGS reads to the human reference genome has been implemented [52-58]; all however face a trade-off between accuracy and speed. The often-ignored limitation of the current aligners is their intended use for the alignment of sequence reads generated from

normal specimens. These tools are not yet optimized for the alignment of reads generated from cancer samples which could potentially contain large insertions and deletions, as well as evidence of other structural variations such as duplications, translocations and gene fusions. In addition, as the sequencing technologies improve and the reads become longer, the majority of the tools for alignment of short reads will not be applicable anymore. There will be a need for specialized software to optimally map longer reads, perhaps containing indels or other structural variations, to the reference genome.

An alternative option to the alignment process is *de novo* assembly of sequence reads. Such approach allows for the identification of highly diverged DNA regions in the sequenced sample compared with the reference genome. The techniques used in assembling longer reads produced by the early sequencing technologies, such as Sanger, generally involved finding areas of overlap between reads and extending those into longer contigs. Shorter reads and higher coverage produced by the NGS technologies however make such algorithms computationally inefficient, if not unfeasible. Currently, the more widely used assemblers make use of the de Bruijn graph data structure where all possible substrings of size *k* are stored in the nodes of the graph and each edge indicates an overlap of size *k*-1 between the two connecting nodes [59-62]. Traversing such a graph built from raw sequence reads will yield a collection of contigs representing the sample's sequence. *De novo* assembly techniques are not yet as computationally efficient as alignment of reads directly to the reference genome and hence not yet as widely used. Currently, the genomic analysis of a cancer and its matched normal tissue

involves separate alignment of each sample; this is followed by variant calling and the identification of somatic mutations in the tumor tissue. With advances in assembly algorithms as well as increase in read length and insert sizes of paired-end libraries, it is conceivable that *de novo* assembly of tumor and normal genomes will eliminate the need for the alignment process. As a result, this approach can provide more comprehensive insights into each individual's unique genomic landscape and pave the way for more personalized diagnosis and treatment options.

1.4.2 Discovery of Point Mutations

The alignment and/or assembly results are subsequently explored for the presence of various types of somatic mutation including single nucleotide variants. The majority of the early SNV detection tools [63-65] rely on setting arbitrary thresholds for variables such as sequence coverage, read mapping quality, base quality and distance between mismatched bases in order to filter out technical noise and identify the positions that show true variability from the reference. These tools however are best suited for the analysis of normal samples and detection of germline variations where, for example, a heterozygote SNV would be expected to have variant allele frequency of 50% while in a homozygote position the variant base would be observed at 100% frequency. When analyzing tumor samples, contamination with adjacent normal tissue, the presence of multiple clonal populations within the tumor, as well as tumor aneuploidy can result in single nucleotide variants that are observed at any allele frequency.

Probability-based models designed specifically for the detection of variants in cancer samples have been developed; these identify the most likely genotype at each position based on a probabilistic model for allelic distribution [66,67]. Dependence of all these tools on separate analysis of cancer and normal samples followed by their pair-wise subtraction has however deemed them as suboptimal in detecting somatic mutations. Recent developments in simultaneous analysis of matched sample pairs have resulted in more confident somatic mutation calls by calculating the likelihood of genotype differences between the two genomes, at all locations [63,68-71]. These algorithms allow for the detection of true somatic mutations which lack strong support in the tumor sequence data and distinguish them from false positive calls with weak support in the normal sequence data. Current state of cancer genomics requires the verification of computationally detected variant calls in their corresponding specimens using orthogonal methods. In the near future, such verification may no longer be needed should advances in sequencing technologies and analysis tools lead to near optimal quality of reads and genotype calls.

1.4.3 Identification of Indels

Detecting small insertions and deletions from NGS short read products has proved more challenging than detecting single nucleotide variants. This is mainly attributed to the limitations of current aligners, which by default allow a set number of small mismatches between a read and the reference, typically with no gaps, leading to misalignment or no alignment of reads

spanning indels. Parameters such as the number of reads supporting an indel, mapping and base qualities as well as presence or absence of homopolymer regions should be taken into account when estimating the true positive probabilities [63,64]. Dindel [72], the 1000 Genomes project indel-caller [73], uses local realignment of reads to increase the accuracy of indel detection rate. Dindel accepts a list of potential indels and SNP calls as input, identifies all candidate haplotypes surrounding these sites and realigns reads to all the candidates in order to identify true events [72]. One limitation of Dindel, however, is its dependence on the sensitivity of the aligner, which provides the initial list of potential insertion and deletions. Indels, having the potential to alter or completely eliminate a protein's function, are the second most abundant type of variation in the human genome after SNVs [74]. The majority of specialized software for indel detection, including the above-mentioned tools, rely on separate analysis of cancer and matched normal tissues and hence have less than optimal sensitivity and specificity. More recent efforts have resulted in the development of robust probabilistic algorithms for the detection of somatic indels from paired specimens [75] and as a result a more accurate analysis of malignant genomes.

1.4.4 Structural Variation Detection

Structural alterations including large insertions and deletions, duplications, inversions, translocations and gene fusions have been associated with various cancer types [76]. Before the advent of NGS technologies, cytogenetics, karyotyping and fluorescent *in situ* hybridization, as

well as array-bases techniques such as SNP arrays and array comparative genomic hybridization, were used in detecting large SVs. However, the emergence of next generation sequencing technologies and the corresponding analysis tools has enabled the detection of various SVs including copy-neutral events and the corresponding break points at a much higher resolution and with greater accuracy.

Paired-end sequencing protocols, where the two ends of a single DNA molecule are read, allow the detection of SVs in the genomic data; since the order and orientation of read pairs and the insert size distribution are known, any deviation from these expectations in the alignment might suggest a variation in the sample. Several tools have been developed which detect read pair anomalies and infer specific SVs in genomic [77-81] and transcriptomic [82-84] datasets. However, we now know that the majority of structural variations are found in duplicated regions of the genome [85,86], regions that pose the most difficulty for the alignment process. As a result, alignment-based SV detection may result in many false positives while missing true events. An alternative to examining the alignment data for anomalies is to assemble the sequence reads *de novo* and compare the resultant contigs with the reference genome [87] or more accurately to the *de novo* assembled matched normal genome; such *de novo* assembly techniques can also detect fusion transcripts [87,88]. As the reads get longer, the assembly of individual genomes becomes more feasible and detection of SVs will have higher sensitivity and specificity.

Large deletions and amplifications, at times encompassing chromosome arms or whole chromosomes, lead to changes in number of gene copies and in some cases their expression levels. These structural alterations are often collectively referred to as copy number variations. Given the assumptions that the whole genome is sampled uniformly and that the reads are generated with equal probability, depth of coverage can serve as a quantitative measure of copy number [89,90]. These assumptions are not strictly correct however. GC content, for instance, introduces bias during the sequencing experiment [91] while challenges such as alignment of short reads to repetitive regions of the reference genome leads to computational biases. Various techniques have been employed in identifying somatic CNVs by correcting for these deviations from the expected distribution and by directly comparing tumor and matched normal datasets [32,92-95]; additionally, tools capable of distinguishing the somatic events that are unique to different subclones in the tumor [96-98] can be very valuable in guiding therapeutic decision-making given the propensity of tumor subclones to become resistance disease in the course of treatment.

1.4.5 Expression Analysis

High-throughput sequencing of the complete transcriptome offers a few advantages over the more traditional means of expression analysis such as oligo-nucleotide microarray technologies. All expressed entities including novel transcripts, novel isoforms and non-human transcripts are sampled in these surveys of the whole transcriptome as opposed to microarray experiments,

which are restricted to known genes and annotations. Digital analysis of the transcriptome also increases both specificity and sensitivity; the high coverage that can be achieved through these experiments enables the identification of genes with even the lowest expression levels. Identifying differentially expressed genes or specific isoforms between malignant and normal states can reveal pathways which when altered might lead to tumorigenesis. Differential expression analysis can also identify subtypes of a disease and subsequently aid in finding diagnostic and prognostic markers [99]. A slew of software including several R packages (R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/) is available for differential expression analysis of RNA-seq data [100].

Expression analysis is not restricted to the cell's messenger RNA. Small non-coding transcripts such as miRNAs can also be subjected to high-throughput sequencing and analysis. Integration of protein-coding gene expression profiles with miRNA expression, promoter methylation and copy number variation data can provide indications as to which genes are silenced and thus may function in tumor suppression, and which are overexpressed and might be acting as oncogenes.

1.5 Data Interpretation and Integration

When interpreting genomic data, it is imperative to be aware of potential confounding factors. Presence of circulating tumor cells in a normal blood sample, normal cell contamination in a tumor sample or a heterogeneous tumor specimen with several sub-clonal populations can lead to false positive and false negative mutation calls. Bioinformatic algorithms have been developed to estimate and correct for the amount of normal contamination and to more sensitively determine the copy number variant regions (CNANorm) [101], SNVs (MutationSeq) [70], or regions with loss of heterozygosity (APOLLOH) [95]. The next step following the computational discovery of candidate mutations typically entails verification of those events in their corresponding sample(s). This step identifies the variant calls that were falsely identified as somatic due to sequencing or computational errors. Mutation verification usually involves the amplification of the potential variant site in both cancer and matched normal tissues using PCR techniques followed by Sanger or next-generation sequencing.

Over the span of just a few years, the cancer genomics community has made great progress in developing algorithms and software for detecting various types of mutations from short read datasets. As improvements are made to sequencing technologies and detection tools, the most challenging task becomes mining the large and diverse mutational profiles that are generated in even a single patient experiment for mutations that contribute to disease initiation and progression. The list of putative cancer-related somatic mutations can be refined using publicly

available reference databases. These include repositories where variations in the healthy population are curated such as dbSNP [102], Database of Genomic Variants [103] and 1000 Genomes Project [73], as well as databases where known cancer genes and their mutations are stored. Examples of such include: the COSMIC (Catalogue of Somatic Mutations in Cancer) [104], an open source database containing somatic mutations and copy number alterations associated with cancers; OMIM (Online Mendelian Inheritance in Man), which collates information on familial cancer genes and susceptibility loci; Cancer Gene Census [76], which catalogues all genes shown to be causally implicated in cancers; and the Mitelman database of chromosome aberrations and gene fusions in cancer [105]. The assumption when using these databases is that variations that are commonly found in the general population are less likely to contribute to diseases such as cancer while genes recurrently mutated in various cancer types are potential tumor suppressors and oncogenes (Figure 1.2).

The number and profile of somatic mutations demonstrate great variability in different cancers [106]. High number of mutations is seen in malignancies such as melanomas [107] whereas some pediatric cancers exhibit a very low number of alterations [108]. Regardless, all somatic mutations can be categorized into 'drivers', changes that are responsible for disease pathogenesis and tumor evolution, or 'passengers' which are simply the by-product of the unstable cancer genome and provide no growth advantage to tumor cells [106]. Distinguishing these two types of mutational entities is critical given that passenger mutations play no functional role in disease initiation, progression or maintenance and thus treatment(s) targeting

them may prove ineffective. Computational techniques are developed that aim at distinguishing drivers and passengers *in silico* prior to the more labor-intensive and time-consuming procedures of functional validation in the wet lab. Since it is believed that mutations that result in changes in protein structure and function are more likely to act as cancer drivers, the majority of the community's focus to date has been on mutations that affect protein-coding regions of the genome. A common computational strategy in examining the functional role of a somatic mutation is to determine its location with respect to functional domains and key amino acid residues in the protein product using resources such as UniProt. Software tools such as PolyPhen [109], MutationAssessor [110] and SIFT [111] use evolutionary conservation of gene sequences as well as homology to provide a likelihood score for the deleterious effect of a point mutation on protein structure and function. Genes with higher number of somatic mutations than would be expected by chance alone are likely contributors to disease phenotype given the general assumption that driver mutations provide growth advantage for the tumor and hence must be under evolutionary positive selection while the passengers are less likely to be selected for [112,113]. Several factors such as gene length, background mutational rate of a particular tumor and a particular region of the genome affected by gene replication timing, for instance, influence the number of somatic mutations in a gene [114]; several algorithms have been developed to control for these variables and to distinguish driver and passenger mutations [112,113].

Integration of all somatic mutation calls and expression data from a cancer sample plays an important role in creating a molecular pathway hypothesis of aberrations driving the tumor. Biology of a cell is complex and involves processes and controls of those processes on the genomic, epigenomic, transcriptomic and proteomic levels (Figure 1.3). Molecules in the cell are not isolated, but are part of a collective system of interacting parts [115], and aberrations in one molecule can perturb the whole system. When examining a cohort of samples, data integration and identifying commonly altered pathway(s) often becomes more significant than pinpointing recurrently mutated gene(s). Different samples could have mutations in various genes, all contributing to one biological process which when disturbed leads to tumorigenesis. The Cancer Genome Atlas pilot project, for instance, uncovered core mutated pathways in glioblastomas by integrating sequence data, gene expression, copy number variation as well as epigenetic assessments [116]. Data integration in these cohorts can also identify genes, which may be frequently altered through multiple mechanisms (point mutations, structural disruption, loss of copy, or hypermethylation of the promoter) but would not otherwise be identified through separate analysis of each data type.

1.6 Thesis Chapter Summaries

In this thesis, I have coupled massively parallel sequencing approaches, bioinformatic analyses and subsequent verification and validation steps to study papillary, oncocytic and anaplastic thyroid carcinomas, benign thyroid nodules and parathyroid carcinoma. These studies have

provided detailed accounts of the altered genomes of these endocrine tumors and have shed light on potentially perturbed pathways and tumorigenesis mechanisms in these diseases. In each of the following research chapters, the analysis methods and results, limitations and strengths of such approaches and future directions are discussed for each disease type. Chapter 2 describes an in-depth analysis of multiple recurrences of a parathyroid tumor, providing the first such analysis in the literature and a first look at the genome of this tumor. The research aim in Chapter 3 was to comprehensively characterize Hürthle cell or oncocytic thyroid carcinomas on the whole genome scale. The analysis led to the identification of recurrent mutations in the tumor suppressor gene *MEN1* in a subset of these cancers. Chapter 4 provides the first whole genome and transcriptome analysis of the rare but extremely aggressive anaplastic thyroid carcinoma. The genomic analysis revealed extensive amount of changes in the number of chromosomal and gene copies. Although no recurrent mutations, besides TP53 loss of function, were identified in the four studied genomes, recurrent alterations of the epigenetic machinery and gene fusions involving known cancer genes were found in these tumors. Finally in Chapter 5, transcriptomic profiles of papillary thyroid carcinoma, the most abundant subtype of thyroid cancer, and benign thyroid nodules are described. These tumors are found to have a very low mutation rate and very few copy number changes; however, one benign tumor demonstrated the loss of TP53 and vast changes in copy number likely due to the loss of this tumor suppressor. Such molecular signatures have only been associated with anaplastic cancers and thus the genomic analysis of this benign tumor might have unveiled a precursor tumor for an aggressive disease. The transcriptomic landscapes of benign and

malignant tumors are also compared in an attempt to identify molecular signatures that are able to discriminate between malignant and benign tumors.

Table 1.1 Advantages and disadvantages of different data types in cancer genomics

Data	Analysis Type	Advantages	Disadvantages
Whole Genome Shotgun Sequencing	- CNVs - Indels - SNPs - SVs	- Comprehensive interrogation of mutations	- Most expensive - No information on expression status
Whole Exome Capture Sequencing	- Indels - SNPs	- Cost efficient	- Restricted to known annotations - Detects only small coding mutations
Whole Transcriptome Shotgun Sequencing	- Expression - Indels - SNPs - SVs	- Cost efficient - Digital gene expression - Detects novel events - Gene fusion discovery	- Detects only expressed alterations

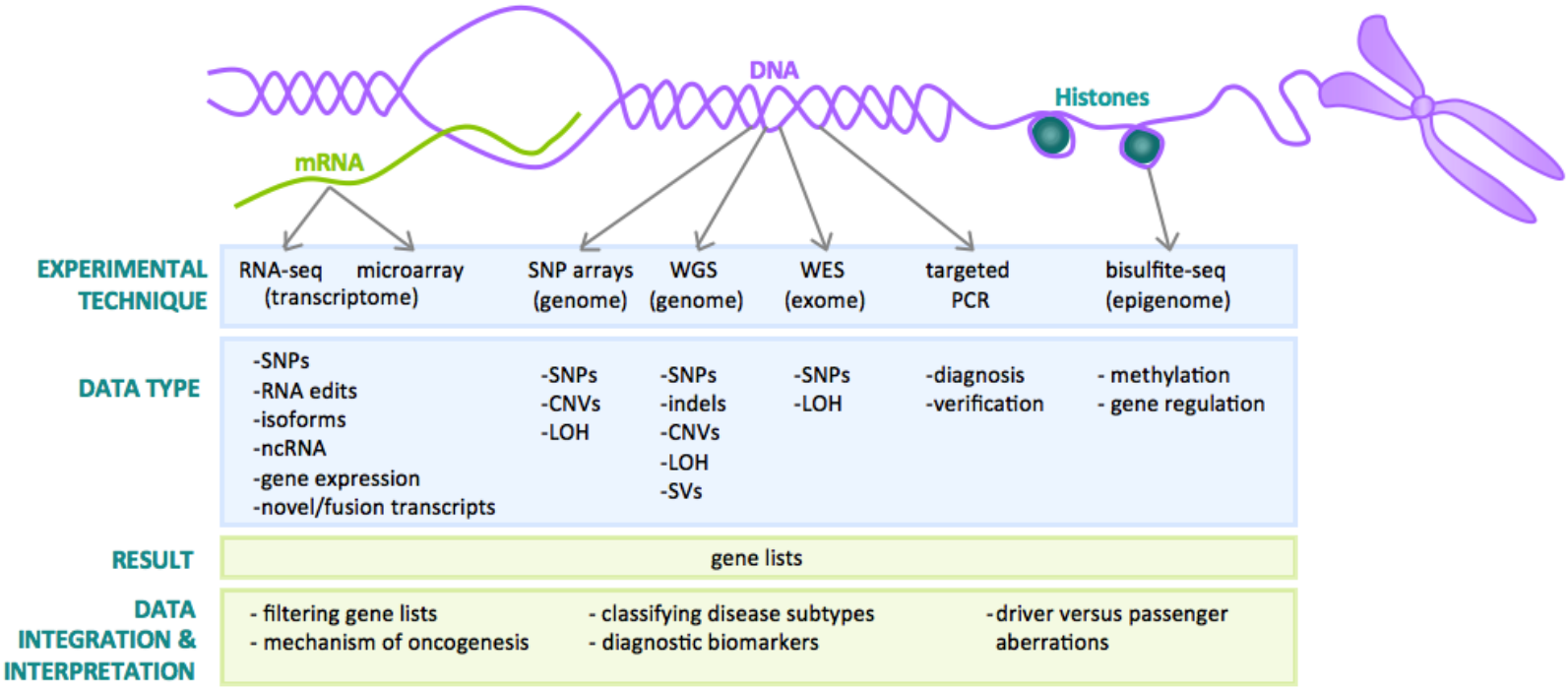

Figure 1.1 Applications of high-throughput sequencing technologies
Through such applications, the genome, the epigenome and the transcriptome can be examined in great detail, providing a comprehensive picture of the state of health or any alterations leading to disease. Such experiments allow for the identification of both small and large variations in individual samples

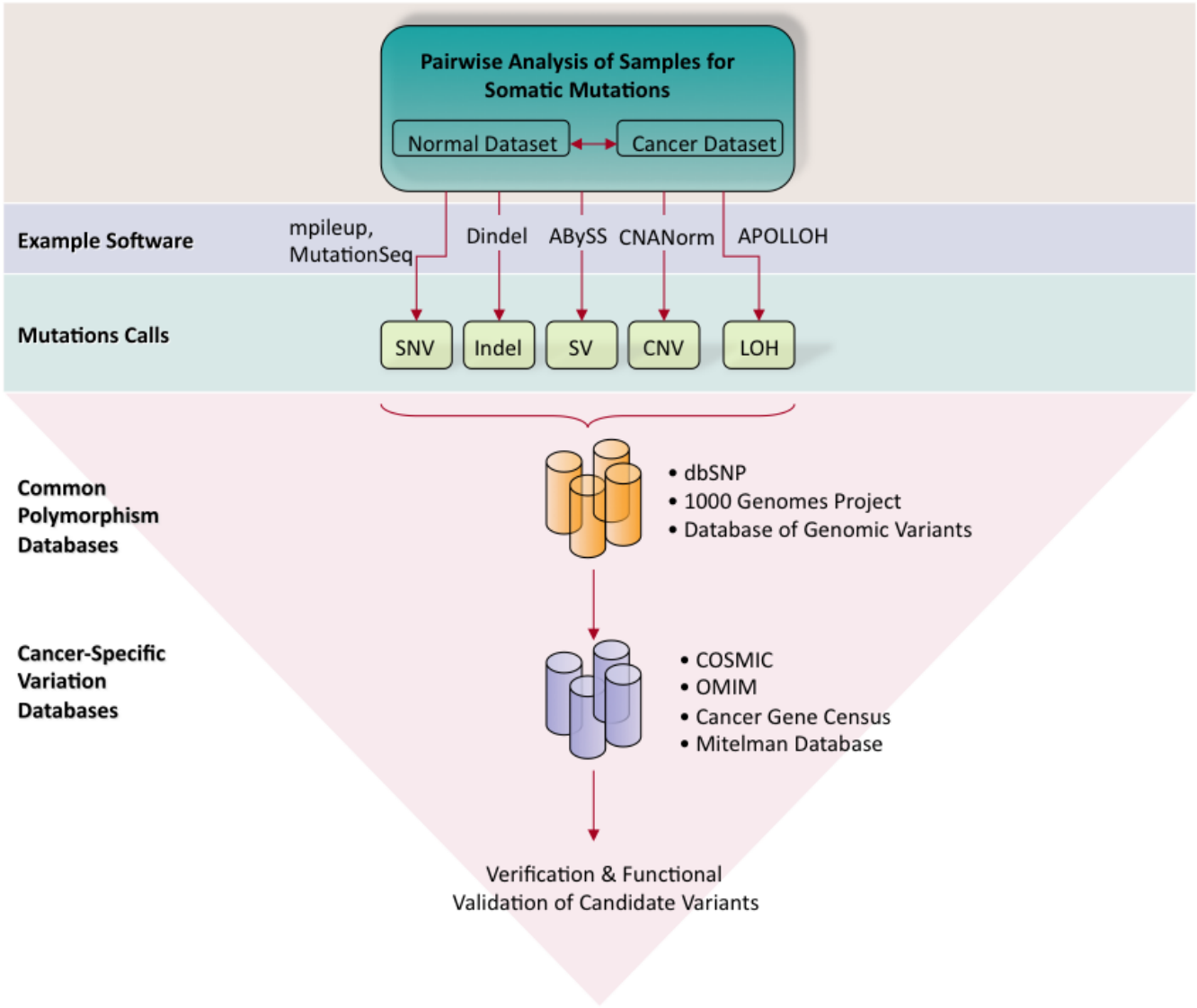

Figure 1.2 Identifying cancer-specific somatic alterations
Filtering the identified somatic mutations in a cancer sample using publicly available databases of common genetic polymorphisms such as dbSNP, 1000 Genomes project and the Database of Genomic Variants as well as those of known cancer-specific variants including COSMIC, OMIM, Cancer Gene Census and Mitelman databases can narrow down the potentially long list of candidates to the most likely drivers

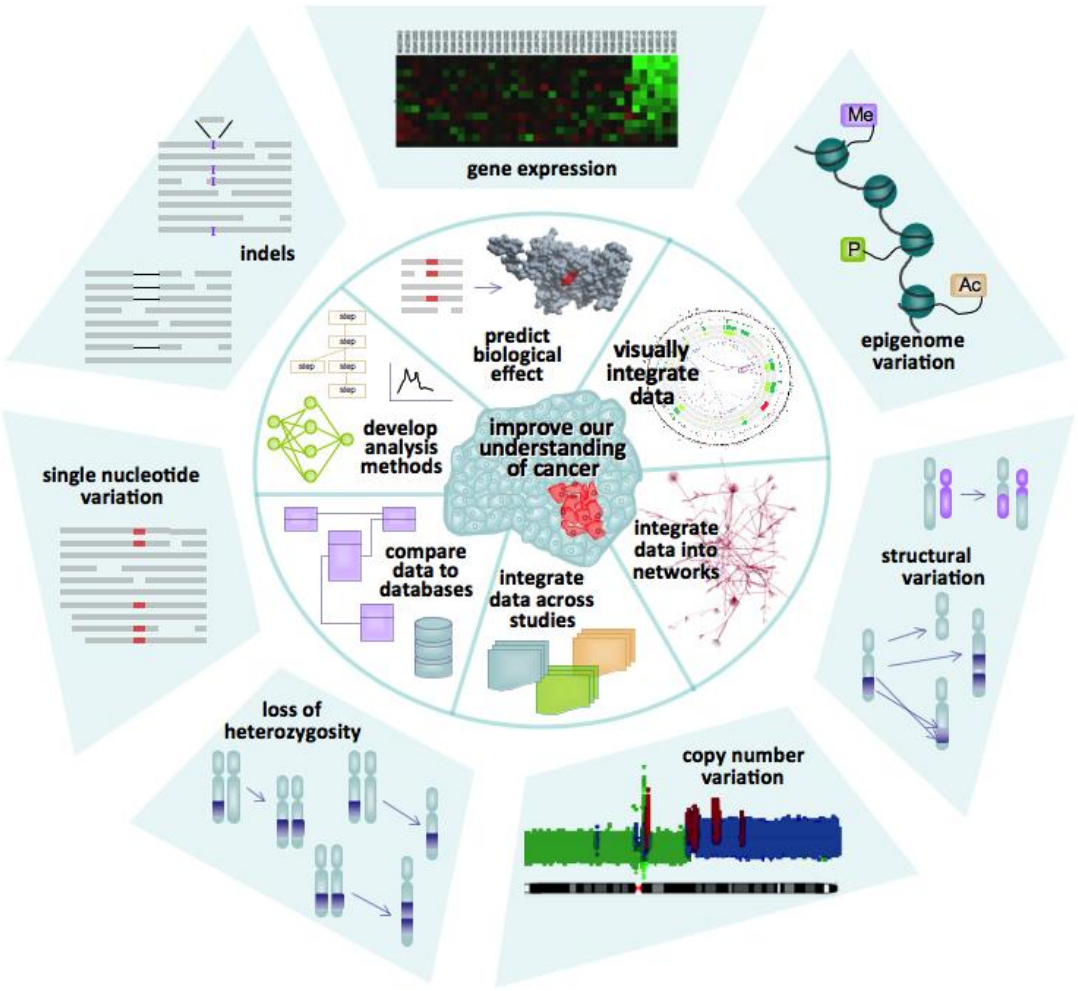

Figure 1.3 Data integration
Identifying the perturbed pathways and networks that are driving the disease is integral to better understanding of the causes of cancer, exploring potential therapeutic targets and predicting drug response. Integration of data generated from both normal and cancer samples promises the unbiased examination of a cell's interconnected network of molecules

Chapter 2: Complete Genomic Landscape of a Recurring Sporadic Parathyroid Carcinoma[2]

2.1 Introduction

The parathyroid glands are important endocrine glands that regulate the serum calcium level through secretion of parathyroid hormone (PTH). PTH binds to type 1 parathyroid hormone receptors on target organs such as bones, kidneys and intestine, and results in an influx of calcium into the blood stream. PTH production and secretion is negatively regulated through binding of calcium to calcium-sensing receptors (CaSRs) that are located on the surface of parathyroid cells [117,118]. Thus, any deviation in the level of secreted PTH, and as a direct result the blood calcium level, may negatively impact multiple body systems. Primary hyperparathyroidism (PHPT), that has an incidence of 1-3 per 1,000 population, causes fatigue, weakness, depression, bone disease, nephrolithisis, pancreatitis, and peptic ulcer disease [27,119]. The majority of cases of PHPT are sporadic and only about 5% are associated with hereditary syndromes such as multiple endocrine neoplasia type 1 and 2A (MEN1, MEN2A), familial isolated hyperparathyroidism (FIHP) and hyperparathyroidism-jaw tumor syndrome

[2] A version of this chapter has been published, and the author contributions are provided in the Preface as per the University of British Columbia PhD thesis guidelines: Katayoon Kasaian, Sam M Wiseman, Nina Thiessen, Karen L Mungall, Richard D Corbett, Jenny Q Qian, Ka Ming Nip, Ann He, Kane Tse, Eric Chuah, Richard J Varhol, Pawan Pandoh, Helen McDonald, Thomas Zeng, Angela Tam, Jacquie Schein, Inanc Birol, Andrew J Mungall, Richard A Moore, Yongjun Zhao, Martin Hirst, Marco A Marra, Blair A Walker, and Steven JM Jones. (2013). Complete genomic landscape of a recurring sporadic parathyroid carcinoma. Journal of Pathology, 230: 249–260. doi: 10.1002/path.4203. Copyright by Wiley.

(HPT-JT) [27]. While virtually all cases of PHPT are the result of parathyroid adenomas or hyperplasia, a small proportion (<1%) are due to parathyroid carcinoma (PC) [117].

PC is an extremely uncommon endocrine malignancy. Unlike PHPT, cases of PC occur with equal frequency in men and women, and at the time of diagnosis patient age ranges from 12 to 90 years (mean age at 44-48 years) [120]. The reported 10-year overall survival rates for PC vary from 49-77% [121-124]. However, 40-60% of patients develop disease recurrence that is difficult to manage [122,124]. The majority of studies reporting on PC are generally limited to small retrospective case studies. Young age at diagnosis, female gender and absence of metastases have all been identified as favorable prognostic factors [123]. The majority of PCs are functional, producing high levels of PTH, and therefore patients tend to present with high blood calcium levels [125]. Mortality and morbidity from PC are usually attributable to the high level of PTH and calcium, rather than the tumor burden itself [124]. Currently the ideal treatment of PC is *en bloc* surgical resection of the tumor and adjacent grossly involved neck structures that may include the ipsilateral thyroid lobe, with grossly clear margins, and taking great care to avoid tumor spillage. Chemotherapy has proven largely ineffective for PC [120,125] and radiation therapy is of limited benefit [123]. PC patients usually suffer from complications of disease that may be neurological, cardiac, renal and skeletal. Therefore, the goals of treating PC patients are elimination of all detectable disease and control of the metabolic complications of the cancer. Prescribing bisphosphonates and calcimimetic agents help in controlling the calcium level [120].

The etiology of sporadic PC is largely unknown. Molecular profiling of familial cases has identified a few candidate genes. Bi-allelic inactivating mutations of the tumor suppressor gene *HRPT2/CDC73* are observed in some parathyroid tumors, mainly those associated with HPT-JT and FIHP [27,120,126-130]. Loss-of-function mutations of the tumor suppressor gene *MEN1* [131-133] and activating mutations of the *RET* proto-oncogene [134] are also associated with benign tumors of the parathyroid in MEN1 and MEN2A patients, respectively. The *CCND1 (cyclin D1)* proto-oncogene, first characterized in a parathyroid adenoma [135], is overexpressed in the majority of PCs [136]. An inversion of chromosome 11 leading to the fusion of *PTH* and *cyclin D1* was found to put this oncogene under the control of the tissue-specific and highly active *PTH* regulatory elements [28,29] causing its overexpression. The detected mutations and alterations in the oncogenes *CCND1* and *RET,* and tumor suppressors *MEN1* and *HRPT2* do not represent cancer-specific states and do not account for all cases of PC. The objective of the current study is to identify novel mutations and altered pathways in a single case of recurring sporadic PC.

The patient is a fit and active 76-year old male who initially presented in March 2005 with a greater than 30 year history of nephrolithiasis that required multiple urological procedures. He also complained of significant musculoskeletal and bony pains. He had no personal or family history of hyperparathyroidism, MEN, or any other endocrine disorder. At presentation his laboratory results were: serum calcium level 3.72 (normal 2.10-2.60 mmol/L), albumin level 40 g/L (normal 34-48 g/L), creatinine 151 umol/L (normal 30-130 umol/L), ionized calcium level

1.97 mmol/L (normal 1.17-1.29 mmol/L), and PTH 72.2 (normal 1.3-6.8 pmol/L). A sestamibi scan suggested a parathyroid carcinoma was located inferior to the right lobe of the thyroid gland. He subsequently underwent a focused parathyroidectomy, utilizing adjunctive intraoperative radioguidance and PTH measurement, and what grossly appeared to be a well-circumscribed parathyroid tumor was removed intact from just inferior to the right lobe of the thyroid gland. Pathological evaluation of the parathyroid tumor described thickened capsule and thick broad fibrous bands. There also was evidence of capsular and vascular invasion, and the tumor was diagnosed as a PC. Postoperatively his PTH and calcium levels normalized. In 2009, however, they began to rise suggesting that the PC had recurred. Evaluation by CT scan of the neck, chest, abdomen, and pelvis, sestamibi scan, MRI and ultrasound of the neck, and selective venous sampling for PTH, all suggested a local recurrence in the right central neck. On November 25, 2009 he underwent a re-exploration of the right central neck with removal of the right thyroid lobe, and also a right central neck dissection with skeletonization of recurrent laryngeal nerve and removal of neck lymph node levels V and VI, including all grossly recurrent PC. The pathology confirmed that PC recurrence was resected. Postoperatively his calcium and PTH levels normalized, but in 2010 they once again began to climb, suggesting another PC recurrence. Repeat imaging suggested this recurrence was also local. On October 27, 2010 he underwent re-exploration of the right central neck with removal of recurrent PC and the right recurrent laryngeal nerve that was grossly invaded by cancer. Postoperatively his calcium and PTH levels normalized and he was also treated with external beam radiation therapy. In early 2012 his calcium and PTH levels yet again began to climb, and repeat imaging suggested

another local recurrence. On March 15, 2012 he underwent re-exploration of his right central neck. Recurrent PC and scar tissue was removed, and postoperatively his calcium and PTH levels again normalized (Figure 2.1). Recently these have again begun to rise, he has refused further surgical intervention, and he is being managed medically with both cinacalcet and regular pamidronate infusions. The formalin fixed paraffin embedded (FFPE) primary cancer specimen, the flash frozen first and second PC recurrences, and patient's blood specimens were analyzed using high-throughput sequencing. The parathyroid specimens were classified according to the World Health Organization criteria. The tumor specimens were collected as part of a research project approved by the University of British Columbia and the British Columbia Cancer Agency Research Ethics Board and are in accordance with the Declaration of Helsinki. Informed consent was obtained from the patient for profiling the tumor using RNA-seq as well as whole genome sequencing. Our protocol stipulates that these data will not be released into the public domain but can be made available via a tiered-access mechanism to named investigators of institutions agreeing by a materials transfer agreement that they will honor the same ethical and privacy principles required by our center.

2.2 Methods

2.2.1 DNA Sequencing

Whole genome sequencing of the first PC recurrence and matched blood specimens was performed by Illumina (Inc.); 100 bp paired-end reads were generated using the PCR-free protocol on HiSeq machines. The subsequent sequencing was performed at the British Columbia Cancer Agency Genome Sciences Centre using Illumina HiSeq2000 technologies following our established protocols (Figure 2.1). Briefly, for RNA-seq analysis, RNA was extracted from 15 x 20 µm sections cut from both recurrent samples using MACS mRNA isolation kit (Miltenyi Biotec), resulting in 5-10 µg of DNase I-treated total RNA as per the manufacturer's instructions. Double-stranded cDNA was synthesized from the purified poly(A)$^+$ RNA using the Superscript Double-Stranded cDNA Synthesis kit (Invitrogen) and random hexamer primers (Invitrogen) at a concentration of 5 µM. The cDNA was fragmented by sonication and a paired-end sequencing library prepared following the Illumina paired-end library preparation protocol. Cluster generation and sequencing were performed on the Illumina HiSeq2000 following the manufacturer's recommended protocol. 75bp paired-end reads were generated for these two libraries (Table 2.1).

A Whole genome shotgun library was constructed from the 7-year old primary FFPE sample, using a modified version of our standard protocol as follows. Tumor DNA was extracted from

formalin-fixed, paraffin-embedded thyroid sections according to Qiagen's Allprep DNA/RNA FFPE Kit protocol (Qiagen Inc, Toronto, Ont.). Two micrograms of extracted DNA were sheared for 55 seconds using a Covaris E210 focused ultra-sonicator (Covaris Inc., Woburn, Mass.) at 20% Duty cycle, 5% Intensity, and 200 Cycles per burst. The sheared products were separated on an 8% Novex TBE gel (Invitrogen Canada, Inc., Burlington, Ont.) and the 200 to 300 bp size fraction was excised and eluted into 300 µl of elution buffer containing 5:1 (vol/vol) LoTe (3mM Tris-HCl, pH7.5, 0.2mN EDTA)/7.5 M ammonium acetate. The elute was purified from the gel slurry by centrifugation through a Spin-X centrifuge tube filter (Fisher Scientific Ltd., Nepean, Ont.), and EtOH precipitated. A small gap paired-end library was constructed from the purified DNA following Illumina's protocol (Illumina Inc., USA). Cluster generation and sequencing were performed on the Illumina HiSeq2000 following the manufacturer's recommended protocol. 100bp paired-end reads were generated (Table 2.1). All genotype data have been deposited at the European Genome-phenome Archive (EGA, http://www.ebi.ac.uk/ega/) under accession number EGAS00001000484.

2.2.2 Sequence Data Alignment and Analysis

Using the Burrows-Wheeler Alignment (BWA, version 0.5.7), sequence reads were aligned to the human reference genome (hg19/GRCh37), or in the case of RNA-seq, to a genome file that was augmented with a set of all exon-exon junction sequences [56]. The exon-exon junction sequences and their corresponding coordinates were defined based on annotations of any transcripts in UCSC known genes, Ensembl (version 54) or the Refseq database (as downloaded

from the UCSC genome browser on March 2009). After the alignment, the junction-aligned reads that mapped to exon-exon junctions were repositioned as large-gapped alignments in the genome based on the coordinates of the exons that were used to construct the junction sequences. Reads that aligned to junctions with insufficient overhang past the splice junction sites were changed to soft-clipped un-gapped genomic alignments. Candidate single nucleotide variations (SNVs) in the primary tumor, first recurrence and the blood genomes as well as variations in both transcriptomes were identified using SAMtools; for matched genomic datasets (primary vs. blood and recurrence vs. blood), the mpileup paired option was used [64]. Variants with CLR (phred log ratio of genotype likelihood) >= 20 were used as input into MutationSeq [70]. MutationSeq simultaneously examines features from both tumor and normal genomes and assigns each variant a probability score indicating the degree of confidence that the mutation is indeed somatic. The variant calls with probability >= 0.5 were manually inspected in the integrative genomics viewer (IGV) [137]. Any SNV at sites assessed as being polymorphisms (SNPs) were disregarded, including variants matching a position in dbSNP [102] or 1000 Genomes project [138]. For paired samples with matched normal DNA sequence, all variants with evidence in the constitutional DNA were considered germline variants and were no longer considered. For the purposes of identifying structural variations such as translocations, inversions and duplications, we analyzed the sequence data using a *de novo* assembly approach. Genome and transcriptome sequence reads from the first recurrence as well as the RNA-seq data from the second recurrence were assembled and analyzed using ABySS [62] and trans-ABySS [87,139]. All variants detected as somatic and not common

polymorphism sites were verified in the original tumor sample using Sanger sequencing and verified as being somatic using DNA from the patient's peripheral blood. Copy number variation (CNV) and loss of heterozygosity (LOH) analyses were performed using HMMcopy and APOLLOH software, respectively [95]. HMMcopy corrects for GC content bias as well as high mappability regions, while APOLLOH segments the genome to regions of LOH accounting for normal tissue contamination. These results were graphed using the Circos tool [140]. An expression profile was derived based on the RNA-seq datasets. In the absence of RNA from matched normal tissue, we took a similar approach to Jones *et al.* [141] in conducting the differential expression analysis. We compared the expression of genes in the two parathyroid transcriptome libraries against a compendium of 19 normal transcriptome libraries from the Illumina Body Map 2.0 project (available from ArrayExpress, query ID: E-MTAB-513) [142]. Sixteen different tissue types are included in the compendium (no parathyroid tissue-derived libraries was available) (Table 2.2). This approach allows for discovering tumor specific changes in expression and thus provides a better understanding of the mechanism of the disease as well as opportunities in identifying relevant therapeutic interventions. Number of reads per kilobase of exon model per million mapped reads (RPKM value) [143] calculated for each protein coding gene as annotated in Ensembl (v54) [144] were used as a measure of expression. Differential expression analysis was done using outlier statistics and fold change comparison between the parathyroid sample RPKM and the compendium's mean RPKM for each gene. Overexpressed genes were defined as having a Benjamini and Hochberg [145] corrected outlier *P*-value < 0.05 and fold change > 2. Genes with an uncorrected outlier *P*-value < 0.1 and fold change < -2 were considered

underexpressed. Pathways enriched for the differentially expressed genes were identified using IPA (Ingenuity Systems, www.ingenuity.com).

2.2.3 Validation of Putative Somatic Variants Using Sanger Sequencing

Primers were designed for all 23 potential somatic SNVs (Table 2.3); due to the degraded nature of FFPE DNA, shorter amplicons were designed for the primary sample. Forward and reverse primers were tailed with T7 and M13Reverse 5' priming sites, respectively. PCR conditions were an initial denaturation of 98°C for 30 seconds, followed by 35 cycles of 98°C for 10 seconds, 69°C for 15 seconds and 72°C for 11 seconds, and a final extension at 72°C for 10 minutes. PCR was set up using Phusion polymerase (Fisher Scientific, catalogue # F-540L) according to manufacturer's specifications. Amplified regions of interest were sequenced using the Sanger technology. The sequencing reactions consisted of 35 cycles of 96°C for 10 seconds, 43°C (for M13Reverse) or 48°C (T7) for 5 seconds and 60°C for 3 minutes and were analyzed using an AB 3730XL DNA sequencer.

All detected structural variations were also verified in the original tumor sample and validated as being somatic by comparing to DNA extracted from patient's blood. PCR primers were designed to amplify 450-1000bp regions around each breakpoint; pairwise mixture of primers from corresponding genes was used to examine the presence of novel fusion events (Table 2.4). Forward and reverse primers were tailed with T7 and M13Reverse 5' priming sites, respectively.

PCR conditions were similar to above. All capillary traces were visually inspected to confirm the presence of novel events in the tumor and their absence from the germline.

2.3 Results

2.3.1 Single Nucleotide Variations

Twenty-three SNVs were confirmed as somatic events (Table 2.5). 15 were detected in both the primary and the recurrent genomes, 7 were only found in the relapse and 1 mutation in the *PIK3CA* gene was found only in the primary tumor. The degraded nature of the DNA from the FFPE sample may explain the lower number of unique variations found in the primary, but we note that the presence and absence of the *PIK3CA* mutation in the primary and recurrent samples, respectively, were verified using Sanger sequencing. This might be suggestive of a role for *PIK3CA* in tumor initiation but not maintenance or of tumor heterogeneity.

Three of the twenty-three SNVs have been observed and verified as somatic in cancers other than parathyroid carcinomas (http://www.sanger.ac.uk/genetics/CGP/cosmic/) [146]. *PIK3CA* E545K (COSM763), *mTOR* L2334V (COSM462591) and *THRAP3* R101* (COSM186652) mutations are seen in 816, 1 and 2 cancers, respectively. *PIK3CA* encodes the p110α catalytic subunit of PI3K, a lipid kinase with an important role in signaling pathways, and as a result in regulating cell growth and proliferation [147]. The observed *PIK3CA* p.Glu545Lys somatic mutation,

although a well-characterized activating mutation [148], has never been previously observed in PCs [146]. Loss of this mutation from the dominant clone in the recurrence underscores the necessity for temporal monitoring of tumors on the molecular level as changes in their mutational profile, in this case in the absence of any chemotherapy or radiation therapy, will likely affect the targeted treatment options.

Other well-characterized cancer genes with mutations are *mTOR*, *CDKN2C/p18* and *MLL2/KMT2D*. Deregulation of mTOR signaling pathway is seen in various cancers and currently mTOR inhibitors are utilized to treat solid tumors [149]. The L2334V *mTOR* mutation is situated in the kinase catalytic domain of the protein. Although the activating status of this specific change is not known, clusters of other activating mutations in this same domain have been observed in cancers such as large intestine adenocarcinoma and renal cell carcinoma [150,151]. *CDKN2C/p18-INK4C* plays a crucial role in regulating cell cycle progression by inhibiting activation of cyclin-dependent kinases 4 and 6 (CDK6/4) [152] and hence could lead to suppression of tumorigenesis. Although mutations of *p18* in cancer are rare [146], its loss of expression and function has been observed in various types of tumors [153-158]. The frequent loss of 1p in parathyroid tumors and the potential for a tumor suppressive activity in the region led Tahara *et al* to examine 25 parathyroid adenomas for mutations, specifically in *p18* [159]; no mutations were found. This may suggest a valuable utility for *p18* as a discriminatory marker between parathyroid carcinoma and adenomas. MLL2/KMT2D is a member of the SET family of proteins with histone 3 lysine 4 methyltransferase activity, playing a pivotal role in regulating

active chromatin states and epigenetic regulation of gene transcription [160,161]. Somatic mutations of this putative tumor suppressor are observed in various cancer types [162-167]. *MLL2/KMT2D* inactivating mutations are found to be the cause of Kabuki syndrome [160], one case of which has been observed in a patient with familial hypocalciuric hypercalcemia (FHH), a clinically benign and related phenotype to primary HPT resulting from heterozygous inactivating mutations of *CaSR* [118,168].

THRAP3/TRAP150, a member of the thyroid hormone receptor-associated protein (TRAP) complex [169], has lost a copy in the recurrent specimen while acquiring a truncating mutation in the remaining copy (Table 2.5 and Figure 2.2). THRAP3 is also a member of the spliceosome [170] and has been shown to act as an activator of pre-mRNA splicing and to participate in post-transcriptional mRNA degradation through its C-terminus [171]. In addition, THRAP3 is found in the SNARP complex, thought to specifically regulate *cyclin D1* RNA stability and expression [172-174]. A recent study found strong phosphorylation of THRAP3 residues Ser-210, 211, 399, 406 and 408 (all deleted in this patient) in response to DNA damage, and hence proposed a potential role for this protein in the DNA damage response pathway [175]. This study also demonstrated higher cell sensitivity to DNA damaging agents when THRAP3 was depleted using siRNAs. This may imply a role for THRAP3 in driving parathyroid carcinomas perhaps through regulating *CCND1* expression and the level of mRNA degradation in the cell. It is noteworthy that deletion or disruption of THRAP3 is observed in other cancers such as oral squamous cell carcinomas [176] and liver cancers [177]. The identified truncating mutation is present in both

recurrent transcriptomes, indicating that non-sense-mediated decay does not affect this mutation.

2.3.2 Structural Variants

Two inter-chromosomal translocations and one inversion were detected in the genomic data; none of which produced an expressed chimeric transcript (Figures 2.2, 2.3 and Tables 2.6, 2.7). All events were validated as somatic using the patient's peripheral blood and Sanger sequencing. To date, no fusion events involving these genes have been reported in any cancer types, including PC [146].

The fusion of the 5'-UTR region of the *PLD1* gene to *AGBL1* might result in the deletion of key *PLD1* regulatory elements such as sequences recognized by DNA binding proteins while leaving the conserved domains of the protein intact, leading to alterations in expression level but not function [178]. Differential expression of *PLD1* could in turn play an important role in parathyroid tumorigenesis since this gene is an upstream regulator of mTOR signaling [179,180] and has been implicated in signal transduction, membrane trafficking, transformation, and cytoskeletal reorganization [181-183]. In addition, PLD1 is an important regulator of intracellular trafficking of the parathyroid hormone ligand-receptor complex after its internalization into target cells [184] and thus alterations in gene expression can affect the downstream PTH signaling pathways.

Fusion of *SKP2* and *BC033837* leads to loss of *SKP2* exon 10. S-phase kinase-associated protein 2 (*SKP2*) is a member of E3 ligase complex regulating the cell cycle through ubiquitin-mediated proteolysis of cyclins and CDKs, specifically p27, and thus acts as an oncogene [185,186]. The deleted exon does not encompass the F-box conserved domain but results in the deletion of leucine-rich repeats 9 and 10. Although *SKP2* and *p27* mRNA levels show similar levels to those of normal controls, loss of leucine-rich repeats might prevent protein-protein interactions and as a result disrupt vital signal transduction pathways in the cell [187,188].

The inversion in chromosome 15 leads to the fusion of first 8 exons of *AKAP13* to exon 18 of *DMXL2*. DMXL2, consisting of 12 WD repeats, plays an important role in Notch signaling [189] while AKAP13 regulates multiple signal transduction pathways including MAPK and estrogen receptor signaling [190-192]. AKAP13 also binds the regulatory subunit of protein kinase A through its RII-binding region and regulates the Rho/Rac GTPase cycle with its dbl oncogene homology (DH) and pleckstrin homology (PH) domains, thus coordinating these two signaling pathways [193,194]. The identified inversion in this case leads to the fusion of RII-binding domain of AKAP13 to the C-terminal WD repeats of DMXL2. Loss of DH and PH domains and the N-terminal WD repeats from AKAP13 and DMXL2, respectively, may disrupt signal transduction pathways important for cell cycle progression and cellular growth.

2.3.3 Copy Number Variants

Similar to previous reports [120], the PC specimen presented with a large number of both arm-level and focal changes in copy number. The arm-level changes included gain of 1q and loss of all or large stretches of sequence from 1p, 3q, 4q, 7p, 11q, 15q, 17p and 22q (Figure 2.2). Focal changes included two homozygously deleted areas on 1p, a homozygous loss on 3q, 22q and Xq, multiple heterozygously deleted regions on 19p and a single heterozygous loss on 5p, 11p, 12q and 21q. Small regions of gain and amplification are observed on 3q, 7p, 7q, 11q (encompassing *CCND1*), 14q, 15q, 16q, 19p, 22q and Xp. Coordinates for focal regions of loss and gain are listed in Table 2.8. Loss of 1p is common in PCs, previously observed in 40% of carcinomas and only 10% of adenomas [195]; we did not identify large areas of gain with the exception of 1q, a carcinoma specific event [118,195].

The degraded nature of DNA extracted from the 7-year old primary FFPE specimen made the CNV and LOH analyses challenging (Figure 2.4). All observed arm-level changes in the recurrence were present in the primary PC except for the loss of 4q. Although no somatic mutations were found in this region, inactivation of the remaining allele of a putative tumor suppressor via promoter hypermethylation or mutations of regulatory elements may be responsible for the progression of the disease. We also observed a large region of loss on 5p in the primary PC specimen that was not present in the relapse; we ruled out allele-specific

amplification of the region in the recurrent sample using APOLLOH (Figure 2.5). Distinguishing true focal changes from the background noise was not feasible in the primary PC specimen.

2.3.4 Analysis of Differential Transcript Abundance

In the first recurrence specimen, 2 and 1339 genes showed under- and overexpression, respectively, while 1 and 1581 genes were under- and overexpressed in the second recurrence. Overlap between the two samples included 1173 overexpressed genes (Kendall's tau =0.899, *p* =< 2.2e-16 for all protein-coding genes).

Among the top 25% differentially expressed genes are *PTH*, *CCND1* and interestingly *CDKN2A* and *CASR*. Perhaps not surprisingly, the *PTH* gene has the highest expression level in both recurrent specimens. Over 90% of PCs show overexpression of *cyclin D1* [136] which in association with cyclin-dependent kinases plays an integral role in regulating the cell-cycle machinery by driving the progression through the G1/S checkpoint [196]. *CCND1* amplification and overexpression are also hallmarks of other cancers [197-202]. Similar to *p18, CDKN2A/p16* is a member of the INK4 cyclin dependent kinase inhibitors (CDKIs) and it acts in parallel to *p18* to inhibit activation of cyclin-dependent kinases 4 and 6 (CDK4/6). Although we see loss of one copy of *p18* and a somatic missense mutation in the remaining copy, a remnant of a tumor suppressor activity, the parallel pathway through *p16* shows overexpression. To date, no somatic mutations of *CASR* but its downregulation have been reported in parathyroid tumors [119]. The overexpression of *CASR* might be attributed to the lack of comparable matched

normal parathyroid tissue even though *CASR* is expressed in other tissue types included in the compendium such as kidney, thyroid, lung and liver. Pathway analysis using differentially expressed genes identified 26 and 19 statistically significant pathways in the first and second recurrence specimens, respectively. These included known cancer-related Wnt/β-catenin, ErbB2-ErbB3 signaling, mismatch repair and Notch signaling pathways. Aberrations in the WNT/β-catenin signaling pathway have previously been suggested to be a cause of a subset of PCs [203]. Pathway analysis using the genes overexpressed exclusively in one recurrent transcriptome showed overexpression of mostly metabolism pathways in the second recurrence. However, cancer genes such as *AKT2* and *ERBB2* were overexpressed in the first recurrence only, perhaps driving the overexpression of pathways such as PI3K-AKT-mTOR.

2.4 Discussion

Despite population-based studies showing a 60% increase in the PC incidence rate in the United States between 1988 and 2003 [123], its etiology has remained largely unknown. The diagnosis of PC is seldom made preoperatively due to a high prevalence of benign disease, as well as a lack of molecular profiling tools that could potentially assist with diagnosis [204]. The overlap in pathological characteristics of benign parathyroid pathology and PC also means features such as local invasion, and the development of local recurrence or distant metastasis, are required for definite histopathological diagnosis of malignancy [127]. These characteristics however are present at a more advanced stage of disease when a cure is less likely. A comprehensive analysis of the molecular profile of PC can aid in not only identifying sensitive diagnostic tools,

but also novel therapeutic options [205]. In this report, we have examined the complete genome and transcriptome of a PC, made a comparison to the primary tumor's genome and identified novel somatic mutations in PC.

The *MEN1* tumor suppressor gene is mutated in a subset of parathyroid tumors [206-208]. Its product, a nuclear protein called menin, is believed to play a role in transcriptional regulation of gene expression, perhaps through modification of chromatin structure [209]. Several DNA-binding transcription factors [210-213] and chromatin-modifying proteins including MLL2 are also shown to interact with menin [214-216]. The histone methyltransferase complex that consists of menin, MLL2 and ASH2L trithorax family members methylates histone H3 on lysine 4, and acts as a transcriptional activator. The activity of the complex however is lost in tumors harboring menin mutations [215]; as a result, the epigenomic regulatory role of MEN1 might be responsible for its tumor suppressive activities [215]. Since H3K4 methylation is typically associated with an active transcription state [217], menin could enforce its tumor suppressive activity through activating important regulatory elements within the cell.

The *CDKN2C/p18* gene shows LOH in this patient with the remaining allele containing a missense mutation. In mice, haploinsufficiency of *p18* causes increased sensitivity to chemical carcinogens and leads to spontaneous pituitary tumors and lymphomas [218]. Other evidence for the role of *p18* as a tumor suppressor in endocrine tissues include the presence of germline mutations in cases of MEN1 with no *MEN1* mutations [219] and reduced expression of p18 in

benign parathyroid tumors [220]. Mutations in *CDKN1B/p27Kip1*, a member of another distinct CDKI, are observed in cases of sporadic parathyroid adenomas [221]. *p27* germline mutations are also found in patients with pituitary and parathyroid tumors lacking MEN1 mutations [222]. Similar to p18, loss of p27 leads to enlargement of organs in mice as well as the development of pituitary adenomas [223-225]. Knockout of both *p18* and *p27* leads to the development of tumors in multiple endocrine glands including the parathyroids, a phenotype similar to multiple endocrine neoplasia syndromes [226]. Since mutations of these CDKIs lead to malignancies of endocrine glands and *MEN1* is a known tumor suppressor whose loss leads to tumors of the endocrine organs, there is a possibility that these two processes are related [227].

An in vitro study using mouse embryonic fibroblasts has suggested that menin regulates the expression of *p18* and *p27* by directly binding to these loci and by recruiting MLL, a close homolog of MLL2, to the promoter of these cell cycle regulators [228]. Thus, loss of function of either MLL or menin results in reduction of H3K4 trimethylation and down-regulation of *p18* and *p27* expression [228,229]. H3K4 methylation of p27 and p18 promoters by menin maintains the transcription of these two cell cycle regulators and as a result prevents the formation of endocrine tumors [229]. Menin also forms a complex with MLL2/KMT2D and menin point mutations have been found to prevent complex methyltransferase activity [215]. The somatically acquired mutations in *MLL2* and *p18* in this patient may be driving the malignant phenotype through the same pathway that would otherwise be disrupted through the loss of MEN1.

Somatic mutations of *mTOR* and *PLD1* upstream of *CCND1* may also contribute to the development of PC. Increased activation of the mTORC1 complex can up-regulate cell cycle regulators such as *MYC* and *cyclin D1* [149-151,230]. mTOR is also a member of the mTORC2 complex that, in association with protein rictor, phosphorylates Akt at Ser473 [231]. *PLD1* hydrolyzes phosphatidylcholine to produce phosphatidic acid (PA), a lipid second messenger and regulator of cell signaling pathways. PA is required for stabilization and activation of both mTOR complexes and their downstream effectors [180,231,232]. As a result, alterations of *PLD1* can affect the mTOR signaling pathway [149,233]. The loss of the 5' UTR region of the gene in this patient suggests a loss of its regulatory elements, and as a consequence may lead to deregulated production of PA in the cells. This is of critical importance when considering therapeutic options for this patient or others with a similar molecular phenotype. PA not only interacts with and activates mTORC1 through binding to its FRB domain, but also competes with rapamycin for binding to this domain, and as a result elevated levels of PA confer resistance to rapamycin treatment [232,234]. PA has also been shown to have a more stable interaction with mTORC2 complex that plays an important role in cancers via phosphorylating Akt, hence blockers of both mTOR complexes along with lowering the PA level may prove to be the most effective means of blocking the mTOR signaling pathway in cancers [231].

The current study, to the best of our knowledge, is the first to profile the complete genomic and transcriptomic landscape of a PC and is also unique in defining somatic mutations in known cancer genes such as *p18*, *MLL2*, *PIK3CA* and *mTOR* that have never been previously identified

in PCs. The high frequency of 1p loss in PC [195] has led to the search for a tumor suppressor in this region. Both *p18* and *THRAP3* serve as candidate tumor suppressors in our case; the observed truncating mutation in the only copy of *THRAP3* is especially intriguing. These identified genomic alterations in PC, and the pathways they affect, could potentially be exploited as markers for diagnosis, and also as potential targets for therapy. Any clinical application of these novel observations will require the functional annotation of the identified mutations.

Table 2.1 Sequence libraries read statistics

	Total Number of Reads	Number of Aligned Reads	Average Coverage
Primary genome	378325412	257948721	8.1
Recurrence genome	1380449244	1213268909	41.3
Blood genome	1249701624	1123730025	38.4
First recurrence transcriptome	256888116	210608444	-
Second recurrence transcriptome	159972160	142737727	-

Table 2.2 Nineteen Illumina Body Map 2.0 project libraries and their tissue types

Lib ID	Protocol	Pathology	Tissue
HCT20142	RNA-seq	normal	kidney
HCT20143	RNA-seq	normal	heart
HCT20144	RNA-seq	normal	liver
HCT20145	RNA-seq	normal	lung
HCT20146	RNA-seq	normal	lymph node
HCT20147	RNA-seq	normal	prostate
HCT20148	RNA-seq	normal	skeletal muscle
HCT20149	RNA-seq	normal	white blood cells
HCT20150	RNA-seq	normal	ovary
HCT20151	RNA-seq	normal	testes
HCT20152	RNA-seq	normal	thyroid
HCT20158	RNA-seq	normal	adipose
HCT20159	RNA-seq	normal	adrenal
HCT20160	RNA-seq	normal	brain
HCT20161	RNA-seq	normal	breast
HCT20162	RNA-seq	normal	colon
HCT20170	RNA-seq	normal	16 Tissues mixture
HCT20172	RNA-seq	normal	16 Tissues mixture
HCT20173	RNA-seq	normal	16 Tissues mixture

Table 2.3 Primers for verification of 23 putative somatic mutations

Genomic position of each mutation is indicated; given the degraded nature of DNA extracted from the FFPE sample, smaller amplicons were designed for the primary specimen compared with the flash frozen recurrent (recur) and blood samples

Chr	Position	Sample	Forward Primer	Reverse Primer	Amplicon Size
1	11177077	recur & blood	AATTAAATTACTCACCTATCTCCCAGGC	GAGGCTGCAGTGAGCCAAGATAG	257
1	11177077	primary	AATTAAATTACTCACCTATCTCCCAGGC	ATAGCACCACTGCCTTCCAGC	238
1	36752132	recur & blood	GTGGGCGTAACAGAGGCTTTTATC	GCCGGCTATCCTTAGAAGAGGAC	306
1	36752132	primary	GTGGGCGTAACAGAGGCTTTTATC	AAGAACGGGAGGATGAGGAGC	228
1	51436102	recur & blood	AATCACGTGTGAATCGAGGGG	TGTGCATTGACGTTTACATTATTTTGC	297
1	51436102	primary	CCAGATTAACCATCCCAGTCCTTC	TGTGCATTGACGTTTACATTATTTTGC	227
1	153747974	recur & blood	AACCAGACGGTGCCGATAGAG	ATGAGAAAGGTGTGCGCGG	287
1	153747974	primary	GTTTCTGCTCTCCGCCCG	TAGCAACAGCAGCAGGGGC	225
1	159284298	recur & blood	GAAGTACATGGGGGTGTGAAGATG	CCTTGGTCCAGCCTACTCTGTTTC	272
1	159284298	primary	GAAGTACATGGGGGTGTGAAGATG	CCAGCTCTTTCAAACCTAGACCTACC	215
1	248263363	recur & blood	CCTGTGGCCGAGTCCTATTTG	ATCCCAAACACTCTCCTCATAGCC	271
1	248263363	primary	CCTGTGGCCGAGTCCTATTTG	CAGGCTGTAGATAATGGGATTGAGC	227
2	58311236	recur & blood	TTGTCTGCAGCTTTCCCCAC	GGGTAATAGCTGGCAATACAGAAAAAC	249
2	58311236	primary	TTTTTGTCTGCAGCTTTCCCC	AGAAAAACAAAGGTTCTCCAGTTTTAATG	233
3	120319986	recur & blood	TTTAGGAAAATCCTGCCTTGCTTC	AAGGTGGGAGGGGCACTG	258
3	120319986	primary	TTTAGGAAAATCCTGCCTTGCTTC	CAGCTGGTCCCTAGAAAGTGAACC	214
3	178936091	recur & blood	ACACGAGATCCTCTCTCTGAAATCAC	GCATTTAATGTGCCAACTACCAATG	277
3	178936091	primary	ACACGAGATCCTCTCTCTGAAATCAC	TTTCCACAAATATCAATTTACAACCATTG	243
5	54570746	recur & blood	AATGAGATCCAGGTTGATTTTATGAGG	CCACTATTAAAAAGTATACGCTTCCTGTATTTTAAC	287
5	54570746	primary	AATGAGATCCAGGTTGATTTTATGAGG	TTTGAATGAATGCAAATACCGCC	217
5	112227760	recur & blood	CAGGAGAAGAAAGAAAAAGAGGAAGC	CCATTCCAAACGTTGTAAACATGC	256
5	112227760	primary	CAGGAGAAGAAAGAAAAAGAGGAAGC	GGACTAGATGTTGGGAAATTATGTTTACG	215
6	56883260	recur & blood	TTATTTCCCAATACGGATGATGTTTC	CACTCCCAGTTCTCCCCTTTTG	284

Chr	Position	Sample	Forward Primer	Reverse Primer	Amplicon Size
6	56883260	primary	TTATTTCCCAATACGGATGATGTTTC	TGGGTCACTAGTATTTTCTAGTAGAGTGATTGG	235
8	22078960	recur & blood	CATGAGCTGGTGCCCACTG	TCTACTGCATGTACACGGCCTACC	251
8	22078960	primary	CATGAGCTGGTGCCCACTG	ACTACGCCATCCTGGTGCTG	227
9	138714298	recur & blood	CTCTATGTCCACCACATCCGAGTC	CCACCGAGACAGGACCACTG	248
9	138714298	primary	TATGTCCACCACATCCGAGTCAG	CGCAGGAGAGGTCTGTGGTG	214
12	49444944	recur & blood	CTCAGGGGACAGATGCGATTC	GCAGGCTGAGGAGCCACAC	274
12	49444944	primary	GACAGATGCGATTCCTCAGGC	GAGGAGCCACACTTGTCCCC	206
12	53701335	recur & blood	AGAACAGGTGGTGGCCCTG	TGGCTGCTGTATGGCTCTGC	272
12	53701335	primary	AGAACAGGTGGTGGCCCTG	CCTTCTCTGCAGGGCTGGTC	209
14	45432103	recur & blood	GGTAGTGATGAGAAGCGGCTCTG	GAAGCAGTAGTGCTGTGGAAGCTC	274
14	45432103	primary	TAGTGATGAGAAGCGGCTCTGC	TATAAGCGTTCTCAGCACCTCTCC	214
16	1877300	recur & blood	GAGTGGGGAAAGAACATCGTCTG	TATCCAGGCACAGGGCATAGC	265
16	1877300	primary	GAGTGGGGAAAGAACATCGTCTG	ACGTAGTCCATGGCCGCAG	239
16	9024162	recur & blood	AACCACAGGCACTTACCATCCTC	TTTTTGGATTAGAATGAATCTTTATTGTGG	265
16	9024162	primary	AACCACAGGCACTTACCATCCTC	GGAATTTATTAATGATTAGGCAACATCAAAAG	228
17	75398219	recur & blood	ACCCAACTCCACCCCACC	CCTCTTGAGCCCGAACCG	289
17	75398219	primary	ACCCAACTCCACCCCACC	GATGTCAATGGACAGCTCAGTGC	229
18	2775787	recur & blood	AGAGCTGCGATGGTTATTTCTTGG	TTCAAGTCAATTCTCTGTTTTAGGTATCTTTTAG	277
18	2775787	primary	TTGACAGTTTATTTTTAATTCATGTGTTTCAG	GACTACACAGTCCATGTCACTTGCC	211
18	3880013	recur & blood	GAGTGGTGCGACAGCGAGTC	AGGCTGCAGGAAGCAGAGATG	279
18	3880013	primary	GAGTGGTGCGACAGCGAGTC	ACCTAATTTCCAAGAGAAATGTTAACGAC	206
X	90691393	recur & blood	GAGCAACTTTCACCAATGTTTTCG	AGGCGTTCAATTTTCTTCTGTGC	259
X	90691393	primary	TGAAGGAACTTTTCTGTGAATATGGG	AGGCGTTCAATTTTCTTCTGTGC	199

Table 2.4 Primers used for verification of putative somatic structural variants

Region	Forward Primer	Reverse Primer
AGBL1 breakpoint	CACTTGGATTTTCTCTCTTCTTTTCTTG	CAGAACATTCTAACCAATAACCACAGAATATG
PLD1 breakpoint	AGAGTTATCGAACCCTAATAACTCCACC	AGGATGTCTCATGACAGTAACAGAATAAGAG
AGBL1-PLD1 fusion point	AGGATGTCTCATGACAGTAACAGAATAAGAG	CAGAACATTCTAACCAATAACCACAGAATATG
BC033837 breakpoint	CTTGGCCCACAGTTCCTCTCTC	GCGGCTATAAACTCTAGTCCTGCC
SKP2 breakpoint	CCAGGAAACTTGAAGTGTAATTGGG	GTGTTTCCCAAAGGAAAGATGGAC
BC033837-SKP2 fusion point	GTGTTTCCCAAAGGAAAGATGGAC	GCGGCTATAAACTCTAGTCCTGCC
DMXL2 breakpoint	TGATCTTCATAGCTCTGTGGTATCTTTG	GGCAAGAAAAAGTGTTGTTGAAGG
AKAP13 breakpoint	TGAATTCAAATACCTCTGTTTCTTAGTACTCC	TCGCAGCTAGAGATAAATTACATGGTTC
DMXL2-AKAP13 fusion point	TGATCTTCATAGCTCTGTGGTATCTTTG	TGAATTCAAATACCTCTGTTTCTTAGTACTCC

Table 2.5 List of verified novel somatic point mutations in the parathyroid carcinoma samples
**Repeated attempts to amplify the region around these three SNVs using PCR failed in the primary tumor, potentially due to the degraded nature of FFPE DNA. As a result the status of these three mutations in the primary sample is not verified despite the strong support from the next-generation sequence data

Chr	Position	Reference Base	Variant Base	Amino Acid Change	Zygosity State	Gene	Genomic Sample with Mutation	Variant Allele Present in 1st Recurrence	Variant Allele Present in 2nd Recurrence	Gene Expression Fold Change: 1st Recurrence vs Compendium	Gene Expression Fold Change: 2nd Recurrence vs Compendium
1	11177077	A	C	L2334V	Hemizygous	mTOR	primary & recurrence	Yes	Yes	-1.5	-1.8
1	36752132	C	T	R101*	Hemizygous	THRAP3	primary & recurrence	Yes	Yes	-12.3	-8.3
1	51436102	T	G	L21R	Hemizygous	CDKN2C	primary & recurrence	Yes	Yes	3.8	4.4
1	153747974	A	G	M48V	Heterozygous	SLC27A3	primary & recurrence	Yes	Yes	-1.2	-1.1
1	159284298	C	T	R51H	Heterozygous	OR10J3	primary & recurrence	No	No	1	1
1	248263363	C	A	S229*	Heterozygous	OR2L13	primary & recurrence	No	No	-1.1	-1.1
2	58311236	A	G	N50S	Heterozygous	VRK2	primary & recurrence	Yes	Yes	-1.4	-1.5
3	120319986	A	G	Y70C	Hemizygous	NDUFB4	recurrence**	Yes	Yes	1.1	1.4
3	178936091	G	A	E545K	Heterozygous	PIK3CA	primary	No	No	-1.2	-1.5
5	54570746	C	T	M840I	Heterozygous	DHX29	primary & recurrence	Yes	Yes	1.2	1.2
5	112227760	C	T	Q142*	Heterozygous	ZRSR1	recurrence**	Yes	Yes	-2.1	-1.9
6	56883260	A	G	I252V	Heterozygous	BEND6	primary & recurrence	No	No	-3.1	-3
8	22078960	C	T	G300E	Heterozygous	PHYHIP	recurrence	No	No	-8.2	-7.3
9	138714298	C	T	D737N	Heterozygous	CAMSAP1	recurrence	Yes	Yes	-1.1	-1.2
12	49444944	C	A	C841F	Heterozygous	MLL2	primary & recurrence	Yes	Yes	1.8	1.3
12	53701335	G	A	P527S	Heterozygous	AAAS	primary & recurrence	Yes	Yes	2.4	3.1
14	45432103	C	T	S160L	Heterozygous	FAM179B	recurrence	Yes	Yes	-1.3	-1.3
16	1877300	G	A	G24R	Heterozygous	FAHD1	primary & recurrence	Yes	Yes	1.7	1.9
16	9024162	C	A	D58Y	Heterozygous	USP7	primary & recurrence	Yes	Yes	-1.5	-1.5
17	75398219	T	A	L52H	Heterozygous	SEPT9	primary & recurrence	Yes	Yes	1.8	1.6
18	2775787	A	C	H1744P	Heterozygous	SMCHD1	recurrence **	Yes	Yes	-1.9	-1.9
18	3880013	G	A	A19V	Heterozygous	DLGAP1	primary & recurrence	No	No	-3.9	-3.7
X	90691393	C	T	R273*	Heterozygous	PABPC5	recurrence	No	No	-1.8	-1.7

Table 2.6 Somatic gene fusions in the parathyroid genome

Three somatic gene fusions were detected in the parathyroid genome (coordinates are based on the hg19/GRCh37 assembly). The expression status of affected genes in both transcriptomes is listed; none showed over- or under-expression compared with the normal compendium

Event	Type	Breakpoint	Gene	Expressed in 1st Recurrence	Expressed in 2nd Recurrence
PLD1-AGBL1 Fusion	Translocation	chr15:87238736	AGBL1	No	No
		chr3:171477831	PLD1	No	No
BC033837-SKP2 Fusion	Translocation	chr22:49971207	BC033837	No	No
		chr5:36177713	SKP2	Yes	Yes
AKAP13-DMXL2 Fusion	Inversion	chr15:51791474	DMXL2	Yes	Yes
		chr15:86131718	AKAP13	Yes	Yes

Table 2.7 Sequence of the assembled genomic contigs providing support for the structural events

Event Type	Genes	Contig Sequence
Translocation	AGBL1-PLD1	TAAAAAAGTCAAATATACTTCTTGGCTTCTACTAAAACCTCCTTTC ATTATTTGCATGACTATAATGGTCTATCAATTTTATGTATCTTTTC AAAGAACCAGCTTTTTGTTTCATTTATTTTTGT
Translocation	SKP2-BC033837	AGTTAATCCGAAAAATTTGGAAAGAAAAAAAAAAAAAGACACTA ACCCACATTGGGTCTGCCTGGCTGAATGGGTCCCGACGGCTCTG ACGGCTCCCCACACCCCTGCCCTGTGGGCCATGCT
Inversion	DMXL2- AKAP13	CCTTAGAAAGGGAAGGAAAAAACTCACATCCTTGAATTCAAATAC CTCTGTTTCTTAGTACTCTATTTCTGATGATGTTTTTTGTTCACC AACTGTAATTCAAGATGGTGGCTTATTTGAGGCTG

Table 2.8 Coordinates for focal-level copy number changes in relapse sample
Gain=1 extra copy, Amplified= 2 extra copies, Highly Amplified= 3 or more extra copies

Chromosome	Start Position	End Position	Event Type
1	13002001	13393000	Homozygous Deletion
1	16866001	16991000	Homozygous Deletion
1	25588001	25665000	Homozygous Deletion
3	162512001	162626000	Homozygous Deletion
3	164685001	166006000	Amplified
5	35531001	36178000	Heterozygous Deletion
7	21641001	21739000	Amplified
7	22076001	22502000	Amplified
11	34703001	34847000	Heterozygous Deletion
11	63032001	64095000	Heterozygous Deletion
11	69043001	69450000	Heterozygous Deletion
11	69450001	69493000	Highly Amplified
11	69493001	69592000	Heterozygous Deletion
12	55197001	56680000	Heterozygous Deletion
14	22474001	22987000	Gain
15	72979001	73074000	Amplified
16	70894001	71201000	Gain
19	235001	392000	Heterozygous Deletion
19	3415001	5386000	Heterozygous Deletion
19	11123001	12066000	Amplified
19	12066001	14696000	Heterozygous Deletion
19	14696001	18207000	Amplified
19	18605001	24513000	Heterozygous Deletion
19	24594001	24630000	Heterozygous Deletion
21	19715001	19758000	Heterozygous Deletion
22	39358001	39389000	Homozygous Deletion
22	47861001	48270000	Highly Amplified
22	48270001	48592000	Amplified
22	48592001	48673000	Highly Amplified
22	48894001	49111000	Amplified
22	49364001	49971000	Heterozygous Deletion
X	143141001	143619000	Homozygous Deletion

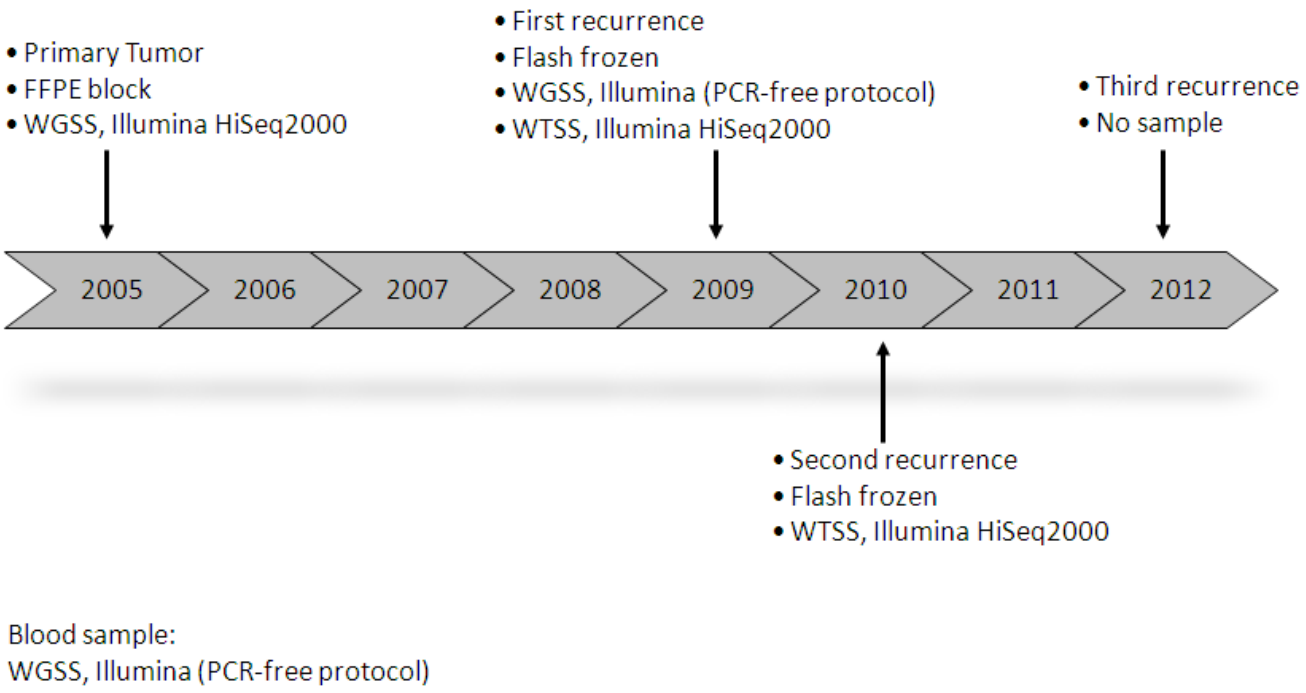

Figure 2.1 Patient history
Timeline of the patient's disease history and the sequencing experiments performed on each sample. WGSS: whole genome shotgun sequencing, WTSS: whole transcriptome shotgun sequencing, FFPE: formalin-fixed paraffin-embedded

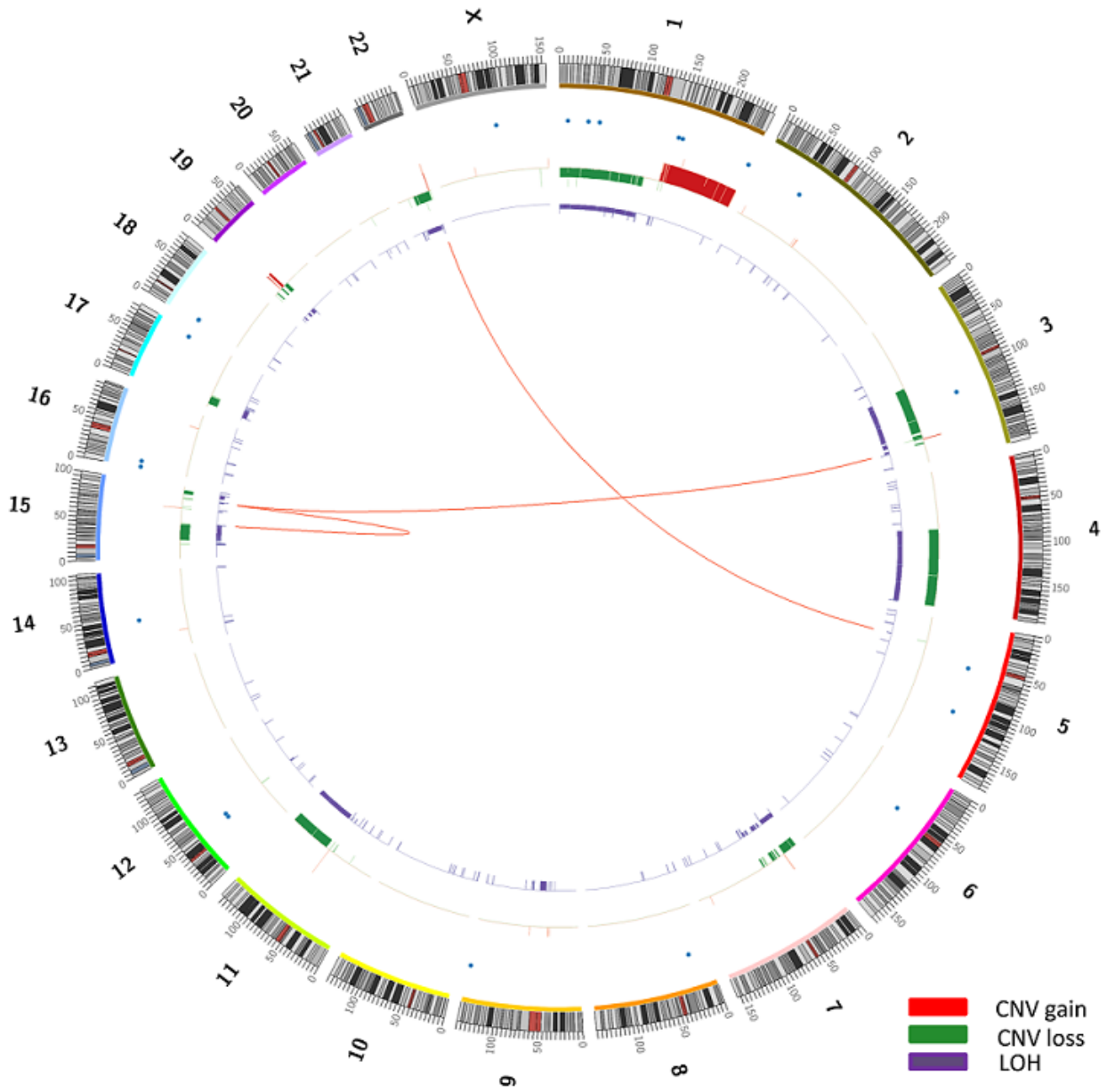

Figure 2.2 Somatic alterations
Regions of CNV and LOH, somatic SNVs and SVs identified from the recurrent genome are depicted. From the outer circle inward: somatic single nucleotide variants (blue dots), regions of copy number gain (red) and loss (green), regions of loss of heterozygosity (purple) and large structural events (red lines)

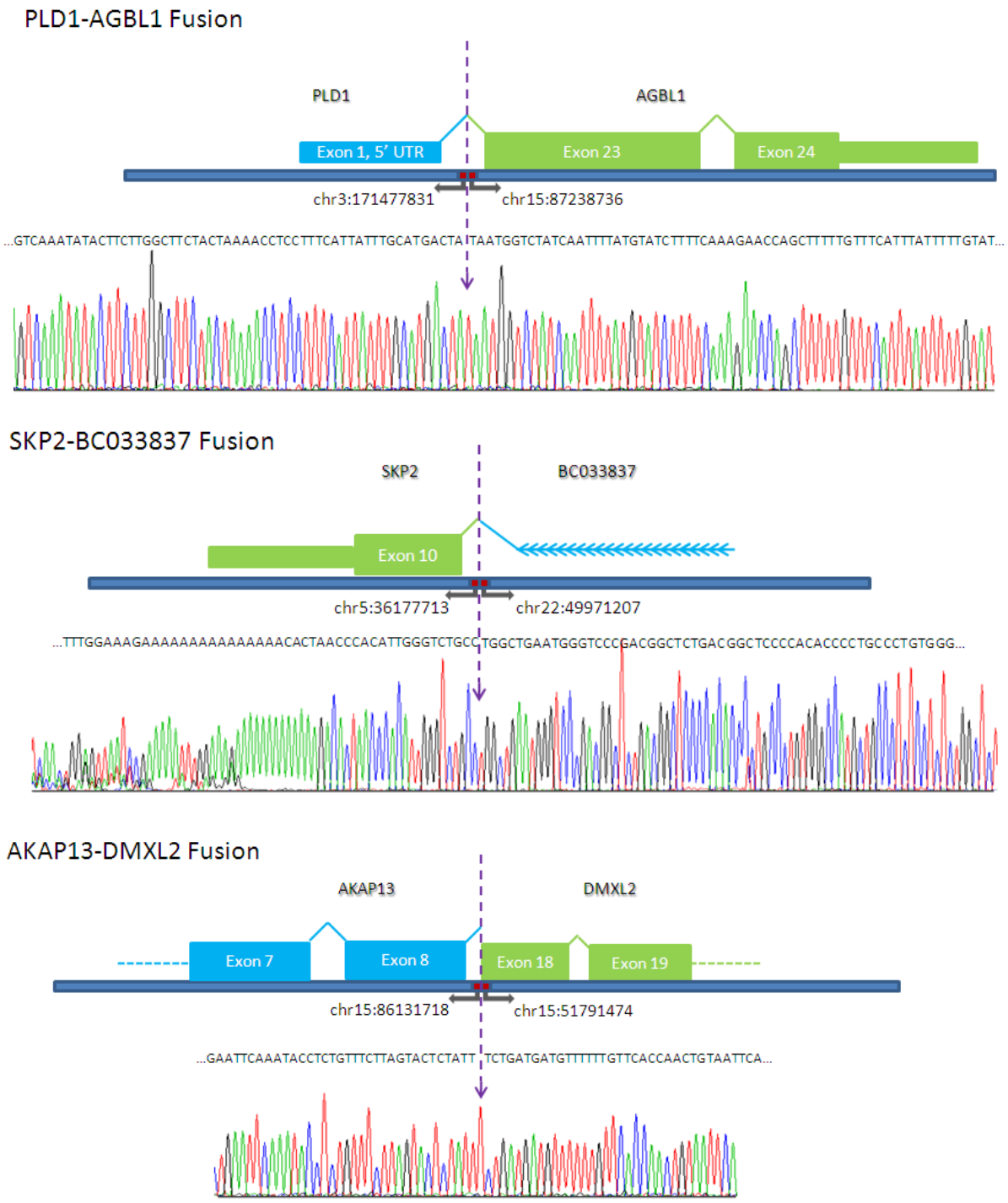

Figure 2.3 Somatic structural variants
Schematic diagrams of the 3 structural variants identified in the recurrent whole genome data. Sanger sequence traces of the validation experiments demonstrate the novel sequences at the fusion breakpoint

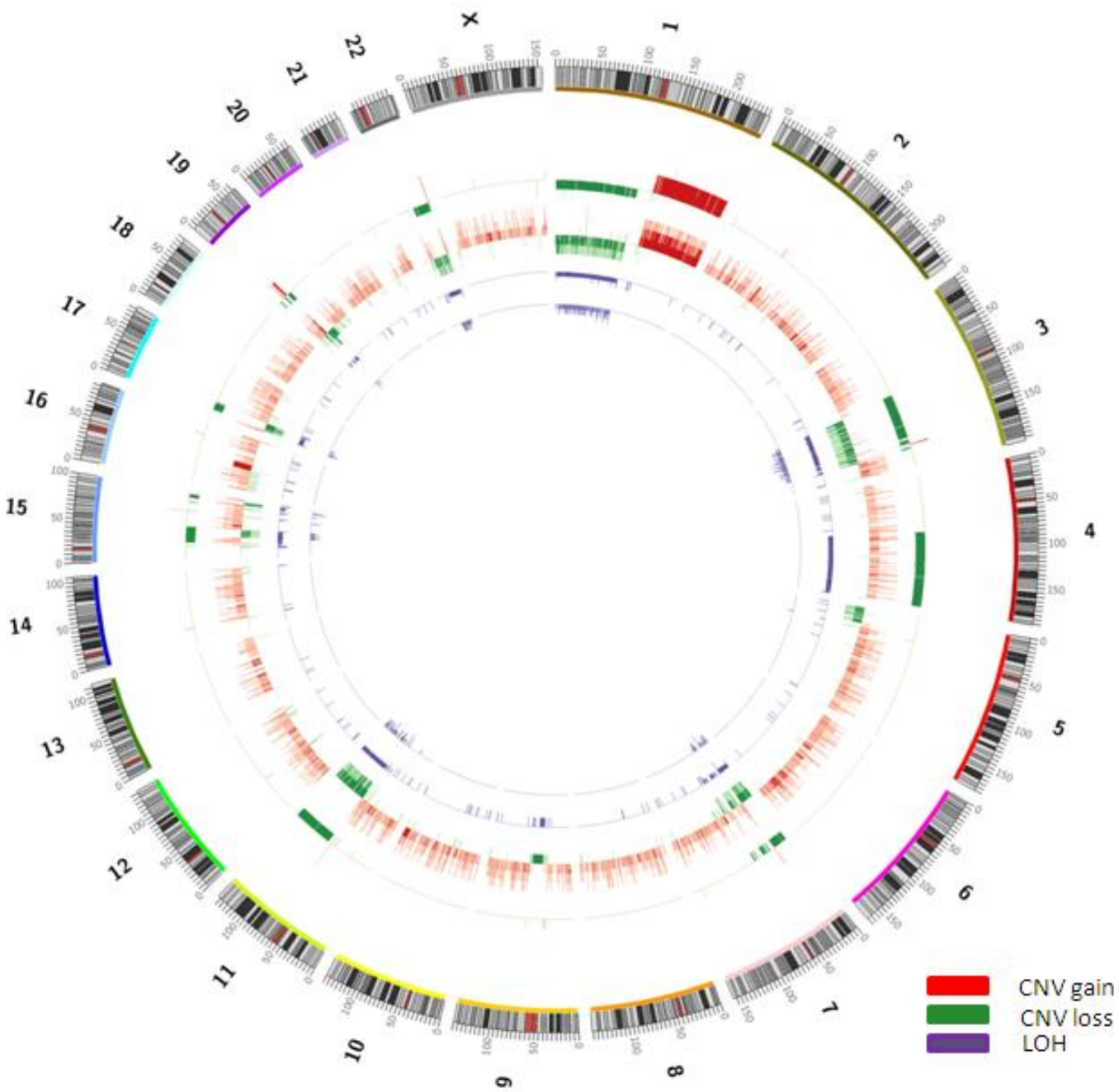

Figure 2.4 Primary and relapse specimens CNV and LOH regions
Tracks from the outer circle inward are relapse CNV, primary CNV, relapse LOH and primary LOH. The formalin-fixed paraffin-embedded relapse sample demonstrated a profile with a much higher background noise due to the degraded nature of preserved DNA

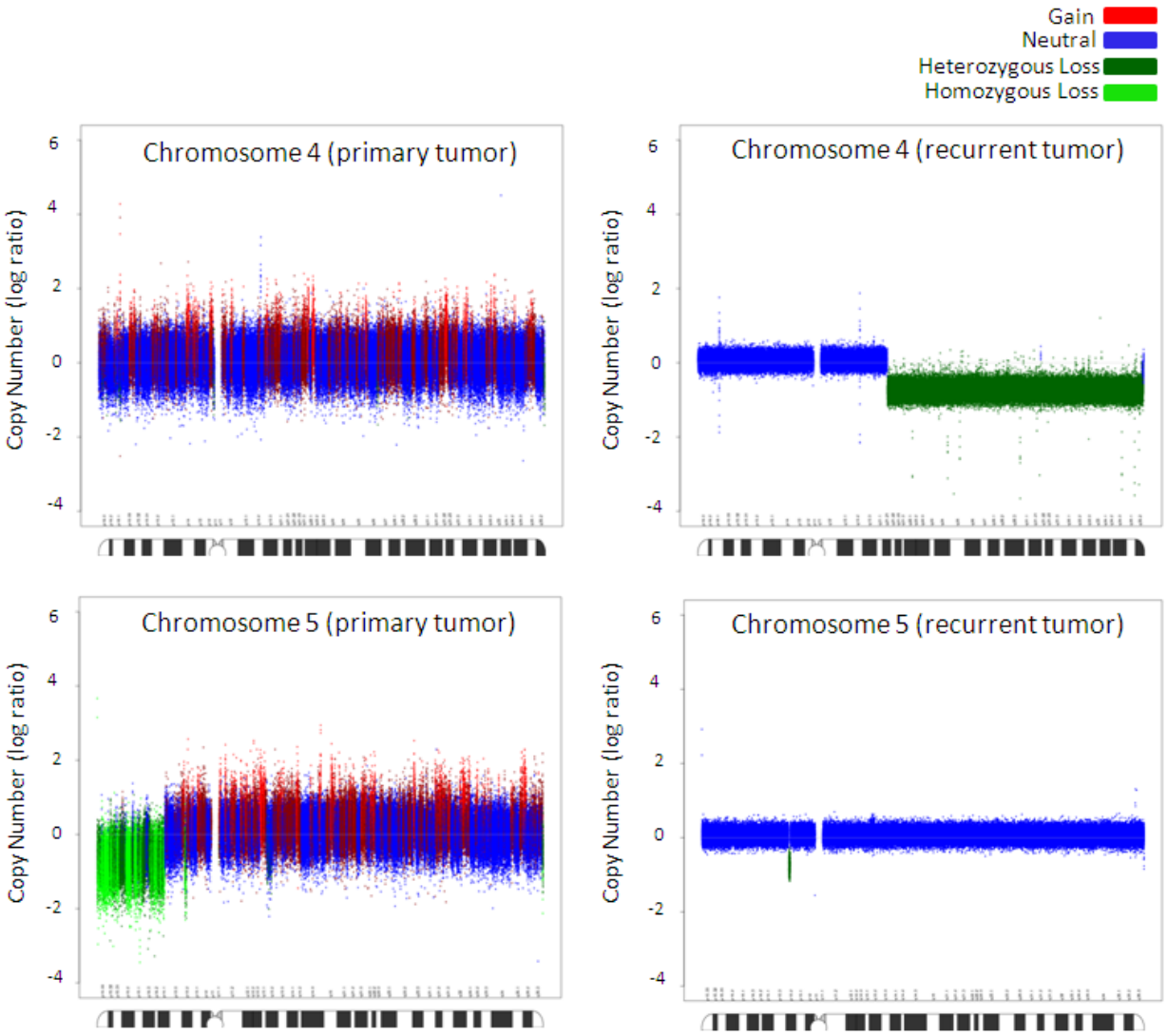

Figure 2.5 Primary and relapse specimens CNV comparison
The most prominent copy number differences between the genomes of the primary and first relapse samples are depicted

Chapter 3: *MEN1* Mutations in Hürthle cell (Oncocytic) Thyroid Carcinoma[3]

3.1 Introduction

Hürthle or oncocytic cells of the thyroid are follicular-derived cells with a large nucleus, prominent nucleolus and an abnormal accumulation of mitochondria resulting in a distinct granular appearance on histology sections [19]. Although oncocytes can be found in various metabolically active tissues such as kidney, parathyroid, salivary and adrenal glands, they are more commonly found in the thyroid and are believed to be the result of metaplastic changes in the epithelial cell linings of thyroid follicles [19,20].

Nodules consisting of 75% or greater oncocytic cells are categorized as Hürthle cell neoplasms; those demonstrating capsular or vascular invasion or presence of distant metastasis are diagnosed as malignant tumors, rendering fine-needle aspiration cytology as an inadequate technique for diagnosis of Hürthle cell malignancies [19]. Hürthle cell thyroid carcinoma, also known as oncocytic thyroid carcinoma, is considered an oncocytic variant of follicular thyroid cancers (FTCs) by some [235] while others regard it as a separate subtype of differentiated

[3] A version of this chapter has been published, and the author contributions are provided in the Preface as per the University of British Columbia PhD thesis guidelines: Katayoon Kasaian, Ana-Maria Chindris, Sam M Wiseman, Karen L Mungall, Thomas Zeng, Kane Tse, Jacqueline E Schein, Michael Rivera, Brian M Necela, Jennifer M Kachergus, John D Casler, Andrew J Mungall, Richard A Moore, Marco A Marra, John A Copland, E Aubrey Thompson, Robert C Smallridge, Steven JM Jones. (2015). *MEN1* Mutations in Hürthle Cell (Oncocytic) Thyroid Carcinoma. Journal of Clinical Endocrinology and Metabolism. 2015 Apr;100(4):E611-5. doi: 10.1210/jc.2014-3622. Epub 2015 Jan 27. Copyright by Endocrine Society.

thyroid cancers (DTCs) [236]. Nonetheless, it is treated according to the same established guidelines for papillary and follicular neoplasms, namely surgical removal of all or part of the gland, radioactive iodine treatment and to a lesser extent with chemotherapy and radiation treatment [236]. Oncocytic thyroid carcinoma is a rare entity accounting for only 3 to 7% of DTCs, yet it demonstrates more aggressive behavior with five-year survival rates ranging between 50% and 60% [19,237]. Demographic comparison of 3,311 Hürthle cell thyroid carcinoma patients with 59,585 individuals diagnosed with papillary or follicular carcinomas from The Surveillance, Epidemiology, and End Results (SEER) database between 1988 and 2009 showed higher prevalence of oncocytic carcinomas among older men who generally present with larger tumors, more advanced disease and demonstrate lower disease-specific survival [237].

The uncommon occurrence of this subtype of thyroid cancer, similar to that of parathyroid carcinoma, has hindered the complete characterization of this malignancy on the molecular level; genetic changes associated with oncocytic thyroid carcinoma and their roles in tumorigenesis are not entirely understood. Here we report the profile of two tumors on the whole genome scale in addition to identification of recurrent inactivating mutations in the tumor suppressor gene *MEN1* in five of 74 patients diagnosed with Hürthle cell thyroid carcinoma.

3.2 Materials and Methods

3.2.1 Study Samples

Biopsy specimens for whole genome sequencing experiments were collected from two patients diagnosed with Hürthle cell carcinoma. One tumor specimen was obtained from a 58-year old male with a 2 cm right carcinoma with focal extrathyroidal extension (T3N0M0). The patient subsequently developed liver metastases for which he received chemotherapy followed by resection of the liver lesion. The primary tumor underwent whole genome sequencing. The second patient was a 55-year old female with 2.9 cm left Hürthle cell thyroid carcinoma with perithyroidal soft tissue involvement (T3N0M0). The patient developed neck recurrence for which she had cervical re-exploration; this was followed by mediastinal lymph node recurrence. The metastasis specimen obtained from the lymph node underwent whole genome sequencing. Neither patient had a family history of cancer, multiple endocrine neoplasia syndrome or involvement of multiple organ systems. Subsequently, DNA extracted from formalin-fixed paraffin-embedded (FFPE) specimens of 72 oncocytic thyroid carcinoma patients (6 accompanied by matched adjacent normal tissue), 5 individuals diagnosed with Hürthle cell adenoma and one Hürthle cell carcinoma cell line, XTC.UC1, were included in the validation experiment. The tumor specimens were collected as part of a research project approved by the University of British Columbia Cancer Agency Research Ethics Board and Mayo Clinic Institutional Review Board and are in accordance with the Declaration of Helsinki. The tumor samples were classified according to the World Health Organization criteria. The data are

consented for research reports and scientific publications. The protocol to be followed requires that these datasets will not be released into the public domain but can be made available via a tiered-access mechanism to named investigators of institutions agreeing by a materials transfer agreement that they will honor the same ethical and privacy principles required by the British Columbia Cancer Agency Research Ethics Board.

3.2.2 DNA Sequencing

DNA extracted from the two frozen tumor tissues and the blood samples were subjected to high-throughput whole genome sequencing using locally established sequencing protocols. Biopsy specimens were embedded in Tissue-Tek O.C.T. (optimal cutting temperature) compound (Sakura Finetek USA, Inc.) and sectioned for DNA extraction. Using 1ug DNA each from the tumor and blood, four whole genome libraries were constructed using a modified version of Illumina TruSeq PCR free protocol (FC-121-3001). In brief, 1ug genomic DNA was sheared for 45 sec, duty cycle 10%, intensity 5 burst per second 200 using Covaris E210, to an average of 400bp. NEB Paired-End Sample Prep Kit (New England Biolabs, USA) was used in library construction. Following the end repair reaction, a size selection was done using Ampure XP bead (Beckman-Coulter, USA). The sample:bead ratio is 110:27 for upper cut and 137:15 for lower cut respectively. The resulting size selected fraction, 300-500bp, was A-tailed, and ligated to Illumina TruSeq adapters. The PCR-free libraries were cleaned up with Ampure XP beads and quantified by qPCR assay using the KAPA SYBR FAST qPCR kit (Kapa Biosystems (Pty) Ltd, South Africa). Paired-end 100bp reads were generated on Illumina HiSeq2500 sequencers following

the manufacturer's protocol with minor variations. Software version HCS1.5.8 was utilized (Table 3.1).

Sanger sequencing was subsequently used as an orthogonal technique for the verification of somatic *MEN1* mutations in both patients using the forward primer 5'-GGCTCAGAGTTGGGGGACTA-3' and the reverse primer 5'-CGGGAGTCCAAGCCAGAG-3', spanning both mutations. Following the verification experiment, the protein coding region of *MEN1* was sequenced by exon-tiling with 17 amplicons; given the degraded nature of DNA extracted from FFPE samples, the primers were designed to produce smaller amplicons with lengths ranging in 162bp to 249bp (Table 3.2). Primers were designed with the Primer3 software [238] with a GC clamp and an optimal Tm of 64°C to ensure specificity. Primers were tested using a combination of UCSC's in-silico PCR tool aligned against the reference human genome and custom in-house scripts to verify that all exons of *MEN1* were covered by an Illumina MiSeq 250bp paired end read. The primers were tagged with Illumina adapters to enable a direct sequencing approach that precludes the need for adapter ligation during sample preparation. The Illumina adapter tags are as follows: 5'- CGCTCTTCCGATCTCTG on the forward amplicon primer and 5'- TGCTCTTCCGATCTGAC on the reverse amplicon primers. The standard PCR conditions used were an initial denaturation of 98°C for 30 seconds, followed by 35 cycles of 98°C for 10 seconds, 68°C for 15 seconds and 72°C for 8 seconds, and a final extension at 72°C for 10 minutes. PCR was set up using Phusion polymerase (Fisher Scientific, catalogue # F-

540L) according to manufacturer's specifications. One amplicon in the set with a high GC content of 69%

(5-'cgctcttccgatctctgCAGAAAATGCTCCACGAAGCC-3' and

5'-tgctcttccgatctgacGTGGAACCTTAGCGGACCCTG-3') required alternate PCR conditions of 98°C for 30 seconds, followed by 35 cycles of 98°C for 10 seconds, 69°C for 15 seconds and 72°C for 8 seconds, and a final extension at 72°C for 10 minutes. The PCR for this GC rich amplicon was set up using Phusion according to manufacturer's specifications with GC buffer and addition of betaine to 1M final concentration. Amplicons were pooled by template for direct-sequencing sample preparation. Sample preparation involved a second round of amplification using Phusion DNA polymerase with 6 cycles using PE primer 1.0-DS (5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCTG-3') and a custom PCR Primer (5'-CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGAC-3') that contains a unique six-nucleotide 'index' shown here as N's. DNA quality was assessed using an Agilent DNA 1000 series II assay (Agilent, Santa Clara CA, USA) and DNA quantity was measured using a Quant-iT dsDNA HS assay kit on a Qubit fluorometer (Life Technologies, Grand Island, NY, USA). The indexed libraries were pooled together and sequenced on the Illumina MiSeq platform with paired-end 250bp reads using v2 reagents. An in-house generated PhiX control library was spiked in to the samples at 20% molar ratio as a sequencing control. Performing the validation experiments on MiSeq instruments rather than Sanger sequencing that was done for parathyroid mutation validation in Chapter 2 allowed for identification of low

allele frequency mutations that would have otherwise been missed. All genomic and targeted sequencing datasets have been deposited at the European Genome-phenome Archive (EGA, http://www.ebi.ac.uk/ega/) under accession number EGAS00001000940.

3.2.3 Bioinformatic Analysis

Sequence reads from the whole genome libraries were aligned to the human reference genome (build hg19) using the Burrows-Wheeler Alignment (BWA) tool [56]. The tumor's genomic sequence was compared to that of patient's constitutive DNA to identify somatic alterations. Regions of copy number variation (CNV) and loss of heterozygosity (LOH) were identified using Hidden Markov model-based approaches HMMcopy and APOLLOH [95], respectively. Single nucleotide mutations were identified using a probabilistic joint variant calling approach utilizing SAMtools and Strelka [64,75]. Small insertions and deletions (indels) were identified using Strelka [75]. *De novo* assembly and annotation of genomic data using ABySS [62] and Trans-ABySS [87,139] were used to identify small indels, structural variants and fusion genes. Sequence reads from the targeted validation experiment were aligned to the same reference genome but using BWA-MEM given the longer read lengths [56]. Variants were called with the same pipeline as above, skipping read depth and duplicate read filtration steps. These reads were also *de novo* assembled using ABySS to provide supporting evidence for the variants identified through alignment-based techniques.

3.3 Results

Copy number analysis revealed large regions of somatic copy number alteration in both tumors. The metastatic tumor had a change in every chromosome with 1 copy loss of chromosomes 1, 2, 3, 4, 6, 8, 9, 11, 14, 15, 16, 21 and X while the remaining chromosomes had gained extra copies. The primary tumor had a striking profile, showing copy-neutral LOH of chromosomes 1, 2, 3, 4, 6, 8, 9, 11, 13, 14, 15, 17 and 22 (Figures 3.1, 3.2 and 3.3). Loss of heterozygosity while maintaining two chromosomal copies is likely the result of chromosomal amplification of a mostly haploid genome during the evolution of the tumor. Copy number data also points to the high amplification of mitochondrial genome confirming the increase in mitochondrial numbers in both tumors. *De novo* assembly of sequence reads found no gene fusions in these tumors. The profile of large genomic alterations in these thyroid oncocytic tumors demonstrated a very different profile compared to the parathyroid cancer genome (Figure 2.2). Such extensive regions of copy number alteration and loss of heterozygosity were not observed in the parathyroid cancer; however, regions of focal copy number change in the parathyroid, perhaps also facilitating the acquirement of gene fusions by introducing genomic breakpoints, were not present in these oncocytic tumors.

We identified 51 and 157 somatic single nucleotide variants (SNVs) and indels in the primary and metastatic genomes, respectively (Tables 3.3 and 3.4). Of particular interest was a splice site mutation in *EWSR1* in the primary tumor, a heterozygous missense mutation in *BRCA1* and

a hemizygous frame-shift deletion in the DNA mismatch repair gene *MSH2* in the metastatic tumor. Although this *MSH2* frame-shift deletion is recorded in dbSNP Clinical Channel (rs63751463), its clinical significance is unknown. Mismatch repair (MMR) deficiency including loss of function of the *MSH2* gene have been observed in various cancers and are often associated with hypermutated and microsatellite unstable phenotypes [239,240]. Higher number of somatic mutations and small insertions and deletions, particularly those in microsatellites, may as a result indicate an MMR deficiency [241]. Over 3 times more somatic mutations were identified in the genome of the metastatic tumor when compared with the primary tumor. While only 3% of these mutations in the primary tumor were short indels, over 25% of those in the metastatic tumor were, perhaps suggesting an MMR deficiency in this tumor. None of the primary tumor indels affected the microsatellites and only a small fraction of those in the metastatic tumor (0.38%) led to uncorrected insertions and deletions in microsatellites. Overall, oncocytic tumors demonstrate a much more unstable genome than that of parathyroid cancer with higher number of small mutations and large CNVs.

Mutational analysis of the two genomic datasets revealed mutations of multiple endocrine neoplasia 1 gene (*MEN1*) in both tumors; *MEN1* was the only shared mutated gene in these specimens and it harbored two distinct somatic single nucleotide deletions. The primary tumor showed a homozygous deletion of a single base leading to a shift in the reading frame and as a result the deletion of amino acids 592 to the end of MEN1. The metastatic tumor showed a hemizygous deletion of a single base causing a frame-shift in MEN1 starting from amino acid

521 in exon 10. This particular deletion has previously been identified in the germline of patients diagnosed with multiple endocrine neoplasia type I disorder [242] and also in 3 large intestine and 1 liver carcinoma samples (COSM1355794) [146]. Both deletions were present at close to 100% allele frequency after correction for normal tissue contamination. The presence of these two somatic mutations was subsequently verified using Sanger sequencing. As mentioned in chapter 2, *MEN1* loss of function mutations are frequently found in parathyroid tumors of patients with MEN1 disorder or those with familial isolated hyperparathyroidism; however, tumors of the thyroid, including oncocytic benign or malignant variants, are not seen in these patients. It is intriguing that while we did not found such mutations in sporadic parathyroid carcinoma, they are likely the drivers of thyroid oncocytic cancer. MEN1 protein may play essential roles in maintaining various normal endocrine functions and it will be of great interest to define its specific role in different endocrine glands.

Targeted sequencing of the validation cohort on MiSeq instruments provided a high depth of sequence coverage of the *MEN1* gene (Figure 3.4). The majority of tumor specimens did not have matched adjacent normal tissues to distinguish somatic and germline mutations; however, given the mutational profiles of the original two tumors and the known tumor suppressive role of *MEN1* [131], we examined the samples for the presence of likely loss-of-function mutations such as nonsense SNVs and indels. We found small deletions in 3 patients diagnosed with oncocytic carcinoma in the validation cohort with mutational allele frequencies of 17.8%, 10.7% and 3.2%. Adjacent normal tissue was available for one of these 3 and it did not harbor the

MEN1 mutation. We also found small deletions in an additional 4 oncocytic thyroid carcinoma patients with allele frequency levels (1.5%, 1.4%, 1.3% and 1%) close to the sequencing technology's inherent error rate and hence we have not included these in the population's mutational rate estimate despite a previous publication reporting one of these as a somatic mutation in sporadic parathyroid adenomas [243] (Figure 3.5 and Table 3.5). All the above-mentioned deletions, including those found at low allele frequencies, were identified through both alignment-based and *de novo* assembly-based variant calling methods. No loss-of-function mutations were observed in the oncocytic cancer cell line XTC.UC1, benign Hürthle cell tumors or the normal specimens.

3.4 Discussion

Hürthle cell thyroid carcinoma, a rare entity accounting for only 2-3% of all thyroid cancers, often presents in a metastatic setting and hence has a poor prognosis [244]. Genomic studies of oncocytic thyroid cancers are limited and the molecular aberrations driving these carcinomas are not entirely understood. Activating *NRAS* mutations, frequently found in follicular thyroid cancers [244], were identified in 3 of 27 oncocytic carcinomas [21] leading perhaps to the conclusion that a subset of these malignancies might be derived from FTCs. It has been suggested that in contrast to these Hürthle cell variants of papillary or follicular carcinomas, in "true" or "primary" Hürthle cell tumors the aberrations leading to oncocytic phenotype occur prior, rather than subsequent, to neoplastic change(s) such as *NRAS* mutations [245]. It is

conceivable that these different Hürthle cell carcinoma subtypes are subjected to distinct oncogenesis mechanisms. In this study, we aimed to provide molecular profiles of those Hürthle cell malignancies that lack the most commonly mutated genes in other subtypes of thyroid cancers. We found *MEN1* mutations in three of 72 (4.2%) patients diagnosed with Hürthle cell thyroid carcinoma. In the initial two tumors that underwent whole genome sequencing, complete loss of MEN1 through the loss of the remaining wild type allele was observed. In agreement with previous reports on the tumor suppressive role of *MEN1* [131], the mutational profile of these tumors, namely homozygous and hemizygous frameshift deletions, has lent itself to a strong argument for a likely causative role of this tumor suppressor gene in Hürthle cell thyroid malignancies.

Hürthle cells are generally found in tissues with low proliferative index, which accumulate excess mitochondria over long periods of time [245]. Mutations leading to decreased mitochondrial function might cause an increase in their number to compensate for the loss of the machinery indispensable for cellular energy production, hence causing the granular appearance of oncocytic tumors [20,245,246]. Disruptive mutations of mitochondrial DNA (mtDNA) have been described in Hürthle cell thyroid cancers [246]; however, mtDNA alterations are not restricted to tumors of the thyroid and are found in a variety of oncocytic and non-oncocytic cancers [245]. It is unclear what function, if any, these mutations might play in initiating a state of malignancy. Aberrations of the energy-producing organelles and their decreased efficiency seem unlikely to have a causative role in oncogenesis; they could,

however, serve as risk factors. We found considerable copy number alterations in the genomes of both tumors, revealing extensive regions of either hemizygosity or copy-neutral LOH. Haploidization, in some cases followed by endoreduplication, has previously been reported in recurrent oncocytic follicular carcinomas leading to the hypothesis that mitochondrial mutations may lead to loss of large regions of the genome as an energy-conserving mechanism and a survival mode [247]. Mitochondrial aberrations followed by haploidization are thus likely to provide a ripe opportunity for the second hit in a tumor suppressor gene, such as *MEN1*, to pave the way for the initiation of tumorigenesis.

Germline loss-of-function mutations of *MEN1* are recognized as the single predisposing event in both familial and sporadic cases of multiple endocrine neoplasia type 1 (MEN1) disorder. Clinical features of MEN1 can be diverse. Affected individuals develop tumors of two or more endocrine organs with the majority found in the parathyroid, pancreatic islets, duodenal endocrine cells and anterior pituitary; although most are benign nodules, some, particularly tumors of the pancreas, thymus and bronchi, can become metastatic [131,248]. A wide array of mutations including frame-shift, nonsense, missense and in-frame deletions throughout *MEN1*, with no distinct hotspot mutations, along with loss of the wild type allele have been identified in association with this syndrome [131,132,249]. Moreover, somatic inactivating mutations of both *MEN1* alleles are found in sporadic endocrine tumors unrelated to multiple endocrine neoplasia type 1 disorder; these include benign parathyroid tumors [206], carcinoids tumors of the lung [250], gastrinomas and insulinomas [251].

In contrast to its prominent role in benign endocrine tumorigenesis, *MEN1* has not been implicated as the main driver of malignancy in any cancer type. A query of the cBioPortal database [252] which contains mutational data from several large-scale studies including published and ongoing work from The Cancer Genome Atlas project revealed a low mutation rate in *MEN1* (Figure 3.6). To date, the only malignancy with higher *MEN1* mutation rate than that found by the current study is adrenocortical carcinoma, where approximately 9% of tumors have *MEN1* mutations. This query also included 401 cases of papillary thyroid carcinoma from The Cancer Genome Atlas project, where only 1 patient was found to have a *MEN1* mutation. Pathogenic implications of *MEN1* mutations in thyroid tumorigenesis have been uncertain. Thyroid adenomas can accompany other endocrine tumors in a small percentage of MEN1 cases [131] and there are three case reports of MEN1 patients with papillary thyroid carcinoma diagnosis [253-255]. However, all three found the carcinoma to be a separate entity from the MEN1-related endocrine tumors. No mutations and/or LOH of *MEN1* were observed indicating that this gene is not etiologically related to papillary thyroid carcinomas [254]. Kim *et al* have described a patient who presented with several MEN1-related clinical features along with Hürthle cell thyroid carcinoma but had no germline mutations of *MEN1* [256]. A report of an atypical MEN1 patient with Hürthle cell adenoma by Pinna *et al* identified a germline heterozygote missense mutation in *MEN1* [257]. However, such mutational profile does not match that of an expected loss-of-function mutation in a tumor suppressor gene.

MEN1 encodes for a 615 amino acid-long tumor suppressor protein called menin. Although ubiquitously expressed, the exact mechanism by which this nuclear protein [258] leads to uncontrolled growth and division of mostly endocrine cells is not understood. *MEN1* is involved in several key cellular processes. This gene plays a role in transcriptional regulation by control of chromatin remodeling through interaction with histone deacetylases [214] and in complex with MLL/SET histone methyltransferases [215,216]; these are possibly accomplished by direct interaction of MEN1 with several transcription factors [215,249]. This protein is also found to interact with the tumor metastasis suppressor NM23 and to bind *hTERT* promoter and directly repress its expression [215]. Menin can also inhibit cell proliferation [259] and induce apoptosis [260]. It maintains DNA integrity and is involved in DNA damage repair pathways [261]. Chromosomal instability due to loss of MEN1 might have resulted in the vast amount of copy number changes observed in the genomic datasets (Figure 3.1). Loss of chromosomal integrity, copy number alterations and premature centromere division are observed in MEN1 patients harboring *MEN1* mutations but not in unaffected individuals or those affected but with wild type *MEN1* [262]. Although, no genotype-phenotype correlations have been established in MEN1 patients with *MEN1* mutations [249], the identified mutations in this study could shed light on the protein domains important for initiation of malignancy.

The current study is perhaps underestimating the prevalence of *MEN1* mutations in the oncocytic thyroid carcinoma population. Intra-tumor heterogeneity, frequently observed in most cancers but specifically in oncocytic carcinomas [247] and low tumor content, a common

characteristic of tissue cores obtained from FFPE specimens, might have resulted in underestimating *MEN1* mutation rate. In addition, we cannot rule out the presence of UTR, intronic or promoter alterations or epigenomic silencing of the gene. Nonetheless, this study implicates *MEN1* in the pathogenesis of a subset of Hürthle cell thyroid carcinomas. Further mutational analyses in this rare cancer type, preferably using micro-dissected regions of flash frozen tissues, are warranted and promise to aid in unraveling the mechanism of disease initiation and progression.

Table 3.1 Sequencing libraries read statistics

	Total Number of Reads	Number of Aligned Reads	Average Coverage
Primary tumor genome	1472791858	1093559232	37.8
Blood genome	1632289136	1285598519	43.9
Metastatic tumor genome	1462312890	1142572224	39
Blood genome	1685907542	1289286505	44.1

Table 3.2 List of primers used for the validation experiment

Targeted Region (build hg19)	Amplicon Size (bp)	Forward Primer Sequence	Reverse Primer Sequence
chr11:64571741-64571986	246	GCTCAGAGTTGGGGGACTAAGG	CACTTTCCAGAGTGAGAAGATGAAGG
chr11:64571843-64572091	249	GTGTAGTCACTAGGGGTGGACACTTTC	GAAGCCTCCTGGGACTGTCG
chr11:64572036-64572270	235	CCGTGCTGCCACCTTCAG	ATAGTGAGCCGAGAGGCCGAG
chr11:64572133-64572324	192	CTTGTCCAGTGCTGGCTTCTTG	AACCTTGCTCTCACCTTGCTCTC
chr11:64572427-64572646	220	CTGGGCCAGAAAAGTCTGACAAG	CTCCAGGACCCTGAGTGCTTC
chr11:64572523-64572726	204	CTAGGGACTGCACAAGAAAGGTGG	CTCTGCTAAGGGGTGAGTAAGAGACTG
chr11:64573066-64573283	218	AGGTGGGAGGCTGGACACAG	AGACCCCTTCAGACCCTACAGAGAC
chr11:64573633-64573865	233	GACGAGGGTGGTTGGAAACTG	GATCCTCTGCCTCACCTCCATC
chr11:64574388-64574597	210	AACACACAAAGTTCTCTTCTCATCTGC	GCAGCCTGAATTATGATCCTTTCC
chr11:64574568-64574729	162	TACCTAGGAAAGGATCATAATTCAGGC	CTGTTCCGTGGCTCATAACTCTCTC
chr11:64575005-64575245	241	CCATTGGCTCAGCCCTCAC	GAAGACAGAAGAGCCCCTTTTCC
chr11:64575319-64575491	173	TACTACAGTATGAAGGGGACAAGGCTG	GCCCTGTCTGAGGATCATGC
chr11:64575425-64575612	188	AGGTGACCTCAGCTGTCTGCTC	AAGCACAGAGGACCCTCTTTCATTAC
chr11:64577042-64577267	226	TCACAAGGCTTACAGTTCTTAAAAGGG	CTATCCTCGAGAAGGGGGTGTCTC
chr11:64577205-64577450	246	CATATGACATCGGAGACCTTCTTCAC	GGAGCATTTTCTGGCTGTCAAC
chr11:64577252+64577487	236	CCCTTCTCGAGGATAGAGGGACAG	CGGACCTGGTGCTCCTTTC
chr11:64577438-64577682	245	CAGAAAATGCTCCACGAAGCC	GTGGAACCTTAGCGGACCCTG

Table 3.3 Tumor 1 (primary tumor) somatic SNVs & indels

Chr	Position	Reference Allele	Alternate Allele	dbSNP/ COSMIC ID	Effect Type	AA Change	Gene	EnsEMBL Gene ID
1	151131447	C	A		Non-synonymous	p.Q92K	TNFAIP8L2	ENSG00000163154
1	152287809	G	A	rs138819199, COSM170933	Non-synonymous	p.R42W	FLG	ENSG00000143631
1	153946171	G	A		Non-synonymous	p.E325K	CREB3L4	ENSG00000143578
2	114256859	C	T	rs189095552, COSM228316	Non-synonymous	p.P9L	FOXD4L1	ENSG00000184492
2	15770132	G	C		Non-synonymous	p.E664Q	DDX1	ENSG00000079785
2	179476159	C	A		Non-synonymous	p.A14365S	TTN	ENSG00000155657
2	210860209	C	T		Non-synonymous	p.H3199Y	UNC80	ENSG00000144406
3	180693943	G	A	rs149239435	Non-synonymous	p.E492K	FXR1	ENSG00000114416
3	47161718	G	T		Non-synonymous	p.P1470T	SETD2	ENSG00000181555
3	53752750	C	T		Non-synonymous	p.S507F	CACNA1D	ENSG00000157388
5	114860337	G	C		Non-synonymous	p.L508V	FEM1C	ENSG00000145780
6	107070781	T	C		Non-synonymous	p.E113G	RTN4IP1	ENSG00000130347
6	168276091	G	A		Non-synonymous	p.E218K	MLLT4	ENSG00000130396
8	144547945	G	C		Non-synonymous	p.S750C	ZC3H3	ENSG00000014164
8	30701727	C	G	COSM1229015	Non-synonymous	p.E1603Q	TEX15	ENSG00000133863
9	127215413	G	A		Non-synonymous	p.C146Y	GPR144	ENSG00000180264
11	21250971	G	T		Non-synonymous	p.G507V	NELL1	ENSG00000165973
11	61290678	TCA	-		Codon deletion	p.MK325K	SYT7	ENSG00000011347
11	64571880	T	-		Frameshift	p.M592fs	MEN1	ENSG00000133895
13	37602394	G	A		Stop gained	p.Q349*	FAM48A	ENSG00000102710
14	105359869	C	T		Non-synonymous	p.R1350C	KIAA0284	ENSG00000099814
14	32562460	C	T	COSM1369573	Non-synonymous	p.S862L	ARHGAP5	ENSG00000100852
15	75982058	T	A	rs147116973, COSM1317755	Non-synonymous	p.R450W	CSPG4	ENSG00000173546
16	2028379	G	C		Non-synonymous	p.V734L	TBL3	ENSG00000183751
16	28331953	C	T		Non-synonymous	p.A329V	SBK1	ENSG00000188322
16	77246033	G	A		Non-synonymous	p.E144K	SYCE1L	ENSG00000205078
17	15469301	G	A		Non-synonymous	p.P93L	CDRT1	ENSG00000181464
17	16455231	G	T		Non-synonymous	p.T742N	ZNF287	ENSG00000141040
17	38176635	G	C		Non-synonymous	p.F163L	MED24	ENSG00000008838
18	50923805	C	T		Non-synonymous	p.T939M	DCC	ENSG00000187323
19	1061865	G	A		Non-synonymous	p.E1850K	ABCA7	ENSG00000064687
19	13264012	G	C		Non-synonymous	p.Q4H	IER2	ENSG00000160888
19	17922780	G	A		Non-synonymous	p.R323Q	B3GNT3	ENSG00000179913
19	38993270	G	A		Non-synonymous	p.D2580N	RYR1	ENSG00000196218
19	407478	G	A		Non-synonymous	p.A295V	C2CD4C	ENSG00000183186
19	48248829	G	C		Non-synonymous	p.G5R	GLTSCR2	ENSG00000105373

Chr	Position	Reference Allele	Alternate Allele	dbSNP/ COSMIC ID	Effect Type	AA Change	Gene	EnsEMBL Gene ID
19	51631292	G	A		Non-synonymous	p.V368I	SIGLEC9	ENSG00000129450
19	53668842	G	A		Stop gained	p.Q301*	ZNF665	ENSG00000197497
19	56369839	C	G		Non-synonymous	p.D360E	NLRP4	ENSG00000160505
19	8399903	C	G		Non-synonymous	p.D270H	KANK3	ENSG00000186994
20	62562911	G	A		Non-synonymous	p.G196E	DNAJC5	ENSG00000101152
22	18165995	T	C	rs2587070	Non-synonymous	p.I46T	BCL2L13	ENSG00000099968
22	29687589	G	C		Splice Site Donor		EWSR1	ENSG00000182944
22	50521523	G	A		Non-synonymous	p.A86V	MLC1	ENSG00000100427
X	14863335	G	A		Stop gained	p.Q524*	FANCB	ENSG00000181544
X	20028971	C	T	COSM1119018	Non-synonymous	p.E717K	MAP7D2	ENSG00000184368
X	48558600	G	A		Non-synonymous	p.R95K	SUV39H1	ENSG00000101945
X	67937574	A	C		Non-synonymous	p.N193T	STARD8	ENSG00000130052
X	90691444	C	T	COSM1469839	Non-synonymous	p.R290W	PABPC5	ENSG00000174740
MT	10726	G	A		Non-synonymous	p.G86D	MT-ND4L	ENSG00000212907
MT	5224	G	A		Non-synonymous	p.G252E	MT-ND2	ENSG00000198763

Table 3.4 Tumor 2 (metastatic tumor) somatic SNVs & indels

Chr	Position	Reference Allele	Alternate Allele	dbSNP/ COSMIC ID	Effect Type	AA Change	Gene	EnsEMBL Gene ID
1	107867300	A	G		Non-synonymous	p.T215A	NTNG1	ENSG00000162631
1	152186042	A	G	rs12751022, COSM1127393	Non-synonymous	p.L2688S	HRNR	ENSG00000197915
1	171756912	G	A		Non-synonymous	p.R228Q	METTL13	ENSG00000010165
1	175957501	C	-		Frameshift	p.C632fs	RFWD2	ENSG00000143207
1	177030270	C	A		Stop gained	p.E139*	ASTN1	ENSG00000152092
1	1961632	G	A		Non-synonymous	p.A424T	GABRD	ENSG00000187730
1	203452824	G	A		Non-synonymous	p.R171Q	PRELP	ENSG00000188783
1	223568289	G	A		Non-synonymous	p.R491H	C1orf65	ENSG00000178395
1	23111042	G	A		Non-synonymous	p.R95H	EPHB2	ENSG00000133216
1	2428962	A	AT		Frameshift	p.E713fs	PLCH2	ENSG00000149527
1	32263872	C	T	rs199645489	Non-synonymous	p.R694Q	SPOCD1	ENSG00000134668
2	103149136	C	CA		Frameshift	p.Q796fs	SLC9A4	ENSG00000180251
2	190593424	A	T		Non-synonymous	p.M1024L	ANKAR	ENSG00000151687
2	239010763	C	T		Non-synonymous	p.A159V	ESPNL	ENSG00000144488
2	242098728	T	C		Non-synonymous	p.I135T	PPP1R7	ENSG00000115685
2	242186517	TGTT	-	COSM1637486	Frameshift	p.K590fs	HDLBP	ENSG00000115677
2	47698146	AG	-	rs63751463	Frameshift	p.E569fs	MSH2	ENSG00000095002
3	48464254	G	A		Stop gained	p.Q404*	PLXNB1	ENSG00000164050
4	145040874	G	T	rs56077914	Non-synonymous	p.S66Y	GYPA	ENSG00000170180
4	42553229	C	T		Non-synonymous	p.V515M	ATP8A1	ENSG00000124406
5	131008245	TCTC	-		Frameshift	p.R602fs	FNIP1	ENSG00000217128
5	154300947	A	G		Non-synonymous	p.V473A	GEMIN5	ENSG00000082516
5	179291023	G	A	rs200438741	Non-synonymous	p.R1043C	TBC1D9B	ENSG00000197226
5	179545655	C	T	COSM172724	Non-synonymous	p.R346H	RASGEF1C	ENSG00000146090
5	180377473	A	T		Non-synonymous	p.I514F	BTNL8	ENSG00000113303
5	38425171	T	G		Non-synonymous	p.F362C	EGFLAM	ENSG00000164318
5	59895033	G	A		Non-synonymous	p.R433C	DEPDC1B	ENSG00000035499
5	65349266	A	-		Frameshift	p.N708fs	ERBB2IP	ENSG00000112851
6	108066295	G	T		Non-synonymous	p.N180K	SCML4	ENSG00000146285
6	157099985	G	A		Non-synonymous	p.G308S	ARID1B	ENSG00000049618
6	158923780	C	T		Non-synonymous	p.R1029W	TULP4	ENSG00000130338
6	161470602	G	A		Non-synonymous	p.R433Q	MAP3K4	ENSG00000085511
6	33289553	GCTTCCTC TG	-		Frameshift	p.A59fs	DAXX	ENSG00000204209
6	36168806	A	G		Non-synonymous	p.N236S	BRPF3	ENSG00000096070
6	637806	G	A		Stop gained	p.R5*	EXOC2	ENSG00000112685
6	72806843	G	A		Non-synonymous	p.R146H	RIMS1	ENSG00000079841
7	106301314	CGCCGC	-		Codon deletion	p.RRR8R	CCDC71L	ENSG00000253276

Chr	Position	Reference Allele	Alternate Allele	dbSNP/ COSMIC ID	Effect Type	AA Change	Gene	EnsEMBL Gene ID
7	149545229	G	A	rs370372278	Non-synonymous	p.R216Q	ZNF862	ENSG00000106479
7	150884183	C	T		Non-synonymous	p.G12E	ASB10	ENSG00000146926
7	5541344	C	T		Non-synonymous	p.G186S	FBXL18	ENSG00000155034
7	73922429	C	T	rs201130740, COSM1091662	Non-synonymous	p.R7C	GTF2IRD1	ENSG00000006704
7	88963606	C	A		Non-synonymous	p.A437E	ZNF804B	ENSG00000182348
7	99689305	G	A		Non-synonymous	p.G293S	COPS6	ENSG00000168090
8	106814997	C	T	rs200840311	Non-synonymous	p.P896L	ZFPM2	ENSG00000169946
8	121237414	GAA	-		Codon deletion	p.E609del	COL14A1	ENSG00000187955
8	144942300	C	T	rs191533123	Non-synonymous	p.G1708S	EPPK1	ENSG00000227184
8	30546709	G	A	rs144377982	Non-synonymous	p.P337L	GSR	ENSG00000104687
8	75926268	G	A	rs150601296	Non-synonymous	p.C186Y	CRISPLD1	ENSG00000121005
9	125140838	G	A		Non-synonymous	p.R113H	PTGS1	ENSG00000095303
9	129870571	C	T		Non-synonymous	p.R147H	ANGPTL2	ENSG00000136859
9	134504534	G	A		Non-synonymous	p.A266V	RAPGEF1	ENSG00000107263
9	15191186	A	G		Splice Site Donor		TTC39B	ENSG00000155158
9	38577960	T	-		Frameshift	p.K811fs	ANKRD18A	ENSG00000180071
10	100148175	C	T		Non-synonymous	p.M461I	PYROXD2	ENSG00000119943
10	102766392	C	T	rs191391710	Non-synonymous	p.R493W	LZTS2	ENSG00000107816
10	116730191	G	GC	COSM1345961	Frameshift	p.L199fs	TRUB1	ENSG00000165832
10	118320015	C	A		Non-synonymous	p.S383Y	PNLIP	ENSG00000175535
10	118618628	A	-		Frameshift	p.K207fs	ENO4	ENSG00000188316
10	127455291	C	CT		Frameshift	p.K550fs	MMP21	ENSG00000154485
10	129905384	G	A		Non-synonymous	p.R1214W	MKI67	ENSG00000148773
10	13213237	G	A	rs374155592	Non-synonymous	p.R108Q	MCM10	ENSG00000065328
10	27369086	A	-		Frameshift	p.F254fs	ANKRD26	ENSG00000107890
10	31809248	C	T	COSM185473	Non-synonymous	p.R329W	ZEB1	ENSG00000148516
10	51465612	C	T		Non-synonymous	p.V282I	AGAP7	ENSG00000204169
10	5494839	C	T	rs373461269	Stop gained	p.R130*	NET1	ENSG00000173848
10	79572115	C	T	COSM1349235	Non-synonymous	p.R1349H	DLG5	ENSG00000151208
11	118627930	TCT	-		Codon deletion	p.KI353I	DDX6	ENSG00000110367
11	120673479	C	T	rs145641056	Non-synonymous	p.R54C	GRIK4	ENSG00000149403
11	126162666	G	A	rs199545047	Non-synonymous	p.R121Q	TIRAP	ENSG00000150455
11	130284488	C	T		Non-synonymous	p.G502R	ADAMTS8	ENSG00000134917
11	32676507	T	-	COSM1353548	Frameshift	p.A220fs	CCDC73	ENSG00000186714
11	45241261	G	A		Non-synonymous	p.R232H	PRDM11	ENSG00000019485
11	45832632	G	A		Non-synonymous	p.G268S	SLC35C1	ENSG00000181830
11	60665363	G	C		Non-synonymous	p.Q458E	PRPF19	ENSG00000110107
11	64572093	G	-	COSM1355794	Frameshift	p.R521fs	MEN1	ENSG00000133895
12	108140181	C	T		Non-synonymous	p.A383T	PRDM4	ENSG00000110851

Chr	Position	Reference Allele	Alternate Allele	dbSNP/ COSMIC ID	Effect Type	AA Change	Gene	EnsEMBL Gene ID
12	122018739	C	CT		Frameshift	p.K27fs	KDM2B	ENSG00000089094
12	13061424	C	A		Non-synonymous	p.L81I	GPRC5A	ENSG00000013588
12	39726186	G	A	rs142292357, COSM170145	Stop gained	p.R948*	KIF21A	ENSG00000139116
12	39760190	C	T	COSM938947	Non-synonymous	p.E289K	KIF21A	ENSG00000139116
12	50480543	G	A		Non-synonymous	p.R138H	SMARCD1	ENSG00000066117
12	52184279	G	A		Non-synonymous	p.R1465H	SCN8A	ENSG00000196876
12	54448978	A	G		Non-synonymous	p.T262A	HOXC4	ENSG00000198353
12	56514899	C	T		Stop gained	p.R185*	ZC3H10	ENSG00000135482
12	9251262	G	A	COSM1181114	Non-synonymous	p.R598C	A2M	ENSG00000175899
13	49075950	G	A	rs376690025	Non-synonymous	p.P391L	RCBTB2	ENSG00000136161
14	23896932	C	T		Non-synonymous	p.G584S	MYH7	ENSG00000092054
14	70926261	T	-	COSM1370843	Frameshift	p.L684fs	ADAM21	ENSG00000139985
14	76241852	G	A	COSM1371183	Non-synonymous	p.R721Q	TTLL5	ENSG00000119685
15	56134324	C	T		Non-synonymous	p.R549Q	NEDD4	ENSG00000069869
15	77236167	A	C		Non-synonymous	p.E172D	RCN2	ENSG00000117906
16	1306802	A	G	rs2401930	Non-synonymous	p.I87V	TPSD1	ENSG00000095917
16	16225734	G	A		Non-synonymous	p.R1303Q	ABCC1	ENSG00000103222
16	1869125	C	T	rs200743955	Non-synonymous	p.V130M	HAGH	ENSG00000063854
16	29818842	G	-		Frameshift	p.G224fs	MAZ	ENSG00000103495
16	31336597	G	A		Non-synonymous	p.V793M	ITGAM	ENSG00000169896
16	85682289	A	AC	COSM1380255	Frameshift	p.V123fs	KIAA0182	ENSG00000131149
17	10404654	G	A	rs192282019, COSM1609829	Non-synonymous	p.R1171W	MYH1	ENSG00000109061
17	17931610	G	A		Non-synonymous	p.A87V	ATPAF2	ENSG00000171953
17	2579802	A	-		Frameshift	p.S304fs	PAFAH1B1	ENSG00000007168
17	36830102	G	-		Frameshift	p.P216fs	C17orf96	ENSG00000179294
17	4098254	C	T	rs375434076	Non-synonymous	p.R464Q	ANKFY1	ENSG00000185722
17	41245098	C	A		Non-synonymous	p.G770V	BRCA1	ENSG00000012048
17	4720549	G	A	rs371269918	Non-synonymous	p.V604I	PLD2	ENSG00000129219
17	4793023	C	T		Non-synonymous	p.R438W	MINK1	ENSG00000141503
17	48632895	C	T	rs145914453	Non-synonymous	p.R701C	SPATA20	ENSG00000006282
17	7221410	C	T	COSM283215	Non-synonymous	p.R1345H	NEURL4	ENSG00000215041
17	72878722	C	T		Non-synonymous	p.R159H	FADS6	ENSG00000172782
17	73732682	G	A		Non-synonymous	p.V633M	ITGB4	ENSG00000132470
17	76167829	G	A		Stop gained	p.W192*	SYNGR2	ENSG00000108639
17	80197898	C	T	rs373977034	Non-synonymous	p.R413H	CSNK1D	ENSG00000141551
18	47462659	G	A	rs121908105	Non-synonymous	p.R656C	MYO5B	ENSG00000167306
19	1229906	G	A		Non-synonymous	p.R484C	C19orf26	ENSG00000099625
19	12774217	C	T		Non-synonymous	p.D275N	MAN2B1	ENSG00000104774

Chr	Position	Reference Allele	Alternate Allele	dbSNP/ COSMIC ID	Effect Type	AA Change	Gene	EnsEMBL Gene ID
19	19655611	G	A		Non-synonymous	p.E753K	CILP2	ENSG00000160161
19	21992318	AA	-		Frameshift	p.F174fs	ZNF43	ENSG00000198521
19	35940962	C	T		Stop gained	p.Q116*	FFAR2	ENSG00000126262
19	41081377	G	A		Non-synonymous	p.A2533T	SPTBN4	ENSG00000160460
19	4179203	G	A		Non-synonymous	p.T92M	SIRT6	ENSG00000077463
19	42224095	A	G		Non-synonymous	p.N580S	CEACAM5	ENSG00000105388
19	47735847	T	A		Non-synonymous	p.M5L	BBC3	ENSG00000105327
19	50775820	G	A	rs374956489	Non-synonymous	p.R1067H	MYH14	ENSG00000105357
19	56113688	G	-		Frameshift	p.V72fs	ZNF524	ENSG00000171443
19	56424537	G	A		Stop gained	p.Q216*	NLRP13	ENSG00000173572
19	56443533	G	-		Frameshift	p.Q49fs	NLRP13	ENSG00000173572
19	56671223	G	A		Non-synonymous	p.G212R	ZNF444	ENSG00000167685
19	58982236	G	C	rs145109076, COSM418737	Non-synonymous	p.R126P	ZNF324	ENSG00000083812
19	58982527	G	C		Non-synonymous	p.R223P	ZNF324	ENSG00000083812
19	6429784	G	A		Non-synonymous	p.A192V	SLC25A41	ENSG00000181240
19	7543215	G	A	rs138259370	Non-synonymous	p.T159M	PEX11G	ENSG00000104883
19	9018191	T	C		Non-synonymous	p.M12583V	MUC16	ENSG00000181143
20	43837307	C	T		Stop gained	p.R457*	SEMG1	ENSG00000124233
20	44444208	C	T		Non-synonymous	p.S43L	UBE2C	ENSG00000175063
20	45633591	C	T		Non-synonymous	p.R56C	EYA2	ENSG00000064655
20	49226196	G	A		Stop gained	p.Q160*	FAM65C	ENSG00000042062
20	57484421	G	A	rs121913495, COSM27895	Non-synonymous	p.R201H	GNAS	ENSG00000087460
20	58547177	C	T	rs112379790, COSM192808	Non-synonymous	p.T131M	CDH26	ENSG00000124215
20	60942085	A	C		Non-synonymous	p.S73A	LAMA5	ENSG00000130702
21	15013879	A	G		Non-synonymous	p.N583D	POTED	ENSG00000166351
21	44841004	C	T		Non-synonymous	p.V212M	SIK1	ENSG00000142178
22	22277475	G	A		Non-synonymous	p.P452L	PPM1F	ENSG00000100034
22	25750724	T	G		Non-synonymous	p.H165P	LRP5L	ENSG00000100068
22	31999747	G	A		Non-synonymous	p.R608Q	SFI1	ENSG00000198089
22	36900599	G	A		Non-synonymous	p.R248C	FOXRED2	ENSG00000100350
22	38121787	C	T	rs201142573	Non-synonymous	p.S1075L	TRIOBP	ENSG00000100106
22	39770342	G	A	rs145517070	Non-synonymous	p.G41S	SYNGR1	ENSG00000100321
22	50615970	G	A		Non-synonymous	p.V277M	PANX2	ENSG00000073150
22	50898007	C	T		Non-synonymous	p.V1194M	SBF1	ENSG00000100241
X	103495261	C	T		Non-synonymous	p.R290H	ESX1	ENSG00000123576
X	118774741	G	A		Non-synonymous	p.P234L	SEPT6	ENSG00000125354
X	152818575	G	A	COSM1117369	Non-synonymous	p.E636K	ATP2B3	ENSG00000067842

Chr	Position	Reference Allele	Alternate Allele	dbSNP/ COSMIC ID	Effect Type	AA Change	Gene	EnsEMBL Gene ID
X	153050278	G	GC		Frameshift	p.A443fs	SRPK3	ENSG00000184343
X	153175478	C	T	COSM1117650	Non-synonymous	p.A740T	ARHGAP4	ENSG00000089820
X	153676859	G	A		Non-synonymous	p.G157E	FAM50A	ENSG00000071859
X	1719571	A	T		Non-synonymous	p.Q391L	AKAP17A	ENSG00000197976
X	50055632	G	T		Non-synonymous	p.E1141D	CCNB3	ENSG00000147082
X	85212923	G	A	rs132630266, COSM1201068	Stop gained	p.R293*	CHM	ENSG00000188419

Table 3.5 Clinical characteristics of 7 patients from the validation cohort with MEN1 mutations

Patient ID	MEN1 mutation & Allele Frequency	Outcome	Tumor Type	Gender	Age at Surgery	Date of Surgery	Surgery Type	T Stage	N Stage	M Stage	Stage	Additional Therapies
2370	p.EAAEAE468E AF=10.7%	Alive with stable disease	widely invasive	M	57	9-Jul-02	Total TDX	T3	N0	M0	III	RAI
1193	p.V178fs AF=3.2%	Alive with no evidence of disease	minimally invasive	F	48	8-Dec-08	Lobectomy	T1b	Nx	M0	X	-
4244	p.I252fs AF=17.8%	Death due to disease	widely invasive	M	61	15-Sep-89	Lobectomy	Tx	Nx	M0	X	RAI
9492	p.P498fs AF=1.3%	Death due to disease	widely invasive	F	46	12-Dec-00	Completion TDX	T4a	N1b	M0	IVA	RAI
8673	p.G111fs AF=1.4%	Dead due to other causes	widely invasive	M	76	23-Feb-05	Total TDX	T4a	Nx	M0	X	RAI
6933	p.AAEA469A AF=1%	Alive with progressive disease	widely invasive	F	53	15-Apr-05	Near total TDX	T2	Nx	M0	X	RAI
6230	p.T215fs AF=1.5%	Alive with stable disease	widely invasive	M	59	8-May-00	Near total TDX	T2	Nx	M0	X	RAI

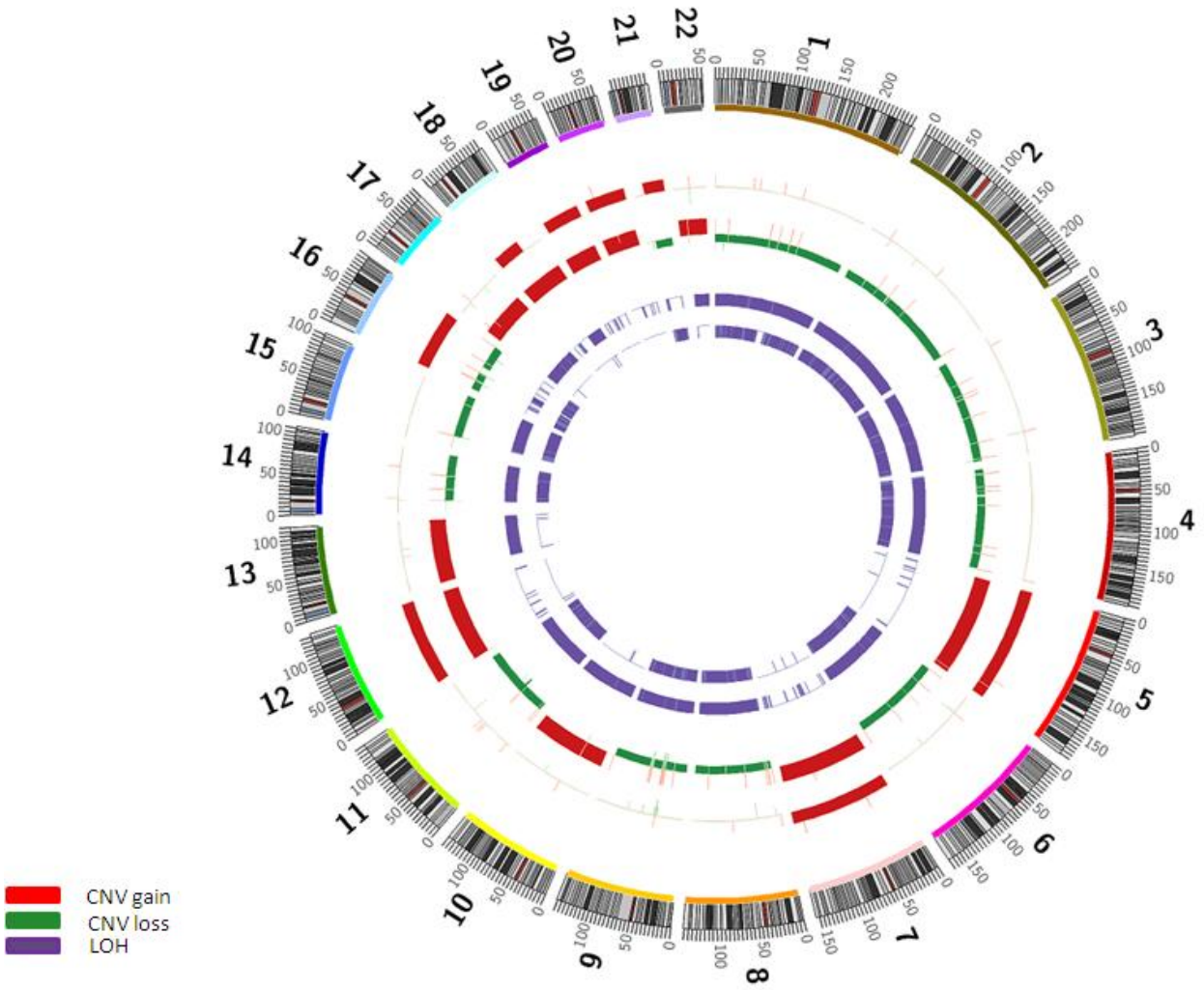

Figure 3.1 CNV and LOH regions in two Hürthle cell thyroid tumors
From the outer ring in: primary tumor CNV, (unrelated) metastatic tumor CNV, primary tumor LOH and metastatic tumor LOH. Both tumors demonstrate large regions of copy number change; the primary tumor has gained extra copies of chromosomes 5, 7, 12, 16, 18p, 19, 20, 21 and X. The metastatic tumor demonstrates a much higher number of CNV changes including one copy loss of chromosomes 1, 2, 3, 4, 6, 8, 9, 11, 14, 15, 16, 21q and X and gain of extra copies of the rest of the genome. Both tumors also show extensive regions of loss of heterozygosity

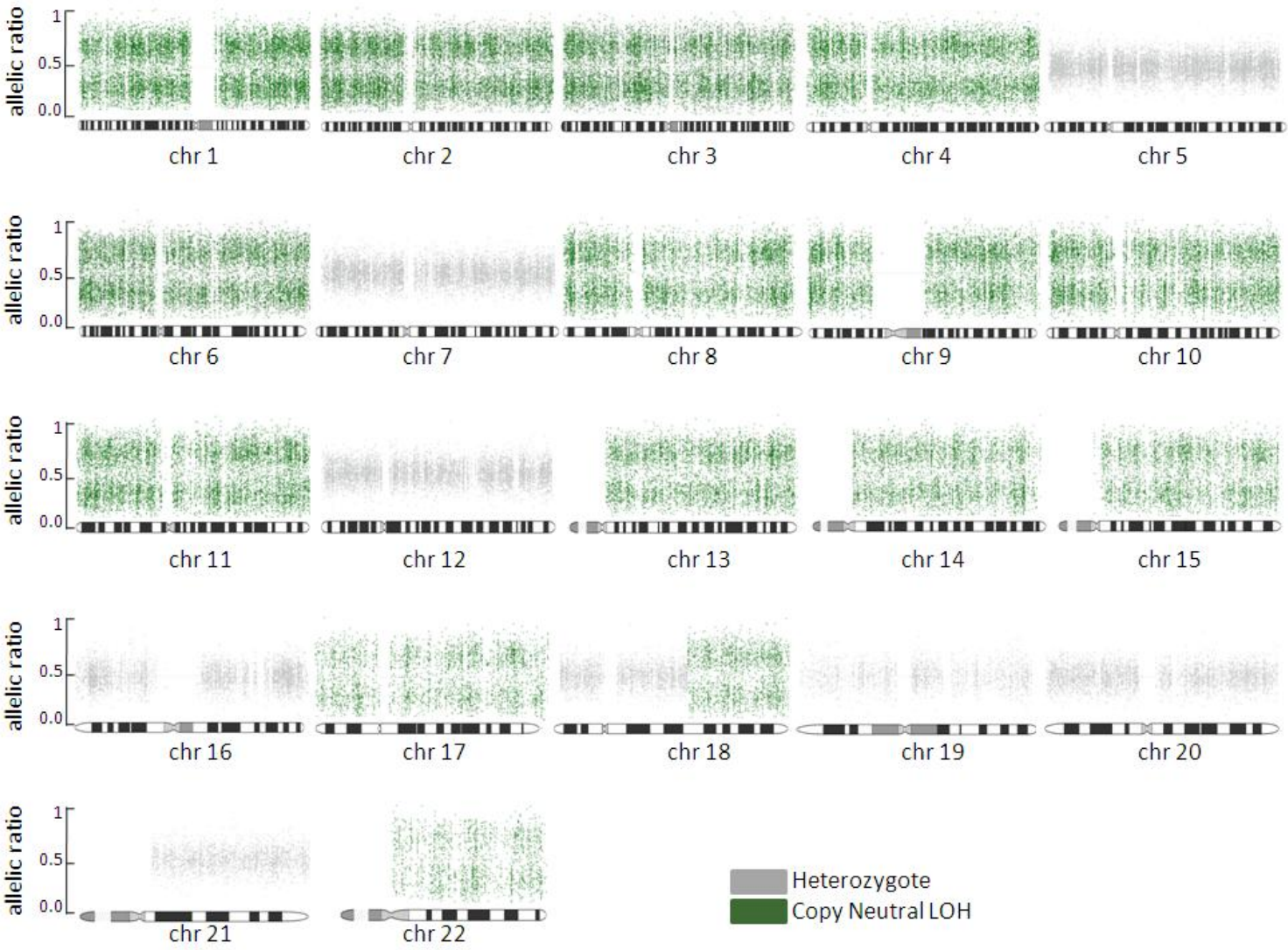

Figure 3.2 B-allele frequency plots for the primary tumor
This genome demonstrates large regions of copy-neutral loss of heterozygosity

Figure 3.3 B-allele frequency plots for the metastatic tumor
This tumor demonstrates large regions of loss of heterozygosity associated with loss of copy number

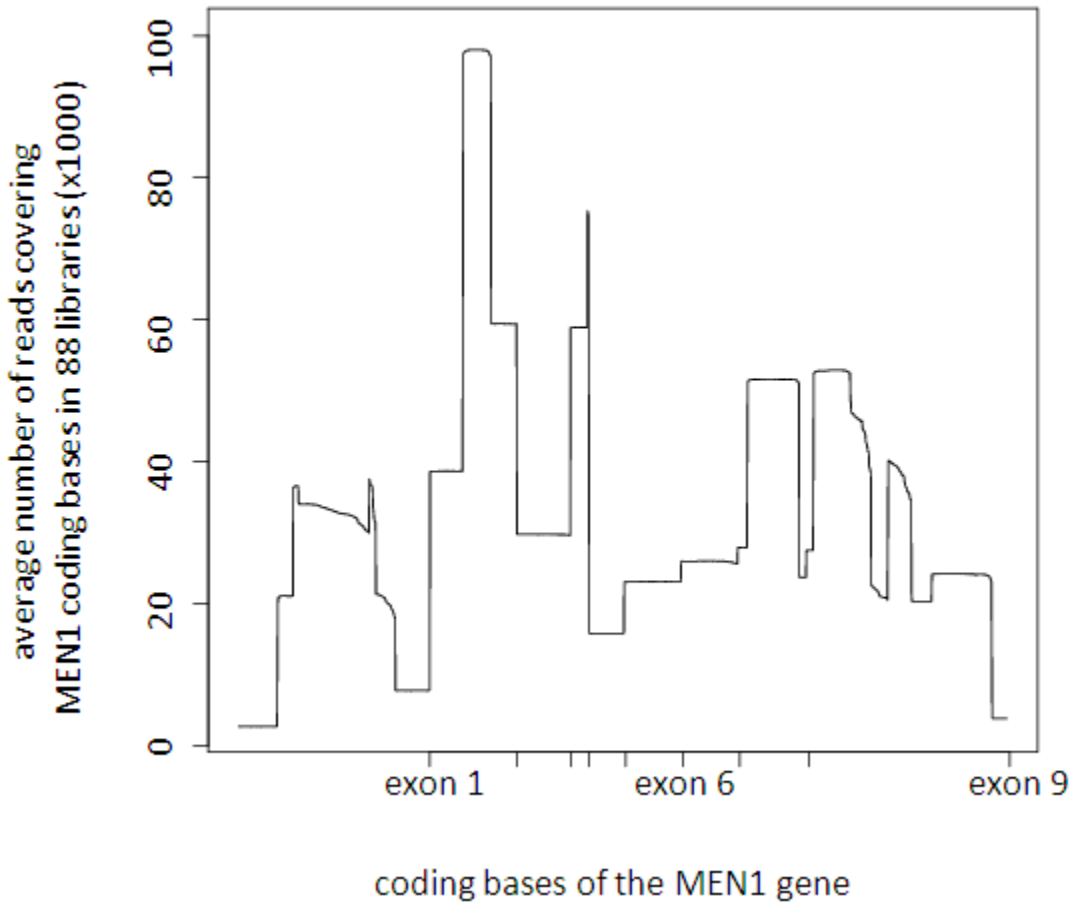

Figure 3.4 Average coverage over MEN1 coding bases in validation experiment libraries
Part of exon 1 is composed of a high GC-content region and as a result deemed more difficult to amplify (methods). Nonetheless, on average over 3,000 sequence reads were produced per base in this region

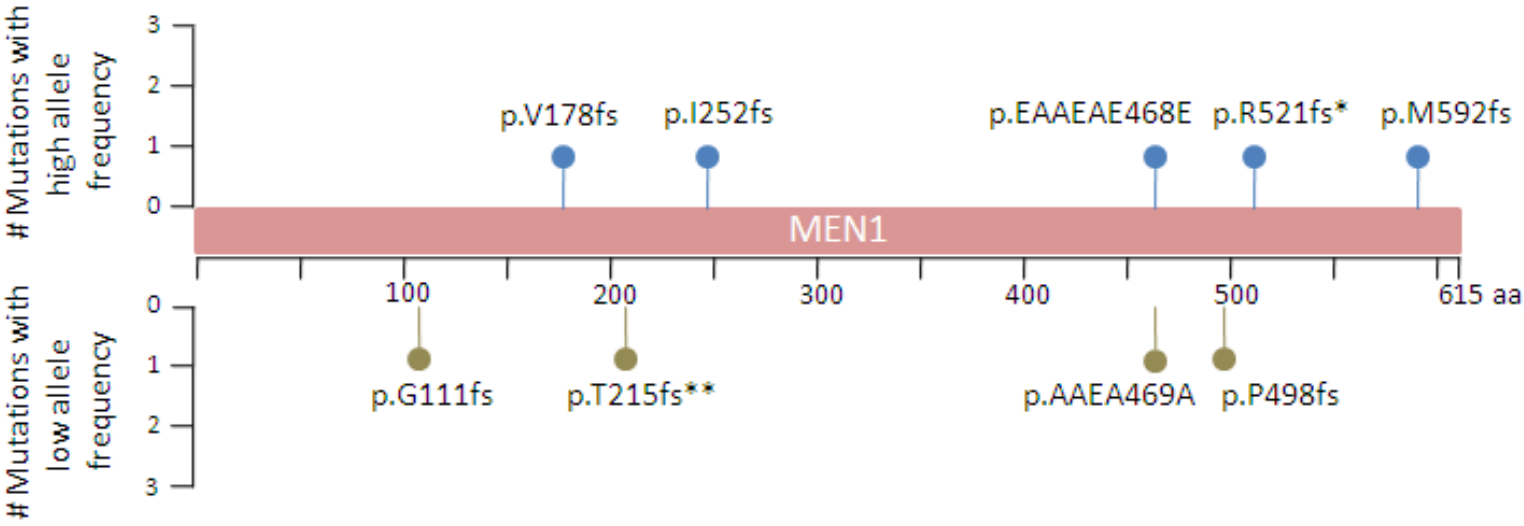

Figure 3.5 The identified mutations throughout MEN1
While 5 mutations were found at high allele frequencies and thus are high-confident calls, extra 4 mutations were found to have low allele frequencies between 1 and 2%.
* This mutation has previously been described in patients diagnosed with MEN1 disorder [242]. It has also been detected in 3 TCGA large intestine carcinoma specimens and 1 liver carcinoma (COSM1355794).
** This mutation has previously been found in sporadic parathyroid adenomas [243] (COSM255131).

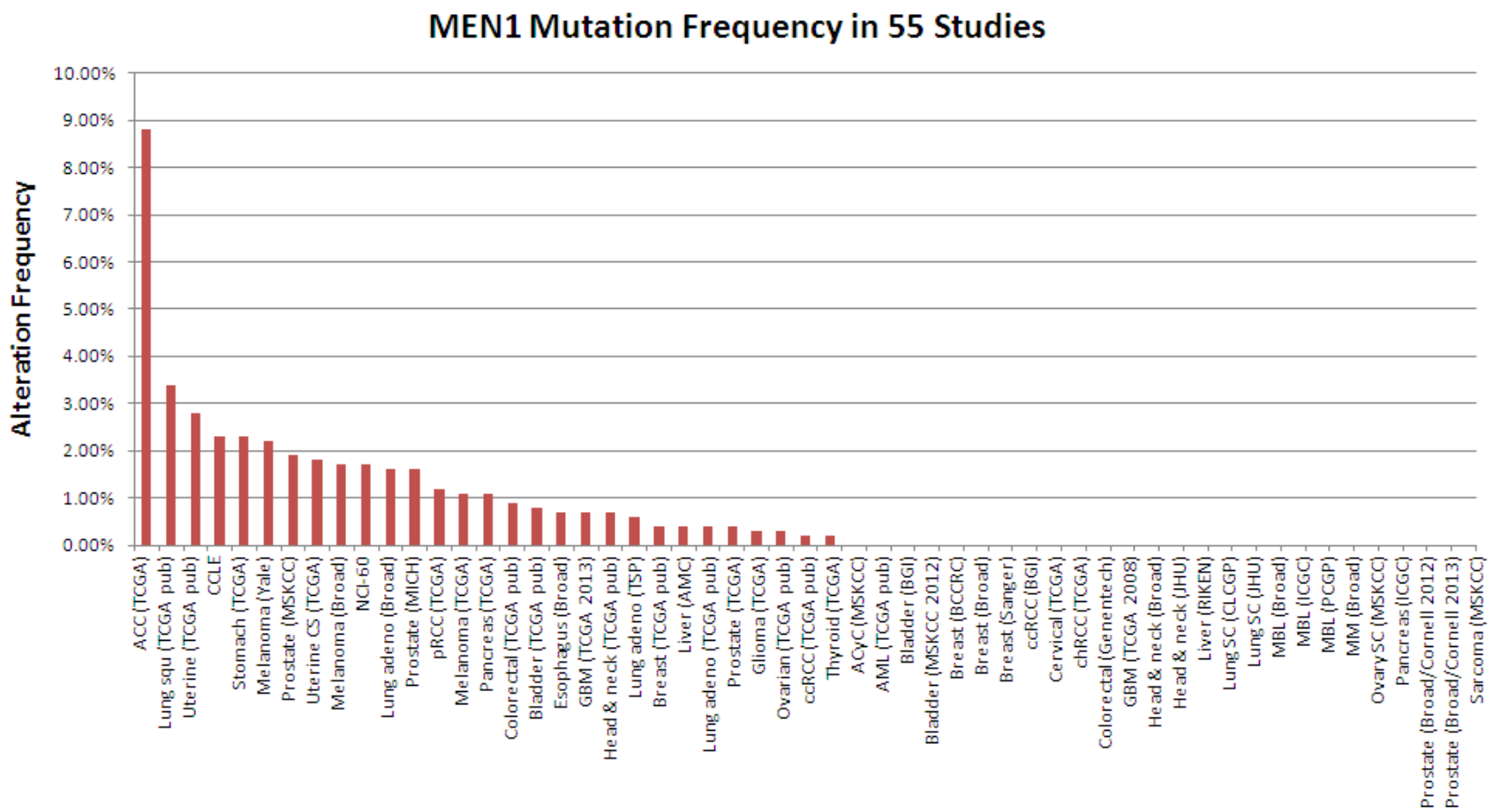

Figure 3.6 MEN1 mutation frequency in 55 cancer studies

Data was extracted from cBioPortal database on August 22, 2014. ACC: Adrenocortical Carcinoma, Lung squ: Lung Squamous Cell Carcinoma, CCLE: Cancer Cell Line Encyclopedia, Uterine CS: Uterine Carcinosarcoma, NCI-60: NCI-60 Cell Lines, Lung adeno: Lung Adenocarcinoma, pRCC: Kidney Renal Papillary Cell Carcinoma, GBM: Glioblastoma, ccRCC: Kidney Renal Clear Cell Carcinoma, ACyC: Adenoid Cystic Carcinoma, AML: Acute Myeloid Leukemia, chRCC: Kidney Chromophobe, Lung SC: Lung Squamous Cell Carcinoma, MBL: Medulloblastoma, MM: Multiple Myeloma, Ovary SC: Small Cell Carcinoma of the Ovary. The particular study associated with each disease is indicated in the parenthesis

Chapter 4: The Genomic and Transcriptomic Landscape of Anaplastic Thyroid Cancer: Implications for Therapy[4]

4.1 Introduction

Anaplastic thyroid carcinoma (ATC) is an uncommon malignancy that accounts for only 1-2% of thyroid cancers and yet it is responsible for 14-39% of all thyroid cancer related deaths [263,264]. Dedifferentiation of thyroid follicular cells in the course of tumor evolution results in this most aggressive form of thyroid cancer and one of the deadliest of all adult solid malignancies with 68.4% and 80.7% mortality rates at 6 and 12 moths, respectively [264]. A study of 516 patients from 12 population-based cancer registries recorded in the Surveillance, Epidemiology and End Results database between 1973 and 2000 found that diagnosis made before the age of 60, confined disease to the thyroid and treatment with surgical resection and external beam radiation therapy are associated with better, but still dismal, survival in ATC patients [264]. Though aggressive multimodal treatment strategies may achieve better survival for those patients who present with fewer disease risks, for those with worse prognosis and extensive local and distant involvement at diagnosis, such treatments could worsen quality of life [265]. No effective or standard therapy for the treatment of anaplastic thyroid cancer exists;

[4] A version of this chapter has been submitted for publication, and the author contributions are provided in the Preface as per the University of British Columbia PhD thesis guidelines: Katayoon Kasaian, Sam M Wiseman, Blair A Walker, Jacqueline E Schein, Yongjun Zhao, Martin Hirst, Richard A Moore, Andrew J Mungall, Marco A Marra, and Steven JM Jones. (2015). The Genomic and Transcriptomic Landscape of Anaplastic Thyroid Cancer: Implications for Therapy.

several clinical trials involving a small number of patients have failed to demonstrate any prolonged response and the use of chemotherapeutics such as doxorubicin and paclitaxel has not shown any significant survival benefits [264,265]. Multikinase inhibitors have more recently been used in the treatment of advanced and refractory thyroid cancers, and although some of these result in objective responses and can improve survival in select patients with differentiated thyroid cancers (DTC), the response of ATCs has been less consequential [263].

The rare occurrence of ATC and the rapid death and short follow-ups as a result of its aggressive progression have made it challenging to study the biology of the disease or to conduct clinical trials where responses to novel therapies can be examined [266]. Retrospective studies of small cohorts of patients have found anaplastic thyroid carcinoma to be a heterogeneous disease on the molecular level, rendering it impossible to define a common and specific route of oncogenic transformation and thus to identify effective therapeutics [267]. Mutations of various pathways including MAPK, PI3K and Wnt have been described as potential drivers of this malignancy [244,267]. While some of these molecular signatures are shared with the less lethal DTCs (Chapter 5), suggesting their progression to ATC through step-wise accumulation of mutations and tumor evolution [266], dedifferentiation of preexisting benign nodules and DTCs are not the only means of disease development and at least a subset of ATCs may arise *de novo* [267].

Tumor-derived cell lines provide an alternative to studying patient specimens when profiling rare tumors and these can facilitate the investigation of therapeutic effectiveness in pre-clinical

settings. Schweppe and colleagues have reported on cross-contamination and mislabeling concerns in 40% of thyroid cancer cell lines that have been used in over 200 published studies [268,269]. They have clearly emphasized the need for detailed characterization of all thyroid-derived, including ATC-derived, cell lines. In this study, we describe the genomic and transcriptomic profiles of 1 primary ATC and 3 authenticated anaplastic thyroid cancer cell lines [269]. Those profiles augmented by the transcriptomes of 4 additional and unique cell lines [268] were compared to 58 pairs of papillary thyroid carcinoma (PTC) and matched normal tissue transcriptomes from The Cancer Genome Atlas (TCGA) study [270]. To the best of our knowledge, this is the first report of whole genome and transcriptome analyses of anaplastic thyroid cancer, allowing for the identification of regions of copy number alteration and large structural events at the base level resolution.

4.2 Materials and Methods

4.2.1 Study Specimens

Excision biopsy of a primary and treatment-naive anaplastic thyroid carcinoma tumor and peripheral blood sample were collected from a 63-year old male at the time of palliative thyroidectomy; the patient lacking prior personal or family history of thyroid disease or cancer and radiation exposure presented with lung metastasis. He provided written informed consent for the complete genomic profiling of his specimens; these were collected as part of a research project approved by the British Columbia Cancer Agency's Research Ethics Board and are in

accordance with the Declaration of Helsinki. In addition, 3 authenticated ATC cell lines, THJ-16T, THJ-21T and THJ-29T [269], obtained from the Mayo Clinic (Jacksonville, FL) and 4 unique cell lines [268], ACT-1 and T238 from Dr. R. Schweppe at the University of Colorado (Denver, Colorado) and C643 and HTh7 from Dr. N.E. Heldin at the Karolinska Institute (Uppsala, Sweden), were evaluated in this study.

4.2.2 Library Preparation and Sequencing

DNA from the ATC tumor, the matched peripheral blood specimen, and THJ-16T, THJ-21T and THJ-29T cell lines were subjected to whole genome sequencing; 100 bp paired-end sequence reads were generated on Illumina HiSeq2500 instruments following the manufacturer's protocol with minor variations. In addition, 75 bp paired-end transcriptome sequence reads were produced for the tumor and all 7 cell lines (Table 4.1). The aligned sequence datasets have been deposited at the protected European Genome-phenome Archive (EGA, http://www.ebi.ac.uk/ega/) under accession number EGAS00001001214.

Tumor biopsy specimen collected from the patient was embedded in Tissue-Tek O.C.T. (optimal cutting temperature) compound (Sakura Finetek USA, Inc.) and sectioned for DNA extraction. Using 1ug DNA each from the tumor and blood and 3 cell lines, THJ-16T, THJ-21T and THJ-29T, five whole genome libraries were constructed using a modified version of Illumina TruSeq PCR free protocol (FC-121-3001). In brief, 1ug genomic DNA was sheared for 45 sec, duty cycle 10%,

intensity 5 burst per second 200 using Covaris E210, to an average of 400bp. NEB Paired-End Sample Prep Kit (New England Biolabs, USA) was used in library construction. Following the end repair reaction, a size selection was done using Ampure XP bead (Beckman-Coulter, USA). The sample:bead ratio was 110:27 for upper cut and 137:15 for lower cut, respectively. The resulting size selected fraction, 300-500 bp, was A-tailed, and ligated to Illumina TruSeq adapters. The PCR-free libraries were cleaned up with Ampure XP beads and quantified by qPCR assay using the KAPA SYBR FAST qPCR kit (Kapa Biosystems (Pty) Ltd, South Africa). Paired-end 100 bp reads were generated on Illumina HiSeq2500 instruments following the manufacturer's protocol with minor variations. Software version HCS1.5.8 was utilized.

For whole transcriptome sequencing, RNA was extracted from 7 cell lines using MACS mRNA isolation kit (Miltenyi Biotec), resulting in 5-10 µg of DNase I-treated total RNA as per the manufacturer's instructions. Double-stranded cDNA was synthesized from the purified poly(A)$^+$ RNA using the Superscript Double-Stranded cDNA Synthesis kit (Invitrogen) and random hexamer primers (Invitrogen) at a concentration of 5 µM. The cDNA was fragmented by sonication and a paired-end sequencing library prepared following the Illumina paired-end library preparation protocol. Cluster generation and sequencing were performed on the Illumina HiSeq instruments following the manufacturer's recommended protocol, producing 75bp paired-end non-stranded whole transcriptome sequence data. One transcriptome library from the tumor was constructed using 3ug RNA by following the strand specific RNA-Seq protocol [271], with a few modifications. Briefly, PolyA+ RNA was purified using the MultiMACS

mRNA isolation kit on the MultiMACS 96 separator (Miltenyi Biotec, Germany). The eluted PolyA+ RNA was ethanol precipitated and re-suspended in 10µL of DEPC treated water. First-strand cDNA was synthesized from the purified polyA+ RNA using the Superscript cDNA Synthesis kit (Life Technologies, USA) and random hexamer primers at a concentration of 5µM along with a final concentration of 1ug/ul Actinomycin D. The second strand cDNA was synthesized following the Superscript cDNA Synthesis protocol by replacing the dTTP with dUTP in dNTP mix, allowing the second strand to be digested by UNG (Uracil-N-Glycosylase, Life Technologies, USA) post adapter ligation to achieve strand specificity. Library construction was carried out following a modified version of the Illumina paired end library protocol using the NEB Paired-End Sample Prep Kit (New England Biolabs, USA), the adapter-ligated products were purified using Ampure XP beads and digested with UNG (1U/ul) at 37°C for 30 min followed by deactivation at 95°C for 15 min. The digested cDNA was purified using Ampure XP beads, and then PCR-amplified with Phusion DNA Polymerase (Thermo Fisher Scientific Inc. USA) using Illumina's PE primer set, with cycle condition 98˚C 30 sec followed by 10 cycles of 98˚C 10 sec, 65˚C 30 sec and 72˚C 30 sec, and then 72˚C 5min. Paired-end 75bp reads were generated on Illumina HiSeq2500 following the manufacturer's protocol with minor variations. Software version HCS1.5.8 was utilized.

4.2.3 Sequence Data Analysis

Sequence reads from the whole genome libraries were aligned to the human reference genome (build GRCh37) using the Burrows-Wheeler Alignment (BWA) tool [56]. The tumor's genomic sequence was compared to that of patient's constitutive DNA to identify somatic alterations. Regions of copy number variation (CNV) and loss of heterozygosity (LOH) were determined using Control-FREEC [32]. This software does not require a matched normal tissue input; given the lack of such controls for the ATC cell lines, unlike the analyses done in Chapters 2 and 3, Control-FREEC was used. *De novo* assembly and annotation of genomic data using ABySS and Trans-ABySS [62] were used to identify small insertions and deletions (indels) and larger structural variants (SVs) including translocations, inversion and duplications leading to gene fusions; identified SVs were verified using an orthogonal alignment-based detection tool, BreakDancer [80]. Single nucleotide variants (SNVs) and indels in the tumor/normal pair were identified using a probabilistic joint variant calling approach utilizing SAMtools and Strelka [64,75]. Variants in the unpaired cell line genomic data were identified using SAMtools [64]; the indel lists for these samples were refined to include only those events that were also called through *de novo* assembly.

Sequence reads from the transcriptome libraries were aligned to the human reference genome (build GRCh37) using TopHat [272] with Ensembl gene model annotation file on the -G parameter. The reference sequence and the corresponding annotation files were provided by

Illumina's iGenome project and downloaded from the TopHat homepage (http://tophat.cbcb.umd.edu/igenomes.shtml). Quantification of gene expression was accomplished using HTSeq [273] in intersection-nonempty mode and excluding reads with quality less than 10, all subsequent analyses were run using only the count values for the protein-coding elements. Fifty-eight pairs of papillary thyroid carcinoma and matched normal tissue transcriptomes from The Cancer Genome Atlas project [270] were used for differential gene expression analysis. To ensure consistent analysis, raw sequence reads were downloaded from the Cancer Genome Hub and processed using the analysis pipeline described above. Protein-coding gene read counts were used as input into the R package edgeR [274] for differential gene expression analysis. Single-sample gene set enrichment analysis (ssGSEA) [275] was performed for each of the 8 transcriptomes to elucidate the oncogenic profiles enriched in each library when compared with normal thyroid tissue expression profiles. Structural variants were identified using *de novo* assembly-based approach employing ABySS and Trans-ABySS [62] and the alignment-based SV detection tool Minimum Overlap Junction Optimizer (MOJO) (https://github.com/cband/MOJO).

4.3 Results

4.3.1 Single Nucleotide Variants and Indels

Twenty-four somatic SNVs and indels were identified in the tumor's genome including heterozygous BRAF p.V600E and TP53 p.Y163C mutations. All three cell lines had *TP53*

homozygous nonsense or missense mutations with known pathogenic alleles. Other variants related to tumor biology included a homozygous BRAF p.V600E mutation in THJ-21T and heterozygous and homozygous frame-shift deletions of HDAC10 (p.H134Tfs) and CDKN2A (p.Q70Sfs), respectively, in THJ-29T. Additionally, THJ-16T harbored a heterozygous activating mutation in PIK3CA (p.E545K), a variant of unknown significance in RET (p.E90K) and a homozygous frame-shift deletion (p.S799Ffs) in EP300. Alterations of *TP53* and *BRAF* were the only recurrent events and no mutations of the previously described ATC genes including *H-*, *K-*, *N-RAS*, *CTNNB1*, *IDH1*, *ALK*, *PTEN*, *APC*, or *AXIN1* [244,276,277] were identified in these specimens. This is likely due to a small number of samples examined here and the infrequent mutations of these genes in the overall ATC population [244]. The numbers of small mutations in ATCs is comparable to that of parathyroid cancer (Chapter 2) and lower than those we observed in oncocytic thyroid tumors (Chapter 3). Due to small sample sizes, no general conclusions can be drawn; however, the number of small somatic mutations in endocrine gland malignancies does not seem to correlate with disease aggressiveness or prognosis.

4.3.2 Copy Number Variants

Evaluation of the copy number status and single nucleotide allele frequencies of the genomic data revealed extensive regions of gene copy loss and gain and the presence of triploid genomes in all 4 samples (Figure 4.1), consistent with previous observations of aneuploidy in the majority of ATCs [278]. Large-scale copy number changes have also been described in ATCs

[263] and are a hallmark of the progression from the mostly “quiet” differentiated cancers [270] to the aggressive and lethal ATCs. Although the tumor and the cell lines showed variable regions of copy number alterations, a 26 Mb minimal region on 5p, encompassing 196 genes, and the long arm of chromosome 20 showed gain of extra gene copies in all samples (Figure 4.1). High-level and recurrent amplifications of 5p and chromosome 20 have been reported in studies utilizing comparative genomic hybridization in studying ATCs [277] indicating that genes located in these regions might play an important role in ATC tumor initiation and/or progression. The 5p region includes proto-oncogenes such as *FGF10* and *SKP2*, mTOR signaling pathway members *RICTOR* and *PRKAA1*, in addition to *IL7R*, *OSMR*, *LIFR*, *PRLR* and *GHR*, all receptors involved in JAK-STAT and the downstream PI3K-Akt pathways. Anti-apoptotic and cell cycle genes *BCL2L1*, *YWHAB*, *E2F1* and *AURKA*, proto-oncogenes *PLCG1* and *STK4* and chromatin remodeling genes *ASXL1*, *CHD6* and *DNMT3B* have all gained extra copies through the amplification of 20q. Noteworthy observations of copy number change included the presence of 15 copies of each of *KDR/VEGFR1*, *KIT* and *PDGFRA* in a region of focal amplification on chromosome 4 in THJ-29T cell line. THJ-21T showed a region of high amplification on chromosome 11 leading to the accumulation of 25 copies of each of *BIRC2*, *BIRC3*, *MMP1*/*3*/*7*/*8*/*10*/*13*/*27* and *YAP1*; this cell line also had a complete loss of a small region on chromosome 9 encompassing *SMARCA2*, a member of the SWI/SNF complex, and *GLIS3*, a transcription factor implicated in the development and normal functioning of the thyroid.

4.3.3 Structural Variants

The study specimens were found to have anywhere between 1 to 32 structural variants (Figure 4.2A and Tables 4.2-4.9). On average, these numbers are higher than that in parathyroid carcinoma (Chapter 2), thyroid oncocytic cancer (Chapter 3) and, as discussed in the following chapter, in benign thyroid tumors. Hence, gene fusions may play an essential role in ATC tumorigenesis. Expressed in-frame gene fusions involving at least one proto-oncogene have been described in various cancers and are shown to be the driver of malignant phenotype, at times as the only such event in the tumor. We identified instances of these fusions in the genomes of THJ-16T and THJ-29T cell lines and the tumor (Figure 4.2B). These included an MKRN1-BRAF fusion in THJ-16T; the fusion product has lost the N terminal regulatory region of BRAF while retaining its kinase domain, hence likely leading to the constitutive activation of the kinase. A fusion of these two genes was also found in 1 TCGA PTC sample (0.2% population frequency) [270]. A reciprocal fusion between chromosomes 7 and 10 led to an in-frame fusion of FGFR2 and OGDH in THJ-29T, retaining the growth factor receptor's kinase domain. Two TCGA PTC cases were also reported to have *FGFR2* gene fusions with *VCL* and *OFD1* as partners [270]. *FGFR2* is found fused to various genes in different cancers where the fusion partners facilitate its constitutive activation through providing dimerization domains [279]. Sensitivity to FGFR inhibitors have been observed in patients harboring *FGFR2* fusions with the same breakpoint as that found in the THJ-29T ATC cell line [279] and thus testing for these fusions might provide a tractable therapeutic option for a subset of patients diagnosed with anaplastic

thyroid cancer. We also identified a translocation between chromosomes 16 and 18 in the tumor, fusing the proto-oncogene *SS18* and *SLC5A11*. SS18 (also known as SYT) is commonly found fused to one of SSX1, SSX2 or SSX4 in synovial sarcomas [280]. SS18 is a subunit of the SWI/SNF complex [281] and hence plays a major role in transcriptional regulation of the cell. It also interacts with various members of chromatin remodeling complexes such as SMARCA2, SMARCA4 [280] and EP300 [282] through its conserved N-terminal SNH domain that is found to be indispensible for the transforming ability of SS18-SSX onco-protein [280]. Although the fusion partner, SLC5A11, is distinct from that observed in sarcomas, it is likely that this fusion has transforming potential in ATCs. Only the last 8 residues of SS18 are deleted in its fusion to SSX genes and the mere deletion of these same 8 amino acids in the absence of a fusion partner was shown to disrupt the normal function of the protein [282]. Loss of SS18 C-terminal might be sufficient for tumorigenesis or that a yet unknown function of SLC5A11 may lead to the malignant transformation. In addition to the above potentially oncogenic fusions, gene members of the axon guidance pathway, recurrently altered in pancreatic cancer [283], were also found to be involved in multiple fusions: *CADM2-EPHA3* fusion in the tumor's genome, fusion of chromosome 19 to *SLIT1* on chromosome 10 in the THJ-21T genome and *SRGAP3-SETD5* fusion in THJ-29T.

4.3.4 Analysis of Differential Transcript Abundance

Despite the heterogeneous molecular profile of ATCs evident from the lack of commonly mutated genes and oncogenic fusions, the transcriptomic analysis of the tumor and all 7 cell lines showed consistent up- and down-regulation of several genes when compared to the compendium of normal thyroid tissue transcriptomes. Overexpressed genes included focal adhesion, cytoskeleton and ECM-receptor interaction pathway genes such as *ITGA3*, *ITGB1*, *FLNA*, *ACTN1*, and *CD44* indicating alterations of genes involved in regulation of normal cell shape and migration. Cancer-related genes with significant up-regulation in all ATCs included *MYC*, *mTOR*, *PRKCA* and *TGFB1* (Figure 4.3B). The down-regulated genes included thyroid differentiation signature genes such as *TG*, *TTF1*, *TSHR* and *TPO* (Figure 4.4) in addition to the tumor suppressor *FHIT*. Genes believed to be cancer drivers and to serve as drug targets in other malignancies showed consistent down-regulation in anaplastic thyroid cancer; these included *ERBB4*, *NTRK2*, *FGF7* and *MAPK10* (Figure 4.5). Differential gene expression analysis of the ATC cohort against the TCGA normal transcriptomes using edgeR found 840 and 574 genes to be down- and up-regulated in ATCs, respectively (Benjamini-Hochberg $P < 0.05$ and fold change > 4 or < -4); similar analysis yielded 605 and 419 down- and up-regulated genes in ATCs when compared to PTCs. Pathway analysis of these differentially expressed genes showed ECM-receptor interaction, focal adhesion, endocytosis, cell cycle, p53 signaling, ErbB signaling and general cancer pathways to be up regulated in ATCs. Common down-regulated networks

included tight junctions, cell adhesion molecules and various metabolism pathways (Figure 4.3A).

Tumor genomes frequently show a vast amount of copy number change and aneuploidy. As these can be the side effect of the altered cell cycle machinery and disease progression rather than its driver(s), all copy number changes may not contribute to changes in gene expression levels. Integrative analysis of CNV and expression datasets thus allowed for the identification of correlated changes of these variations in all 4 specimens. Cell cycle kinase *AURKA* and the transcription factor *E2F1*, both located on chromosome 20 with gain of copies, also showed overexpression providing additional evidence for the deregulation of cell cycle control in ATCs. Overexpression of aurora kinase A is believed to be the cause of vast chromosomal abnormalities in ATCs given its key regulatory role in mitotic cell division, chromosome segregation and cytokinesis through association with centrosomes and the mitotic spindle [267,284]. Several investigational drugs with inhibitory effect on AURKA are under study and these might serve as promising therapeutics in ATCs. It is however imperative to demonstrate the high expression of these kinases as the driver of malignancy rather than just a by-product of the high rate of cell division in cancers particularly ATCs [285]. Similarly, tissue transglutaminase gene (*TGM2*) has gained extra copies in all samples and also shows overexpression compared with normal thyroid tissue and PTCs. Over-activation of *TGM2* in ATCs correlates with its observed over-expression in pancreatic cancer, another aggressive human malignancy with mortality rates close to 100%. TGM2 over-expression leads to tissue invasion, metastasis and

chemotherapeutic resistance in cancers of the pancreas [286] and is shown to protect these cancer cells from autophagy leading to growth advantage and resistance to chemotherapy [286]. TGM2 may as a result serve as a direct drug target where its blockage leads to autophagic cell death.

4.4 Discussion

Anaplastic thyroid cancer is an extremely aggressive malignancy with dismal prognosis that has had little change in its 4-month median survival rate over the past 50 years [277]. Similar to the case we genomically profiled, the majority of ATC patients present with a rapidly growing neck mass often causing dyspnea, dysphagia and at times vocal cord paralysis [285]. The extremely poor prognosis of ATC is reflected by the current American Joint Committee on Cancer staging system for thyroid cancer in which individuals with anaplastic histopathology, regardless of extent of disease, are classified as having stage IV disease [264]. There are currently no standard therapies for the treatment of anaplastic thyroid cancer as its rarity and rapidly fatal course have made it difficult to study large cohorts of patients and to conduct randomized clinical trials [287]. Doxorubicin is the most commonly used chemotherapeutic agent for the treatment of progressive and metastatic ATC, but has little impact on survival, with a partial response rate estimated to be 10-30%; if administered in combination with cisplatin, it may have slightly higher efficacy [287,288]. Multimodal treatments comprised of surgical resection, external beam radiation therapy and systemic therapy have been associated with increased

survival in some patients [263] though often only effective in managing uncommonly diagnosed localized ATCs [284]. Individual responses to targeted therapies including multi-kinase inhibitors have been reported [289-291], however, no single agent has shown significant improvement in progression-free survival in the setting of a clinical trial and thus none has gained approval for routine clinical use. Phase II trials of Pazopanib [292], imatinib [293], gefitinib [294], axitinib [295] and sorafenib [296,297] in small patient cohorts showed limited or negligible activity. This is despite some of these agents, such as sorafenib, resulting in objective response and receiving approval for the treatment of advanced DTCs.

The important role of increased endothelial cell proliferation and angiogenesis in thyroid cancer progression and maintenance is well recognized [295], and consequently the majority of the tested compounds are aimed at blocking these signaling pathways. The expression of some of the intended targets of these drugs by our ATC specimens, and the 58 pairs of PTC and normal thyroid tissues, are depicted in Figure 4.5. The majority of these drug targets, including *FGFR1*, *2*, *3* and *4*, *VEGFR1*, *2* and *3*, *PDGFRA*, *PDGFRB*, *KIT* and *RET*, show similar or lower expression in ATCs compared with both normal tissues and PTCs. The extent of messenger RNA expression might not be an accurate estimate of the protein level in the cell, and over-activation of a kinase is not captured on the transcript level, nonetheless, mRNA is an intermediary information molecule and its amount in the cell serves as a surrogate for protein expression levels. Based on the current differential mRNA expression analysis none of the multi-kinase inhibitors with observed response in DTCs would have an effect on the survival of ATC patients;

this is in agreement with the failure of all tested compounds to date and has implications in the development of future clinical trials. Lenvatinib has recently gained approval for the treatment of refractory DTCs, but the first described trial for its use in the treatment of 9 ATC patients showed only a median progression-free survival of 5.5 months [298]. We predict, based on the current study, that lenvatinib would not result in prolonged response in ATCs given the lower expression of all its targets (vascular endothelial growth factor receptors 1,2, and 3, fibroblast growth factor receptors 1, 2, 3 and 4, platelet-derived growth factor receptor alpha, RET and KIT) in ATC specimens (Figure 4.5). Generally, inhibitors of growth factors and their receptors appear to have a very limited effect on the survival of ATC patients. A similar lack of inefficacy is also found when using vascular disrupting agents. A single agent trial of the fosbretabulin (also known as combretastatin A-4 phosphate) or its combination use with carboplatin/paclitaxel in a cohort of patients, although showed some clinical activity, had no effect on progression-free survival [299,300].

Analysis of genomic and transcriptomic datasets in this study allowed for identification of potential new drug targets. *TRIP13* has gained extra copies in all specimens as a result of the 5p gain described above. This gene and its binding partner PRKDC promote non-homologous end joining (NHEJ) in cancer cells resulting in chemoresistance in head and neck malignancies where inhibitors of NHEJ, such as Nu7026, are believed to re-sensitize cells to cisplatin [301]. Both *TRIP13* and *PRKDC* show very high expression in the ATCs we studied and could serve as novel targets for therapy. The mTOR signaling pathway is also a putative target and inhibitors such as

everolimus may show efficacy in ATC. Mutations of the pathway genes including *mTOR* and the tumor suppressor *TSC2* have been previously described in ATC [276,289] and a dramatic and long-lasting response to everolimus in an ATC patient with a truncating mutation in *TSC2* was reported [289]. Though no mutations were identified in the current study, a high level expression of *mTOR* and its downstream effector *HIF1A* was observed, thus raising the possibility for the use of mTOR inhibitors (Figure 4.3B). Overexpression of *mTOR* or loss of *TSC2*, its negative regulator, through promoting the transcriptional level of *HIF1A* leads to increased angiogenesis that is sensitive to rapamycin treatment [302]. Given that overexpression of vascular growth factor receptors are not likely to directly lead to increased angiogenesis in ATCs, mTOR signaling emerges as a key angiogenesis driving pathway in this cancer. The effect of everolimus on 5 ATC cell lines including HTh7 and C643 were tested by Papewalis and colleagues [303]. They found that both cell lines responded to therapy with HTh7 exhibiting a much higher sensitivity when compared to known responding lymphoma cell lines. The alterations of mTOR pathway and its potential role in parathyroid tumorigenesis were also described in Chapter 2, suggesting a primary role for this signaling pathway in endocrine function. Prior to embarking on clinical trials, further *in vitro* and *in vivo* studies are needed to elucidate the mechanism of response and resistance to targeted therapeutics such as mTOR inhibitors.

A successful evolutionary history for cancer requires rapid and dynamic changes in the blueprint of the cell. Through providing a larger pool of possible mutational targets, recurrent

hits to specific cellular machineries or pathways, rather than the same gene, can accelerate the success of the cancer in overcoming its host defenses. We found alterations of the epigenetic machinery in all 4 ATC specimens with genome sequence data. A translocation of SS18, a member of SWI/SNF complex [281] in the tumor, homozygous frame-shift deletion in the histone acetyltransferase EP300 and a fusion of methyl CpG binding protein MECP2 and F8 in THJ-16T cell line, complete loss of SMARCA2, another member of the SWI/SNF complex and interacting partner of SS18 [281], in THJ-21T, a heterozygous frame-shift deletion in the histone deacetylase HDAC10 and a gene fusion of the transcriptional repressor and member of the SWI/SNF complex BCL11A [281] and GRIP2 in THJ-29T. The *FGFR2-OGDH* fusion in THJ-29T is, in addition to the involvement of the growth factor receptor, intriguing considering the role of OGDH in the control of metabolism and cellular epigenetic state. OGDH is a metabolic enzyme of the tricarboxylic acid (TCA) cycle and a subunit of the complex which converts 2-oxoglutarate, product of IDH, to succinate, substrate of SDH. Mutations of IDH1 and IDH2 as well as those in SDH have been observed in numerous cancers and found to cause global epigenetic changes in the tumor [304,305]. 2-oxoglutarate is required for the normal functioning of chromatin-modifying enzymes such as UTX, JARID1C and TET2 [305] and succinate acts as an inhibitor of DNA and histone demethylases [304]; changes in their cellular concentration as a result of OGDH translocation can in turn alter the epigenomic state of ATC cells. Further evidence for the potential role of epigenomic deregulation in ATC came from single-sample GSEA. Top 20% most enriched oncogenic signatures in each of the 8 transcriptome libraries were identified and those shared in two or more libraries are plotted in

Figure 4.3C. Top signatures enriched with over- and under-expressed ATC genes included genes that were up- and down-regulated, respectively, upon knockdown of *BMI1* or *PCGF2* or both genes [306]. BMI1 and PCGF2 are members of the Polycomb group of transcriptional regulators which control the expression of, among others, genes involved in ECM remodeling, cell adhesion and integrin-mediated signaling pathways [306], all of which demonstrated deregulation in ATCs. It is conceivable that understanding the effect of epigenetic changes in anaplastic thyroid cancer could pave the way for the development and application of novel therapeutics in this aggressive solid tumor. Histone deacetylase inhibitor valproic acid, for instance, increases the effect of both doxorubicin and paclitaxel in ATC cells [277] providing *in vitro* experimental evidence for a driving role of deregulated epigenetic control in ATC. Epigenetic alterations in ATCs in addition to mutations of *MLL2* in parathyroid carcinoma (Chapter 2) and *MEN1* mutations in oncocytic thyroid cancer (Chapter 3), both members of the histone methyltransferase complex, may be indicative of a role the epigenetic alterations play in all endocrine tumorigenesis.

In this study, we profiled the molecular alterations of several anaplastic thyroid carcinoma specimens including unique and authenticated ATC cell lines. Given the heterogeneous genomic profiles of these samples and the low frequency of recurrent mutations, studies involving larger cohorts of cases through multi-institutional collaborations are required to identify genes at the "long tail" of the mutational spectrum, and to decipher the underlying biology of the disease. Furthermore, lack of common targetable oncogenic mutations, observed responses to targeted

therapies in other cancer types harboring the same aberrations as those found in at least a small subset of ATCs [279], and clinical responses to targeted therapies described in individual ATC patients [289-291] calls for a more genotype-driven approach to diagnosis and treatment of this rare and rapidly fatal cancer. With recent advances in molecular and information technology alike, it is anticipated that sequencing-based clinical tests provide the ability to comprehensively assay the large number of diverse and complex mutational forms that can arise, hence facilitating routine application of precision oncology in the clinic.

Table 4.1 Sequence libraries read statistics

	Total Number of Reads	Number of Aligned Reads	Average Coverage
Tumor genome	1022984062	887257120	30.5
Blood genome	999542818	872101026	29.9
THJ-16T genome	1063849260	944842455	32.4
THJ-21T genome	1215949294	1057883434	36.2
THJ-29T genome	1273726388	1124785419	38.5
Tumor transcriptome	310755118	295756636	-
THJ-16T transcriptome	178466960	160865258	-
THJ-21T transcriptome	179493758	167089410	-
THJ-29T transcriptome	166304638	150904101	-
ACT-1 transcriptome	169675398	152479566	-
T238 transcriptome	228841146	196149944	-
C643 transcriptome	164143700	146172923	-
HTh7 transcriptome	182831104	160580587	-

Table 4.2 List of somatic SVs in the tumor

Event Type	Breakpoint Coordinates (chr:pos)	Genes	Dataset
translocation	16:29755668|18:36982773	BOLA2,NA	genome
duplication	3:85605596|3:89300765	CADM2,EPHA3	genome
inversion	9:20380439|9:35743408	MLLT3,GBA2	genome
translocation	16:26167689|18:23620965	NA,SS18	genome
translocation	16:24873489|18:23622703	SLC5A11,SS18	genome
translocation	16:24873922|18:23632588	SLC5A11,SS18	transcriptome

Table 4.3 List of SVs in THJ-16T cell line

Event Type	Breakpoint Coordinates (chr:pos)	Genes	Dataset
deletion	7:140339236|7:140484371	DENND2A,BRAF	genome
duplication	X:153312442|X:154169354	MECP2,F8	genome
duplication	X:153357642|X:154159951	MECP2,F8	transcriptome
duplication	7:140159131|7:140484340	MKRN1,BRAF	genome
duplication	7:140159507|7:140482957	MKRN1,BRAF	transcriptome

Table 4.4 List of SVs in THJ-21T cell line

Event Type	Breakpoint Coordinates (chr:pos)	Genes	Dataset
inversion	15:78475563|15:79249350	ACSBG1,NA	genome
translocation	1:49740819|22:34381101	AGBL4,NA	genome
duplication	11:108142376|11:112351888	ATM,NA	genome
duplication	11:101946694|11:106151630	C11orf70,NA	genome
inversion	11:115069069|11:129563504	CADM1,NA	genome
duplication	5:122755261|5:122822216	CEP120,NA	genome
translocation	3:3050371|3:177932944	CNTN4,NA	genome
inversion	5:148925846|5:177497743	CSNK1A1,NA	genome
inversion	9:37739931|9:37753561	FRMPD1,NA	genome
duplication	11:120691806|11:122939218	GRIK4,NA	genome
translocation	6:29814678|12:133066751	HLA-H,NA	genome
deletion	10:121572626|10:121591129	INPP5F,MCMBP	genome
inversion	11:102666101|11:102755581	MMP1,NA	genome
inversion	11:101607373|11:106685771	NA,GUCY1A2	genome
inversion	9:37749768|9:37777310	NA,TRMT10B	genome
translocation	6:43962453|8:14835894	NA,SGCZ	genome
duplication	1:48685420|1:48712126	NA,SLC5A9	genome
inversion	22:21079831|22:34070927	PI4KA,LARGE	genome
translocation	10:72640780|16:34298022	SGPL1,NA	genome
duplication	8:19251056|8:19274690	SH2D4A,CSGALNACT1	genome
translocation	10:98931589|19:7716903	SLIT1,NA	genome
inversion	11:101424515|11:102934901	TRPC6,DCUN1D5	genome
inversion	9:37169126|9:37764686	ZCCHC7,TRMT10B	genome
duplication	17:57915656|17:57970686	VMP1,RPS6KB1	transcriptome
translocation	12:133721111|15:63190850	ZNF10,NA	transcriptome

Table 4.5 List of SVs in THJ-29T cell line

Event Type	Breakpoint Coordinates (chr:pos)	Genes	Dataset
translocation	1:55030044|22:19435942	ACOT11,NA	genome
translocation	1:55029949|7:36106893	ACOT11,NA	genome
inversion	8:41658930|8:42009254	ANK1,NA	genome
translocation	2:60745077|3:14544960	BCL11A,GRIP2	genome
translocation	3:8671456|15:24393332	SSUH2,NA	genome
inversion	3:7749290|3:82978863	GRM7,NA	genome
translocation	16:19773458|16:83946365	IQCK,MLYCD	genome
translocation	16:19773338|16:83945972	IQCK,MLYCD	transcriptome
duplication	20:33051633|20:33442216	ITCH,GGT7	genome
inversion	15:37381985|15:40139066	MEIS2,GPR176	genome
translocation	3:154811040|X:130141822	MME,NA	genome
translocation	3:564457|18:25560906	NA,CDH2	genome
inversion	4:57763192|4:57915169	NA,IGFBP7	genome
translocation	14:40057342|15:28265969	NA,OCA2	genome
translocation	3:8446180|15:102006040	NA,PCSK6	genome
inversion	4:56021378|4:57852745	NA,POLR2B	genome
duplication	8:40804685|8:41142233	NA,SFRP1	genome
translocation	8:93134086|19:19220429	NA,SLC25A42	genome
inversion	5:123822053|5:136833667	NA,SPOCK1	genome
translocation	7:44679455|10:123240245	OGDH,FGFR2	genome
translocation	7:44684926|10:123243212	OGDH,FGFR2	transcriptome
translocation	2:179191810|3:63693746	OSBPL6,NA	genome
translocation	3:9449750|14:41243421	SETD5,NA	genome
deletion	20:48506228|20:61944892	SLC9A8,COL20A1	genome
duplication	3:9218482|3:9463500	SRGAP3,SETD5	genome
translocation	21:32810936|X:31501405	TIAM1,DMD	genome
translocation	2:175510803|18:47452566	WIPF1,MYO5B	genome
translocation	2:175510843|5:114425634	WIPF1,NA	genome
translocation	3:2942489|4:54119104	CNTN4,SCFD2	genome
translocation	3:2944560|4:54139993	CNTN4,SCFD2	transcriptome
translocation	16:87450678|20:57361072	ZCCHC14,NA	genome
translocation	16:87451066|20:57357892	ZCCHC14,NA	transcriptome
translocation	5:136832346|6:73764725	SPOCK1,KCNQ5	genome
translocation	5:136602744|6:73751785	SPOCK1,KNCQ5	transcriptome
translocation	3:8115946|14:53144141	NA,ERO1L	genome
translocation	3:8148799|14:53145152	NA,ERO1L	transcriptome
deletion	X:54222315|X:54471569	WNK3,TSR2	genome
deletion	4:6878218|4:6991990	KIAA0232,TBC1D14	genome

Table 4.6 List of SVs in ACT1 cell line

Event Type	Breakpoint Coordinates (chr:pos)	Genes	Dataset
inversion	11:71112918|11:71640170	NA,RNF121	transcriptome

Table 4.7 List of SVs in C643 cell lines

Event Type	Breakpoint Coordinates (chr:pos)	Genes	Dataset
deletion	1:165470864|1:214724526	LOC400794,PTPN14	transcriptome
translocation	1:25072116|1:222761907	CLIC4,TAF1A	transcriptome
inversion	1:8877219|1:12438482	RERE,VPS13D	transcriptome
inversion	1:167042776|1:198126406	GPA33,NEK7	transcriptome
duplication	5:110712558|5:139574237	CAMK4,CYSTM1	transcriptome
translocation	1:45363116|1:224868728	EIF2B3,CNIH3	transcriptome
duplication	1:17380443|1:173495853	SDHB,SLC9C2	transcriptome
translocation	2:24985645|4:82380668	NCOA1,RASGEF1B	transcriptome
duplication	1:1509858|1:154942675	SSU72,SHC1	transcriptome
duplication	1:39768596|1:90152170	MACF1,LRRC8C	transcriptome
inversion	1:16174645|1:94057950	SPEN,BCAR3	transcriptome
inversion	1:40420840|1:42800730	MFSD2A,FOXJ3	transcriptome
deletion	5:112043579|5:140358534	APC,PCDHA1	transcriptome
translocation	1:15665977|1:205418996	FHAD1,NA	transcriptome

Table 4.8 List of SVs in HTh7 cell line

Event Type	Breakpoint Coordinates (chr:pos)	Genes	Dataset
deletion	1:1354929|1:9262638	ANKRD65,NA	transcriptome
duplication	5:38151367|5:38925473	NA,OSMR	transcriptome
translocation	5:43644904|7:92546151	NNT,NA	transcriptome
inversion	22:39317220|22:50905767	NA,SBF1	transcriptome
translocation	1:60538342|5:38556436	C1orf87,LIFR	transcriptome
inversion	11:74209114|11:74500744	MGC12965,RNF169	transcriptome
translocation	7:19555924|9:133738422	NA,ABL1	transcriptome
translocation	7:156589083|8:41393879	LMBR1,GINS4	transcriptome
duplication	7:91924203|7:94285892	ANKIB1,PEG10	transcriptome
duplication	7:91794294|7:94285892	AL133568,PEG10	transcriptome
duplication	5:43067189|5:43388578	NA,CCL28	transcriptome
inversion	1:58971732|1:60223603	DAB1,FGGY	transcriptome

Table 4.9 List of SVs in T238 cell line

Event Type	Breakpoint Coordinates (chr:pos)	Genes	Dataset
duplication	18:55233679|18:56205456	FECH,ALPK2	transcriptome
deletion	1:162551249|1:224318174	UAP1,FBXO28	transcriptome
inversion	18:53128250|18:54547175	TCF4,WDR7	transcriptome
duplication	18:52544798|18:54483375	RAB27B,WDR7	transcriptome
inversion	1:150772185|1:150998149	CTSK,PRUNE	transcriptome
inversion	19:58500089|19:59070776	ZNF606,LOC100131691	transcriptome
translocation	3:1189611|18:55919286	CNTN6,NEDD4L	transcriptome
inversion	18:53565658|18:55711940	NA,NEDD4L	transcriptome
inversion	18:55628638|18:56001124	NA,NEDD4L	transcriptome

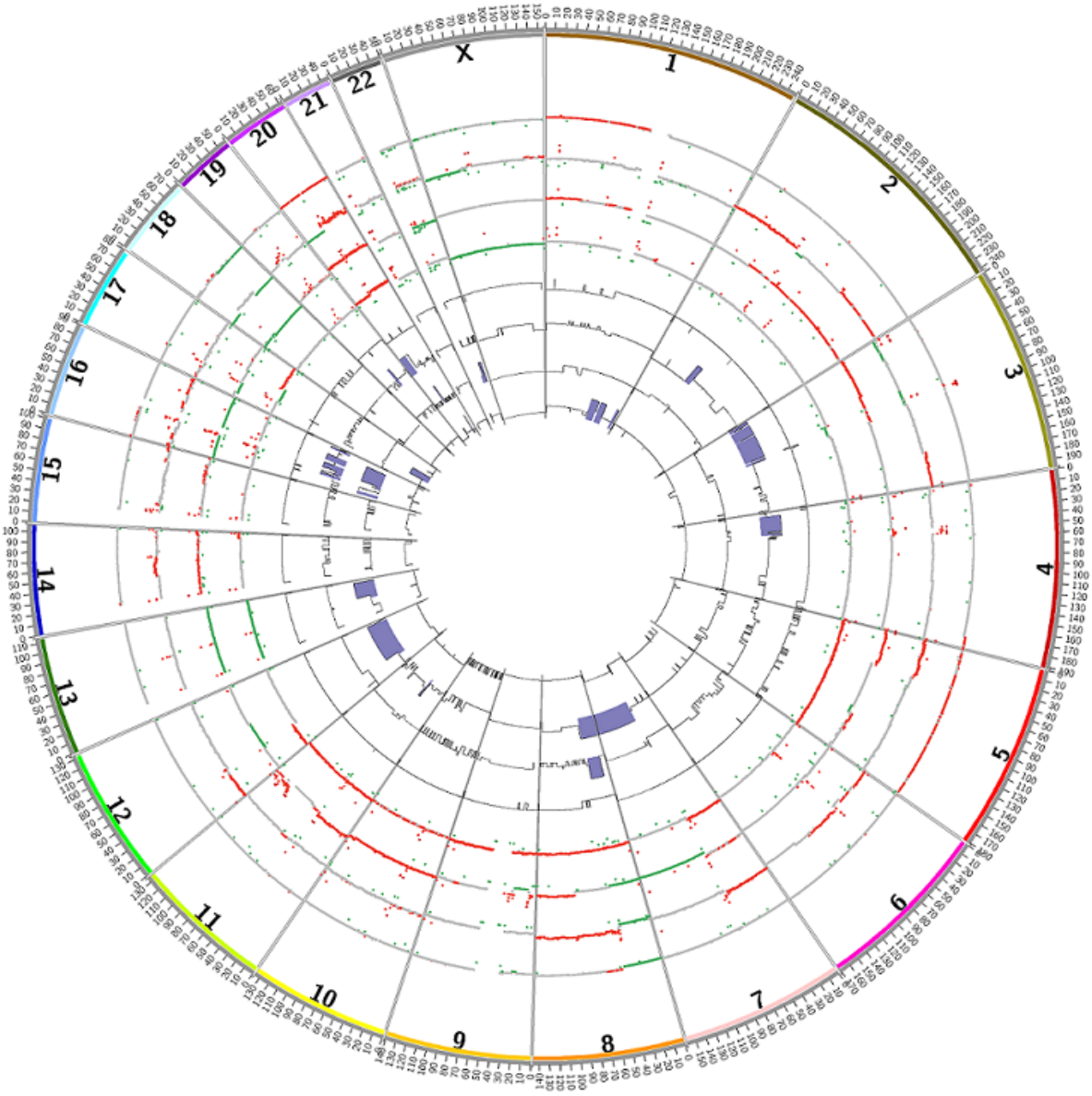

Figure 4.1 CNV regions in sequenced genomes
A circos plot depicting, from the outer ring inward, tumor CNV, THJ-29T CNV, THJ-21T CNV, THJ-16T CNV, tumor LOH, THJ-29T LOH, THJ-21T LOH and THJ-16T LOH. Red and green CNV regions illustrate the regions of copy gain and loss, respectively. The LOH tracks illustrate the B Allele Frequencies (BAF) ranging from 0.5 to 1. Those regions with BAF >= 0.9 are highlighted in purple. Regions of 5p and 20q showed recurrent copy gain in all samples

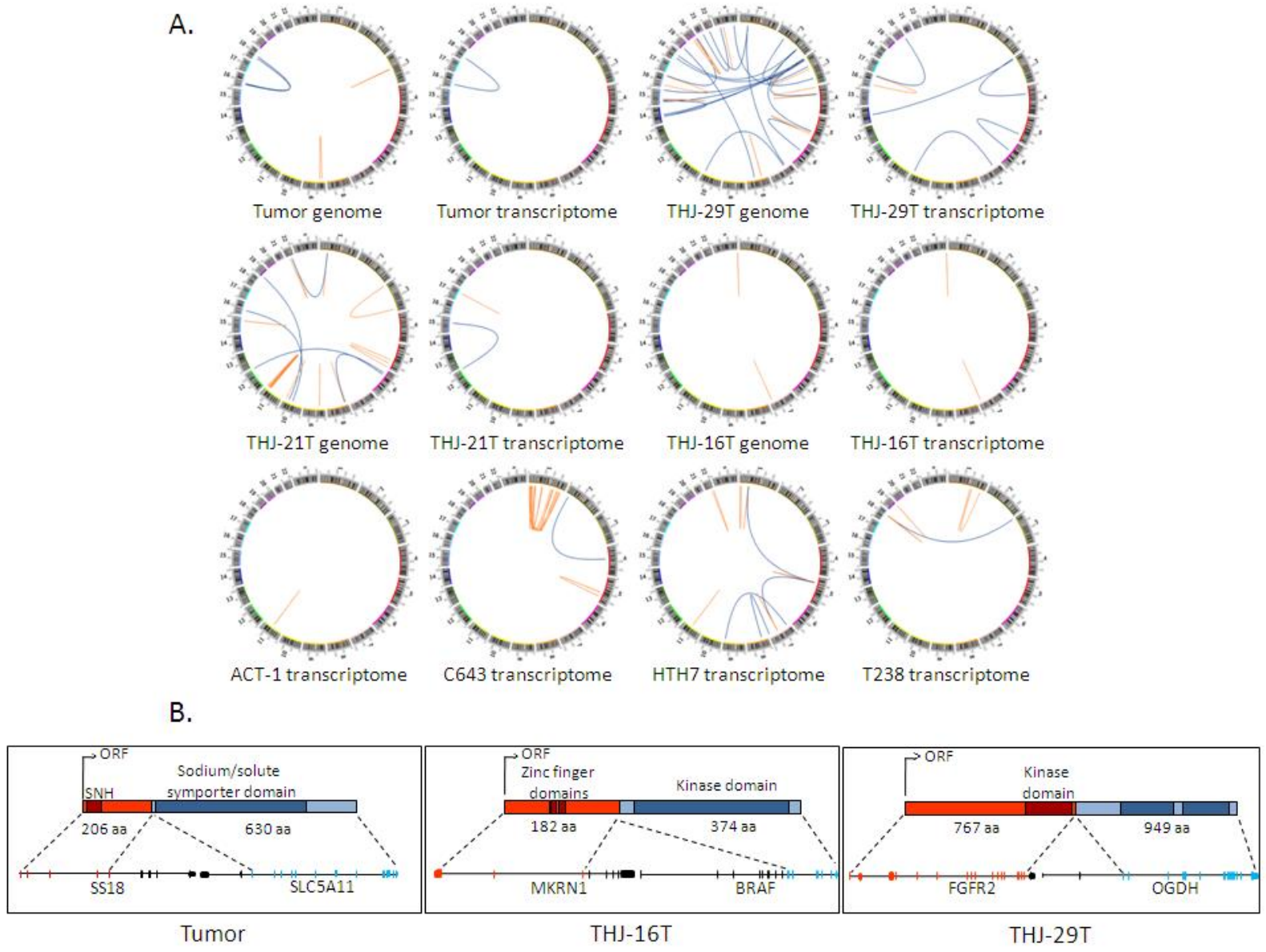

Figure 4.2 Structural variants in ATCs

A. Structural variants identified in the genomic and transcriptomic datasets B. Detailed structure of the potentially oncogenic fusions: SS18 (transcript: ENST00000415083)/SLC5A11 (transcript: ENST00000347898) fusion in the tumor, MKRN1 (transcript: ENST00000255977)/BRAF (transcript: ENST00000288602) fusion in THJ-16T cell line and FGFR2 (transcript: ENST00000358487)/OGDH (transcript: ENST00000222673) fusion in THJ-29T cell line

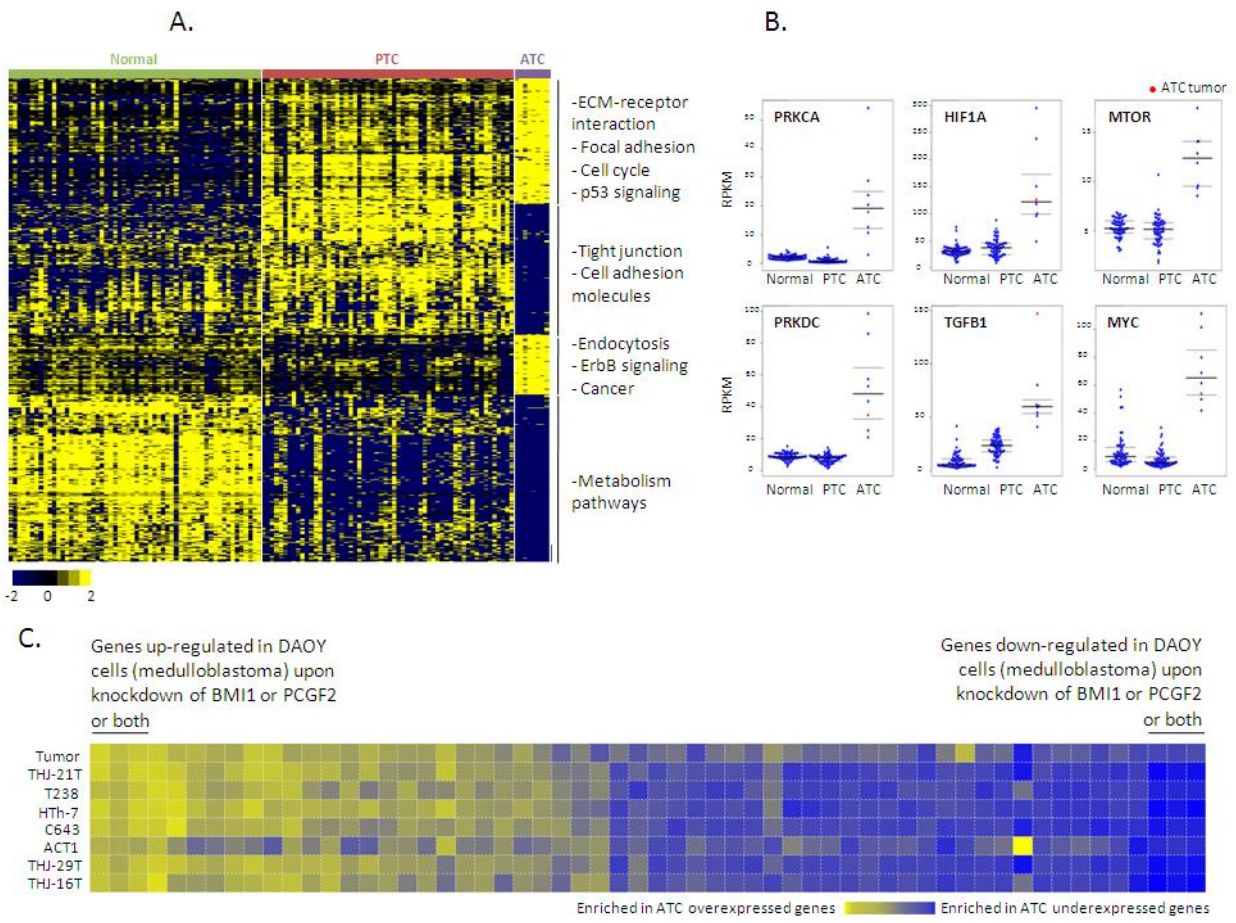

Figure 4.3 ATC expression analyses

A. Samples were ordered on the basis of pathology and 1647 significantly expressed genes in 58 TCGA normal thyroid tissue transcriptomes, 58 TCGA papillary thyroid cancer transcriptomes and 8 anaplastic thyroid cancer transcriptomes were clustered B. The expression levels (RPKM=reads per kilobase per million mapped reads) of select genes in the TCGA and ATC specimens are plotted C. ssGSEA was performed for all 8 transcriptome libraries using fold changes in expression of each gene (ATC expression/average expression in 58 normal libraries) in order to identify enriched oncogenic signatures. Top 20% most enriched signatures that were shared in two or more libraries are plotted. The molecular signatures enriched with up- and down-regulated ATC genes included genes that were up- and down-regulated upon knockdown of BMI1 or PCGF2 or both genes

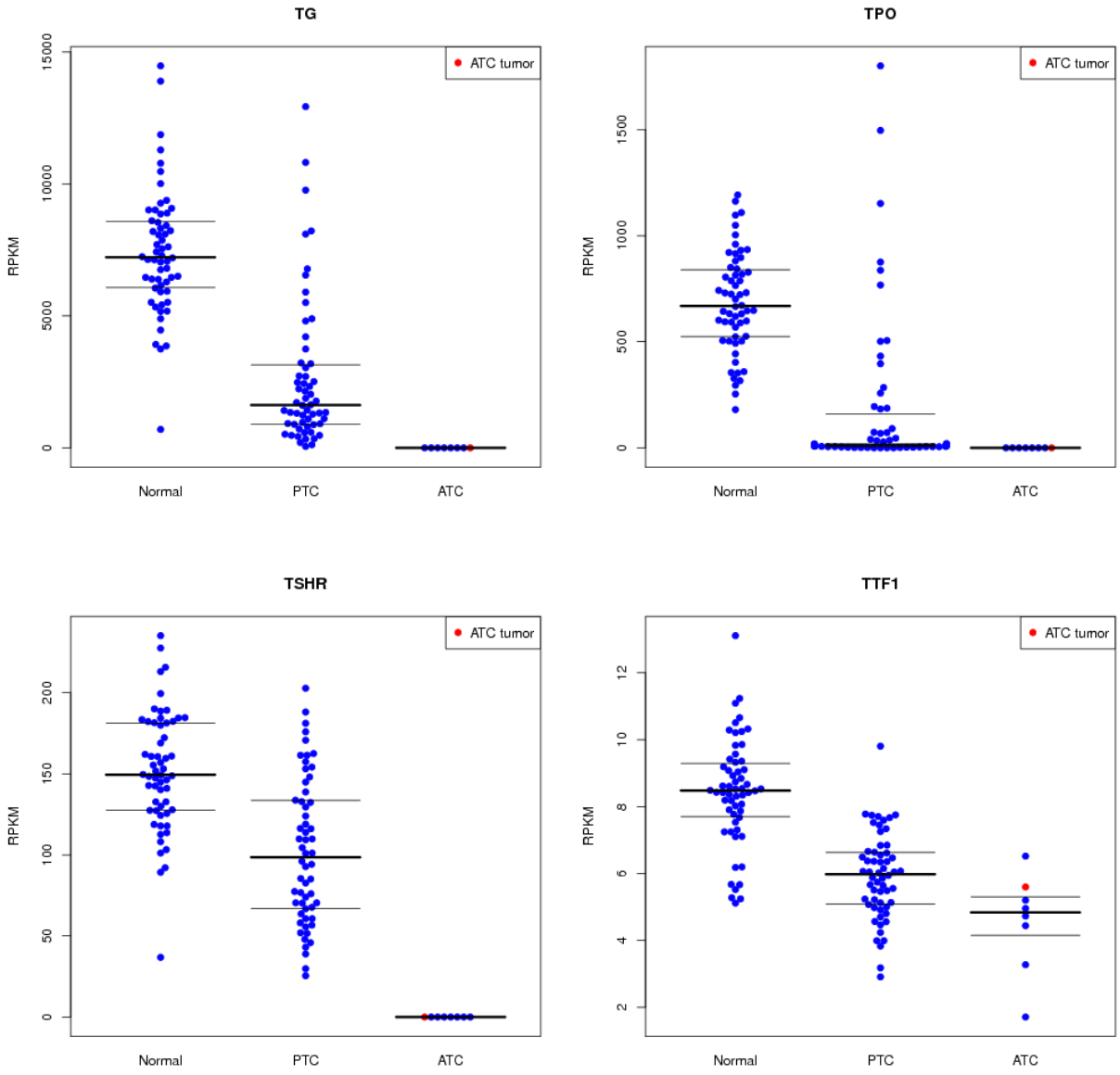

Figure 4.4 Down-regulation of thyroid differentiation marker genes in ATCs

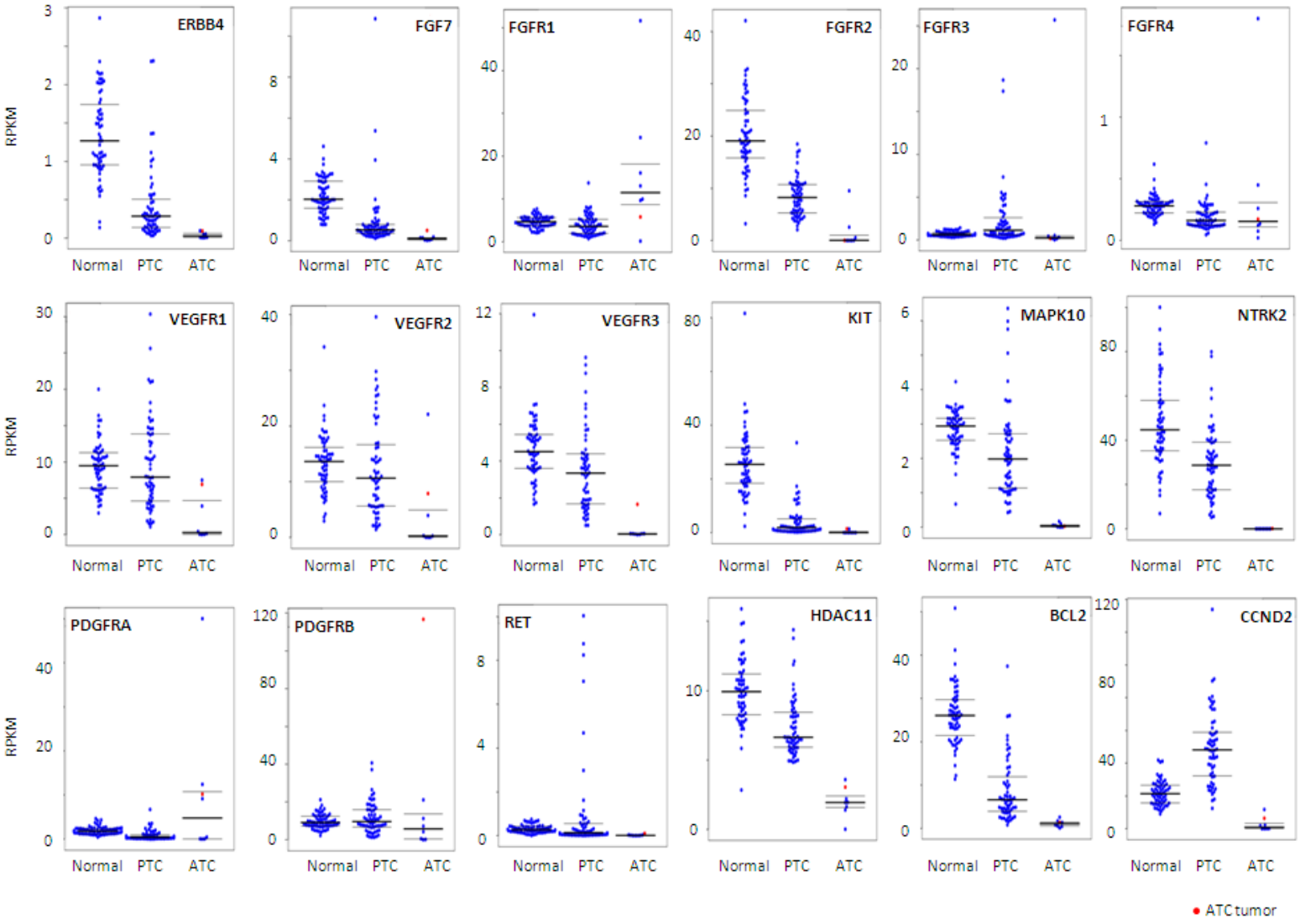

Figure 4.5 Down-regulation of potential cancer drivers and drug targets in ATCs

Chapter 5: Molecular Profiling of Papillary Thyroid Carcinoma and Benign Thyroid Nodules[5]

5.1 Introduction

Approximately 5% of the population has palpable thyroid disease and by ultrasound examination, over 50% of the population will be diagnosed with thyroid nodules [6]. In 20-35% of cases, preoperative diagnosis by fine needle aspiration (FNA) biopsy is inconclusive and so a large proportion of individuals with indeterminate FNAs undergo thyroid surgery as a diagnostic procedure for cancer [307]. After surgery, over 80% of suspicious tumors are found to be benign nodules [23]. Hence, there is a great need for robust diagnostic markers that can improve the ability of fine needle aspiration biopsy to discriminate between benign and malignant nodules, reducing the number of surgeries that are needlessly undertaken. Over 90% of all thyroid malignancies are those referred to as differentiated thyroid cancer (DTC), the

[5] Portions of this chapter have either been published or are in preparation for submission; the author contributions are provided in the Preface as per the University of British Columbia PhD thesis guidelines: Section 5.3.1, "Cav1 and Gal3 Immunohistochemical Analysis", was published as Jay Shankar, Sam M Wiseman, Fanrui Meng, Katayoon Kasaian, Scott Strugnell, Alireza Mofid, Allen Gown, Steven JM Jones, and Ivan R Nabi. (2012). Coordinated expression of galectin-3 and caveolin-1 in thyroid cancer. Journal of Pathology, 228: 56–66. doi: 10.1002/path.4041. Copyright by Wiley. Section 5.2.1, "Prognostic Significance of Papillary Thyroid Carcinoma Presentation Mode", has been published as Heywood Choi, Katayoon Kasaian, Adrienne Melck, Kaye Ong, Steven JM Jones, Adam White, Sam M Wiseman. (2015). Papillary Thyroid Carcinoma: Prognostic Significance of Cancer Presentation. American Journal of Surgery, doi: 10.1016/j.amjsurg.2014.12.047. Copyright by Elsevier. Section 5.2.2, "Prognostic Significance of Tumor Laterality in Papillary Thyroid Cancer", is in preparation for submission: Sarah E Moore, Katayoon Kasaian, Steven JM Jones, Adrienne Melck, and Sam M Wiseman. (2015). Papillary Thyroid Cancer: Epidemiology of Bilateral Disease. Whole genome and transcriptome studies of benign thyroid nodules and papillary thyroid carcinoma described in sections 5.4 and 5.5 are based on unpublished work.

group of cancers derived from follicular cells of the thyroid gland; these include the papillary, follicular and Hürthle cell (Chapter 3) thyroid carcinomas with the papillary thyroid carcinoma (PTC) accounting for the majority of DTCs [9]. Unlike Hürthle cell (Chapter 3) and anaplastic thyroid cancers (Chapter 4), which do not pose any diagnostic challenges, discriminating PTCs from benign nodules is less simple. In the following described studies, we aimed to identify diagnostic and prognostic markers for PTCs and to examine and compare the genomic profiles of benign thyroid nodules and PTCs with the aim of defining the spectrum of mutations and genetic alterations accruing during the development of these tumors and of identifying ways these can be utilized for diagnostic purposes.

5.2 Prognostic Factors for Papillary Thyroid Cancer

We performed two studies using a prospectively maintained database of papillary thyroid carcinoma patients in order to identify potential correlations between the mode of disease presentation and prognosis in addition to any associations between disease bilaterality and prognosis. These studies have the potential to stratify patients to those with low and high risk of disease recurrence and metastasis before surgery and hence facilitate treatment decision-making.

5.2.1 Prognostic Significance of Papillary Thyroid Carcinoma Presentation Mode

The aim of this study was to make a comparison of prognosis between patients who presented with symptomatic disease and those who were diagnosed incidentally either through routine physical examinations or through imaging performed for unrelated purposes such as chest x-ray. We hypothesized that the increasing number of radiological studies that are being performed would lead to a greater number of diagnosed PTCs; many of these likely represent over-diagnosis, given the rising incidence of PTCs with stable mortality rates [308].

We conducted a retrospective cohort study utilizing a prospectively maintained thyroid cancer database from the St Paul's Hospital, Vancouver, British Columbia, Canada. This database contained clinical and pathological information of thyroid cancer patients treated surgically between 2000 and 2013. The patient charts were reviewed to identify the initial event leading to cancer diagnosis. These events were categorized as follows:

1. Incidental imaging detection group: the detection of thyroid nodule by imaging performed for indications unrelated to the thyroid mass.
2. Incidental physical examination detection group: the thyroid nodule detected by a clinician during an evaluation for complaints not related to a thyroid mass.

3. Non-incidental detection group: the patient presents with complaints possibly related to the thyroid mass such as dysphagia, dysphonia, neck pain, self-detection of a neck mass or self-requested screening for thyroid cancer.

MACIS (metastasis, age, completeness of resection, invasion, and size) scoring system reflecting the 20-year disease-specific survival [309], developed at the Mayo Clinic and widely utilized as a measure of papillary thyroid cancer prognosis, was used as a measure of prognosis comparison between the incidental and non-incidental diagnosed PTCs. The twenty-year disease specific survival rate is 99% for MACIS score <6, 89% for MACIS score 6-6.99, 56% for MACIS score 7-7.99, and 24% for MACIS score >8 [309]. Significant associations between the type of PTC presentation and MACIS score, as well as with each component of the MACIS score, were assessed using Pearson chi-squared or Fisher's exact test, where appropriate. Scripts written in the R programming language (version 3.1.1, R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria) were used for these analyses. *P* values were corrected for multiple testing using the Benjamini–Hochberg (BH) correction method [145]. All statistical tests were two-tailed and a *P* value of less than 0.05 was considered statistically significant.

168 PTC patients met study inclusion criteria and made up the study population. There were 126 women and 42 men in this study cohort. 28 (17%) patients had incidental imaging that led to their PTC detection, 60 (36%) patients had their PTC detected incidentally during a physical examination by a physician, and 80 (47%) of PTC patients presented with complaints related to

a thyroid mass. There was no significant difference in gender and whether PTC presented incidentally or symptomatically. The distribution of MACIS scores for patients in the incidental imaging PTC detected group was: <6 (85%), 6-6.99 (4%), 7-7.99 (7%), and ≥8 (4%). The distribution of MACIS scores for patients in the incidental physical examination PTC detected group was: <6 (78%), 6-6.99 (13%), 7-7.99 (7%) and ≥8 (2%). The distribution of MACIS scores for patients presented with complaints related to thyroid mass was: <6 (90%), 6-6.99 (5%), 7-7.99 (4%) and ≥8 (1%). The difference in the proportion of patients in each MACIS group amongst the 3 clinical presentation categories was not statistically significant. Each individual component of the MACIS score (presence of distant metastases, patient age, completeness of cancer resection, cancer invasion, and tumor size) was also examined with respect to the three presentation categories, with no significant differences between them. Data was also analyzed after subdividing patients into groups presenting prior to or subsequent to 2009, representing groups who were diagnosed before or after a major increase in the use of medical diagnostic technologies, with no significant differences observed between the two time periods with respect to PTC presentation and cancer prognosis.

Our study suggests that regardless of the mode of presentation, the disease specific survival did not differ between patients. These findings support the current practice of disregard for the diagnostic event when a fine needle aspiration biopsy is recommended [236].

5.2.2 Prognostic Significance of Tumor Laterality in Papillary Thyroid Cancer

The standard of care for patients diagnosed with PTC is to perform a total thyroidectomy even in cases of unilateral disease; however, the extent of surgery, especially in low-risk individuals, has been extensively debated [7] and some experts support that lobectomy may be suitable for individuals at low risk of developing local or distant metastasis [310]. While known complications of total thyroidectomy (permanent hypoparathyroidism, recurrent laryngeal nerve damage and vocal cord paresis) are uncommon, they are not negligible [7] and therefore may affect the decision for pursuing a more aggressive surgical course. The aim of this study was to examine the correlations between disease laterality and MACIS score and to identify which, if any, clinicopathological factors are associated with bilateral thyroid cancer. If unilateral disease is found to pose a lower risk to the individual, less extensive surgery could be performed in these cases.

We reviewed data for 203 patients with papillary thyroid cancer who were treated with either total thyroidectomy or completion thyroidectomy at St. Paul's Hospital, Vancouver, British Columbia, Canada, between 2000 and 2012. All patients in the cohort presented with one or more thyroid nodules and had an ultrasound-guided FNA biopsy. Patients were then referred to one of three head and neck surgeons and following resection, specimens were analyzed and a formal diagnosis of PTC was given. It is the standard of care at St. Paul's Hospital to perform total thyroidectomy in the setting of PTC. In the event of indeterminate FNA results, thyroid

lobectomy may be performed; patients then proceed onto completion thyroidectomy if PTC is identified on pathology. Demographic factors of gender and age and histopathologic characteristics including disease in the contralateral lobe (bilaterality), presence of multifocal disease, size of tumor, presence of extrathyroidal invasion, vascular invasion, nodal or distant metastases and completeness of resection were recorded. A Pearson chi-squared or Fisher's exact test, where appropriate, was used to determine if differences between the bilateral and unilateral groups were significant with respect to clinicopathological characteristics. *P* values were corrected for multiple testing using the Benjamini-Hochberg correction [145]. All statistical tests were two-tailed and a *P* value of < 0.05 was considered statistically significant. Multivariate logistic regression was also performed, considering all covariates. All analyses were done using scripts written in the R programming language (version 3.1.1, R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria).

Only 2 of the studied variables demonstrated a correlation with bilaterality – smaller tumor size ($P < 0.0001$) and presence of vascular invasion ($P < 0.0001$). The rest of the clinicopathological variables and the MCIS score did not correlate with laterality with any significance. Eighty two (40.4%) patients had bilateral disease, and 121 (59.6%) had unilateral disease; the analysis demonstrated that all 121 patients with unilateral disease had tumor sizes > 1 cm; however, of the 82 patients with bilateral disease, 22 (26.8%) had a tumor size ≤ 1 cm. A total of 16 patients (19.5%) with bilateral disease had evidence of vascular extension compared to only 2 (1.7%) patients with unilateral disease. MACIS scores for each patient were calculated and stratified

into categories of low grade (score < 6), middle-low grade (6-6.99), middle-high (7-7.99) or high grade (≥ 8). Most patients had low MACIS scores (< 6) for both bilateral disease (65 patients, 82.3%) and unilateral disease (92 patients, 76.7%). When scores were created as a binary variable (<7 vs ≥7), the number of patients with higher scores did not differ significantly between the bilateral and unilateral groups – 7 (8.9%) and 11 (9.2%), respectively.

Multifocality is defined as the presence of more than one site of disease and by definition all patients with bilateral disease were considered to have multifocal disease. By comparison, only 3 patients in the unilateral group (2.5%) showed any evidence of multifocal disease within the same lobe. Multifocality was further studied to determine if the main site of multifocal disease in bilateral cases was ipsilateral or contralateral to the dominant tumor. One of 82 patients with bilateral disease was not included in this analysis since detailed pathology reports were lacking. Interestingly, the majority of patients in the bilateral cohort had their main site of multifocality within the lobe ipsilateral to that of the dominant tumor ($P = 0.03$). This finding may suggest that these tumors tend to spread locally first within the ipsilateral lobe before appearing in the contralateral lobe, thereby giving rise to bilateral disease. This observation would also support the notion of step-wise accumulation of mutations and tumor evolution over time rather than the appearance of multiple unrelated tumors in the gland.

Overall, our results demonstrate very few correlations between poor prognostic factors and bilateral disease. Bilaterality was associated with smaller tumor size and vascular invasion; the

association between bilaterality and vascular invasion suggests that they are possibly more locally aggressive within the gland itself, rather than having a higher preponderance towards systemic spread. This may also be supported by the fact that cases of bilateral disease demonstrated a significantly higher incidence of multifocality within the ipsilateral lobe. The association with smaller tumor size suggests that bilaterality is an early event in thyroid tumor progression, and perhaps earlier than other forms of tumor progression. It is possible that bilateral disease occurs because of vascular invasion, indicating that it may result from intra-thyroidal metastases. However, whether this is the case, or whether bilaterality is a result of multiple primary tumors requires further study to understand the molecular behavior of these lesions.

5.3 Diagnostic Markers for Papillary Thyroid Carcinoma

Tissue microarray (TMA) analyses of several markers, alone or combination were performed on a large cohort of benign thyroid nodules and PTCs with the aim of finding diagnostic biomarkers for routine use in the clinic. TMA construction, staining and scoring were performed as described before [311,312]. Briefly, two sets of benign thyroid lesion TMAs, one composed of 100 specimens and the other of 236, and two sets of malignant TMAs, one with 99 DTCs and the other 242, were prepared from archival pathology specimens of patients. Clinical, cytologic and pathologic data were available for all specimens and our Institutional Research Ethics Boards had approved the use of all tumors and clinical data for this study. A Leica microtome (Leica Microsystems, Richmond Hill, Ontario, Canada) was used to cut serial 4-µm sections from

the TMA blocks that were transferred onto adhesive-coated glass slides for immunohistochemistry. Sections were then de-paraffinized and antigen retrieval was performed. Antibodies were optimized for thyroid tissue according to the manufacturer's instructions and appropriate positive and negative controls were used for each antibody. Two pathologists, blinded to all clinical data, examined the stained TMA sections at high-power magnification to determine the proportion of cells expressing the markers. The scoring systems used were based on previously published reports of immunohistochemical studies evaluating these markers, and are summarized in Table 5.1. The correlation of clinicopathological characteristics (patient age, gender, tumor size, presence or absence of vascular invasion, completeness of cancer resection, presence of extrathyroidal extension, American Joint Committee on Cancer T, N and M stages and MACIS score) with expression or co-expression of markers, and the significance of marker expression in malignant versus benign tissues was assessed using contingency table statistics (Pearson χ^2 or the Fisher exact test, where appropriate for categorical variables and the Mann-Whitney U test for continuous variables) using scripts written in the R programming language (version 2.4.1, R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria). The analysis was run with both categorical and semi-quantitative marker scorings. Two marker score categories were analyzed; in ''grouping 1,'' marker scores were grouped as either ''negative'' (score = 0) or ''positive'' (score >= 1). In ''grouping 2,'' marker scores were grouped as either ''negative/weak'' (score = 0 or 1) or ''moderate/strong'' (score >= 2). *P* values were corrected for multiple testing using the Benjamini-Hochberg (BH) correction [145]; all tests were 2-tailed and a *P* value of less than 0.05

was considered statistically significant.

Alterations of normal cell shape and scaffolding and cell anchorage to the extracellular matrix are considered the main events leading to tumor metastasis. Expression of several proteins which function in maintaining normal cellular structure including Cav1, Gal3 and CK19 were tested in the above described TMAs. Cav1 is a membrane protein and a major component of caveolae; in addition, it is involved in a variety of cellular signal transduction pathways and can act as both a tumor suppressor and an oncogene depending on the tissue type [313]. This protein was shown to promote cell polarization and promote focal adhesion turnover in thyroid cells and hence may be a key player in inducing metastasis in thyroid tumors [314]. CK19 is a member of the keratin family of proteins and is a constituent of the cytoskeleton [315]; its overexpression has been associated with various cancers and it may serve as a diagnostic or prognostic marker for thyroid cancer [316]. Gal3 is a carbohydrate binding galectin, which forms an extracellular lattice facilitating receptor tyrosine kinase signaling [314]. Given the important roles of MAPK and PI3K signaling pathways in thyroid cancer initiation and progression [317] and the contributing role of Gal3 to PTC phenotype maintenance [318], this protein was considered as a promising diagnostic marker candidate for thyroid cancer.

5.3.1 Cav1 and Gal3 Immunohistochemical Analysis

The study cohort TMAs were composed of human specimens from 100 benign thyroid lesions (26 follicular adenomas, 54 goiters, three cases of Hashimoto's thyroiditis, 10 Hürthle cell adenomas, four hyperplastic nodules and three cases of lymphocytic thyroiditis) and 99 sporadic DTCs (90 PTCs, six follicular thyroid cancers and three Hürthle cell carcinomas). A significantly higher proportion of DTCs either expressed Gal3, alone or in conjunction with Cav1, compared to benign lesions (Table 5.2). Individually, Gal3 and Cav1 were expressed in 83.7% and 51.5% of DTC cases, demonstrating significantly increased expression in DTCs (Gal3, 83.7% versus 5.05%, $P < 0.001$; Cav1, 51.5% versus 10.1%, $P < 0.001$) compared to benign thyroid lesions. Co-expression of Gal3 and Cav1 was significantly increased in DTCs compared to benign thyroid lesions, and the majority of Cav1-expressing DTCs also expressed Gal3. Overall, the utility of Gal3 and Cav1 co-expression for clinical diagnostic purposes has an accuracy, sensitivity, specificity and precision of 74.5%, 48.98%, 100% and 100%, respectively. Evaluation of the clinicopathological characteristics of the DTC cohort and the expression of Gal3 and Cav1 showed a statistically significant correlation only between Gal3 and Cav1 expression and papillary DTC pathology. The extended *in vitro* study, which followed the TMA analysis demonstrated the coordinated expression of these two proteins to be a major player in driving the papillary thyroid cancer cell migration [314].

5.3.2 CK19 and Gal3 Immunohistochemical Analysis

The co-expression of CK19 and Gal3 proteins and their diagnostic utilities were examined in DTCs and benign thyroid tissues. The correlation of clinicopathological characteristics with the expression of these markers in the DTC specimens was also studied. The study cohort included 236 patients diagnosed with benign thyroid lesions and 254 patients diagnosed with thyroid malignancies. After excluding the 12 patients with medullary thyroid cancer (MTC) that is derived from a distinct cell type than follicular cells, 478 patients were remained in the study cohort. The expression of both Gal3 and CK19 was found to be higher in DTCs compared with benign lesions (Gal3, 77.4% versus 6.5%, $P < 0.001$; CK19, 74.4% versus 11.2%, $P < 0.001$). Higher proportion of malignant samples also showed co-expression of Gal3 and CK19 (Table 5.3). The utility of Gal3 and CK19 co-expression for diagnostic purposes has an accuracy, sensitivity, specificity and precision of 81.9%, 66.1%, 98.3% and 97.5%, respectively. Evaluating the correlation between the clinicopathological characteristics of the DTC cohort and the expression of CK19 and Gal3 showed that the expression of CK19 in the absence of Gal3 (CK19+Gal3-) demonstrates correlation with the absence of lymph node metastasis ($P < 0.001$) and N0 stage ($P < 0.001$); the expression of Gal3 alone (CK19-Gal3+) does not show any correlation with any clinicopathological characteristics. However, the co-expression of CK19 and Gal3 (CK19+Gal3+) shows correlation with papillary thyroid cancer pathology ($P < 0.000$), the presence of lymph node metastasis ($P < 0.000$), extra thyroidal extension ($P = 0.004$), smaller

tumor size (P = 0.005) and N1 stage (P < 0.000). This is indicative of perhaps a role for Gal3 in tumor aggressiveness but only in synergy with other cellular players such as CK19.

5.4 Whole Genome Profiling of Benign Thyroid Nodules

In this study, we profiled the common benign thyroid nodules on the whole genome scale with the aim of identifying alterations that might play a causal role in benign tumorigenesis. It is not yet understood if step-wise accumulation of mutations in benign tumors could and do lead to a malignant state. *RET/PTC* rearrangements and *RAS* mutations, both found in 10-20% of PTCs, are also observed in 10-45% and 20-40% of thyroid adenomas, respectively [9]. This might be suggestive of a pre-cancerous state in at least a subset of benign tumors giving rise to follicular variants of PTCs and FTCs. This is the first study to date to provide a comprehensive genomic profile of benign thyroid tumors; whole genome sequencing allowed us to identify regions of copy number loss and gain with base-level precision. Novel large-scale rearrangements and gene fusions were identified through both *de novo* and alignment-based methods.

5.4.1 Materials and Methods

5.4.1.1 Study Samples

Biopsy specimens for whole genome sequencing experiments were collected from three patients diagnosed with benign thyroid nodules; a 67-year old female diagnosed with follicular adenoma (F67FA), a 46-year old female diagnosed with follicular adenoma (F46FA) and a third tumor from a 55-year old male diagnosed with goiter (M55G). Adjacent matched normal tissue was also collected from each patient for sequencing and these served as the control specimens. The tumor samples were collected as part of a research project approved by the British Columbia Cancer Agency's Research Ethics Board are in accordance with the Declaration of Helsinki. The tumor samples were classified according to the World Health Organization criteria.

5.4.1.2 DNA Sequencing

DNA extracted from the frozen tumor and normal tissues were subjected to high-throughput whole genome sequencing using locally established sequencing protocols. Biopsy specimens were embedded in Tissue-Tek O.C.T. (optimal cutting temperature) compound (Sakura Finetek USA, Inc.) and sectioned for DNA extraction. Using 1ug DNA each from each sample, six whole genome libraries were constructed using a modified version of Illumina TruSeq PCR free protocol (FC-121-3001) as described in Chapter 3. Paired-end 100bp reads were generated on

Illumina HiSeq2500 sequencers following the manufacturer's protocol with minor variations. Software version HCS1.5.8 was utilized (Table 5.4).

5.4.1.3 Bioinformatic Analysis

Sequence reads from the whole genome libraries were aligned to the human reference genome (build hg19) using the Burrows-Wheeler Alignment (BWA) tool [56]. The tumor's genomic sequence was compared to that of normal tissue DNA to identify somatic alterations. Regions of copy number variation (CNV) and loss of heterozygosity (LOH) were identified using Hidden Markov model-based approaches HMMcopy and APOLLOH [95], respectively. Single nucleotide mutations were identified using a probabilistic joint variant calling approach utilizing SAMtools and Strelka [64,75]. Small insertions and deletions (indels) were identified using Strelka [75]. *De novo* assembly and annotation of genomic data using ABySS [62] and Trans-ABySS [87,139] were used to identify small indels, structural variants and fusion genes.

5.4.2 Results

The F67FA, F46FA and M55G tumors harbored 22, 8 and 10 somatic SNVs and indels, respectively (Tables 5.5, 5.6 and 5.7). These indicate a much lower mutation rate in benign thyroid nodules compared with malignant tumors derived from the follicular cells of the gland

(Chapters 3 and 4). No recurrent mutations or those in known cancer drivers were identified. Of particular interest however were a frameshift insertion in *RPTOR*, a member of the mTOR signaling pathway and an inhibitor of the mTOR kinase, a non-sense mutation in *TRIM16*, a regulator of the cell cycle and cell proliferation inhibitor [319] and a two-codon deletion in ribunuclease 3 domain of *DICER1* in F67FA; a splice site donor mutation of *KIF1B*, potentially implicated in neural crest-derived tumors such as pheochromocytomas and neuroblastomas [320] was identified in F46FA. It is noteworthy that the two follicular adenomas demonstrated such mutations but not the goiter specimen indicating that perhaps these diseases although all falling under the umbrella of benign thyroid disease are subjected to different disease pathways. Small single nucleotide mutations of the thyroglobulin gene (*TG*) have previously been found in patients with nodular goitre [321,322]. Thyroglobulin acts as a substrate for the synthesis of thyroid hormones T3 and T4. Although M55G did not harbor any somatic SNVs or indels in *TG*, *de novo* assembly of the sequence data revealed the presence of a translocation between chromosomes 8 and 19, with one of the breakpoints in *TG* gene (Table 5.8) potentially leading to the loss of function of this gene. Table 5.8 also lists 5 large-scale structural events that were identified in F67FA including a gene fusion between *DICER1* and *NTNG2*. As mentioned above, this patient also harbored an in-frame deletion in *DICER1*, a gene with integral role in miRNA processing and synthesis. The indel and the translocation breakpoint are situated over 53Kb apart and hence it is not possible to deduce if they affect one or both alleles of the gene from the sequence data.

The copy number and loss of heterozygosity analyses revealed a striking difference between these three benign tumors. Although F46FA and M55G had relatively quiet genomes with 1 copy loss of 2q and 9q in F46FA and a small region of 1 copy loss in 2p in M55G (Figures 5.1 and 5.2), the F67FA tumor demonstrated several large regions of gene copy loss and gain (Figure 5.3). These included one copy loss of a region of 1q, 3q, 4p, 15q, 16p, 16q, 17p, 20p, 20q, 22q and the entire chromosome 13; chromosome 18 showed high amplification in this follicular adenoma specimen. Not only is the difference between these benign nodules striking, but also given the very quiet genomes of papillary thyroid carcinomas [270], it was perhaps unexpected for a benign nodule to demonstrate such large-scale alterations with respect to gene copy numbers. It is also intriguing that the two follicular adenomas show very different profiles.

5.4.3 Conclusion

The three benign thyroid nodules demonstrated vast differences in their genomic profiles. While all harbored a small number of somatic SNVs and indels, larger alterations of gene copies were observed. 70 protein coding genes on chromosome 2 had lost one copy in the goiter specimen, 722 genes on chromosome 2 and 541 genes on chromosome 9 lost one gene copy in the follicular adenoma specimen, F46FA; these included tumor suppressors *TSC1* and *PTCH1*. Proto-oncogenes with additional gene copies in F67FA specimen included *SS18*, *BCL2* and *YES1*. Tumor suppressors with 1 copy loss included *BRCA2*, *RBL1, SMARCB1*, *AXIN1*, *TSC2*, *RBL2*, *TP53*, *BUB1B*, *CHEK2* and *NF2* (Figure 5.3). This adenoma specimen, F67FA, and the goiter also

appeared to have lost all mitochondrial content. Partial loss of TP53 in the F67FA follicular adenoma specimen is particularly of interest given that loss of this protein, often through point mutations, is unique to anaplastic forms of thyroid cancer. This was confirmed by our observations in Chapter 4 where *TP53* was the only recurrently mutated gene in the examined ATC samples. The vast amount of changes in copy number might also be explained through the loss of one copy of this tumor suppressor as observed in ATCs and explained in chapter 4. In addition, *SS18* copy gain and loss of *AXIN1* and *TSC2* copies, all found to be mutated in ATCs (see Chapter 4), also suggest that this benign tumor might have been a precursor to an anaplastic thyroid cancer. This patient had undergone a thyroidectomy to remove the benign tumor and as such she is not expected to develop any further disease. However, continued monitoring of this patient would be recommended bases on the genomic analysis.

The Cancer Genome Atlas (TCGA) initiative has conducted an extensive study of over 400 papillary thyroid carcinomas; the biomarker publication has described a very low mutation rate in PTCs with a small percentage of tumors demonstrating somatic copy number variations [270]. Gene fusions involving known PTC drivers such as *RET* and *BRAF* were also identified. Comparison of the genomic data from the current study examining benign nodules with that of the TCGA PTC study revealed no common mutations or gene fusions among these tumors. This observation provides supporting evidence for the hypothesis that not all benign nodules become malignant through accumulation of mutations over time. Alternatively, given the finding from TCGA study that the majority of PTCs only harbor one driver mutation, it is

plausible that acquiring only one such mutation is sufficient for transforming a benign tumor to a papillary malignant phenotype. Profiling larger cohorts of benign tumors will shed light on the association between these various entities and the mode of tumor evolution.

This study was limited in that it only examined 3 specimens of two different pathologies; hence the variables such as age and sex, in addition to pathology, were not controlled for. Another potential limitation of this study was the use of adjacent matched tissue as the normal control. Although two independent pathologists confirmed that these were indeed normal thyroid tissues, it is possible that somatic driver mutations are present in tissues, which appear "normal" to pathologists resulting in incorrect findings and conclusions on somatic alterations.

5.5 Transcriptomic Comparison of Benign Thyroid Nodules and Papillary Thyroid Carcinoma

The transcriptome provides a snapshot of the cell population dynamics in a tumor specimen. Not only expressed mutations such as SNVs and SVs can be detected, changes in the levels of all transcripts can also be identified. Moreover, events such as novel transcripts, splicing and polyadenylation sites can be detected. As a result, we next aimed to compare the transcriptomic landscapes of benign thyroid nodules and papillary thyroid carcinomas with the aim of identifying recurrent mutations, differentially active or silenced pathways and discriminative biomarkers between the two disease groups.

5.5.1 Materials and Methods

5.5.1.1 Study Samples

RNA for the sequencing experiment was extracted from 19 benign thyroid nodules (distinct from the 3 tumors described in section 5.4 above) and 10 papillary thyroid carcinoma specimens. All patients had gone through total or partial thyroidectomy and had provided written informed consent for the complete profiling of their tumor specimens. Table 5.9 lists patient characteristics and tumor pathologies.

5.5.1.2 RNA Sequencing

In order to construct transcriptome libraries, RNA was extracted from 15 x 20 µm sections cut from flash-frozen tissue using MACS mRNA isolation kit (Miltenyi Biotec), resulting in 5-10 µg of DNase I-treated total RNA as per the manufacturer's instructions. Double-stranded cDNA was synthesized from the purified poly(A)$^+$ RNA using the Superscript Double-Stranded cDNA Synthesis kit (Invitrogen) and random hexamer primers (Invitrogen) at a concentration of 5 µM. The cDNA was fragmented by sonication and a paired-end sequencing library prepared following the Illumina paired-end library preparation protocol (Illumina). Cluster generation and sequencing were performed on the Illumina HiSeq2000 following the manufacturer's recommended protocol (Illumina Inc., Hayward, CA) (Table 5.10).

5.5.1.3 Bioinformatics Analysis

The sequence data were aligned to the human reference genome (build hg19) using TopHat 2.0.6 [323]. The reference sequence and the corresponding annotation files were provided by Illumina's iGenome project and downloaded from the TopHat homepage (http://tophat.cbcb.umd.edu/igenomes.shtml). Quantification of gene expression was accomplished using HTSeq-0.5.4p3 in intersection-nonempty mode [324], all subsequent analyses were run using the count values for the protein-coding elements only. The generated read counts were used as input in the R package edgeR v.3.4.0 [274] for differential gene expression analysis; reads with quality less than 10 were discarded from differential expression analyses. *De novo* assembly and annotation of sequence data using ABySS [62] and Trans-ABySS [87,139] were used to identify structural variants and gene fusions. Only those events also identified by the alignment-based fusion detection software Minimum Overlap Junction Optimizer (MOJO) (https://github.com/cband/MOJO) were considered to be true positives.

5.5.2 Results

We sequenced the mRNA of 19 benign and 10 malignant thyroid tumors using next generation sequencing technologies. An average of 160M 75bp paired-end reads were generated for each sample. Sequence reads were aligned to the human reference genome (build hg19); on average 137.5M reads were mapped to the reference for each sample with 84% of read pairs having the

expected insert size and orientation. The only recurrent expressed single nucleotide variant was the BRAF p.V600E mutation in 8 out of 10 papillary cancer samples. In addition, 4 benign tumors harbored activating RAS mutations; a follicular adenoma with p.Q61R NRAS mutation, one with p.Q61R HRAS mutation and two goiters one with p.G13R HRAS mutation and the other p.G12D KRAS activating mutation. No recurrent gene fusions or fusions involving known thyroid or cancer genes were identified. Differential gene expression analysis identified 867 upregulated and 324 downregulated genes with fold change >=4 and <=-4, respectively (Benjamini-Hochberg $P < 0.05$) in papillary carcinomas compared with the benign nodules. A Heatmap of these 1191 differentially expressed genes is depicted in Figure 5.4. Pathway analysis using these genes as input was performed with DAVID (Database for Annotation, Visualization and Integrated Discovery) [325] and the KEGG (Kyoto Encyclopedia of Genes and Genomes) knowledge base [326]. Pathways enriched with the downregulated genes included steroid hormone biosynthesis, hedgehog and PPAR signaling pathways. Those enriched with PTC overexpressed genes included cell adhesion molecules, cytokine-cytokine receptor interaction, chemokine signaling pathway and various networks related to the role of immune system including graft-versus-host disease, type I diabetes mellitus, primary immunodeficiency and autoimmune thyroid disease, to name a few. This is not a surprising finding given the extensive lymphocytic infiltration observed in PTCs and reported in pathology reviews of the samples examined in this study. Figure 5.5 depicts the expression of select differentially expressed genes with the least overlap in expression values between benign and malignant tumors. *CDON* and *SLC4A4* genes show downregulation in PTCs when compared to benign

tumors. CDON is a cell surface receptor that is a member of the immunoglobulin superfamily and its loss of function may play a role in oncogenesis [327]. SLC4A4, a sodium bicarbonate cotransporter, regulates intracellular pH levels [328]. A Polish study found the expression of SLC4A4 to be higher in PTCs than normal thyroid tissue [329] and thus the even higher expression observed in benign tumors of the current study might suggest an active and significant role of this gene in benign tumorigenesis. *CTSH*, *CYP1B1*, *PTPRE* and *RUNX1* showed higher expressions in PTCs compared with benign tumors. All these genes and their protein products have shown associations with malignant phenotypes. Increased expression of *CTSH*, a lysosomal proteinase, was observed to cause prostate cancer cell migration and disease progression [330]. PTPRE is a protein tyrosine phosphatase that, when overexpressed, leads to overexpression and activation of ERK1/2 and AKT in human breast cancer cells [331]. The transcription factor RUNX1 may act as both a tumor suppressor and an oncogene depending on the tissue type; its expression has been correlated with poorer prognosis in triple negative breast cancers [332,333]. CYP1B1 is a member of the cytochrome P450 whose overexpression was found in malignant tumors of various organs such as breast, colon, lung, esophagus, skin, lymph node, brain and testis; the expression was specific to the tumors and missing from the matched adjacent normal tissues [334]. Moreover, upregulation of IL-6, a pro-inflammatory cytokine, in colorectal cancers leads to overexpression of CYP1B1 [335], perhaps providing a suitable diagnostic marker for thyroid cancers which are associated with chronic inflammation. The mRNA expression level of these genes or their protein levels, contingent on further studies, can be used as diagnostic markers for papillary thyroid cancer.

De novo assembly of sequence reads allowed for identification of events such as novel 5' and 3' splice sites leading to skipped exons, retained introns and novel transcript start sites or end positions. We identified a novel 5' splice site in *SLC34A2* gene, causing a deletion of 30 amino acids from exon 9 in 6 PTC samples and none of the benign tumors. Figure 5.6 depicts a schematic of this deletion as well as the alignment of assembled contigs from all 6 tumors to the reference genome. Since this novel event appeared to be specific to the malignant samples, we hypothesized that this resultant transcript and in particular the 30-amino acid deleted region can serve as a diagnostic marker. Moreover, SLC34A2 is a membrane transporter molecule with the amino acids 310 to 339 located in the extracellular domain [336] and thus specific antibodies for this region may be of clinical use. PCR primers were designed to amplify and validate this observation (Table 5.11). One pair was designed such that they flanked the deletion; two bands would be expected in this case, one corresponding to the original and longer transcript and one shorter band corresponding to the transcript with the deletion. The other primer pair was designed such that one primer spanned the novel breakpoint joining exons 8 and 9 and the other flanking it. This pair would only result in a band if the novel sequence, which was unique in the human reference genome, was present in the sample. Figure 5.7 is an image of the cDNA PCR products from the 6 PTC samples that were computationally found to have the novel splice site and an additional 6 benign tissues. It is evident from the image that the expression of both wild type and novel *SLC34A2* transcripts are higher in PTCs compared with benign tissues. PCR products in lanes 1 and 2 for each sample (products of flanking primers) are dominated by the wild type and it is not clear if the novel

transcript is present at all. However, PCR products are seen in lane 4 where we would expect to see a band only if the unique and chimeric sequence is present. These observations collectively suggest that the transcript with the 30-amino acid deletion is present in the cell but at a much lower expression level compared with the wild type transcript(s). The low expression level may pose challenges for the utility of antibodies for diagnostic purposes; however, the use of sensitive tools such as digital PCR techniques can provide avenues for novel diagnostic markers such as this *SLC34A2* variant.

5.5.3 Conclusion

Benign thyroid nodules are commonly found in the population but the molecular alterations leading to these tumors are not yet understood. Our analysis of a small group of these nodules revealed a very low mutation rate with mostly quiet genomes harboring minimal copy number changes. Papillary thyroid carcinoma represents the most common form of thyroid cancer; it has a favorable prognosis in the majority of patients with 25-year overall survival rate estimated at 80-90%. Analysis of our in-house data and those of The Cancer Genome Atlas study revealed a very low mutation rate even in these malignant tumors. The lack of such genomic and particularly protein-coding alterations might be indicative of a causative role for the epigenomic processes in thyroid tumorigenesis. Examining the epigenome utilizing the now available technologies with base-level sensitivity such as whole genome bisulfite sequencing, and ChIP-Seq analysis of chromatin markers will be the next step in deciphering the molecular

biology of these tumors. Deregulation of miRNAs is also a well-known contributor to the malignant phenotype in a variety of cancer types. The Cancer Genome Atlas study of papillary thyroid cancers found differential expression of some of these miRNAs between the malignant and normal thyroid tissues as well as between different clusters of PTCs derived based on the methylation profiles and/or mutational spectrums [270]. miRNA sequencing of benign and papillary tumors and their comparison may provide clues about the mechanism of disease initiation and progression.

Despite the mostly great prognosis for patients with thyroid tumors, the immediate clinical need still remains in the process of diagnosis. A large subset of patients with “indeterminate” thyroid cytopathology will undergo thyroidectomy while the histopathology review finds the tumor to be benign after surgery. Although there is some evidence that chronic benign tumors may lead to malignancy over time and particularly to those with the most aggressive behavior such as ATCs, avoiding unnecessary surgeries and monitoring the benign disease over time could eliminate patient anxiety associated with surgery and also lower health care costs.

Table 5.1 Antibody characteristics and the scoring system used for each marker

Marker Name	Isotype	Company	Antigen Retrieval	Concentration	Localization	Scoring System
Caveolion 1 (CAV1)	Rabbit polyclonal	Santa Cruz Biotechnology	Heat induced	1:1000	Membrane	3+=>75% of cells positive 2+=26-75% of cells positive 1+=5-25% of cells positive 0=<5% of cells positive
Cytokeratin 19 (CK19)	IgGk1	Dako	PC8 citrate	1:100	Cytoplasm	3+=>75% of cells positive 2+=26-75% of cells positive 1+=5-25% of cells positive 0=<5% of cells positive
Galectin-3 (Gal3)	IgG1	Vector Laboratories, Burlingame, California	S20 EDTA	1:250	Cytoplasm	3+=Strong 2+=Moderate 1+=Weak 0=Negative

Table 5.2 Percentage of benign and DTC samples expressing Cav1, Gal3 and their co-expression

Expression	Benign (%)	DTC (%)
Cav1-positive	10.1	51.5
Gal3-positive	5.1	83.7
Cav1-positive and Gal3-positive	0	49
Cav1-positive and Gal3-negative	10.2	3.1
Cav1-negative and Gal3-positive	5.1	34.7
Cav1-negative and Gal3-negative	84.7	13.3

Table 5.3 Percentage of benign and DTC samples expression CK19, Gal3 and their co-expression

Expression	Benign (%)	DTC (%)
CK19-positive	11.2	74.4
Gal3-positive	6.5	77.4
CK19-negative and Gal3-negative	84	13.8
CK19-negative and Gal3-positive	4.8	11.3
CK19-positive and Gal3-negative	9.5	8.8
CK19-positive and Gal3-positive	1.7	66.1

Table 5.4 Sequence libraries read statistics

	Total Number of Reads	Number of Aligned Reads	Average Coverage
F67FA tumor genome	1297569058	1128514273	38.6
F67FA blood genome	1357133174	1109232518	38
F46FA tumor genome	1302977538	1127562684	38.6
F46FA blood genome	1386382678	1178621340	40.3
M55G tumor genome	1337648720	1131656888	38.6
M55G blood genome	1437793306	1146850574	39.1

Table 5.5 F67FA somatic SVNs and indels

Chr	Position	Reference allele	Alternate allele	dbSNP/COSMIC ID	Effect type	AA change	Gene	EnsEMBL Gene ID
1	169493094	T	C	-	Non-synonymous	N1946S	F5	ENSG00000198734
1	197091560	C	A	-	Non-synonymous	D1186Y	ASPM	ENSG00000066279
4	160266304	C	G	-	Non-synonymous	Q948E	RAPGEF2	ENSG00000109756
5	110440043	G	A	-	Non-synonymous	A300T	WDR36	ENSG00000134987
9	5921851	A	-	-	Frame-shift	L1382Wfs	KIAA2026	ENSG00000183354
10	73498319	C	T	-	Non-synonymous	A1428V	CDH23	ENSG00000107736
11	102826408	G	T	-	Non-synonymous	F9L	MMP13	ENSG00000137745
11	64083331	C	T	rs80310817	Non-synonymous	R388C	ESRRA	ENSG00000173153
12	2566818	G	A	-	Non-synonymous	A235T	CACNA1C	ENSG00000151067
13	111102778	T	G	-	Non-synonymous	L439R	COL4A2	ENSG00000134871
14	95560472	AATTCT	-	-	Codon-deletion	LEF1704F	DICER1	ENSG00000100697
16	74425826	G	A	COSM472094	Non-synonymous	A258T	NPIPL2	ENSG00000196436
16	8994397	C	T	-	Non-synonymous	V668I	USP7	ENSG00000187555
17	15554473	G	A	-	Stop-gained	Q151*	TRIM16	ENSG00000221926
17	40936511	G	C	-	Non-synonymous	G362R	WNK4	ENSG00000126562
17	71334761	G	A	-	Non-synonymous	R1319W	SDK2	ENSG00000069188
17	78935240	-	C	-	Frame-shift	L1061Pfs	RPTOR	ENSG00000141564
18	20793945	G	C	-	Non-synonymous	V12L	CABLES1	ENSG00000134508
18	21419818	G	T	-	Non-synonymous	R1087S	LAMA3	ENSG00000053747
MT	12889	G	A	-	Non-synonymous	A185T	MT-ND5	ENSG00000198786
MT	13069	G	A	COSM488740	Non-synonymous	A245T	MT-ND5	ENSG00000198786
X	103267905	C	G	-	Non-synonymous	V110L	H2BFWT	ENSG00000123569

Table 5.6 F46FA somatic SVNs and indels

Chr	Position	Reference allele	Alternate allele	dbSNP/COSMIC ID	Effect type	AA change	Gene	EnsEMBL Gene ID
1	10381915	G	T	-	Non-synonymous	K740N	KIF1B	ENSG00000054523
1	10381916	G	T	-	Splice-site-donor	-	KIF1B	ENSG00000054523
1	16388642	G	A	rs79991837	Non-synonymous	R74C	FAM131C	ENSG00000185519
1	248487576	T	A	-	Non-synonymous	T99S	OR2M7	ENSG00000177186
11	71238675	C	G	rs200832929	Non-synonymous	S110C	KRTAP5-7	ENSG00000244411
15	92647683	G	A	-	Non-synonymous	R100K	SLCO3A1	ENSG00000176463
X	12937599	C	A	-	Non-synonymous	S147Y	TLR8	ENSG00000101916
X	106888559	C	G	-	Non-synonymous	T128R	PRPS1	ENSG00000147224

Table 5.7 M55G somatic SNVs and indels

Chr	Position	Reference allele	Alternate allele	dbSNP/COSMIC ID	Effect type	AA change	Gene	EnsEMBL Gene ID
4	6107649	C	A	-	Stop-gained	E59*	JAKMIP1	ENSG00000152969
11	65810306	T	A	-	Non-synonymous	E323V	GAL3ST3	ENSG00000175229
12	122691456	G	A	COSM159311	Non-synonymous	D195N	B3GNT4	ENSG00000176383
14	51404528	A	C	-	Non-synonymous	Y91D	PYGL	ENSG00000100504
14	52986004	G	A	-	Non-synonymous	L134F	TXNDC16	ENSG00000087301
16	4702722	C	-	-	Frame-shift	R115Gfs	MGRN1	ENSG00000102858
19	24115932	G	C	-	Non-synonymous	E338D	ZNF726	ENSG00000213967
21	47754549	G	A	rs61735823	Non-synonymous	R169H	PCNT	ENSG00000160299
X	82763964	C	T	-	Non-synonymous	T211M	POU3F4	ENSG00000196767
MT	13039	T	C	-	Non-synonymous	S235P	MT-ND5	ENSG00000198786

Table 5.8 Somatic translocations and gene fusions in F67FA and M55G
No events were found in F46FA. Coordinates are based on the hg19 human genome assembly

Patient	Event	Breakpoint 1	Gene 1	Breakpoint 2	Gene 2
F67FA					
	Translocation	chr15:67795174	-	chr16:6537736	RBFOX1
	Translocation	chr18:77197367	NFATC1	chr22:42642302	-
	Translocation	chr1:244547178	C1orf100	chr20:11928844	-
	Translocation	chr4:17866584	LCORL	chr16:47477978	ITFG1
	Translocation	chr9:135096660	NTNG2	chr14:95614287	DICER1
M55G					
	Translocation	chr8:134009359	TG	chr19:54093641	-

Table 5.9 Characteristics of 19 benign thyroid nodules profiles using RNA-seq

Patient ID	Sex	Age	Index Lesion Pathology	Background Pathology	Papillary Carcinoma Subtype	Hotspot Mutations
WT017	M	44	Follicular Adenoma	Hashimoto's Thyroiditis		-
WT049	F	51	Follicular Adenoma	Lymphocytic Thyroiditis		-
WT075	F	27	Follicular Adenoma	None		NRAS (Q61R)
WT091	F	45	Follicular Adenoma	None		-
WT119	F	32	Follicular and Hurthle Cell Adenoma	None		HRAS (Q61R)
WT015	M	59	Goiter	None		-
WT055	F	78	Goiter	Lymphocytic Thyroiditis		-
WT061	F	52	Goiter	None		HRAS (G13R)
WT079	F	34	Goiter	None		-
WT083	F	65	Goiter	None		-
WT127	M	57	Goiter	None		KRAS (G12D)
WT025	F	55	Goiter	Hurthle Cell Metaplasia		-
WT095	F	44	Goitre	None		-
WT099	M	51	Goitre	None		-
WT077	F	61	Hurtle Cell Adenoma	Goiter		-
WT019	F	50	Hyperplastic Nodule	Goiter		-
WT073	F	51	Hyperplastic Nodule	Lymphocytic Thyroiditis		-
WT037	F	19	Hyperplastic Nodule	None		-
WT051	F	37	Toxic Nodule	Goiter		-
WT107	F	28	Papillary Thyroid Carcinoma	None	Classic	BRAF (V600E)
WT123	F	54	Papillary Thyroid Carcinoma	Follicular Hyperplasia and Chronic Thyroiditis	Follicular	BRAF (V600E)
WT125	F	55	Papillary Thyroid Carcinoma	Hashimoto's Thyroiditis	Classic	BRAF (V600E)
WT045	F	35	Papillary Thyroid Carcinoma	None	Classic	-
WT069	F	74	Papillary Thyroid Carcinoma	Lymphocytic Thyroiditis	Classic	BRAF (V600E)
WT071	F	49	Papillary Thyroid Carcinoma	Hurthle Cell Metaplasia	Classic	BRAF (V600E)
WT033	F	45	Papillary Thyroid Carcinoma	Goiter	Mixed	-
WT001	M	43	Papillary Thyroid Carcinoma	Benign Hyperplastic Nodule	Classic	BRAF (V600E)
WT003	F	55	Papillary Thyroid Carcinoma	Lymphocytic Thyroisitis	Classic	BRAF (V600E)
WT013	F	65	Papillary thyroid carcinoma	Hashimoto's Thyroiditis	Classic	BRAF (V600E)

Table 5.10 Sequence libraries read statistics

Patient ID	Pathology	Total Number of Reads	Number of Aligned Reads
WT015	Benign	197428460	171940362
WT017	Benign	162817354	131851407
WT019	Benign	195772878	174118863
WT025	Benign	178915706	159843334
WT037	Benign	143222902	128256360
WT049	Benign	138657386	120867918
WT051	Benign	162050768	142335041
WT055	Benign	119354060	102832972
WT061	Benign	243675894	205271600
WT073	Benign	135870362	118119142
WT075	Benign	169640988	140124334
WT077	Benign	155734764	133418278
WT079	Benign	134663182	119305555
WT083	Benign	151309892	127780804
WT091	Benign	165329818	142489611
WT095	Benign	126150386	114158421
WT099	Benign	138945768	122671836
WT119	Benign	165395100	142176111
WT127	Benign	186404728	143416691
WT001	Cancer	171851018	149240644
WT003	Cancer	199245096	154792078
WT013	Cancer	101524322	86567402
WT033	Cancer	96628950	78816041
WT045	Cancer	170572728	152940242
WT069	Cancer	145602160	125742598
WT071	Cancer	147139830	124501464
WT107	Cancer	165587530	137770957
WT123	Cancer	189567686	152878294
WT125	Cancer	188854436	153571099

Table 5.11 List of primers used to validate the novel slicing event in SLC35A2
Two sets of flanking primers and 2 sets of spanning ones were designed. One of the primer pairs in red was later found to have been designed incorrectly

Primer Pair Name	Forward Primer	Reverse Primer	Expected WT Amplicon Size (bp)	Expected Novel Amplicon Size (bp)
SV68.flank.001	AAATCAGTGTTGATGGTCTTCTTGATG	AAAGTTATCAGCCAAATTGCAATGAAC	380	290
SV68.flank.002	GAGGTGGAAATTCACAAAGATATGCTG	AAAGTCATCACTAAGCCCTTCACAAAG	285	195
SV68.span.001	CTTGTAGGTCACATTCTTGTTGGTAAAAGT	ATCATAACCCAGCTTATAGTGGAGAGC	0	216
SV68.span.002	GATAAGCCCTCTCAATGGTTATCACG	ACTTTTACCAACAAGAATGTGACCTACAAG	0	364

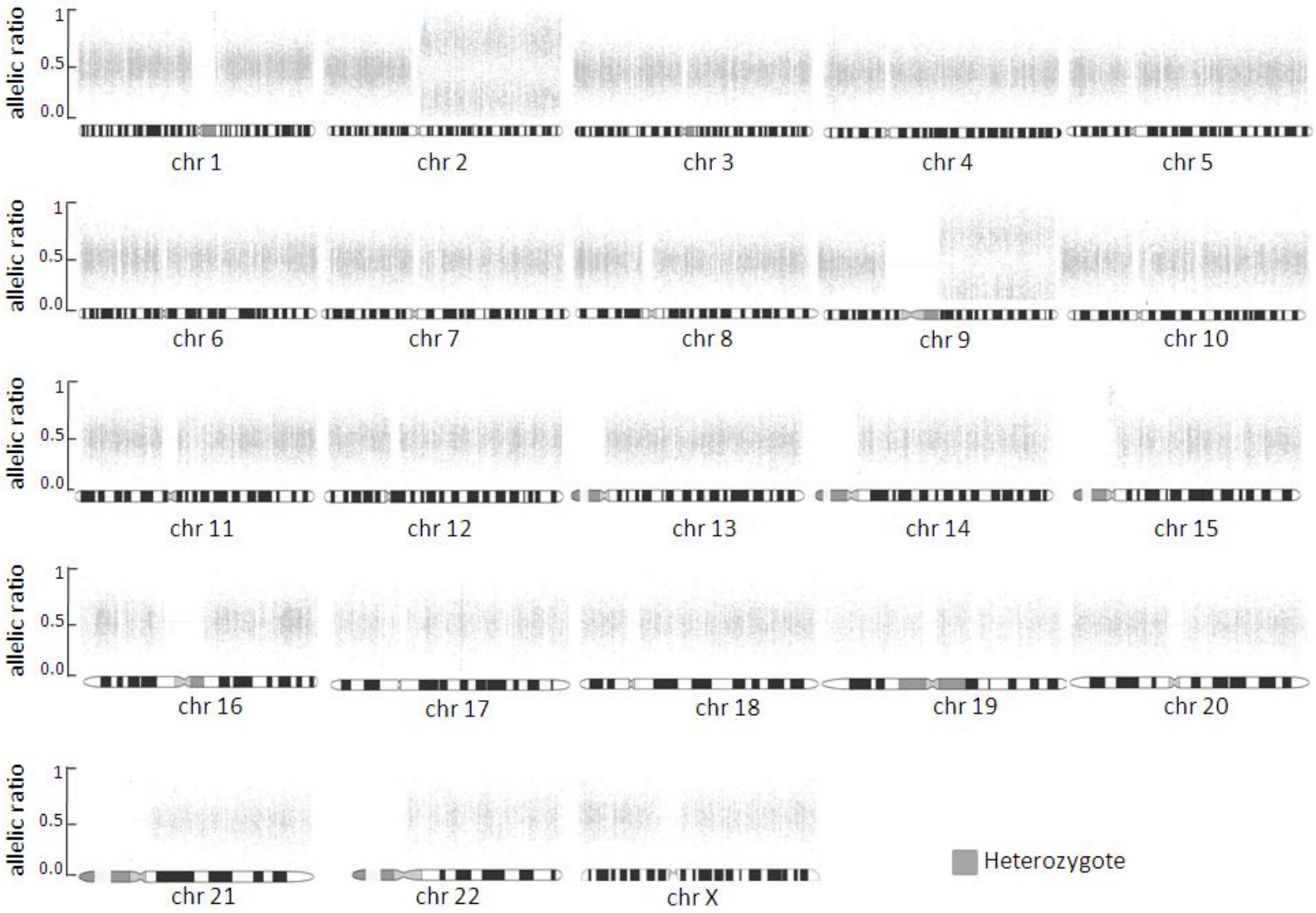

Figure 5.1 B-allele frequency plots for F46FA
Long arms of chromosomes 2 and 9 show wider allele separation due to loss of chromosomal copies

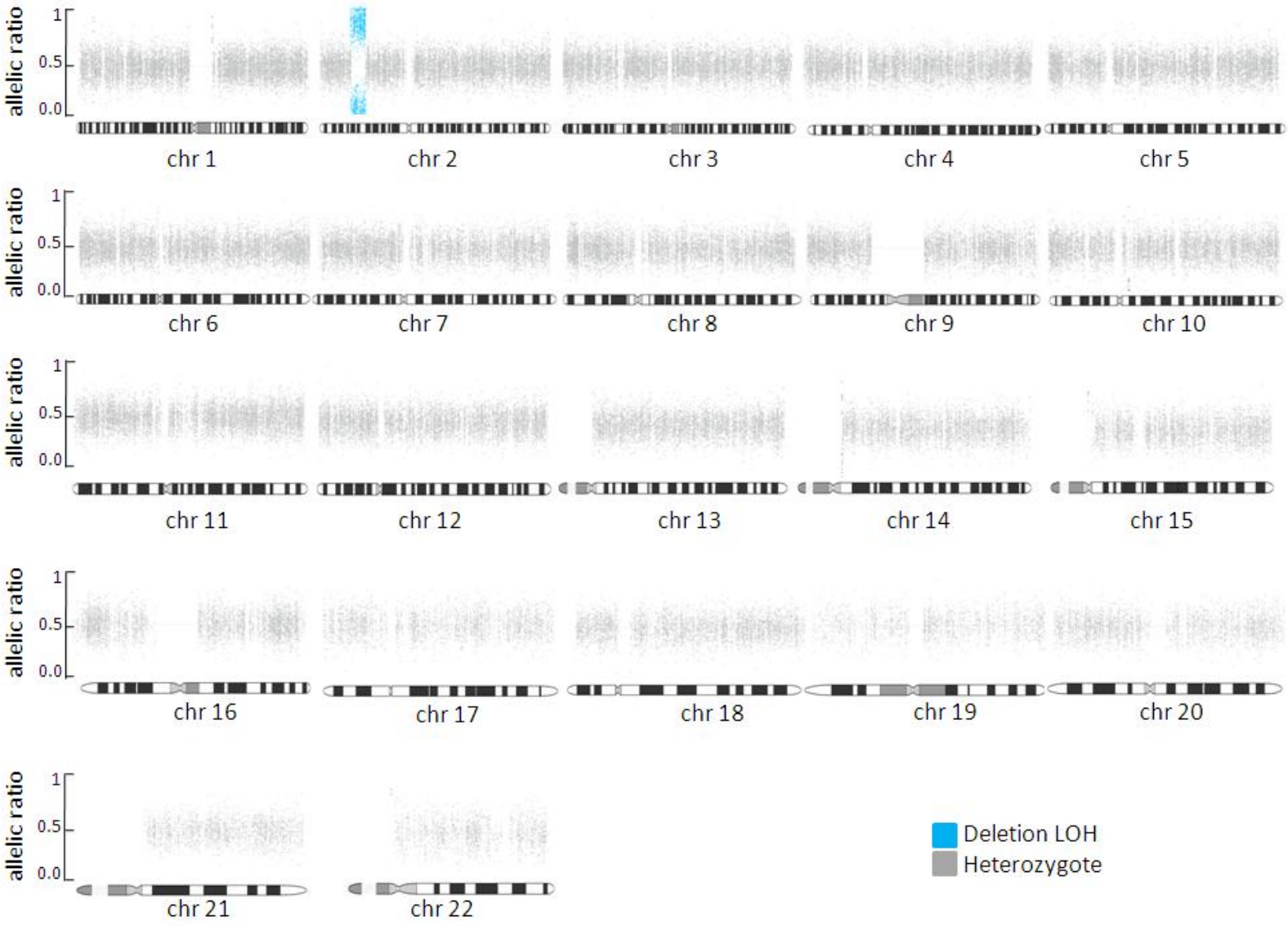

Figure 5.2 B-allele frequency plots for M55G
A small region on the short arm of chromosome 2 shows wider allele separation due to loss of chromosomal copies

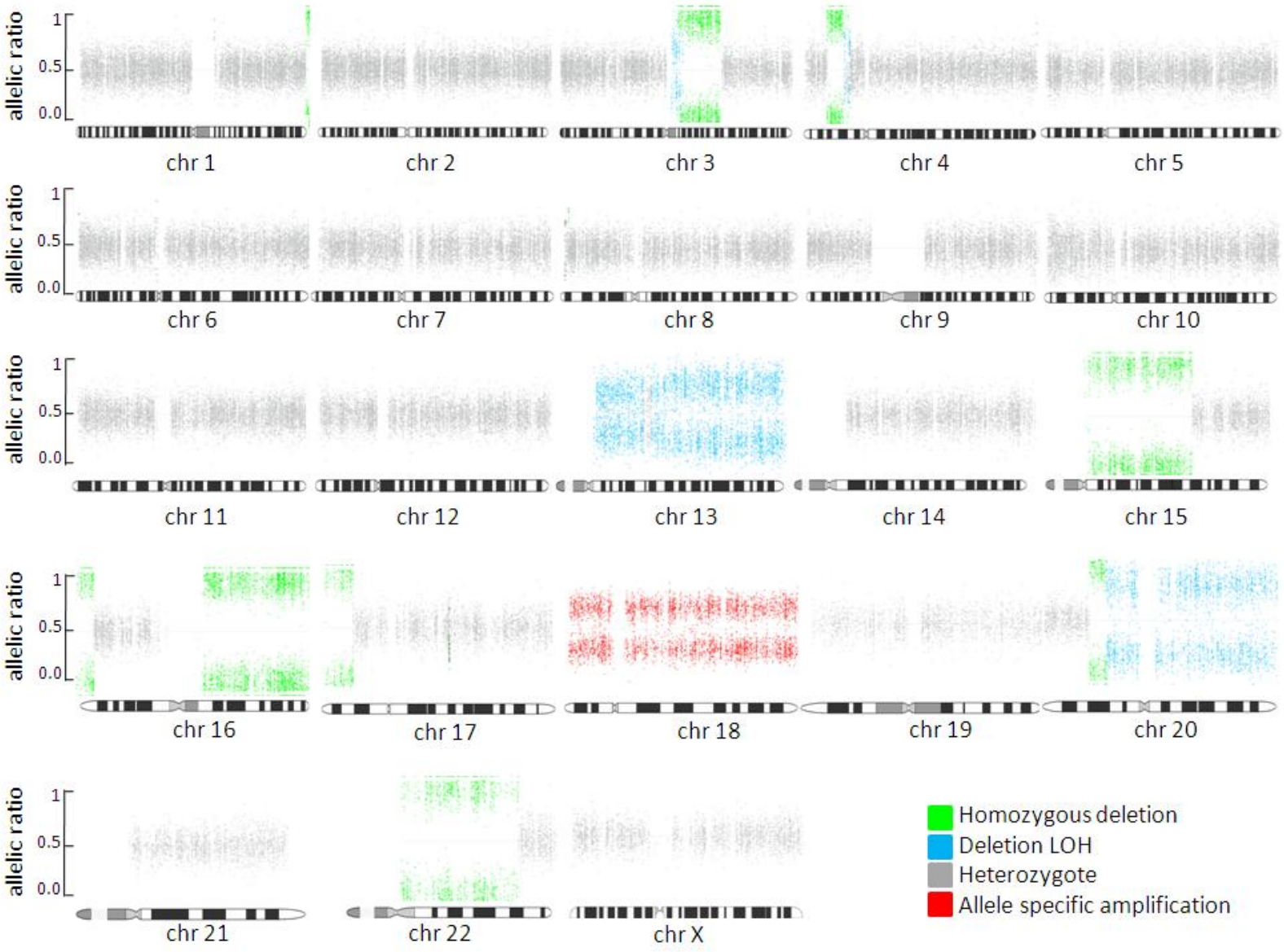

Figure 5.3 B-allele frequency plots for F67FA
Extensive regions of the genome demonstrate allele separation

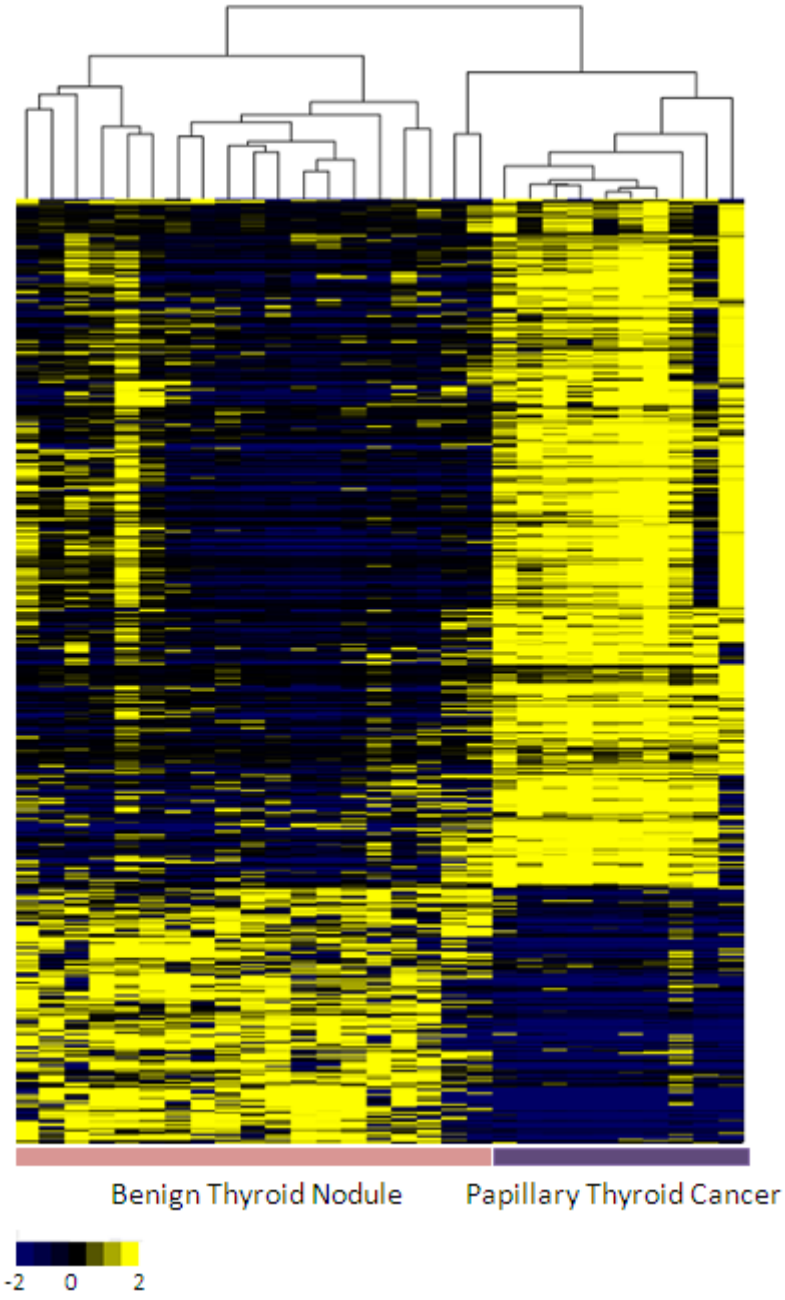

Figure 5.4 Hierarchical clustering of differentially expressed genes
Heatmap demonstrating pairwise complete-linkage hierarchical clustering of 1191 differentially expressed genes in 10 papillary thyroid carcinoma specimens and 19 benign thyroid nodules. Rows (genes) were median-centered and Spearman's rank correlation was used for as the distance measure for both genes and samples

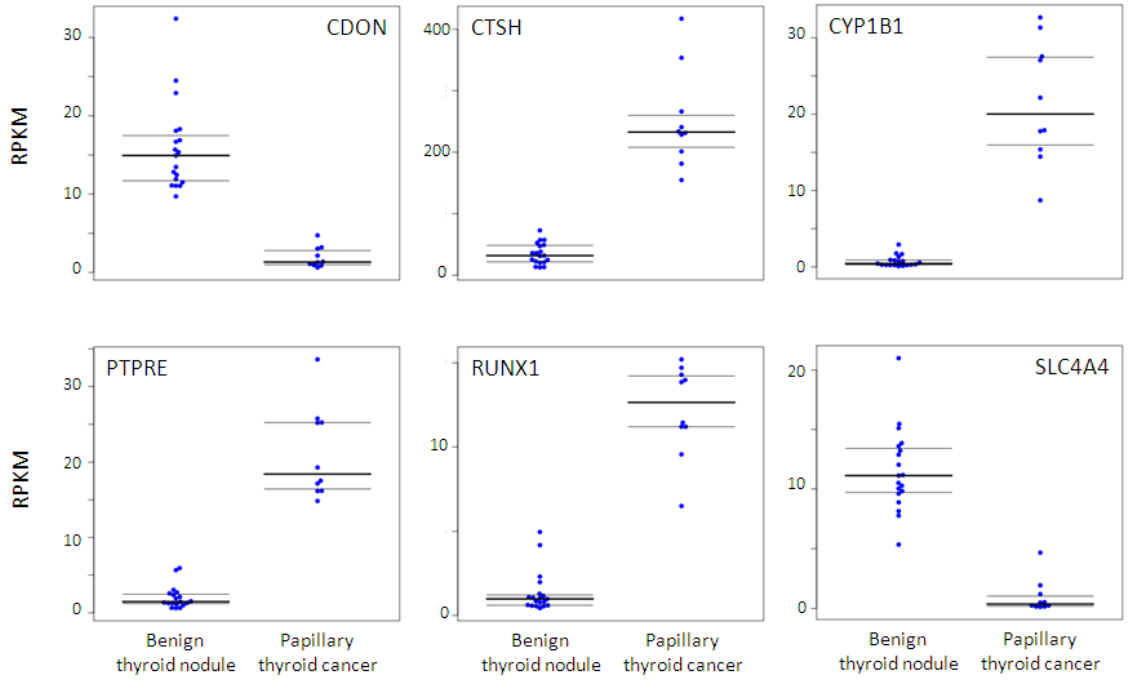

Figure 5.5 Potential biomarkers
The expression (RPKM) of select genes in 19 benign and 10 papillary carcinoma tumors. Targeted RNA expression panels designed for profiling the expression of these genes may have utility for thyroid cancer diagnosis

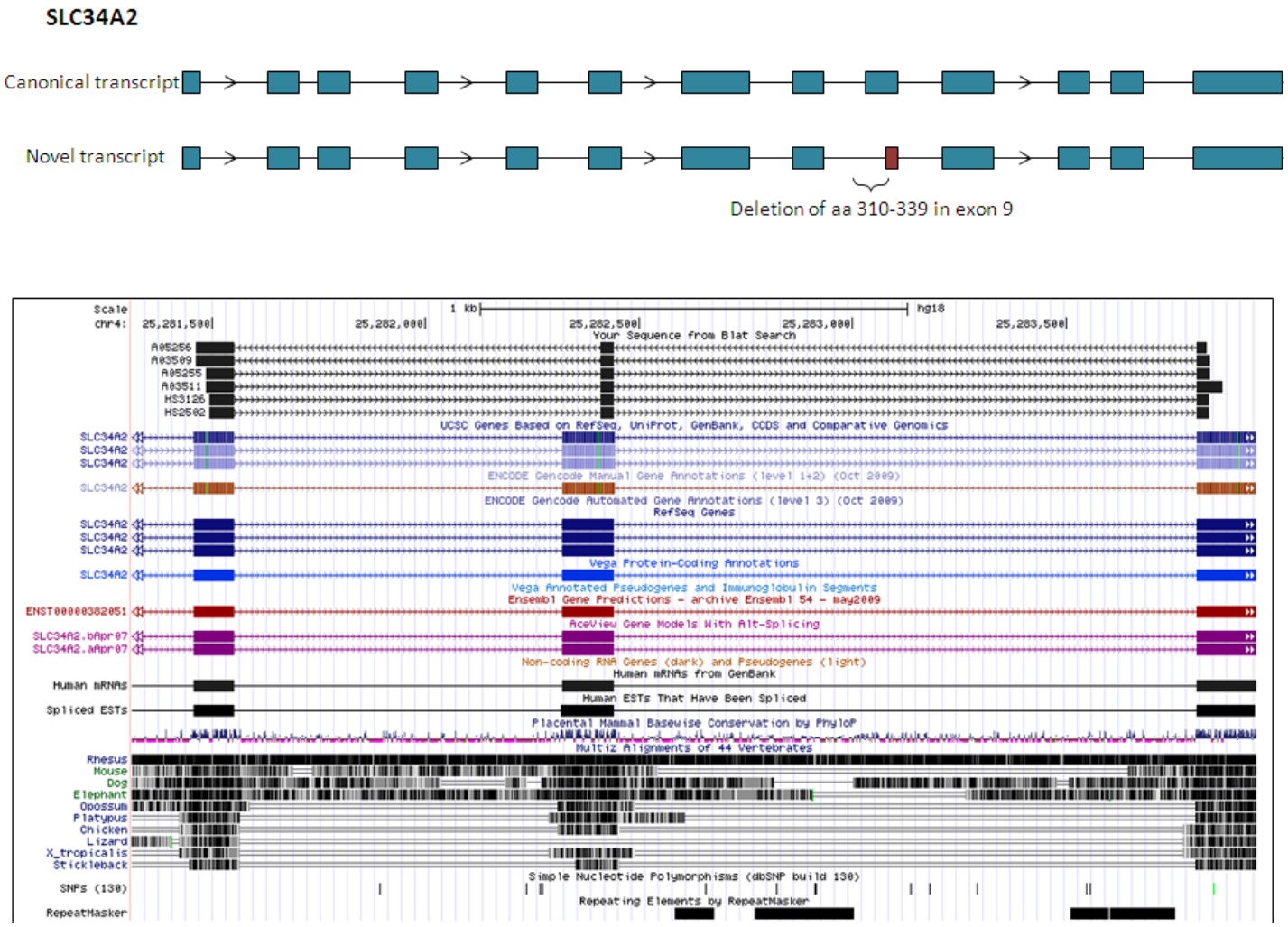

Figure 5.6 Novel splicing event in SLC34A2

Top: A schematic diagram showing the loss of 30 amino acids from exon 9 of SLC34A2 as a result of a novel 5' splice site detected through de novo assembly. Bottom: A screenshot of the alignment of assembled contigs from 6 different papillary thyroid carcinoma specimens to the reference human genome

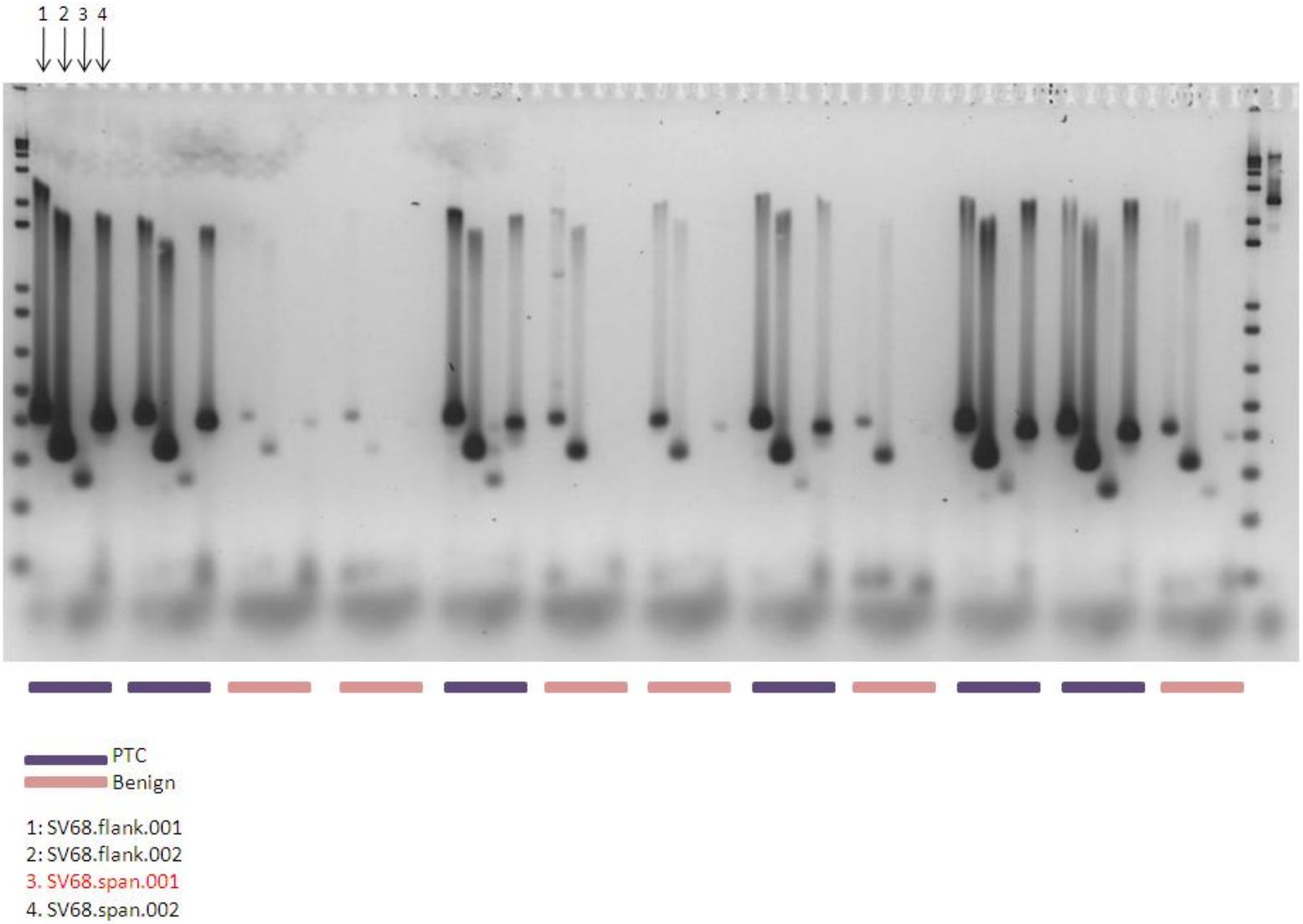

Figure 5.7 SLC34A2 novel splicing event validation
Lanes containing PCR products amplified from 6 papillary thyroid cancers and 6 benign thyroid tumors. Four lanes were run for each sample: lanes 1 and 2 with flanking primer pairs and lane 4 with a spanning primer pair. Lane 3 was run with a primer pair which was later found to be incorrectly designed. The primer sequences are listed in Table 5.11. It is evident when comparing the amount of PCR products in lanes 1 and 2 of all specimens that the SLC34A2 gene, in its wild type form, is much more highly overexpressed in PTCs. This may indicate that the gene itself is a sensitive diagnostic marker for thyroid tumors. The PCR product in lane 4 would only be present in samples with the novel and unique splice site sequence. These are present in all 6 malignant samples and none of the benign specimens. Although the novel sequence is present, we did not observe two clear and separate bands in lanes 1 and 2 as would be expected. This suggests that although the novel sequence is present in malignant samples, it is expressed at a much lower quantity compared with the wild type transcripts

Chapter 6: Conclusion

6.1 Summary

This thesis describes the application of massively parallel and high-throughput sequencing technologies and generation of whole genome and transcriptome datasets in the study of benign and malignant thyroid tumors and those of its neighboring gland, the parathyroid. Both rare and common tumors of the thyroid were studied and despite their shared cell of origin, stark differences in the genomes of these tumors were observed. In general, very few recurrent mutations were found in the tumors studied here. We found the benign tumors to harbor very few mutations and copy number changes; The Cancer Genome Atlas (TCGA) study of papillary thyroid carcinoma (PTC) that was conducted concurrent to this work also found the genomes of PTCs to be very quiet with respect to both small mutations and large copy number variant regions [270]. On the other hand, we found vast amount of copy number alterations and aneuploidy in anaplastic and oncocytic thyroid cancers and those of the parathyroid gland. This work revealed mutations and gene fusions that had not been previously described and thus has shed light on the biology of disease and tumorigenesis pathways in these endocrine organs. Additionally, I have examined and compared the genomic landscapes of various types of tumors that arise from the thyroid gland with the aim of uncovering molecular evidence in support or in disagreement of step-wise transformation of benign to cancerous tumors. Integrative analysis of genomic and transcriptomic datasets were described and shown to increase our

confidence in identifying the altered pathways leading to disease. The continual application of such analysis over time or in studying biopsy specimens from multiple sites in the tumor will allow for deciphering the temporal and spatial evolution of these tumors. In addition, detailed epigenomic profiles of the common thyroid tumors including benign nodules and papillary thyroid carcinomas are being generated at our institution. These experiments include ChIP-Seq and whole genome bisulfite sequencing for identifying select histone marks and methylation status of the complete genome. Proteomic experiments through providing complementary information to genomic, transcriptomic and epigenomic studies promise to unveil the molecular mechanisms leading to thyroid tumorigenesis; however, such experiments still await the more high-throughput and affordable techniques in proteomics. In this final chapter, I will outline the general conclusions from all studies, their strengths and limitations and provide directions for future work.

6.2 Parathyroid Cancer

Parathyroid cancer is an extremely rare disease with the incidence rate of about 1 per million population. In chapter 2, I have described the genomic and transcriptomic analysis of a primary parathyroid tumor and two recurrences of the same tumor from one patient. Retrospective studies in the literature have pointed to high rate of disease recurrence in parathyroid cancer with up to 11 recurrences observed in one patient [122,124]. The molecular comparison of all tumors from the patient in our study revealed no major changes in cancer driver pathways between the different recurrences. Due to the rare occurrence of this malignancy, no

established therapy protocols including cytotoxic chemotherapeutics or targeted therapies are available or recommended. The patient under study only received calcimimetic agents to control his blood calcium level and no other therapeutics were administered. As a result, unlike most cancer patients who receive therapy and malignant tumors that evolve in response to them, minimal adaptations in the genome of this patient's tumor was observed over time. This is evident from the lack of observation of any major differences between the tumors. This may imply that the tumor lacks any need for accumulation of further mutations in order to recur at a later time; however, the observed copy number changes in chromosomes 4 and 5 (Figure 2.5) may point to the contrary. It is postulated in the literature that incomplete removal of malignant tissue or accidental residual disease could lead to future recurrence(s); thus it is recommended that the entire gland and its adjacent tissues be resected when malignancy is established [121]. The molecular profiling of parathyroid cancer in the current study provides supporting evidence for the hypothesis that multiple recurrences of the disease are likely due to the presence of residual disease and not attributed to disease progression and evolution.

Whole genome datasets revealed large areas of copy gain and loss. Thousands of genes incurred changes in copy number; hence, the integrative analysis of the genome data with the transcriptome profiles was used to identify altered pathways. Those genes with gain of copy and over-expression or those which had lost copies and showed lower expressions were considered for network analysis. Gene copy loss and as a result loss of heterozygosity augmented with a truncating mutation in the remaining allele of *THRAP3* pointed to possible

role for this gene in driving parathyroid malignancy. Another interesting observation was the loss of activating PIK3CA point mutation in the recurrent specimen that was originally present in the primary tumor. Loss or gain of activating mutations of this gene during the evolution of the tumor has been described in other cancer types, particularly in breast cancers [337], but never in the absence of therapy. This may indicate that PIK3CA mutations was present in the dominant clone in the primary tumor but only in a minor sub-clone in the recurrent parathyroid tumor suggesting that its activation was not necessary for tumor progression and maintenance but required for tumor initiation. It is also possible that due to random sampling of the tumor, entirely different clonal populations were examined; a comprehensive spatial profiling of both tumors will provide more insight into the role of this oncogene in parathyroid tumorigenesis. Activation of the PI3K/Akt pathway might still be important for tumor progression in the absence of PIK3CA mutation and this need might be met through the activation of the downstream MTOR and its targets without reliance on PIK3CA. This is an important consideration in the case of rare parathyroid cancer; the patient under study harbored an *MTOR* mutation and the inhibitors of this signaling pathway including everolimus are available and approved for use in treating several cancer types. Additional studies, for instance those of immunohistochemical examination of mTOR pathway expression levels, can determine if this pathway is constitutively activated in parathyroid cancer and can as a result lead to the use of the already available targeted drugs in treating parathyroid cancer.

Although this study was the first to examine a parathyroid malignant tumor on the genomic level, it was limited to only a few samples from one patient. No other parathyroid carcinoma specimens were available for inclusion in this study and we do not expect to encounter another case in the near future given the very uncommon nature of this malignancy. Given the heterogeneous nature of cancer and varying mutational profiles in patients with the same cancer type, interpreting the results of this study is limited to this patient and no specific conclusions can be drawn about the disease. A recent study published after our report performed whole exome analysis of 8 parathyroid cancer patients and described novel germline and somatic mutations of *PRUNE2* in 2 patients [338]. Although invaluable, whole exome sequencing experiments examine only the protein-coding regions, a mere 2% of the entire genome. Such approach does not provide a high-resolution view of the regions with copy number change nor does it allow for identification of fusion breakpoints falling in non-coding regions. Collaborative and multi-institutional studies are required to examine larger cohorts of parathyroid cancers on the whole genome and transcriptome scale. These studies will likely have to rely on cohorts of formalin-fixed paraffin-embedded (FFPE) tissues. The study described in this chapter detailed the successful application of whole genome sequencing to both flash frozen tumor tissues as well as the more commonly available FFPE clinical specimens. Future studies of this rare tumor promise to identify recurrent mutations including single nucleotide variants, small insertion and deletions and gene fusions that could serve as therapeutic targets. Only through the use of such mutation-driven treatments, residual disease can be eradicated and recurrent disease prevented.

6.3 Hürthle Cell (Oncocytic) Thyroid Carcinoma

Hürthle cell, also known as oncocytic, thyroid cancer arises from the follicular cells of this endocrine gland and accounts for about 3-5% of all thyroid cancers [19]. No comprehensive molecular profiling of these tumors had been performed and the knowledge about the molecular drivers of this malignancy is very limited. In chapter 3, I described whole genome study of two Hürthle cell thyroid tumors, and the follow-up validation experiment in a larger cohort of patients. Whole genome sequencing revealed large regions of copy change often encompassing whole chromosomes. Common changes between the two tumors included gain of chromosomes 5, 7, 12, 18p, 19 and 20. More intriguingly large regions of the genome showed loss of heterozygosity, at times while maintaining two copies of the chromosome. Both tumors showed loss of heterozygosity of chromosomes 1, 2, 3, 4, 6, 8, 9, 11, 14, 15 and X. Despite vast amount of copy number change, fewer regions of focal gain or loss were identified. No structural variants or gene fusions were found through *de novo* assembly of the raw sequence reads. Collectively, these observations imply a causative role for small alterations such as single nucleotide variants and small insertions and deletions, particularly those affecting genes with loss of heterozygosity, in Hürthle cell thyroid tumorigenesis. We identified two distinct hemi- and homozygous frame-shift deletions in *MEN1* gene in both tumors. This gene has been known to play a key role in endocrine organ tumorigenesis. Its mutation and loss of function is the cause of multiple endocrine neoplasia type I syndrome, which manifests in benign tumors of multiple endocrine organs such as the parathyroid, pancreatic islets, duodenal

endocrine cells and anterior pituitary. Its mutations had not been known to cause thyroid malignancy. Targeted examination of the *MEN1* gene in a larger cohort of Hürthle cell thyroid tumors followed where we identified mutations in an additional 3 samples.

We identified and validated somatic MEN1 frame-shift deletions in the two original flash frozen Hürthle cell tumors but only in 3 of 72 validation cohort specimens (4.2% population frequency). The low frequency of the mutation in the validation cohort was a surprising finding given the presence of mutations in both discovery specimens. Although it is quite possible that MEN1 mutations are only present in a small subset of oncocytic thyroid tumors, it is also likely that these loss-of-function mutations are present in a larger patient subpopulation but our study was underpowered to detect them. A major limitation of this experiment was the use of formalin-fixed paraffin-embedded tissues for the validation experiment while no information regarding the tumor DNA content for these specimens was available. The relatively rare occurrence of this malignancy required that we rely on FFPE specimens collected over many years for the validation experiment. It is well established that over time DNA integrity of these preserved samples is diminished and a rigorous pathology review of the biopsy material is needed in such cases. If serial sections had been made from each of the 72 samples and pathology review of hematoxylin and eosin stained slides had been performed, we would have a more accurate understanding of the phenotype of the validation cohort samples.

Hürthle cells are not unique to the thyroid gland and can be found in other organs with high metabolic rate such as the kidney, parathyroid, salivary and adrenal glands [19,20]. It will be of interest to examine and compare the Hürthle cell tumors from all these organs for the presence of mutations especially those of the *MEN1* gene and identify associations, if any, between these mutations and the specific phenotype of excess mitochondrial accumulation. The distinct haploid genomic profiles of the tumors in this study also raise the question of whether loss of MEN1 protein function in the cell is a causative event for the appearance of haploid genomes. It is also of importance to identify potential links, if any, between benign Hürthle cell tumors and those of malignant tumors; it is not known if benign oncocytic tumors can or would lead to malignant tumors through accumulation of mutations. Genomic profiling of large cohorts of benign and malignant tumors could provide a better understanding of these tumors.

6.4 Anaplastic Thyroid Carcinoma

In chapter 4, I described the genomic profile of the rare and aggressive anaplastic thyroid cancer (ATC). ATCs account for only a small subset of all thyroid cancers but they are responsible for the majority of deaths in patients diagnosed with cancers of this endocrine gland. Anaplastic transformation of follicular thyroid cells leads to un-differentiation and complete loss of all thyroid-specific markers from the cell surface. As a result, not only diagnosis of ATCs, and at times even differentiating them from sarcomas, becomes challenging but their treatment with radioiodine ablation is also impossible due to loss of sodium/iodide cotransporter that is unique to thyroid cells and the target of therapy in papillary carcinomas.

The genomic data revealed aneuploidy with vast areas of copy gain and loss. Recurrent mutations of the epigenetic machinery including novel gene fusions were also observed in all samples and the transcriptome profiles hinted to a potential causative role for epigenetic deregulation in tumorigenesis. We also identified fusions of gene members of the axon guidance pathway in several of the ATC specimens. Deregulation of this pathway and its recurrent alterations have been observed in pancreatic ductal carcinoma, lung, breast, kidney and cervical cancers [283]. This is a pathway with regulatory roles in embryogenesis and it also interacts with and modulates known cancer pathways such as MET and WNT [283]. Evidence is emerging for members of this signaling network as promising drug targets [339]; these might have clinical applications for ATCs. Given that mutations such as BRAF p.V600E and those of RAS family of genes are shared between a subset of PTCs and those of anaplastic cancers, it is believed that some ATCs arise from precursor differentiated thyroid cancers while the rest arise *de novo*. It would be of great interest and of clinical utility to distinguish those PTCs that will eventually become undifferentiated and develop into ATCs; continual monitoring of patients at risk can facilitate early diagnosis and the delivery of more effective treatments. A small cohort of ATCs was examined in this study, however, I found gene fusions involving *FGFR2* and *BRAF* genes that have also been found in less than 1% of the PTC population [270]. Those PTCs harboring oncogenic fusions may demonstrate more aggressive behavior and represent a small subset of papillary cancers that will evolve to ATCs.

The lack of recurrent targetable mutations in ATCs predicts variable and unpredictable responses to one-size-fits-all therapies. This is in agreement with lack of objective responses to therapy in various clinical trials to date and the absence of approved and standard therapies for this cancer. ATC may as a result be a suitable disease candidate for an approach to diagnosis and treatment that is mutation driven and more "personalized". Such oncogenomic efforts have become more commonplace in the past few years and several centers around the world are increasingly utilizing the power of NGS technologies for identifying targeted therapy options in individual patients.

Although this study was the first to provide an in-depth molecular signature of ATCs including several unique and authenticated cell lines, it was limited to a few specimens. Small sample sizes can lead to over-estimation of the true effect of findings while failing to identify all relevant and causative events in this cancer. Multidimensional genomic analyses of a large cohort of anaplastic thyroid cancers, similar to what has been accomplished for papillary thyroid cancer by The Cancer Genome Atlas study, promises to find low frequency DNA mutations and describe alterations of cell's mRNA and miRNA repertoires and the methylome.

6.5 Papillary Thyroid Carcinoma and Benign Thyroid Nodules

An estimated 4% to 7% of the population will develop a clinically significant thyroid nodule during their lifetime. In up to a quarter of cases, preoperative diagnosis by needle biopsy is inconclusive and so a large proportion of individuals undergo thyroidectomy as a diagnostic

procedure for cancer. The molecular mechanisms that drive thyroid tumorigenesis and progression are still poorly understood. Likewise, the molecular causes of benign thyroid nodules have also yet to be elucidated. We performed RNA sequencing of papillary thyroid cancer, the most abundant form of the disease, and benign thyroid nodules using massively parallel sequencing technologies in order to characterize the molecular changes underlying these lesions. Whole genome sequencing of 3 benign nodules and their matched normal tissues were also performed.

Our study demonstrated a very low mutation rate in benign nodules of the thyroid. The Cancer Genome Atlas study of 402 PTC specimens also estimated the mutation rate to be low and around 0.41 nonsynonymous mutations per Mb [270]. Both diseases showed very quiet genomes with very few copy number changes throughout. A few, although non-recurrent, gene fusions were observed in the genomes as described in Chapter 5. The most frequently mutated gene in PTCs was BRAF harboring the p.V600E activating mutation in over 60% of the population. It is believed that common adult epithelial cancers require at least 5 to 7 driver gene mutations to become a malignant mass [340]. Although TCGA study was the first to comprehensively examine a large cohort of PTCs and succeeded in shrinking the percentage of "dark matter" PTCs, tumors with no known driver mutations, to 1.2%, only a very small fraction of tumors were found to have two or more driver mutations. This raises the question of whether there are other mutations, perhaps non-coding and regulatory alterations, that are responsible for these malignancies and which are not identified through the use of the current

technologies. Alternatively, since the majority of PTCs are very indolent tumors and do not become locally aggressive or metastasize to distant organs, it is feasible that the presence for 5-7 driver mutations is not a universal requirement and having merely one driver is sufficient for these nodules to be declared malignant based on current pathological standards. These tumors may remain indolent until they acquire further disease drivers. Although only the genomes of 3 benign nodules were examined in this thesis, no shared mutations with PTCs were identified. It is well recognized that 20-25% of benign nodules harbor *RAS* mutations that are also found in the follicular variants of PTC [244]; however, it is still unknown if a step-wise accumulation of mutations transforms a benign thyroid nodule to a PTC. Studies that will examine PTCs from patients with history of benign nodules will shed light on the evolutionary process in these tumors.

The required whole genome sequence depth and coverage for identifying all variations in a genome has been extensively discussed and updated over the past few years to reflect the advances in technology [341]. Currently, utility of sequencing experiments in clinical oncology mandates a high depth of coverage of at least 80-100x, while genome resequencing experiments can rely on an average depth of 35x [341]. However, although this threshold may suffice when examining a near-normal genome, studying cancers pose unique challenges. Sample and tumor heterogeneity in addition to aneuploid genomes that are observed in close to all cancer specimens require a high depth of sequence coverage to identify all relevant somatic mutations. Due to still substantial cost of whole genome sequencing, 30-40x coverage

was produced for the studies discussed in the thesis (Tables 2.1, 3.1, 4.1 and 5.4). Although these datasets result in reliable identification of structural variants and copy number altered regions at the base-level resolution, they are not as robust in finding all SNVs and indels. Low coverage in the tumor or the matched normal tissue may have resulted in false negative or false positive somatic calls, respectively, while the studied datasets did not have the power to detect subclonal events.

Whole genome, exome and transcriptome profiling of cancer specimens are very powerful in deciphering mutations that are likely to be driver of disease, this is evident from the many discoveries made in just the past few years. However, these studies do not examine or provide any insight into numerous other factors that may be crucial in cancer initiation and progression. One of these concerns the role of the microenvironment surrounding the tumor. If the immediate environment around the newly formed nodule consisting of only a limited number of mutated cells is hostile to its maintenance, the tumor will not be able to progress into a more aggressive form invading local or distant tissues. Small and indolent *in situ* thyroid nodules were found in over 50% of autopsies from patients with clinically normal thyroid gland and the presence of occult thyroid cancer is reported in up to 13% of autopsies [342]. It would be imperative to compare the microenvironment of the small fraction of aggressive PTCs and ATCs with those of the majority, indolent and confined PTCs. This can not only shed light on the biology of the disease but also enable the discrimination of aggressive and non-aggressive tumors early in the course of the disease and hence enable the delivery of more effective

therapies [343].

Another related but important consideration in the pathogenesis of thyroid tumors is the role of immune system in carcinogenesis. Only when the body's defense system fails to recognize and eradicate a nodule, regardless of the tissue and organ of origin, the tumor will have the opportunity to evolve and become invasive. However, through maintaining an inflammatory microenvironment, the immune system may also facilitate the tumorigenesis process. It is not yet understood how the immune system contributes to or prevents the development of thyroid tumors. Cancers can arise if the immune system does not recognize tumor-specific antigens and hence remain inactive. Malignant tumors may develop mechanisms to escape the immune system's inhibitory effect through various processes such as downregulation of antigen presentation, expression of inhibitory molecules, recruitment of suppressor cells or eliminating the need for growth stimulation by developing autocrine signaling [344]. Loss of MHC class I was recently found in a large proportion of PTCs and was shown to be associated with immune escape [345]. Cataloguing all pathways through which the tumor escapes the immune system in thyroid cancer and identifying those that are reversible [345] can facilitate the administration of target therapies. It has also been suggested that thyroid tumors associated with inflammation and higher number of infiltrating lymphocytes have a better prognosis, due perhaps to the early immune response to the tumor [345,346]. Pathology reports of PTC biopsy specimens (Chapter 5) indicated an extensive lymphocytic infiltration in the majority of these tumors. While the presence of white blood cells in and around the tumor might be suggestive of the efforts of the

immune system in eradicating the disease, it may also explain the manifestation of the disease after battle with a long-standing inflammation in the organ. The link between chronic inflammation and cancer has long been established [343] and inflammation as a result of thyroid autoimmune diseases such as Grave's or Hashimoto's thyroiditis may contribute to the progression of cancer in this endocrine gland [347]. Such inflammatory events may also arise in response to the immune system's anti-tumor activity which at times leads to unintended protumor effects [344]. As discussed in Chapter 5 (section 5.5.1.4), cytokine-cytokine receptor interaction and chemokine signaling pathways showed statistically significant upregulation in PTCs compared with benign tumors, these chemicals secreted by the invading leukocytes can help to maintain the malignant phenotype by increasing cell proliferation and angiogenesis [347]. Defining the mechanisms by which chronic inflammation may harm or protect the tumor can guide future therapy options for thyroid cancer. Generally, papillary thyroid cancer has a great prognosis and while immunotherapy can provide a more personalized approach to treatment, more effective stratification of patients based on their immune phenotype prior to radical surgery and total thyroid ablation will have high and immediate impact in the clinic.

References

1. Kilfoy BA, Zheng T, Holford TR*, et al.* International patterns and trends in thyroid cancer incidence, 1973-2002. *Cancer Causes Control* 2009;**20:**525-531.

2. Davies L, Welch HG. Increasing incidence of thyroid cancer in the United States, 1973-2002. *JAMA* 2006;**295:**2164-2167.

3. Colonna M, Guizard AV, Schvartz C*, et al.* A time trend analysis of papillary and follicular cancers as a function of tumour size: a study of data from six cancer registries in France (1983-2000). *Eur J Cancer* 2007;**43:**891-900.

4. Burgess JR. Temporal trends for thyroid carcinoma in Australia: an increasing incidence of papillary thyroid carcinoma (1982-1997). *Thyroid* 2002;**12:**141-149.

5. Li N, Du XL, Reitzel LR*, et al.* Impact of enhanced detection on the increase in thyroid cancer incidence in the United States: review of incidence trends by socioeconomic status within the surveillance, epidemiology, and end results registry, 1980-2008. *Thyroid* 2013;**23:**103-110.

6. Gharib H, Papini E. Thyroid nodules: clinical importance, assessment, and treatment. *Endocrinol Metab Clin North Am* 2007;**36:**707-35, vi.

7. Byrd JK, Yawn RJ, Wilhoit CS*, et al.* Well differentiated thyroid carcinoma: current treatment. *Curr Treat Options Oncol* 2012;**13:**47-57.

8. LiVolsi VA. Papillary thyroid carcinoma: an update. *Mod Pathol* 2011;**24 Suppl 2:**S1-9.

9. Nikiforov YE, Nikiforova MN. Molecular genetics and diagnosis of thyroid cancer. *Nat Rev Endocrinol* 2011;**7:**569-580.

10. Tuttle RM, Leboeuf R, Martorella AJ. Papillary thyroid cancer: monitoring and therapy. *Endocrinol Metab Clin North Am* 2007;**36:**753-78, vii.

11. Schlumberger MJ. Papillary and follicular thyroid carcinoma. *N Engl J Med* 1998;**338:**297-306.

12. Xing M. BRAF mutation in thyroid cancer. *Endocr Relat Cancer* 2005;**12:**245-262.

13. Ezzat S, Zheng L, Kolenda J*, et al.* Prevalence of activating ras mutations in morphologically characterized thyroid nodules. *Thyroid* 1996;**6:**409-416.

14. Begum S, Rosenbaum E, Henrique R*, et al.* BRAF mutations in anaplastic thyroid carcinoma: implications for tumor origin, diagnosis and treatment. *Mod Pathol* 2004;**17:**1359-1363.

15. Lin RY. Thyroid cancer stem cells. *Nat Rev Endocrinol* 2011;**7:**609-616.

16. Donghi R, Longoni A, Pilotti S*, et al.* Gene p53 mutations are restricted to poorly differentiated and undifferentiated carcinomas of the thyroid gland. *J Clin Invest* 1993;**91:**1753-1760.

17. Garcia-Rostan G, Tallini G, Herrero A*, et al.* Frequent mutation and nuclear localization of beta-catenin in anaplastic thyroid carcinoma. *Cancer Res* 1999;**59:**1811-1815.

18. Rodrigues RF, Roque L, Krug T*, et al.* Poorly differentiated and anaplastic thyroid carcinomas: chromosomal and oligo-array profile of five new cell lines. *Br J Cancer* 2007;**96:**1237-1245.

19. Montone KT, Baloch ZW, LiVolsi VA. The thyroid Hurthle (oncocytic) cell and its associated pathologic conditions: a surgical pathology and cytopathology review. *Arch Pathol Lab Med* 2008;**132:**1241-1250.

20. Tallini G. Oncocytic tumours. *Virchows Arch* 1998;**433:**5-12.

21. Ganly I, Ricarte Filho J, Eng S*, et al.* Genomic dissection of Hurthle cell carcinoma reveals a unique class of thyroid malignancy. *J Clin Endocrinol Metab* 2013;**98:**E962-72.

22. Maximo V, Botelho T, Capela J*, et al.* Somatic and germline mutation in GRIM-19, a dual function gene involved in mitochondrial metabolism and cell death, is linked to mitochondrion-rich (Hurthle cell) tumours of the thyroid. *Br J Cancer* 2005;**92:**1892-1898.

23. Cooper DS, Doherty GM, Haugen BR*, et al.* Management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* 2006;**16:**109-142.

24. Popoveniuc G, Jonklaas J. Thyroid nodules. *Med Clin North Am* 2012;**96:**329-349.

25. Mazzaferri EL. Management of a solitary thyroid nodule. *N Engl J Med* 1993;**328:**553-559.

26. Krohn K, Fuhrer D, Bayer Y*, et al.* Molecular pathogenesis of euthyroid and toxic multinodular goiter. *Endocr Rev* 2005;**26:**504-524.

27. Howell VM, Haven CJ, Kahnoski K*, et al.* HRPT2 mutations are associated with malignancy in sporadic parathyroid tumours. *J Med Genet* 2003;**40:**657-663.

28. Arnold A, Kim HG, Gaz RD*, et al.* Molecular cloning and chromosomal mapping of DNA rearranged with the parathyroid hormone gene in a parathyroid adenoma. *J Clin Invest* 1989;**83:**2034-2040.

29. Motokura T, Bloom T, Kim HG*, et al.* A novel cyclin encoded by a bcl1-linked candidate oncogene. *Nature* 1991;**350:**512-515.

30. Boveri T. Zur Frage der Entstehung Maligner Tumoren. . 1914:1.

31. von Hansemann D. Ueber asymmetrische Zelltheilung in epithel Krebsen und deren biologische Bedeutung. Virchow's Arch Path Anat. 1890;119(299).

32. Boeva V, Popova T, Bleakley K*, et al.* Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* 2012;**28:**423-425.

33. Sathirapongsasuti JF, Lee H, Horst BA*, et al.* Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics* 2011;**27:**2648-2654.

34. Chmielecki J, Crago AM, Rosenberg M*, et al.* Whole-exome sequencing identifies a recurrent NAB2-STAT6 fusion in solitary fibrous tumors. *Nat Genet* 2013;**45:**131-132.

35. A gene-centric human proteome project: HUPO--the Human Proteome organization. *Mol Cell Proteomics* 2010;**9:**427-429.

36. Robertson G, Hirst M, Bainbridge M*, et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* 2007;**4:**651-657.

37. Lee TI, Johnstone SE, Young RA. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc* 2006;**1:**729-748.

38. Hirst M, Marra MA. Next generation sequencing based approaches to epigenomics. *Brief Funct Genomics* 2010;**9:**455-465.

39. Wilson IM, Davies JJ, Weber M*, et al.* Epigenomics: mapping the methylome. *Cell Cycle* 2006;**5:**155-158.

40. Jacinto FV, Ballestar E, Esteller M. Methyl-DNA immunoprecipitation (MeDIP): hunting down the DNA methylome. *BioTechniques* 2008;**44:**35, 37, 39 passim.

41. Serre D, Lee BH, Ting AH. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Res* 2010;**38:**391-399.

42. Harris RA, Wang T, Coarfa C*, et al.* Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotechnol* 2010;**28:**1097-1105.

43. Maunakea AK, Nagarajan RP, Bilenky M*, et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 2010;**466:**253-257.

44. Cokus SJ, Feng S, Zhang X*, et al.* Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 2008;**452:**215-219.

45. Lister R, O'Malley RC, Tonti-Filippini J*, et al.* Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 2008;**133:**523-536.

46. Lister R, Pelizzola M, Dowen RH*, et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 2009;**462:**315-322.

47. Meissner A, Gnirke A, Bell GW*, et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* 2005;**33:**5868-5877.

48. Deng J, Shoemaker R, Xie B*, et al.* Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat Biotechnol* 2009;**27:**353-360.

49. Ball MP, Li JB, Gao Y*, et al.* Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* 2009;**27:**361-368.

50. Smith ZD, Gu H, Bock C*, et al.* High-throughput bisulfite sequencing in mammalian genomes. *Methods* 2009;**48:**226-232.

51. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol* 1981;**147:**195-197.

52. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 2008;**18:**1851-1858.

53. Li R, Li Y, Kristiansen K*, et al.* SOAP: short oligonucleotide alignment program. *Bioinformatics* 2008;**24:**713-714.

54. Rumble SM, Lacroute P, Dalca AV*, et al.* SHRiMP: accurate mapping of short color-space reads. *PLoS Comput Biol* 2009;**5:**e1000386.

55. Langmead B, Trapnell C, Pop M*, et al.* Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;**10:**R25.

56. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;**26:**589-595.

57. Li R, Yu C, Li Y*, et al.* SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 2009;**25:**1966-1967.

58. Li R, Li Y, Kristiansen K*, et al.* SOAP: short oligonucleotide alignment program. *Bioinformatics* 2008;**24:**713-714.

59. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008;**18:**821-829.

60. Chaisson MJ, Pevzner PA. Short read fragment assembly of bacterial genomes. *Genome Res* 2008;**18:**324-330.

61. Butler J, MacCallum I, Kleber M*, et al.* ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome Res* 2008;**18:**810-820.

62. Simpson JT, Wong K, Jackman SD*, et al.* ABySS: a parallel assembler for short read sequence data. *Genome Res* 2009;**19:**1117-1123.

63. Koboldt DC, Chen K, Wylie T*, et al.* VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* 2009;**25:**2283-2285.

64. Li H, Handsaker B, Wysoker A*, et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;**25:**2078-2079.

65. McKenna A, Hanna M, Banks E*, et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;**20:**1297-1303.

66. Goya R, Sun MG, Morin RD*, et al.* SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. *Bioinformatics* 2010;**26:**730-736.

67. Li R, Li Y, Fang X*, et al.* SNP detection for massively parallel whole-genome resequencing. *Genome Res* 2009;**19:**1124-1132.

68. Roth A, Ding J, Morin R*, et al.* JointSNVMix: a probabilistic model for accurate detection of somatic mutations in normal/tumour paired next-generation sequencing data. *Bioinformatics* 2012;**28:**907-913.

69. Larson DE, Harris CC, Chen K*, et al.* SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics* 2012;**28:**311-317.

70. Ding J, Bashashati A, Roth A*, et al.* Feature-based classifiers for somatic mutation detection in tumour-normal paired sequencing data. *Bioinformatics* 2012;**28:**167-175.

71. Cibulskis K, Lawrence MS, Carter SL*, et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013;**31:**213-219.

72. Albers CA, Lunter G, MacArthur DG*, et al.* Dindel: accurate indel calls from short-read data. *Genome Res* 2011;**21:**961-973.

73. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010;**467:**1061-1073.

74. Mullaney JM, Mills RE, Pittard WS*, et al.* Small insertions and deletions (INDELs) in human genomes. *Hum Mol Genet* 2010;**19:**R131-6.

75. Saunders CT, Wong WS, Swamy S*, et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 2012;**28:**1811-1817.

76. Futreal PA, Coin L, Marshall M*, et al.* A census of human cancer genes. *Nat Rev Cancer* 2004;**4:**177-183.

77. Korbel JO, Urban AE, Affourtit JP*, et al.* Paired-end mapping reveals extensive structural variation in the human genome. *Science* 2007;**318:**420-426.

78. Korbel JO, Abyzov A, Mu XJ*, et al.* PEMer: a computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data. *Genome Biol* 2009;**10:**R23.

79. Hormozdiari F, Alkan C, Eichler EE*, et al.* Combinatorial algorithms for structural variation detection in high-throughput sequenced genomes. *Genome Res* 2009;**19:**1270-1278.

80. Chen K, Wallis JW, McLellan MD*, et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* 2009;**6:**677-681.

81. McKernan KJ, Peckham HE, Costa GL*, et al.* Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res* 2009;**19:**1527-1541.

82. Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol* 2011;**12:**R72-2011-12-8-r72.

83. McPherson A, Hormozdiari F, Zayed A*, et al.* deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput Biol* 2011;**7:**e1001138.

84. Sboner A, Habegger L, Pflueger D*, et al.* FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. *Genome Biol* 2010;**11:**R104-2010-11-10-r104. Epub 2010 Oct 21.

85. Human Genome Structural Variation Working Group, Eichler EE, Nickerson DA*, et al.* Completing the map of human genetic variation. *Nature* 2007;**447:**161-165.

86. Kidd JM, Cooper GM, Donahue WF*, et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature* 2008;**453:**56-64.

87. Robertson G, Schein J, Chiu R*, et al.* De novo assembly and analysis of RNA-seq data. *Nat Methods* 2010;**7:**909-912.

88. Grabherr MG, Haas BJ, Yassour M*, et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011;**29:**644-652.

89. Xie C, Tammi MT. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics* 2009;**10:**80.

90. Chiang DY, Getz G, Jaffe DB*, et al.* High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nat Methods* 2009;**6:**99-103.

91. Bentley DR, Balasubramanian S, Swerdlow HP*, et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008;**456:**53-59.

92. Yoon S, Xuan Z, Makarov V*, et al.* Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res* 2009;**19:**1586-1592.

93. Campbell PJ, Stephens PJ, Pleasance ED*, et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet* 2008;**40:**722-729.

94. Alkan C, Kidd JM, Marques-Bonet T*, et al.* Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat Genet* 2009;**41:**1061-1067.

95. Ha G, Roth A, Lai D*, et al.* Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. *Genome Res* 2012;**22:**1995-2007.

96. Ha G, Roth A, Khattra J*, et al.* TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome Res* 2014;**24:**1881-1893.

97. Oesper L, Satas G, Raphael BJ. Quantifying tumor heterogeneity in whole-genome and whole-exome sequencing data. *Bioinformatics* 2014;**30:**3532-3540.

98. Fischer A, Vazquez-Garcia I, Illingworth CJ*, et al.* High-definition reconstruction of clonal composition in cancer. *Cell Rep* 2014;**7:**1740-1752.

99. Verhaak RG, Hoadley KA, Purdom E*, et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 2010;**17:**98-110.

100. Rapaport F, Khanin R, Liang Y*, et al.* Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol* 2013;**14:**R95.

101. Gusnanto A, Wood HM, Pawitan Y*, et al.* Correcting for cancer genome size and tumour cell content enables better estimation of copy number alterations from next-generation sequence data. *Bioinformatics* 2012;**28:**40-47.

102. Sherry ST, Ward MH, Kholodov M*, et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;**29:**308-311.

103. Iafrate AJ, Feuk L, Rivera MN*, et al.* Detection of large-scale variation in the human genome. *Nat Genet* 2004;**36:**949-951.

104. Forbes SA, Bhamra G, Bamford S*, et al.* The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet* 2008;**Chapter 10:**Unit 10.11.

105. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* 2007;**7:**233-245.

106. Greenman C, Stephens P, Smith R*, et al.* Patterns of somatic mutation in human cancer genomes. *Nature* 2007;**446:**153-158.

107. Nikolaev SI, Rimoldi D, Iseli C*, et al.* Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma. *Nat Genet* 2011;**44:**133-139.

108. Downing JR, Wilson RK, Zhang J*, et al.* The pediatric cancer genome project. *Nat Genet* 2012;**44:**619-622.

109. Adzhubei IA, Schmidt S, Peshkin L*, et al.* A method and server for predicting damaging missense mutations. *Nat Methods* 2010;**7:**248-249.

110. Reva B, Antipin Y, Sander C. Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biol* 2007;**8:**R232.

111. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res* 2001;**11:**863-874.

112. Greenman C, Wooster R, Futreal PA*, et al.* Statistical analysis of pathogenicity of somatic mutations in cancer. *Genetics* 2006;**173:**2187-2198.

113. Lawrence MS, Stojanov P, Polak P*, et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 2013;**499:**214-218.

114. Vogelstein B, Papadopoulos N, Velculescu VE*, et al.* Cancer genome landscapes. *Science* 2013;**339:**1546-1558.

115. Werner HM, Mills GB, Ram PT. Cancer Systems Biology: a peek into the future of patient care? *Nat Rev Clin Oncol* 2014;**11:**167-176.

116. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 2008;**455:**1061-1068.

117. Marx SJ. Hyperparathyroid and hypoparathyroid disorders. *N Engl J Med* 2000;**343:**1863-1875.

118. Sharretts JM, Simonds WF. Clinical and molecular genetics of parathyroid neoplasms. *Best Pract Res Clin Endocrinol Metab* 2010;**24:**491-502.

119. Haven CJ, van Puijenbroek M, Karperien M*, et al.* Differential expression of the calcium sensing receptor and combined loss of chromosomes 1q and 11q in parathyroid carcinoma. *J Pathol* 2004;**202:**86-94.

120. Sharretts JM, Kebebew E, Simonds WF. Parathyroid cancer. *Semin Oncol* 2010;**37:**580-590.

121. Hundahl SA, Fleming ID, Fremgen AM*, et al.* Two hundred eighty-six cases of parathyroid carcinoma treated in the U.S. between 1985-1995: a National Cancer Data Base Report. The American College of Surgeons Commission on Cancer and the American Cancer Society. *Cancer* 1999;**86:**538-544.

122. Sandelin K, Auer G, Bondeson L*, et al.* Prognostic factors in parathyroid cancer: a review of 95 cases. *World J Surg* 1992;**16:**724-731.

123. Lee PK, Jarosek SL, Virnig BA*, et al.* Trends in the incidence and treatment of parathyroid cancer in the United States. *Cancer* 2007;**109:**1736-1741.

124. Busaidy NL, Jimenez C, Habra MA*, et al.* Parathyroid carcinoma: a 22-year experience. *Head Neck* 2004;**26:**716-726.

125. Wiseman SM, Rigual NR, Hicks WL,Jr*, et al.* Parathyroid carcinoma: a multicenter review of clinicopathologic features and treatment outcomes. *Ear Nose Throat J* 2004;**83:**491-494.

126. Carpten JD, Robbins CM, Villablanca A*, et al.* HRPT2, encoding parafibromin, is mutated in hyperparathyroidism-jaw tumor syndrome. *Nat Genet* 2002;**32:**676-680.

127. Shattuck TM, Valimaki S, Obara T*, et al.* Somatic and germ-line mutations of the HRPT2 gene in sporadic parathyroid carcinoma. *N Engl J Med* 2003;**349:**1722-1729.

128. Cetani F, Pardi E, Borsari S*, et al.* Genetic analyses of the HRPT2 gene in primary hyperparathyroidism: germline and somatic mutations in familial and sporadic parathyroid tumors. *J Clin Endocrinol Metab* 2004;**89:**5583-5591.

129. Bradley KJ, Cavaco BM, Bowl MR*, et al.* Parafibromin mutations in hereditary hyperparathyroidism syndromes and parathyroid tumours. *Clin Endocrinol (Oxf)* 2006;**64:**299-306.

130. Krebs LJ, Shattuck TM, Arnold A. HRPT2 mutational analysis of typical sporadic parathyroid adenomas. *J Clin Endocrinol Metab* 2005;**90:**5015-5017.

131. Chandrasekharappa SC, Guru SC, Manickam P*, et al.* Positional cloning of the gene for multiple endocrine neoplasia-type 1. *Science* 1997;**276:**404-407.

132. Agarwal SK, Kester MB, Debelenko LV*, et al.* Germline mutations of the MEN1 gene in familial multiple endocrine neoplasia type 1 and related states. *Hum Mol Genet* 1997;**6:**1169-1175.

133. Thakker RV, Bouloux P, Wooding C*, et al.* Association of parathyroid tumors in multiple endocrine neoplasia type 1 with loss of alleles on chromosome 11. *N Engl J Med* 1989;**321:**218-224.

134. Pausova Z, Soliman E, Amizuka N*, et al.* Role of the RET proto-oncogene in sporadic hyperparathyroidism and in hyperparathyroidism of multiple endocrine neoplasia type 2. *J Clin Endocrinol Metab* 1996;**81:**2711-2718.

135. Mallya SM, Arnold A. Cyclin D1 in parathyroid disease. *Front Biosci* 2000;**5:**D367-71.

136. Vasef MA, Brynes RK, Sturm M*, et al.* Expression of cyclin D1 in parathyroid carcinomas, adenomas, and hyperplasias: a paraffin immunohistochemical study. *Mod Pathol* 1999;**12:**412-416.

137. Robinson JT, Thorvaldsdottir H, Winckler W*, et al.* Integrative genomics viewer. *Nat Biotechnol* 2011;**29:**24-26.

138. 1000 Genomes Project Consortium, Abecasis GR, Auton A*, et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;**491:**56-65.

139. Birol I, Jackman SD, Nielsen CB*, et al.* De novo transcriptome assembly with ABySS. *Bioinformatics* 2009;**25:**2872-2877.

140. Krzywinski M, Schein J, Birol I*, et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* 2009;**19:**1639-1645.

141. Jones SJ, Laskin J, Li YY*, et al.* Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors. *Genome Biol* 2010;**11:**R82.

142. Asmann YW, Necela BM, Kalari KR*, et al.* Detection of redundant fusion transcripts as biomarkers or disease-specific therapeutic targets in breast cancer. *Cancer Res* 2012;**72:**1921-1928.

143. Mortazavi A, Williams BA, McCue K*, et al.* Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008;**5:**621-628.

144. Hubbard TJ, Aken BL, Ayling S*, et al.* Ensembl 2009. *Nucleic Acids Res* 2009;**37:**D690-7.

145. Benjamini YH, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J Royal Stat Soc* 1995;**57:**289-300.

146. Forbes SA, Bindal N, Bamford S*, et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* 2011;**39:**D945-50.

147. Vivanco I, Sawyers CL. The phosphatidylinositol 3-Kinase AKT pathway in human cancer. *Nat Rev Cancer* 2002;**2:**489-501.

148. Miled N, Yan Y, Hon WC*, et al.* Mechanism of two classes of cancer mutations in the phosphoinositide 3-kinase catalytic subunit. *Science* 2007;**317:**239-242.

149. Laplante M, Sabatini DM. mTOR signaling at a glance. *J Cell Sci* 2009;**122:**3589-3594.

150. Hardt M, Chantaravisoot N, Tamanoi F. Activating mutations of TOR (target of rapamycin). *Genes Cells* 2011;**16:**141-151.

151. Sato T, Nakashima A, Guo L*, et al.* Single amino-acid changes that confer constitutive activation of mTOR are discovered in human cancer. *Oncogene* 2010;**29:**2746-2752.

152. Hirai H, Roussel MF, Kato JY*, et al.* Novel INK4 proteins, p19 and p18, are specific inhibitors of the cyclin D-dependent kinases CDK4 and CDK6. *Mol Cell Biol* 1995;**15:**2672-2681.

153. Sanchez-Aguilera A, Delgado J, Camacho FI*, et al.* Silencing of the p18INK4c gene by promoter hypermethylation in Reed-Sternberg cells in Hodgkin lymphomas. *Blood* 2004;**103:**2351-2357.

154. Leone PE, Walker BA, Jenner MW*, et al.* Deletions of CDKN2C in multiple myeloma: biological and clinical implications. *Clin Cancer Res* 2008;**14:**6033-6041.

155. Wiedemeyer R, Brennan C, Heffernan TP*, et al.* Feedback circuit among INK4 tumor suppressors constrains human glioblastoma development. *Cancer Cell* 2008;**13:**355-364.

156. Uziel T, Zindy F, Xie S*, et al.* The tumor suppressors Ink4c and p53 collaborate independently with Patched to suppress medulloblastoma formation. *Genes Dev* 2005;**19:**2656-2667.

157. Franklin DS, Godfrey VL, Lee H*, et al.* CDK inhibitors p18(INK4c) and p27(Kip1) mediate two separate pathways to collaboratively suppress pituitary tumorigenesis. *Genes Dev* 1998;**12:**2899-2911.

158. Hossain MG, Iwata T, Mizusawa N*, et al.* Expression of p18(INK4C) is down-regulated in human pituitary adenomas. *Endocr Pathol* 2009;**20:**114-121.

159. Tahara H, Smith AP, Gaz RD*, et al.* Parathyroid tumor suppressor on 1p: analysis of the p18 cyclin-dependent kinase inhibitor gene as a candidate. *J Bone Miner Res* 1997;**12:**1330-1334.

160. Ng SB, Bigham AW, Buckingham KJ*, et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 2010;**42:**790-793.

161. Guo C, Chang CC, Wortham M*, et al.* Global identification of MLL2-targeted loci reveals MLL2's role in diverse signaling pathways. *Proc Natl Acad Sci U S A* 2012;**109:**17603-17608.

162. Morin RD, Mendez-Lago M, Mungall AJ*, et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* 2011;**476:**298-303.

163. Pasqualucci L, Trifonov V, Fabbri G*, et al.* Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat Genet* 2011;**43:**830-837.

164. Lohr JG, Stojanov P, Lawrence MS*, et al.* Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc Natl Acad Sci U S A* 2012;**109:**3879-3884.

165. Parsons DW, Li M, Zhang X*, et al.* The genetic landscape of the childhood cancer medulloblastoma. *Science* 2011;**331:**435-439.

166. Pugh TJ, Weeraratne SD, Archer TC*, et al.* Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* 2012;**488:**106-110.

167. Jones DT, Jager N, Kool M*, et al.* Dissecting the genomic complexity underlying medulloblastoma. *Nature* 2012;**488:**100-105.

168. Ho J, Fox D, Innes AM*, et al.* Kabuki syndrome and Crohn disease in a child with familial hypocalciuric hypercalcemia. *J Pediatr Endocrinol Metab* 2010;**23:**975-979.

169. Ito M, Yuan CX, Malik S*, et al.* Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. *Mol Cell* 1999;**3:**361-370.

170. Merz C, Urlaub H, Will CL*, et al.* Protein composition of human mRNPs spliced in vitro and differential requirements for mRNP protein recruitment. *RNA* 2007;**13:**116-128.

171. Lee KM, Hsu I, Tarn WY. TRAP150 activates pre-mRNA splicing and promotes nuclear mRNA degradation. *Nucleic Acids Res* 2010;**38:**3340-3350.

172. Roche KC, Wiechens N, Owen-Hughes T*, et al.* The FHA domain protein SNIP1 is a regulator of the cell cycle and cyclin D1 expression. *Oncogene* 2004;**23:**8185-8195.

173. Witzel II, Koh LF, Perkins ND. Regulation of cyclin D1 gene expression. *Biochem Soc Trans* 2010;**38:**217-222.

174. Bracken CP, Wall SJ, Barre B*, et al.* Regulation of cyclin D1 RNA stability by SNIP1. *Cancer Res* 2008;**68:**7621-7628.

175. Beli P, Lukashchuk N, Wagner SA*, et al.* Proteomic investigations reveal a role for RNA processing factor THRAP3 in the DNA damage response. *Mol Cell* 2012;**46:**212-225.

176. Cha JD, Kim HJ, Cha IH. Genetic alterations in oral squamous cell carcinoma progression detected by combining array-based comparative genomic hybridization and multiplex ligation-dependent probe amplification. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2011;**111:**594-607.

177. Paterlini-Brechot P, Saigo K, Murakami Y*, et al.* Hepatitis B virus-related insertional mutagenesis occurs frequently in human liver cancers and recurrently targets human telomerase gene. *Oncogene* 2003;**22:**3911-3916.

178. Yu JJ, Thornton K, Guo Y*, et al.* An ERCC1 splicing variant involving the 5'-UTR of the mRNA may have a transcriptional modulatory function. *Oncogene* 2001;**20:**7694-7698.

179. Yoon MS, Du G, Backer JM*, et al.* Class III PI-3-kinase activates phospholipase D in an amino acid-sensing mTORC1 pathway. *J Cell Biol* 2011;**195:**435-447.

180. Fang Y, Park IH, Wu AL*, et al.* PLD1 regulates mTOR signaling and mediates Cdc42 activation of S6K1. *Curr Biol* 2003;**13:**2037-2044.

181. Hammond SM, Altshuller YM, Sung TC*, et al.* Human ADP-ribosylation factor-activated phosphatidylcholine-specific phospholipase D defines a new and highly conserved gene family. *J Biol Chem* 1995;**270:**29640-29643.

182. Kam Y, Exton JH. Phospholipase D activity is required for actin stress fiber formation in fibroblasts. *Mol Cell Biol* 2001;**21:**4055-4066.

183. Rizzo MA, Shome K, Watkins SC*, et al.* The recruitment of Raf-1 to membranes is mediated by direct interaction with phosphatidic acid and is independent of association with Ras. *J Biol Chem* 2000;**275:**23911-23918.

184. Garrido JL, Wheeler D, Vega LL*, et al.* Role of phospholipase D in parathyroid hormone type 1 receptor signaling and trafficking. *Mol Endocrinol* 2009;**23:**2048-2059.

185. Carrano AC, Eytan E, Hershko A*, et al.* SKP2 is required for ubiquitin-mediated degradation of the CDK inhibitor p27. *Nat Cell Biol* 1999;**1:**193-199.

186. Nakayama KI, Nakayama K. Ubiquitin ligases: cell-cycle control and cancer. *Nat Rev Cancer* 2006;**6:**369-381.

187. Gay NJ, Packman LC, Weldon MA*, et al.* A leucine-rich repeat peptide derived from the Drosophila Toll receptor forms extended filaments with a beta-sheet structure. *FEBS Lett* 1991;**291:**87-91.

188. Kobe B, Kajava AV. The leucine-rich repeat as a protein recognition motif. *Curr Opin Struct Biol* 2001;**11:**725-732.

189. Sethi N, Yan Y, Quek D*, et al.* Rabconnectin-3 is a functional regulator of mammalian Notch signaling. *J Biol Chem* 2010;**285:**34757-34764.

190. Smith FD, Langeberg LK, Cellurale C*, et al.* AKAP-Lbc enhances cyclic AMP control of the ERK1/2 cascade. *Nat Cell Biol* 2010;**12:**1242-1249.

191. Rubino D, Driggers P, Arbit D*, et al.* Characterization of Brx, a novel Dbl family member that modulates estrogen receptor action. *Oncogene* 1998;**16:**2513-2526.

192. Driggers PH, Segars JH, Rubino DM. The proto-oncoprotein Brx activates estrogen receptor beta by a p38 mitogen-activated protein kinase pathway. *J Biol Chem* 2001;**276:**46792-46797.

193. Sterpetti P, Hack AA, Bashar MP*, et al.* Activation of the Lbc Rho exchange factor proto-oncogene by truncation of an extended C terminus that regulates transformation and targeting. *Mol Cell Biol* 1999;**19:**1334-1345.

194. Diviani D, Soderling J, Scott JD. AKAP-Lbc anchors protein kinase A and nucleates Galpha 12-selective Rho-mediated stress fiber formation. *J Biol Chem* 2001;**276:**44247-44257.

195. Agarwal SK, Schrock E, Kester MB*, et al.* Comparative genomic hybridization analysis of human parathyroid tumors. *Cancer Genet Cytogenet* 1998;**106:**30-36.

196. Sherr CJ. Cancer cell cycles. *Science* 1996;**274:**1672-1677.

197. Schuuring E, Verhoeven E, Mooi WJ*, et al.* Identification and cloning of two overexpressed genes, U21B31/PRAD1 and EMS1, within the amplified chromosome 11q13 region in human carcinomas. *Oncogene* 1992;**7:**355-361.

198. Zhang SY, Caamano J, Cooper F*, et al.* Immunohistochemistry of cyclin D1 in human breast cancer. *Am J Clin Pathol* 1994;**102:**695-698.

199. Nishida N, Fukuda Y, Komeda T*, et al.* Amplification and overexpression of the cyclin D1 gene in aggressive human hepatocellular carcinoma. *Cancer Res* 1994;**54:**3107-3110.

200. Bosch F, Jares P, Campo E*, et al.* PRAD-1/cyclin D1 gene overexpression in chronic lymphoproliferative disorders: a highly specific marker of mantle cell lymphoma. *Blood* 1994;**84:**2726-2732.

201. Yang WI, Zukerberg LR, Motokura T*, et al.* Cyclin D1 (Bcl-1, PRAD1) protein expression in low-grade B-cell lymphomas and reactive hyperplasia. *Am J Pathol* 1994;**145:**86-96.

202. Weinstat-Saslow D, Merino MJ, Manrow RE*, et al.* Overexpression of cyclin D mRNA distinguishes invasive and in situ breast carcinomas from non-malignant lesions. *Nat Med* 1995;**1:**1257-1260.

203. Svedlund J, Auren M, Sundstrom M*, et al.* Aberrant WNT/beta-catenin signaling in parathyroid carcinoma. *Mol Cancer* 2010;**9:**294-4598-9-294.

204. Schulte KM, Talat N. Diagnosis and management of parathyroid cancer. *Nat Rev Endocrinol* 2012;**8:**612-622.

205. Kasaian K, Jones SJ. A new frontier in personalized cancer therapy: mapping molecular changes. *Future Oncol* 2011;**7:**873-894.

206. Heppner C, Kester MB, Agarwal SK*, et al.* Somatic mutation of the MEN1 gene in parathyroid tumours. *Nat Genet* 1997;**16:**375-378.

207. Farnebo F, Teh BT, Kytola S*, et al.* Alterations of the MEN1 gene in sporadic parathyroid tumors. *J Clin Endocrinol Metab* 1998;**83:**2627-2630.

208. Haven CJ, van Puijenbroek M, Tan MH*, et al.* Identification of MEN1 and HRPT2 somatic mutations in paraffin-embedded (sporadic) parathyroid carcinomas. *Clin Endocrinol (Oxf)* 2007;**67:**370-376.

209. Dreijerink KM, Hoppener JW, Timmers HM*, et al.* Mechanisms of disease: multiple endocrine neoplasia type 1-relation to chromatin modifications and transcription regulation. *Nat Clin Pract Endocrinol Metab* 2006;**2:**562-570.

210. Agarwal SK, Guru SC, Heppner C*, et al.* Menin interacts with the AP1 transcription factor JunD and represses JunD-activated transcription. *Cell* 1999;**96:**143-152.

211. Heppner C, Bilimoria KY, Agarwal SK*, et al.* The tumor suppressor protein menin interacts with NF-kappaB proteins and inhibits NF-kappaB-mediated transactivation. *Oncogene* 2001;**20:**4917-4925.

212. Kaji H, Canaff L, Lebrun JJ*, et al.* Inactivation of menin, a Smad3-interacting protein, blocks transforming growth factor type beta signaling. *Proc Natl Acad Sci U S A* 2001;**98:**3837-3842.

213. Lemmens IH, Forsberg L, Pannett AA*, et al.* Menin interacts directly with the homeobox-containing protein Pem. *Biochem Biophys Res Commun* 2001;**286:**426-431.

214. Kim H, Lee JE, Cho EJ*, et al.* Menin, a tumor suppressor, represses JunD-mediated transcriptional activity by association with an mSin3A-histone deacetylase complex. *Cancer Res* 2003;**63:**6135-6139.

215. Hughes CM, Rozenblatt-Rosen O, Milne TA*, et al.* Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. *Mol Cell* 2004;**13:**587-597.

216. Yokoyama A, Wang Z, Wysocka J*, et al.* Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol Cell Biol* 2004;**24:**5639-5649.

217. Santos-Rosa H, Schneider R, Bannister AJ*, et al.* Active genes are tri-methylated at K4 of histone H3. *Nature* 2002;**419:**407-411.

218. Bai F, Pei XH, Godfrey VL*, et al.* Haploinsufficiency of p18(INK4c) sensitizes mice to carcinogen-induced tumorigenesis. *Mol Cell Biol* 2003;**23:**1269-1277.

219. Agarwal SK, Mateo CM, Marx SJ. Rare germline mutations in cyclin-dependent kinase inhibitor genes in multiple endocrine neoplasia type 1 and related states. *J Clin Endocrinol Metab* 2009;**94:**1826-1834.

220. Buchwald PC, Akerstrom G, Westin G. Reduced p18INK4c, p21CIP1/WAF1 and p27KIP1 mRNA levels in tumours of primary and secondary hyperparathyroidism. *Clin Endocrinol (Oxf)* 2004;**60:**389-393.

221. Costa-Guda J, Marinoni I, Molatore S*, et al.* Somatic mutation and germline sequence abnormalities in CDKN1B, encoding p27Kip1, in sporadic parathyroid adenomas. *J Clin Endocrinol Metab* 2011;**96:**E701-6.

222. Pellegata NS, Quintanilla-Martinez L, Siggelkow H*, et al.* Germ-line mutations in p27Kip1 cause a multiple endocrine neoplasia syndrome in rats and humans. *Proc Natl Acad Sci U S A* 2006;**103:**15558-15563.

223. Fero ML, Rivkin M, Tasch M*, et al.* A syndrome of multiorgan hyperplasia with features of gigantism, tumorigenesis, and female sterility in p27(Kip1)-deficient mice. *Cell* 1996;**85:**733-744.

224. Kiyokawa H, Kineman RD, Manova-Todorova KO*, et al.* Enhanced growth of mice lacking the cyclin-dependent kinase inhibitor function of p27(Kip1). *Cell* 1996;**85:**721-732.

225. Nakayama K, Ishida N, Shirane M*, et al.* Mice lacking p27(Kip1) display increased body size, multiple organ hyperplasia, retinal dysplasia, and pituitary tumors. *Cell* 1996;**85:**707-720.

226. Franklin DS, Godfrey VL, O'Brien DA*, et al.* Functional collaboration between different cyclin-dependent kinase inhibitors suppresses tumor growth with distinct tissue specificity. *Mol Cell Biol* 2000;**20:**6147-6158.

227. Wu X, Hua X. Menin, histone h3 methyltransferases, and regulation of cell proliferation: current knowledge and perspective. *Curr Mol Med* 2008;**8:**805-815.

228. Milne TA, Hughes CM, Lloyd R*, et al.* Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors. *Proc Natl Acad Sci U S A* 2005;**102:**749-754.

229. Karnik SK, Hughes CM, Gu X*, et al.* Menin regulates pancreatic islet growth by promoting histone methylation and expression of genes encoding p27Kip1 and p18INK4c. *Proc Natl Acad Sci U S A* 2005;**102:**14659-14664.

230. Engelman JA. Targeting PI3K signalling in cancer: opportunities, challenges and limitations. *Nat Rev Cancer* 2009;**9:**550-562.

231. Toschi A, Lee E, Xu L*, et al.* Regulation of mTORC1 and mTORC2 complex assembly by phosphatidic acid: competition with rapamycin. *Mol Cell Biol* 2009;**29:**1411-1420.

232. Fang Y, Vilella-Bach M, Bachmann R*, et al.* Phosphatidic acid-mediated mitogenic activation of mTOR signaling. *Science* 2001;**294:**1942-1945.

233. Sun Y, Chen J. mTOR signaling: PLD takes center stage. *Cell Cycle* 2008;**7:**3118-3123.

234. Chen Y, Zheng Y, Foster DA. Phospholipase D confers rapamycin resistance in human breast cancer cells. *Oncogene* 2003;**22:**3937-3942.

235. Hedinger C, Williams ED, Sobin LH. The WHO histological classification of thyroid tumors: a commentary on the second edition. *Cancer* 1989;**63:**908-911.

236. American Thyroid Association (ATA) Guidelines Taskforce on Thyroid Nodules and Differentiated Thyroid Cancer, Cooper DS, Doherty GM*, et al.* Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* 2009;**19:**1167-1214.

237. Goffredo P, Roman SA, Sosa JA. Hurthle cell carcinoma: a population-level analysis of 3311 patients. *Cancer* 2013;**119:**504-511.

238. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 2000;**132:**365-386.

239. Wijnen J, van der Klift H, Vasen H*, et al.* MSH2 genomic deletions are a frequent cause of HNPCC. *Nat Genet* 1998;**20:**326-328.

240. Pritchard CC, Morrissey C, Kumar A*, et al.* Complex MSH2 and MSH6 mutations in hypermutated microsatellite unstable advanced prostate cancer. *Nat Commun* 2014;**5:**4988.

241. Lu Y, Soong TD, Elemento O. A novel approach for characterizing microsatellite instability in cancer cells. *PLoS One* 2013;**8:**e63056.

242. Bassett JH, Forbes SA, Pannett AA*, et al.* Characterization of mutations in patients with multiple endocrine neoplasia type 1. *Am J Hum Genet* 1998;**62:**232-244.

243. Cromer MK, Starker LF, Choi M*, et al.* Identification of somatic mutations in parathyroid tumors using whole-exome sequencing. *J Clin Endocrinol Metab* 2012;**97:**E1774-81.

244. Xing M. Molecular pathogenesis and mechanisms of thyroid cancer. *Nat Rev Cancer* 2013;**13:**184-199.

245. Maximo V, Sobrinho-Simoes M. Hurthle cell tumours of the thyroid. A review with emphasis on mitochondrial abnormalities with clinical relevance. *Virchows Arch* 2000;**437:**107-115.

246. Gasparre G, Porcelli AM, Bonora E*, et al.* Disruptive mitochondrial DNA mutations in complex I subunits are markers of oncocytic phenotype in thyroid tumors. *Proc Natl Acad Sci U S A* 2007;**104:**9001-9006.

247. Corver WE, Ruano D, Weijers K*, et al.* Genome haploidisation with chromosome 7 retention in oncocytic follicular thyroid carcinoma. *PLoS One* 2012;**7:**e38287.

248. Giraud S, Zhang CX, Serova-Sinilnikova O*, et al.* Germ-line mutation analysis in patients with multiple endocrine neoplasia type 1 and related disorders. *Am J Hum Genet* 1998;**63:**455-467.

249. Wautot V, Vercherat C, Lespinasse J*, et al.* Germline mutation profile of MEN1 in multiple endocrine neoplasia type 1: search for correlation between phenotype and the functional domains of the MEN1 protein. *Hum Mutat* 2002;**20:**35-47.

250. Debelenko LV, Brambilla E, Agarwal SK*, et al.* Identification of MEN1 gene mutations in sporadic carcinoid tumors of the lung. *Hum Mol Genet* 1997;**6:**2285-2290.

251. Zhuang Z, Vortmeyer AO, Pack S*, et al.* Somatic mutations of the MEN1 tumor suppressor gene in sporadic gastrinomas and insulinomas. *Cancer Res* 1997;**57:**4682-4686.

252. Cerami E, Gao J, Dogrusoz U*, et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012;**2:**401-404.

253. Kim HJ, Park JS, Kim CS*, et al.* A case of multiple endocrine neoplasia type 1 combined with papillary thyroid carcinoma. *Yonsei Med J* 2008;**49:**503-506.

254. Desai D, McPherson LA, Higgins JP*, et al.* Genetic analysis of a papillary thyroid carcinoma in a patient with MEN1. *Ann Surg Oncol* 2001;**8:**342-346.

255. Vortmeyer AO, Lubensky IA, Skarulis M*, et al.* Multiple endocrine neoplasia type 1: atypical presentation, clinical course, and genetic analysis of multiple tumors. *Mod Pathol* 1999;**12:**919-924.

256. Kim YL, Jang YW, Kim JT*, et al.* A rare case of primary hyperparathyroidism associated with primary aldosteronism, Hurthle cell thyroid cancer and meningioma. *J Korean Med Sci* 2012;**27:**560-564.

257. Pinna G, Orgiana G, Carcassi C*, et al.* A novel germline mutation of MEN 1 gene in a patient with acromegaly and multiple endocrine tumors. *J Endocrinol Invest* 2004;**27:**577-582.

258. Guru SC, Goldsmith PK, Burns AL*, et al.* Menin, the product of the MEN1 gene, is a nuclear protein. *Proc Natl Acad Sci U S A* 1998;**95:**1630-1634.

259. Kim YS, Burns AL, Goldsmith PK*, et al.* Stable overexpression of MEN1 suppresses tumorigenicity of RAS. *Oncogene* 1999;**18:**5936-5942.

260. Schnepp RW, Mao H, Sykes SM*, et al.* Menin induces apoptosis in murine embryonic fibroblasts. *J Biol Chem* 2004;**279:**10685-10691.

261. Jin S, Mao H, Schnepp RW*, et al.* Menin associates with FANCD2, a protein involved in repair of DNA damage. *Cancer Res* 2003;**63:**4204-4210.

262. Sakurai A, Katai M, Itakura Y*, et al.* Premature centromere division in patients with multiple endocrine neoplasia type 1. *Cancer Genet Cytogenet* 1999;**109:**138-140.

263. Smallridge RC, Ain KB, Asa SL*, et al.* American Thyroid Association guidelines for management of patients with anaplastic thyroid cancer. *Thyroid* 2012;**22:**1104-1139.

264. Kebebew E, Greenspan FS, Clark OH*, et al.* Anaplastic thyroid carcinoma. Treatment outcome and prognostic factors. *Cancer* 2005;**103:**1330-1335.

265. Sugitani I, Miyauchi A, Sugino K*, et al.* Prognostic factors and treatment outcomes for anaplastic thyroid carcinoma: ATC Research Consortium of Japan cohort study of 677 patients. *World J Surg* 2012;**36:**1247-1254.

266. Wiseman SM, Loree TR, Hicks WL,Jr*, et al.* Anaplastic thyroid cancer evolved from papillary carcinoma: demonstration of anaplastic transformation by means of the inter-simple sequence repeat polymerase chain reaction. *Arch Otolaryngol Head Neck Surg* 2003;**129:**96-100.

267. Ragazzi M, Ciarrocchi A, Sancisi V*, et al.* Update on anaplastic thyroid carcinoma: morphological, molecular, and genetic features of the most aggressive thyroid cancer. *Int J Endocrinol* 2014;**2014:**790834.

268. Schweppe RE, Klopper JP, Korch C*, et al.* Deoxyribonucleic acid profiling analysis of 40 human thyroid cancer cell lines reveals cross-contamination resulting in cell line redundancy and misidentification. *J Clin Endocrinol Metab* 2008;**93:**4331-4341.

269. Marlow LA, D'Innocenzi J, Zhang Y*, et al.* Detailed molecular fingerprinting of four new anaplastic thyroid carcinoma cell lines and their use for verification of RhoB as a molecular therapeutic target. *J Clin Endocrinol Metab* 2010;**95:**5338-5347.

270. Cancer Genome Atlas Research Network. Electronic address: giordano@umich.edu, Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. *Cell* 2014;**159:**676-690.

271. Parkhomchuk D, Borodina T, Amstislavskiy V*, et al.* Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res* 2009;**37:**e123.

272. Kim D, Pertea G, Trapnell C*, et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013;**14:**R36-2013-14-4-r36.

273. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;**31:**166-169.

274. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;**26:**139-140.

275. Subramanian A, Tamayo P, Mootha VK*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;**102:**15545-15550.

276. Kunstman JW, Juhlin CC, Goh G*, et al.* Characterization of the mutational landscape of anaplastic thyroid cancer via whole-exome sequencing. *Hum Mol Genet* 2015;.

277. Smallridge RC, Marlow LA, Copland JA. Anaplastic thyroid cancer: molecular pathogenesis and emerging therapies. *Endocr Relat Cancer* 2009;**16:**17-44.

278. Wiseman SM, Loree TR, Rigual NR*, et al.* Anaplastic transformation of thyroid cancer: review of clinical, pathologic, and molecular evidence provides new insights into disease biology and future therapy. *Head Neck* 2003;**25:**662-670.

279. Borad MJ, Champion MD, Egan JB*, et al.* Integrated genomic characterization reveals novel, therapeutically relevant drug targets in FGFR and EGFR pathways in sporadic intrahepatic cholangiocarcinoma. *PLoS Genet* 2014;**10:**e1004135.

280. Nagai M, Tanaka S, Tsuda M*, et al.* Analysis of transforming activity of human synovial sarcoma-associated chimeric protein SYT-SSX1 bound to chromatin remodeling factor hBRM/hSNF2 alpha. *Proc Natl Acad Sci U S A* 2001;**98:**3843-3848.

281. Kadoch C, Hargreaves DC, Hodges C*, et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat Genet* 2013;**45:**592-601.

282. Eid JE, Kung AL, Scully R*, et al.* p300 interacts with the nuclear proto-oncoprotein SYT as part of the active control of cell adhesion. *Cell* 2000;**102:**839-848.

283. Biankin AV, Waddell N, Kassahn KS*, et al.* Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* 2012;**491:**399-405.

284. Wiseman SM, Masoudi H, Niblock P*, et al.* Anaplastic thyroid carcinoma: expression profile of targets for therapy offers new insights for disease treatment. *Ann Surg Oncol* 2007;**14:**719-729.

285. Nikiforov YE. Editorial: anaplastic carcinoma of the thyroid--will aurora B light a path for treatment? *J Clin Endocrinol Metab* 2005;**90:**1243-1245.

286. Akar U, Ozpolat B, Mehta K*, et al.* Tissue transglutaminase inhibits autophagy in pancreatic cancer cells. *Mol Cancer Res* 2007;**5:**241-249.

287. Shimaoka K, Schoenfeld DA, DeWys WD*, et al.* A randomized trial of doxorubicin versus doxorubicin plus cisplatin in patients with advanced thyroid carcinoma. *Cancer* 1985;**56:**2155-2160.

288. Gottlieb JA, Hill CS,Jr. Chemotherapy of thyroid cancer with adriamycin. Experience with 30 patients. *N Engl J Med* 1974;**290:**193-197.

289. Wagle N, Grabiner BC, Van Allen EM*, et al.* Response and acquired resistance to everolimus in anaplastic thyroid cancer. *N Engl J Med* 2014;**371:**1426-1433.

290. Grande E, Capdevila J, Diez JJ*, et al.* A significant response to sunitinib in a patient with anaplastic thyroid carcinoma. *J Res Med Sci* 2013;**18:**623-625.

291. Rosove MH, Peddi PF, Glaspy JA. BRAF V600E inhibition in anaplastic thyroid cancer. *N Engl J Med* 2013;**368:**684-685.

292. Bible KC, Suman VJ, Menefee ME*, et al.* A multiinstitutional phase 2 trial of pazopanib monotherapy in advanced anaplastic thyroid cancer. *J Clin Endocrinol Metab* 2012;**97:**3179-3184.

293. Ha HT, Lee JS, Urba S*, et al.* A phase II study of imatinib in patients with advanced anaplastic thyroid cancer. *Thyroid* 2010;**20:**975-980.

294. Pennell NA, Daniels GH, Haddad RI*, et al.* A phase II study of gefitinib in patients with advanced thyroid cancer. *Thyroid* 2008;**18:**317-323.

295. Cohen EE, Rosen LS, Vokes EE*, et al.* Axitinib is an active treatment for all histologic subtypes of advanced thyroid cancer: results from a phase II study. *J Clin Oncol* 2008;**26:**4708-4713.

296. Gupta-Abramson V, Troxel AB, Nellore A*, et al.* Phase II trial of sorafenib in advanced thyroid cancer. *J Clin Oncol* 2008;**26:**4714-4719.

297. Savvides P, Nagaiah G, Lavertu P*, et al.* Phase II trial of sorafenib in patients with advanced anaplastic carcinoma of the thyroid. *Thyroid* 2013;**23:**600-604.

298. Takahashi S. Phase II study of lenvatinib (LEN), a multi-targeted tyrosine kinase inhibitor, in patients with all histologic subtypes of advanced thyroid cancer (differentiated, medullary and anaplastic). 2014;.

299. Mooney CJ, Nagaiah G, Fu P*, et al.* A phase II trial of fosbretabulin in advanced anaplastic thyroid carcinoma and correlation of baseline serum-soluble intracellular adhesion molecule-1 with outcome. *Thyroid* 2009;**19:**233-240.

300. Sosa JA, Elisei R, Jarzab B*, et al.* Randomized safety and efficacy study of fosbretabulin with paclitaxel/carboplatin against anaplastic thyroid carcinoma. *Thyroid* 2014;**24:**232-240.

301. Banerjee R, Russo N, Liu M*, et al.* TRIP13 promotes error-prone nonhomologous end joining and induces chemoresistance in head and neck cancer. *Nat Commun* 2014;**5:**4527.

302. Land SC, Tee AR. Hypoxia-inducible factor 1alpha is regulated by the mammalian target of rapamycin (mTOR) via an mTOR signaling motif. *J Biol Chem* 2007;**282:**20534-20543.

303. Papewalis C, Wuttke M, Schinner S*, et al.* Role of the novel mTOR inhibitor RAD001 (everolimus) in anaplastic thyroid cancer. *Horm Metab Res* 2009;**41:**752-756.

304. Gaude E, Frezza C. Defects in mitochondrial metabolism and cancer. *Cancer Metab* 2014;**2:**10-3002-2-10. eCollection 2014.

305. Kaelin WG,Jr, McKnight SL. Influence of metabolism on epigenetics and disease. *Cell* 2013;**153:**56-69.

306. Wiederschain D, Chen L, Johnson B*, et al.* Contribution of polycomb homologues Bmi-1 and Mel-18 to medulloblastoma pathogenesis. *Mol Cell Biol* 2007;**27:**4968-4979.

307. Wiseman SM, Baliski C, Irvine R*, et al.* Hemithyroidectomy: the optimal initial surgical approach for individuals undergoing surgery for a cytological diagnosis of follicular neoplasm. *Ann Surg Oncol* 2006;**13:**425-432.

308. Davies L, Welch HG. Current thyroid cancer trends in the United States. *JAMA Otolaryngol Head Neck Surg* 2014;**140:**317-322.

309. Hay ID, Bergstralh EJ, Goellner JR*, et al.* Predicting outcome in papillary thyroid carcinoma: development of a reliable prognostic scoring system in a cohort of 1779 patients surgically treated at one institution during 1940 through 1989. *Surgery* 1993;**114:**1050-7; discussion 1057-8.

310. Matsuzu K, Sugino K, Masudo K*, et al.* Thyroid lobectomy for papillary thyroid cancer: long-term follow-up study of 1,088 cases. *World J Surg* 2014;**38:**68-79.

311. Wiseman SM, Griffith OL, Melck A*, et al.* Evaluation of type 1 growth factor receptor family expression in benign and malignant thyroid lesions. *Am J Surg* 2008;**195:**667-73; discussion 673.

312. Wiseman SM, Melck A, Masoudi H*, et al.* Molecular phenotyping of thyroid tumors identifies a marker panel for differentiated thyroid cancer diagnosis. *Ann Surg Oncol* 2008;**15:**2811-2826.

313. Goetz JG, Lajoie P, Wiseman SM*, et al.* Caveolin-1 in tumor progression: the good, the bad and the ugly. *Cancer Metastasis Rev* 2008;**27:**715-735.

314. Shankar J, Wiseman SM, Meng F*, et al.* Coordinated expression of galectin-3 and caveolin-1 in thyroid cancer. *J Pathol* 2012;**228:**56-66.

315. Eckert RL. Sequence of the human 40-kDa keratin reveals an unusual structure with very high sequence identity to the corresponding bovine keratin. *Proc Natl Acad Sci U S A* 1988;**85:**1114-1118.

316. Nasser SM, Pitman MB, Pilch BZ*, et al.* Fine-needle aspiration biopsy of papillary thyroid carcinoma: diagnostic utility of cytokeratin 19 immunostaining. *Cancer* 2000;**90:**307-311.

317. Xing M. Genetic alterations in the phosphatidylinositol-3 kinase/Akt pathway in thyroid cancer. *Thyroid* 2010;**20:**697-706.

318. Yoshii T, Inohara H, Takenaka Y*, et al.* Galectin-3 maintains the transformed phenotype of thyroid papillary carcinoma cells. *Int J Oncol* 2001;**18:**787-792.

319. Bell JL, Malyukova A, Kavallaris M*, et al.* TRIM16 inhibits neuroblastoma cell proliferation through cell cycle regulation and dynamic nuclear localization. *Cell Cycle* 2013;**12:**889-898.

320. Schlisio S, Kenchappa RS, Vredeveld LC*, et al.* The kinesin KIF1Bbeta acts downstream from EglN3 to induce apoptosis and is a potential 1p36 tumor suppressor. *Genes Dev* 2008;**22:**884-893.

321. van de Graaf SA, Ris-Stalpers C, Veenboer GJ*, et al.* A premature stopcodon in thyroglobulin messenger RNA results in familial goiter and moderate hypothyroidism. *J Clin Endocrinol Metab* 1999;**84:**2537-2542.

322. Targovnik HM, Rivolta CM, Mendive FM*, et al.* Congenital goiter with hypothyroidism caused by a 5' splice site mutation in the thyroglobulin gene. *Thyroid* 2001;**11:**685-690.

323. Kim D, Pertea G, Trapnell C*, et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013;**14:**R36.

324. Anders S, Pyl PT, Huber W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2014;.

325. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009;**37:**1-13.

326. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;**28:**27-30.

327. Kang JS, Gao M, Feinleib JL*, et al.* CDO: an oncogene-, serum-, and anchorage-regulated member of the Ig/fibronectin type III repeat family. *J Cell Biol* 1997;**138:**203-213.

328. Gorbatenko A, Olesen CW, Boedtkjer E*, et al.* Regulation and roles of bicarbonate transporters in cancer. *Front Physiol* 2014;**5:**130.

329. Galeza-Kulik M, Zebracka J, Szpak-Ulczok S*, et al.* Expression of selected genes involved in transport of ions in papillary thyroid carcinoma. *Endokrynol Pol* 2006;**57 Suppl A:**26-31.

330. Jevnikar Z, Rojnik M, Jamnik P*, et al.* Cathepsin H mediates the processing of talin and regulates migration of prostate cancer cells. *J Biol Chem* 2013;**288:**2201-2209.

331. Nunes-Xavier CE, Elson A, Pulido R. Epidermal growth factor receptor (EGFR)-mediated positive feedback of protein-tyrosine phosphatase epsilon (PTPepsilon) on ERK1/2 and AKT protein pathways is required for survival of human breast cancer cells. *J Biol Chem* 2012;**287:**3433-3444.

332. Ferrari N, Mohammed ZM, Nixon C*, et al.* Expression of RUNX1 correlates with poor patient prognosis in triple negative breast cancer. *PLoS One* 2014;**9:**e100759.

333. Goyama S, Schibler J, Cunningham L*, et al.* Transcription factor RUNX1 promotes survival of acute myeloid leukemia cells. *J Clin Invest* 2013;**123:**3876-3888.

334. Murray GI, Taylor MC, McFadyen MC*, et al.* Tumor-specific expression of cytochrome P450 CYP1B1. *Cancer Res* 1997;**57:**3026-3031.

335. Patel SA, Bhambra U, Charalambous MP*, et al.* Interleukin-6 mediated upregulation of CYP1B1 and CYP2E1 in colorectal cancer involves DNA methylation, miR27b and STAT3. *Br J Cancer* 2014;**111:**2287-2296.

336. Yin BW, Kiyamova R, Chua R*, et al.* Monoclonal antibody MX35 detects the membrane transporter NaPi2b (SLC34A2) in human carcinomas. *Cancer Immun* 2008;**8:**3.

337. Gonzalez-Angulo AM, Ferrer-Lozano J, Stemke-Hale K*, et al.* PI3K pathway mutations and PTEN levels in primary and metastatic breast cancer. *Mol Cancer Ther* 2011;**10:**1093-1101.

338. Yu W, McPherson JR, Stevenson M*, et al.* Whole-exome sequencing studies of parathyroid carcinomas reveal novel PRUNE2 mutations, distinctive mutational spectra related to APOBEC-catalyzed DNA mutagenesis and mutational enrichment in kinases associated with cell migration and invasion. *J Clin Endocrinol Metab* 2015;**100:**E360-4.

339. Mehlen P, Delloye-Bourgeois C, Chedotal A. Novel roles for Slits and netrins: axon guidance cues as anticancer targets? *Nat Rev Cancer* 2011;**11:**188-197.

340. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature* 2009;**458:**719-724.

341. Sims D, Sudbery I, Ilott NE*, et al.* Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* 2014;**15:**121-132.

342. Ross DS. Nonpalpable thyroid nodules--managing an epidemic. *J Clin Endocrinol Metab* 2002;**87:**1938-1940.

343. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med* 2013;**19:**1423-1437.

344. Stewart TJ, Abrams SI. How tumours escape mass destruction. *Oncogene* 2008;**27:**5894-5903.

345. Angell TE, Lechner MG, Jang JK*, et al.* MHC class I loss is a frequent mechanism of immune escape in papillary thyroid cancer that is reversed by interferon and selumetinib treatment in vitro. *Clin Cancer Res* 2014;**20:**6034-6044.

346. Matsubayashi S, Kawai K, Matsumoto Y*, et al.* The correlation between papillary thyroid carcinoma and lymphocytic infiltration in the thyroid gland. *J Clin Endocrinol Metab* 1995;**80:**3421-3424.

347. Fugazzola L, Colombo C, Perrino M*, et al.* Papillary thyroid carcinoma and inflammation. *Front Endocrinol (Lausanne)* 2011;**2:**88.