

©[2019]

Saurabh V. Laddha

**ALL RIGHTS RESERVED**

GENOMIC CHARACTERIZATION OF HUMAN NEUROENDOCRINE TUMORS

By

SAURABH V. LADDHA

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Quantitative Biomedicine

Written under the direction of

Chang S. Chan

And approved by

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

New Brunswick, New Jersey

January 2019

## ABSTRACT OF THE DISSERTATION

### GENOMIC CHARACTERIZATION OF HUMAN NEUROENDOCRINE TUMORS

by SAURABH V. LADDHA

Dissertation Director:

Chang S. Chan

Neuroendocrine tumors (NETs) are a rare, slow growing and biologically poorly understood presenting unique clinical challenges. The majority of NETs are localized to the gastrointestinal tract (predominately small intestinal and pancreas) and lung. NETs are well differentiated and the majority follows a benign course. However, these benign tumors can transform to malignant disease and results in adverse clinical outcome with few therapeutic options. The WHO classification of NETs has evolved over the last two decades but still lack clinical biomarkers for NETs stratifications. With greater awareness of NETs in clinic and improvement in diagnostic imaging techniques, the incidence rate of NETs has increased. Despite the increased incidence rate, the biological knowledge of these NETs is limited.

The central theme of this thesis was to provide greater insights into NET biology including clinically relevant molecular subtypes, tumorigenesis pathways, tumor cell-of-

origins and biomarkers for translation research. Specifically, the thesis focuses on genomic characterization of three major NET types: pancreatic NETs, lung carcinoids and small intestine NETs. The projects discussed in chapter 2, 3 and 4 involved the integration of genomics dataset (DNAseq, RNAseq and DNA CpG methylation) accompanied by clinical information.

In Chapter 2, I discuss the results for pancreatic NETs (PanNETs). We identified two molecular subtypes of PanNETs with distinct genotype and clinical phenotypes. PanNETs with mutation in *ATRX*, *DAXX* or *MEN1* gene (A-D-M mutant subtype) have adverse clinical outcome and resemble the gene expression profile of pancreatic alpha cells. We identified novel gene signature and biomarkers that differentiate PanNETs genotypes and gained an enhanced biological understanding of PanNETs from the cell lineage viewpoint.

In Chapter 3, I discuss the results for lung carcinoids. We identified three novel molecular subtypes (LC1, LC2, and LC3) with distinct clinical phenotypes. The recurrent mutations we identified were enriched for genes involved in covalent histone modification/chromatin remodeling (34.5%) (*MEN1*, *ARID1A*, *KMT2C* and *KMT2A*) as well as DNA repair (17.2%) pathways. We found two biomarkers, *ASCL1* and *S100* that can stratify the three subtypes. *MEN1* mutations were found to be exclusively enriched in subtype LC2. Subtype LC1 and LC3 is predominately found at peripheral and endobronchial lung respectively. Subtype LC3 is diagnosed on average 10 years earlier than LC1 and LC2.

In Chapter 4, I discuss the results for small intestine NETs. We identified two (SINET-A and B) molecular subtypes of SI-NETs using gene expression and genetic dataset with distinct cell-of-origin signature. We identified one copy loss of chromosome 18 (chr18) in 85% (22 of 26) of subtype SINET-B while subtype SINET-A is diploid for chr18. We found that SINET-A subtype may originate from *TPHI*<sup>-</sup>/*REG4*<sup>-</sup> neuroendocrine cells of the small intestine and SINET-B from EC cells, which are *TPHI*<sup>+</sup>/*REG4*<sup>+</sup>. Gene expression profile of two potential biomarkers (*LMX1A* and *ONECUT2*) was found to stratify the two subtypes.

Taken together, this research demonstrates the clinically relevant molecular subtypes of NETs with distinct molecular genotypes, cell lineage and clinical phenotypes. This molecular classification of NET subtypes will improve NETs stratification, and may facilitate the molecular understanding of their pathogenesis and improve clinical management.

## Acknowledgments

I would like to thank Dr. Chang S. Chan, my mentor for giving me a chance to join his lab and perform the cutting edge research in the field of cancer genomics and bioinformatics. I would like to thank him for his intellectual and analytical support, giving me an independence to choose research problems, chance to work on multiple cancer informatics related projects at RCINJ and presenting my results at multiple conferences. His guidance helped me to grow as an independent research scientist.

I would like to thank our collaborators at MSKCC (New York), especially Prof. Laura Tang for consistently giving us experimental support to validate my findings and guidance on clinical understanding of my research.

I would like to thank my committee members, Dr. Eileen White, Dr. Shridar Ganesan and Dr. Michael Verzi for your commitment and time, invaluable comments and suggestions and constructive ideas during this work.

I would like to thank my graduate department, Quantitative Biomedicine at Rutgers, specifically Prof. Gail Arnold for keeping tabs on my progress report every semester.

I also thank my friends, colleagues and faculty members at Rutgers CSCB department with whom I discussed my project multiple times and gave me variety of constructive ideas and insights into different computational methods and statistical approaches to

improve my results. Special thanks Dr. Hossein K, Dr. Gyan B, Dr. Subhajyoti D, Dr. Greg R, Dr. Ying C and Dr. Hua K for all your support, help and criticisms on my work.

Finally, I am thankful for being a part of such an extraordinary family with unconditional love and support – my parents, my sisters and her family, my in-laws and family and my wife, Anchal. Special thanks my father and mother. My mother could not complete her high school and always dreamed of me getting highest education in life. Thank you ma and papa for all things you have done for me. I owe a great debt of gratitude to my wife Anchal for her help, support, patience and unconditional love.

**Dedication**

*To all those  
who suffer because of this disease*

*To Ma, Papa and Anchal*

## Table of Contents

Abstract.....	ii
Acknowledgments.....	v
Dedication.....	vii
Table of Contents.....	viii
List of Tables.....	xi
List of Figures.....	xii
List of Publication.....	xv

### Chapter 1: Introduction and Background

1.1 Neuroendocrine Cells.....	1
1.2 Neuroendocrine Tumor.....	4
1.3 Neuroendocrine Tumor Classifications.....	6
1.4 World Health organization Classifications and Grading.....	6
1.5 Epidemiology of Neuroendocrine Tumor.....	8
1.6 Biomarkers, Diagnosis and Treatment for Neuroendocrine Tumor.....	10
1.7 Genomics of NETs.....	12
1.7.1 Pancreatic NETs.....	14
1.7.2 Lung Carcinoids.....	17
1.7.3 Small Intestine NETs.....	19
1.8 Thesis Overview	
1.8.1 Objectives and Aims.....	22
1.8.2 Multi-Layered NETs dataset and Meta-Analysis for Aim 1 and 2.....	23
1.8.3 Chapter Overview.....	24

<b>Chapter 2: ATRX, DAXX or MEN1 mutant pancreatic neuroendocrine tumors are a distinct alpha-cell signature subgroup (Published in Nature Communications).....</b>	<b>25</b>
---	-----------

2.1 Abstract.....	26
2.2 Introduction.....	26
2.3 Materials and Methods.....	29
2.4 Results.....	36
2.4.1 Patient cohort, clinical annotation, and genotyping for ATRX, DAXX and MEN1	
2.4.2 Gene expression and DNA methylation reveal two subtypes of PanNETs	
2.4.3 A-D-M mutant PanNET gene expression resembles that of alpha cells	
2.4.4 Validation of distinct subtype and alpha cell signature in A-D-M mutant PanNETs	
2.4.5 HNF1A pathway is transcriptionally upregulated in A-D-M mutant PanNETs and alpha cells	
2.4.6 Integrative analysis reveals PDX1 gene is hypermethylated with low expression in A-D-M mutant PanNETs	
2.5 Discussion and Conclusions.....	49
<b>Chapter 3: Integrative Genomic Analysis Identifies Molecular Subtypes of Lung Carcinoids .....</b>	<b>54</b>
3.1 Abstract.....	55
3.2 Introduction.....	55
3.3 Materials and Methods.....	57
3.4 Results.....	60
3.4.1 Patient cohort, clinical annotations and mutational profile of lung carcinoids in a 354-cancer gene panel	
3.4.2 Transcriptome and methylome profiles reveal three distinct subtypes	
3.4.3 Subtype-specific molecular characterization of lung carcinoids	
3.4.4 Independent validation of lung carcinoid classification	

3.4.5 ASCL1 and S100 are novel biomarkers for lung carcinoid subtypes	
3.4.6 Lung carcinoid subtypes have distinct clinical phenotypes	
3.4.7 Cell cycle and mitotic genes are highly expressed in atypical carcinoids	
3.5 Discussion and Conclusions.....	72
<b>Chapter 4: Genomic analyses identify two distinct cell-of-origin molecular subtypes of small intestine neuroendocrine tumors .....</b>	<b>76</b>
4.1 Abstract.....	77
4.2 Introduction.....	77
4.3 Materials and Methods.....	79
4.4 Results.....	81
4.4.1 Loss of Chromosome 18 is frequent genetic alteration in SI-NETs	
4.4.2 Gene expressions reveal two distinct subtypes of SI-NETs	
4.4.3 Subtypes of SI-NETs reveal distinct cell-of-origin signature	
4.4.4 <i>LMX1A</i> and <i>ONECUT2</i> are potential biomarkers to distinguish SI-NET subtypes	
4.5 Discussion and Conclusions.....	85
<b>Chapter 5: Conclusion and Future Directions</b>	
5.1 Thesis Conclusions.....	86
5.2 Work in Progress and Future Directions.....	88
5.2.1 Investigate Single Cell RNA sequencing data of <i>MEN1</i> knockout PanNETs	
5.2.2 Continued Characterization of <i>MEN1</i> mutants PanNETs and identifications of genome wide binding sites of <i>MEN1</i>	
5.3 Concluding Remarks.....	90
<b>References (for All Chapters).....</b>	<b>91</b>

## List of Tables

Table 1-1. Different NE cells across GI-tract and Pancreas with hormonal spectrum.....	2
Table 1-2. WHO Classification of NETs based on Ki67 index and grading system.....	7
Table 1-3. Genomic dataset of LCs, PanNETs and SI-NETs.....	23
Table 3-1b. Clinical features of LC subtypes with median age of diagnosis.....	71
Table 4-1. Chromosome 18 status for SI-NET subtypes.....	81

## List of Figures

### Chapter 1: Introduction

Figure 1-1. The regulatory system shows morphological and functional similarities between neurons and endocrine cells.....	1
Figure 1-2. Enteroendocrine cells of small intestine from single cell RNAseq.....	3
Figure 1-3. NETs classification based on embryological site of origin (foregut, midgut and hindgut).....	6
Figure 1-4. Incidence and prevalence of NETs from SEER (v17) dataset.....	9
Figure 1-5 Treatment for NETs includes surgery, medical and radiological therapies....	11
Figure 1-6. Somatic mutation burden for major human cancer types from TCGA.....	14
Figure 1-7. Mutation profiles of PanNETs.....	15
Figure 1-8. Classification of Lung NETs using morphological and proliferation criteria.	18
Figure 1-9. Genomic landscapes of SI-NETs from primary and metastatic tumors.....	21

### Chapter 2: Pancreatic NETs (published in Nature Communications, see page 25)

Figure 2-1. Mutational landscape of ATRX, DAXX, and MEN1 in PanNETs.....	37
Figure 2-2. A-D-M mutant and WT PanNETs as two distinct gene expression groups....	39
Figure 2-3. Distinct DNA methylation pattern for PanNETs subtypes.....	41

Figure 2-4. A-D-M mutant PanNETs with alpha-cell signature.....	43
Figure 2-5. Validation of A-D-M mutant PanNET and alpha cell signatures.....	45
Figure 2-6. HNF1A pathway with transcriptionally up-regulation in A-D-M mutant panNETs and alpha cells.....	46
Figure 2-7. PDX1 has promoter hypermethylated and lower gene expression in A-D-M mutant panNETs.....	49
 <b>Chapter 3: Lung Carcinoids</b>	
Figure 3-1. Three novel molecular subtype of lung carcinoids with mutational, gene expressions and DNA CpG methylation with distinct clinical features.....	62
Figure 3-2. Heatmap of differentially expressed genes between LC subtypes.....	64
Figure 3-3. Subtype specific molecular characterization of gene expression and DNA methylation profiles.....	66
Figure 3-4. Validation of novel classification of LC using independent collection of LCs from Fernandez-Cuesta et al., 2014.....	68
Figure 3-5. Gene expression and IHC for ASCL1 and S100 biomarker genes.....	69
Figure 3-6. Heatmap of differentially expressed genes between ACs and TCs within LC1 subtypes.....	72

Figure 3-7. Subtype-specific molecular events for LCs based on mutation, gene expression, DNA methylation and clinical informations.....73

**Chapter 4: Small Intestine NETs**

Figure 4-1. Molecular subtypes of SI-NETs using gene expression profile.....82

Figure 4-2. Cell type specific marker for SI-NETs subtypes.....83

Figure 4-3. Gene expression boxplot of LMX1A and ONECUT2 biomarkers.....84

**Chapter 5: Future Directions**

Figure 5-1. Schematic representation of *MEN1* knockout mice model for single cell RNA sequencing from normal islets, hyperplasia and PanNETs.....89

## List of Publication

- 1: Chan CS\*, **Laddha SV\***, Lewis PW, Koletsky MS, Robzyk K, Da Silva E, Torres PJ, Untch BR, Li J, Bose P, Chan TA, Klimstra DS, Allis CD, Tang LH. ATRX, DAXX or MEN1 mutant pancreatic neuroendocrine tumors are a distinct alpha-cell signature subgroup. *Nat Commun.* 2018 Oct 12;9(1):4158. doi: 10.1038/s41467-018-06498-2. PubMed PMID: 30315258.
- 2: Bartucci M, Hussein MS, Huselid E, Flaherty K, Patrizii M, **Laddha SV**, Kui C, Bigos RA, Gilleran JA, El Ansary MMS, Awad MAM, Kimball SD, Augeri DJ, Sabaawy HE. Synthesis and Characterization of Novel BMI1 Inhibitors Targeting Cellular Self-Renewal in Hepatocellular Carcinoma. *Target Oncol.* 2017 Aug;12(4):449-462. PMID: 28589491.
- 3: Paul D, Sinha AN, Ray A, Lal M, Nayak S, Sharma A, Mehani B, Mukherjee D, **Laddha SV**, Suri A, Sarkar C, Mukhopadhyay A. A-to-I editing in human miRNAs is enriched in seed sequence, influenced by sequence contexts and significantly hypoedited in glioblastoma multiforme. *Sci Rep.* 2017 May 26;7(1):2466. doi:10.1038/s41598-017-02397-6. PubMed PMID: 28550310
- 4: Guo JY\*, Teng X\*, **Laddha SV**, Ma S, Van Nostrand SC, Yang Y, Khor S, Chan CS, Rabinowitz JD, White E. Autophagy provides metabolic substrates to maintain energy charge and nucleotide pools in Ras-driven lung cancer cells. *Genes Dev.* 2016 Aug 1;30(15):1704-17. doi: 10.1101/gad.283416.116. Epub 2016 Aug 11. PubMed PMID: 27516533.
- 5: Zhang C, Liu J, Tan C, Yue X, Zhao Y, Peng J, Wang X, **Laddha SV**, Chan CS, Zheng S, Hu W, Feng Z. microRNA-1827 represses MDM2 to positively regulate tumor suppressor p53 and suppress tumorigenesis. *Oncotarget.* 2016 Feb 23;7(8):8783-96. PMID: 26840028.
- 6: Puzio-Kuter AM, **Laddha SV**, Castillo-Martin M, Sun Y, Cordon-Cardo C, Chan CS, Levine AJ. Involvement of tumor suppressors PTEN and p53 in the formation of multiple subtypes of liposarcoma. *Cell Death Differ.* 2015 Nov;22(11):1785-91. Epub 2015 Mar 27. PMID: 25822339
- 7: Wong C, **Laddha SV**, Tang L, Vosburgh E, Levine AJ, Normant E, Sandy P, Harris CR, Chan CS, Xu EY. The bromodomain and extra-terminal inhibitor CPI203 enhances the antiproliferative effects of rapamycin on human neuroendocrine tumors. *Cell Death Dis.* 2014 Oct 9;5:e1450. doi: 10.1038/cddis.2014.396. PubMed PMID: 25299775.
- 8: Zhang C, Liu J, Wang X, Wu R, Lin M, **Laddha SV**, Yang Q, Chan CS, Feng Z. MicroRNA-339-5p inhibits colorectal tumorigenesis through regulation of the MDM2/p53 signaling. *Oncotarget.* 2014 Oct 15;5(19):9106-17. PubMed PMID: 25193859.

Manuscript Submitted or under preparation

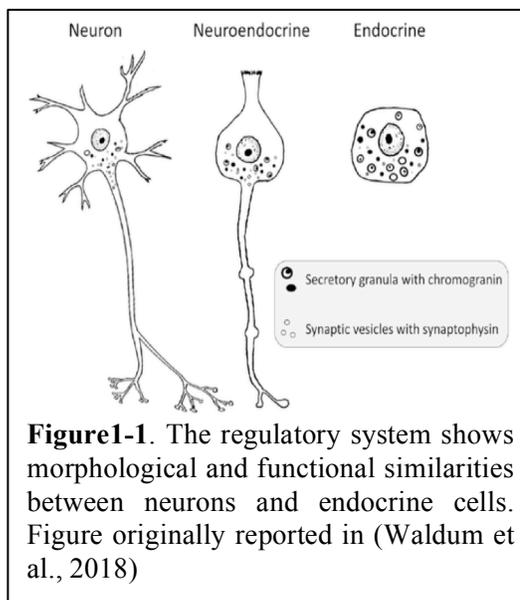
**1. Integrative Genomic Characterization Identifies Molecular Subtypes of Lung Carcinoids.**  
**Saurabh V. Laddha**, Edaise Da Silva, Brian R. Untch, Kenneth Robzyk, Hua Ke, Natasha Rekhtman, John T. Poirier, William D. Travis, Laura H. Tang, Chang S. Chan  
*Submitted to Journal of National Cancer Institute*

**2. Genomic analyses identify two distinct cell-of-origin molecular subtypes of small intestine neuroendocrine tumors**  
**Saurabh V. Laddha**, Kenneth Robzyk, Laura H. Tang, Chang S. Chan\*

## Chapter 1: Introduction and Background

### 1.1 Neuroendocrine Cells

Neuroendocrine (NE) cells are unique and exhibit morphological and physiological characteristics of both “neuro” and “endocrine” cells. This characteristic of “neuro” is mainly based on dense core granules (store monoamines) and “endocrine” for synthesis and secretion (Waldum et al., 2018)(Figure 1-1). NE cells are specialized cells that react to chemical signals (neuronal input) by secreting hormones into the blood (like insulin production by beta cells of pancreas)(table 1-1 for list of NE cells with specific secreted hormone) and control body homeostasis. NE



cells are widely distributed across human body and clustered mainly in endocrine glands. Endocrine glands are aggregated and majorly made up of NE cells (such as pituitary, parathyroid etc) as well as endocrine islet tissue in pancreas and thyroid. NE cells, which are scattered throughout the length of an organs are often called “diffuse neuroendocrine systems” (DNES). These systems are suggested to have differences in terms of embryological origin, DNES are derived from endodermal tissues and endocrine glands are derived from ectodermal tissue(Schimmack, Svejda, Lawrence, Kidd, & Modlin, 2011).

NE cells are majorly found throughout the length of gastrointestinal (GI) tract, pancreas (Islet cells) and thyroid. Within GI tract, 15 different NE cells have been found

secreting wide range of neuropeptides and amines. Small intestine NE cells (majorly enterochromaffin cells) produce serotonin, chromogranin A (CHGA), tachykinins etc.

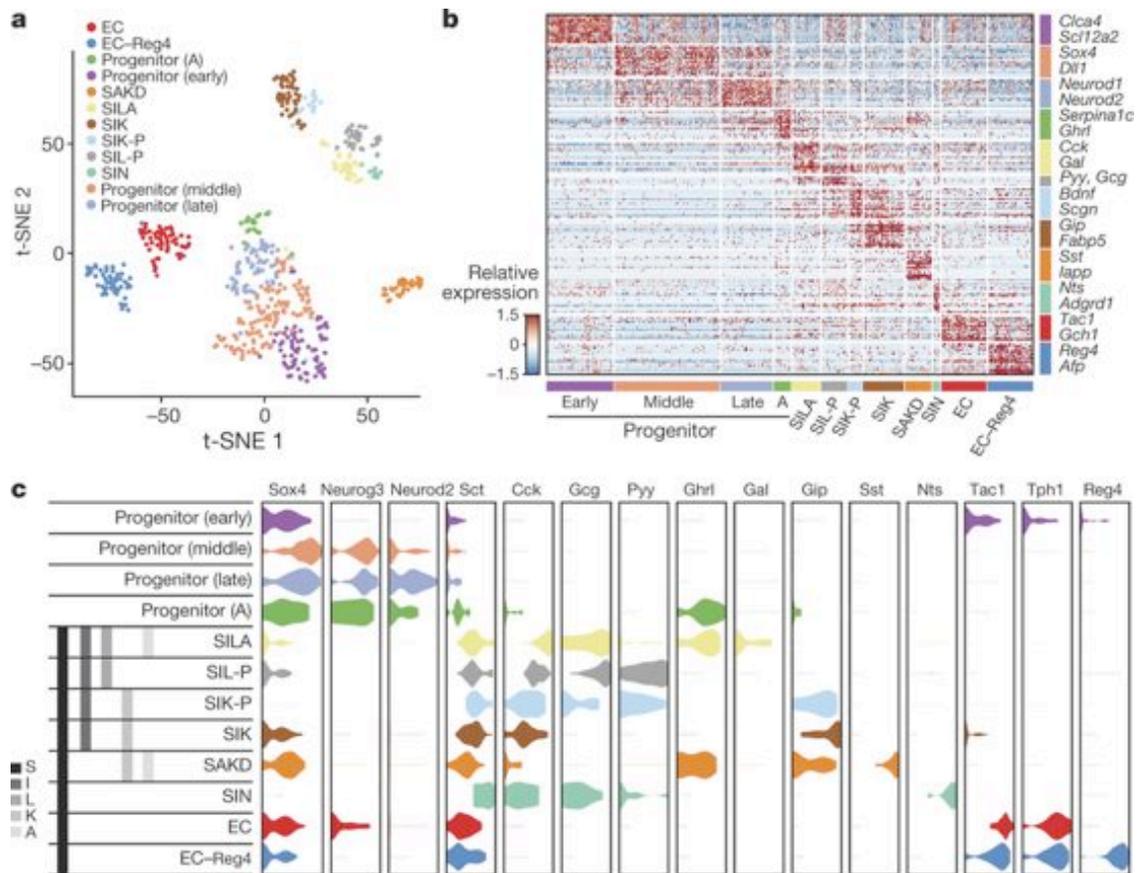
Cell Type	Hormone Produced	Distribution
$\alpha$	Glucagon	Pancreas
$\beta$	Insulin	Pancreas
$\delta$ (D)	Somatostatin	Stomach, small intestine, pancreas, appendix, colorectum
EC	Serotonin	Stomach, small intestine, pancreas, appendix, colorectum
ECL	Histamine	Stomach
G	Gastrin	Stomach, duodenum
I	Cholecystokinin	Small intestine
K	Gastric inhibitory peptide	Small intestine
L	Glucagon-like peptide 1, peptide YY	Rectum, small intestine
M	Motilin	Small intestine
N	Neurotensin	Small intestine
P/D1	Ghrelin	Stomach, small intestine, appendix, colon
PP	Pancreatic polypeptide	Pancreas
S	Secretin	Small intestine, pancreas
VIP	Vasoactive intestinal peptide	Pancreas, stomach, small intestine, appendix, colorectum

Abbreviations: EC, enterochromaffin; ECL, enterochromaffin-like.

**Table 1-1.** Different NE cells across GI-tract and Pancreas with hormonal spectrum. Table originally reported in (J. Y. Kim & Hong, 2016), shows cell-type-specific hormonal secretion across all organs.

Pancreatic islet cells represent ~1% to 2% volume of pancreases and have 5 major known NE cell types: Alpha (produces Glucagon, ~20%), Beta (produces Insulin, ~70%), Delta (produces somatostatin, < 10%), Gamma (produces pancreatic polypeptide, < 5%) and Epsilon (produces Ghrelin, <1%). A pancreatic islet regulates blood glucose level by tracking the existing glucose level in blood. In lung, NE cells exist as clustered (NeuroEpithelial Bodies, NEB) and single cell (Kolschitzky cell) and function as chemoreceptor, immune response and in lung development and maturation(Linnoila, 2006). Enteroendocrine cells (EEC) of small intestine secrets diverse hormones that function as metabolic signal transductions components. EECs are prime sensors of nutrients and microbial metabolites(Haber et al., 2017). Enterochromaffin cells (EC) or Kulchitsky cells, a type of EEC played an important function for gut motility and secretory reflexes. Recent single cell(Haber et al., 2017) RNAseq study of mice small

intestine reveal atleast 12 distinct neuroendocrine cells (Figure1-2) secreting variety of hormones such as serotonin, ghrelin, secretin, proglucagon, somatostatin, neurotensin etc.



**Figure 1-2.** Enteroendocrine cells of small intestine from single cell RNAseq. a) Unsupervised clustering (t-SNE) of 533 EEC cells into 12 clusters b) Subtype signature of EEC across rows and cells across column c) EEC classification based on cell-type-specific hormones expression. Figure originally reported in and adapted from (Haber et al., 2017).

Moreover, four of the 12 cells have expression of enteroendocrine precursor markers (*SOX4*, *NEUROG3* or *NEUROD1*) and other eight represent mature enteroendocrine cells. Interestingly, crossover between these cell types has been established (Gribble & Reimann, 2016). Enteroendocrine cells of small intestine majorly made of enterochromaffin (EC) cells and subdivided into EC and EC-Reg4 based on expression of

*REG4* gene(Haber et al., 2017) (Figure 1-2). Based on *TPHI* and *REG4* gene expression profile, these clusters/cells may mainly be divided into two sub-classes: *TPHI*<sup>+</sup>/*REG4*<sup>+</sup> neuroendocrine EC cells and *TPHI*<sup>-</sup>/*REG4*<sup>-</sup> neuroendocrine cells. The pathways and molecular process regulating EC cell function and growth remains unknown.

NE cells across different organ and within organ have different morphology, suggesting diverse functions among these organs. Characterization of NE progenitor cells in different organ is unmet. Recent single cell genomic applications have leveraged identification of different NE cell populations across different organ (like in pancreases(Muraro et al., 2016), small intestine(Haber et al., 2017) etc.) elucidating transcription profile, cellular heterogeneity and NE cell-type-specific transcription factors and pathways.

## **1.2 Neuroendocrine Tumors**

Tumor originating from hormone producing neuroendocrine cell is defined as Neuroendocrine Tumor (NET) or Neuroendocrine Cancer (NEC). Neuroendocrine cancers or carcinomas are poorly differentiated, aggressive and metastasizes more commonly. NETs are well differentiated and benign in nature (sometime referred as “Carcinoids”). However, these benign tumors can transform to malignant disease and results in adverse clinical outcome(Scherubl et al., 2013). NETs are clinically less challenging as compared to poorly differentiated NECs. NETs may produce higher amount of hormones with respect to its normal physiological amounts, which ultimately causes many symptoms and results in misdiagnosis(Mafficini & Scarpa, 2018). On the other side, most GI-NETs may present as asymptomatic for a long duration or present

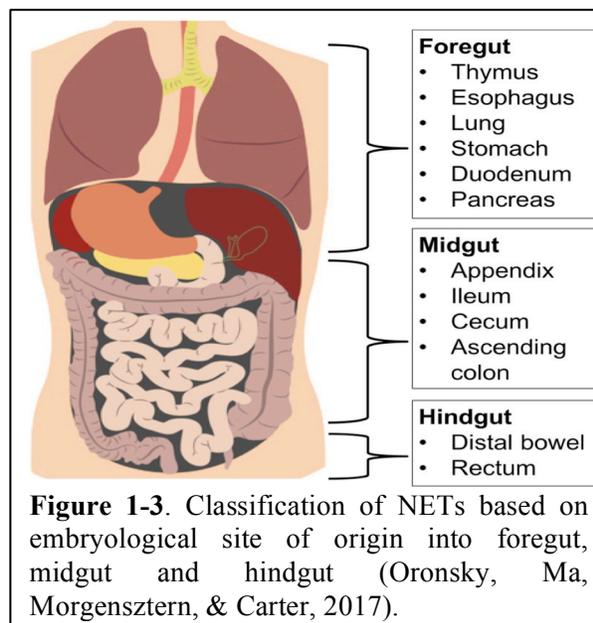
non-specific symptom that results in large fraction of patients with delayed diagnosis and present as advanced diseases or metastasis with mortality rate of 50% (Mafficini & Scarpa, 2018; Scherubl et al., 2013). NETs are also observed throughout human body but majorly found at GI track, pancreas and lung. Because of lack of heterogeneous NET diseases models, most of the NETs research investigated on tumor specimens(Capdevila et al., 2017).

NETs are rare, slow growing, biologically poorly understood and clinically challenging type of primary neoplasms and represent ~2% of all malignancies. NETs are considered as orphan diseases with prevalence of <200,000 in the US(Basu, Sirohi, & Corrie, 2010). NETs are classically defined as “carcinoids” or “carcinoid tumors” and observed body-wide distribution, but these terms do not accurately explain their biology, histopathological differences, or secretory capabilities. The discovery of GI-NETs started in 1870 when Rudolph Heidenhain discovered the NE cells. Later in 1907, pathologist Siegfried Oberndorfer was the first to describe little lesions of small intestine carcinomas and called them as “Carcinoids”. He defined carcinoids as slow growing, less aggressive, benign in nature than adenocarcinoma. In year of 1929, several cases have been observed with metastasis carcinoids(Modlin, Shapiro, & Kidd, 2004) and later it was concluded that NE carcinoids have a malignant potential. Recent efforts have identified genomic and epigenomics alterations in development and progression of NETs but the underlying biological knowledge is limited as compared to its respective tissue cancer.

### 1.3 Neuroendocrine tumor Classifications

Histopathological classification of NETs has proven challenging and often shown to be error-prone in clinical practice (Basu et

al., 2010). Like other cancers, NETs are mainly classified according to the site of origin and whether the tumor over-secreting hormones (functional) or not (non-functional). The majority of NETs are observed in gastrointestinal tract and in lungs (Vinik et al., 2010). In 1963, E.D Williams and M Sandler (Williams &



Sandler, 1963) suggested a classification for GI-NETs based on their embryological gut origin and divided into three main class : foregut, midgut and hindgut (Figure 1-3). Within foregut, pancreatic NETs can be classified as functional (at least five main types) and non-functional based on physiologically hyper-secretion of hormone in blood (Williams & Sandler, 1963).

### 1.4 World Health organization Classifications and Grading

NETs are heterogeneous neoplasms with varied clinical outcome. The classification terminology of NETs (specifically for GI-NETs) has evolved over the last two decades and revisited many times. NE cell neoplasm broadly classified into two fundamental group based on clinical features, histology, proliferation index (mitotic rate): a) Well-

differentiated, low-proliferating NETs or carcinoids and b) poorly differentiated, high proliferative index (like in LCNECs)(Kloppel, 2017). Well-differentiated NETs have better clinical outcome as compared to poorly differentiated NEC (Kloppel, 2017).

Recent classification of NETs is based on histological grade and proliferative index

(Ki67) (Kloppel, 2017). For GI-NETs

using Ki67 index, NETs are graded as

G1 (Ki67 < 3%), G2 (Ki67 3-20%)

and NE carcinomas (NECs) as G3

(Ki67 > 20%)(Kloppel, 2017).

Because of the overlapping Ki67 index

for NETs and NECs in pancreases,

WHO 2017 updated the classification

of Pancreatic NETs with recognition of

heterogeneity of grade three (G3) NETs and included both well-differentiated and poorly

differentiated NETs in G3 category. Expression of *TP53* and *RBI* genes are key

distinguishable markers for G3 NETs from G3 NECs(Kloppel, 2017). Mixed

neuroendocrine-non-neuroendocrine neoplasms (MiNEN) are recently recognized as

heterogeneous group of rare pancreatic NETs that represents a one third of poorly

differentiated NECs in combination to its neuroendocrine components(de Mestier et al.,

2017; Kloppel, 2017). NETs are also classified based on TNM (Tumor-Node-Metastasis)

stage, which mainly includes invasion and metastasis spread. NET classifications are

clinically relevant and significantly impact treatment decisions and prognosis. With

recent molecular and genomic profiling success, systematic characterization and novel

	Ki67, %	Mitotic index, mitoses per 10 high-power fields
Grade 1	<3	<2
Grade 2	3–20	2–20
Grade 3	>20	>20

	Grade	Differentiation
G1 neuroendocrine tumour	Grade 1	Well differentiated
G2 neuroendocrine tumour	Grade 2	Well differentiated
G3 neuroendocrine tumour	Grade 3	Well differentiated
G3 neuroendocrine carcinoma	Grade 3	Poorly differentiated
MiNEN	All grades	Association of a neuroendocrine and a non-neuroendocrine component

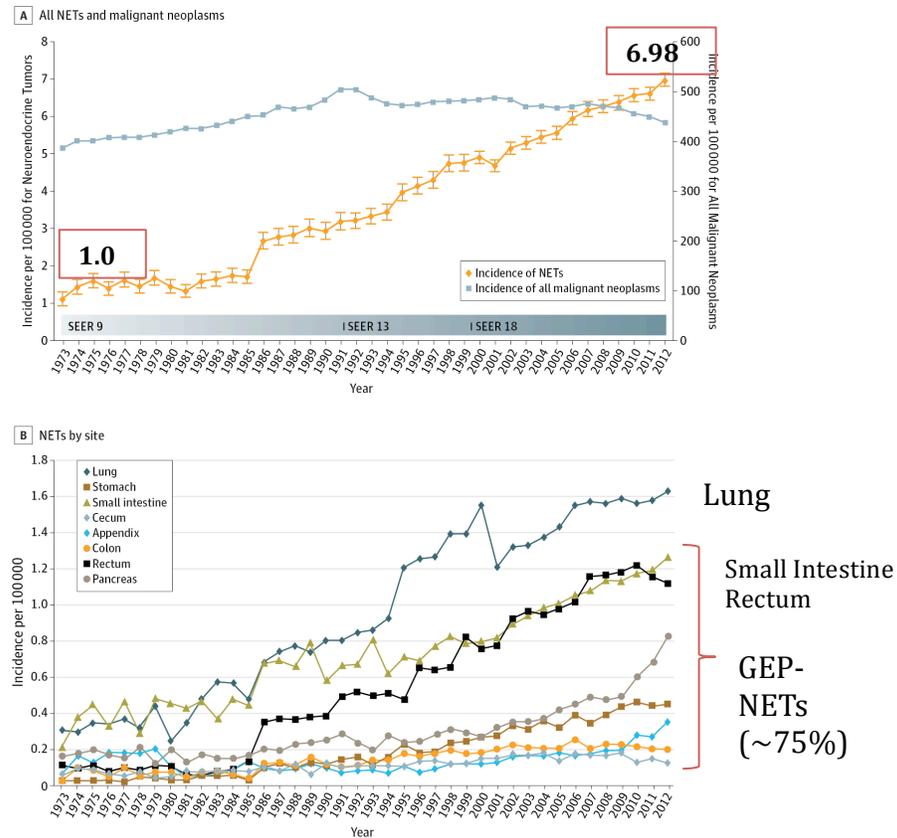
**Table 1-2.** WHO Classification of NETs based on Ki67 index and grading system. Table originally reported in (de Mestier et al., 2017)

molecular markers discovery will improve the classification in combination with histology.

### **1.5 Epidemiology of Neuroendocrine Tumors**

NETs are a rare and slow-growing type of tumor with vague symptoms and are easily misdiagnosed as other condition (such as diarrhea, respiratory symptoms etc.). This misdiagnosis is mainly because of over-secretion of hormone, which leads to other conditions like excess serotonin, which further causes mild (diarrhea) to server symptoms (fever, seizures). Some of the common symptoms of NETs are flushing, diarrhea, high blood pressure, fatigue, abdominal pains, wheezing, coughing, skin lesions etc.

NETs represent 0.5% from all newly diagnosed neoplasms(Taal & Visser, 2004). The diagnosis incidence of NETs is steadily increasing over last 30 years, probably due to new diagnostic techniques and increased awareness(Gustafsson et al., 2008). Each year, approximately 12000 people in the US are diagnosed with a NET and the 5-year survival rate for these tumors varies and depends on several factors, primarily the location of tumor, grade, stage, age and tumorigenesis factor.



**Figure 1-4** NETs Incidence and prevalence from SEER (v17) dataset. a) Annual incidence rate (per 100,000) of NETs (age adjusted) from 1973 to 2012 year, orange line represent NETs and blue line for all other cancers b) NETs by site. Figure adapted from (Dasari et al., 2017).

Recent age, sex and race adjusted data analysis from Surveillance, Epidemiology and End Results (SEER) program, for NETs (from 1973 to 2012 year) revealed increased incidence of NETs by the factor of 6.4 fold (from 1.09 to 6.98 per 100,000 cases)(Dasari et al., 2017)(figure 1-4a) and eight fold more NETs were seen in those with age 65 or more. Among NETs, Lung subtypes (1.49 per 100,000 in the lung) were highest incident rate (figure 1-4b) followed by GI/Pancreatic NETs (3.56 per 100,000 in GI sites) and 0.84 per 100,000 NETs with unknown primary. This increased incidence appeared at all tissue sites, stage and grade with varied overall survival (median 5 year). Moreover, no

difference has been observed for sex (male vs female) from SEER dataset for incidence of NETs.

### **1.6 Biomarkers, Diagnosis and Treatment for Neuroendocrine Tumors**

NETs are originating from hormone producing NE cells and may lead to hyper-secretion of neuropeptides, which often results into hormonal syndromes. Chromogranin A (CgA) is a membrane protein of neurosecretory granules, widely expressed in almost all NE cells and over-expressed in NETs, which is one of the prime clinical biomarkers to identify NETs (in serum as well as immunohistochemistry (IHC) tumor markers) (Wiedenmann & Huttner, 1989), but not useful to stratify subtypes of NETs. However, increased levels of CgA have been associated with shorter survival and may correlate with tumor progression (Arnold et al., 2008). Neuron-Specific Enolase (NSE) is also expressed in NETs but is often elevated only in high grade or poorly differentiated tumors and is less commonly used as compared to CgA marker in NETs identification (Vinik et al., 2009). The diagnosis of NETs is not commonly straightforward and is not easily considered for differential diagnosis. Several different imaging and IHC markers have been utilized in clinics (such as computer tomography (CT), magnetic resonance imaging (MRI), 5-hydroxytryptophan, somatostatin receptor imaging, serotonin, synaptophysin, NSE, gonadotropin hormone, peptide YY etc.) for NET diagnosis. However, CgA marker is specific, sensitive and widely used for NETs but not perfect (Vinik et al., 2009) and new biomarkers for NETs are needed to provide better diagnostics and prognostic information, specifically for subtypes of NETs, unknown primary NETs, well-differentiated NETs (Vinik et al., 2009). Treatment decision for NETs depends on

multiple factors, such as: primary site, metastasis spread, grade, functional NETs vs non-functional NETs, hereditary NETs, other health conditions, NETs subtype and mutational background. Current treatment strategies include, surgery, medical and radiological therapies (Figure 1-5).

Surgery, one of the first lines of treatment for most of the well differentiated NETs(Welin

et al., 2009). Plasma-CgA level

measurement after surgery is

strongly recommended as follow-up

twice a year and ultrasonography

once a year(Welin et al., 2009) for

asymptomatic NETs. For localized

NETs with stage 1 and 2, radical

surgery has been recommended(Jann

et al., 2011). PanNETs with MEN1

mutations, surgery is frequently not

as successful as cases without *MEN1*

mutations(Frost et al., 2018) mainly

because, *MEN1* mutated panNETs

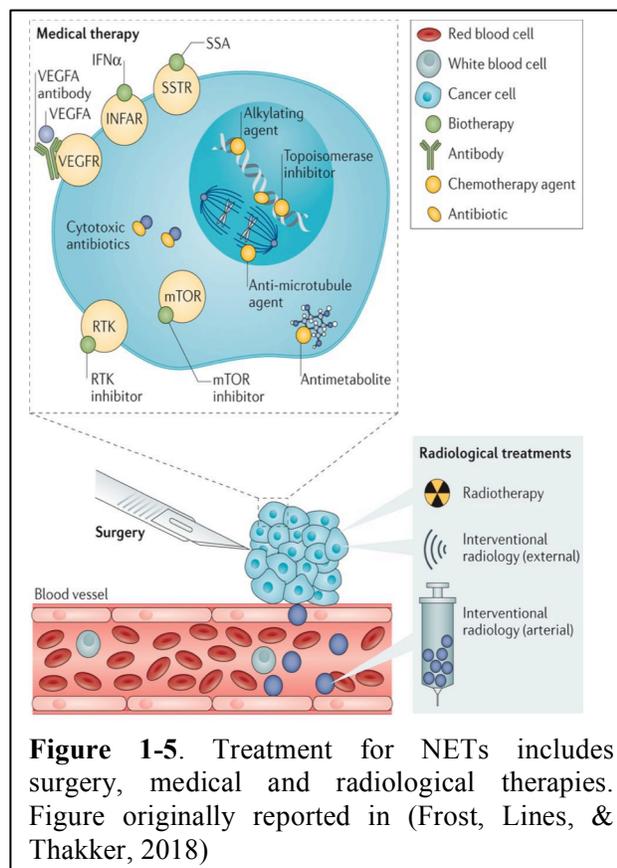
are usually has multiple tumors and metastatic disease might be present(Anlauf et al.,

2008). Medical therapies(Frost et al., 2018) mainly include biotherapies and

chemotherapy. Biotherapies for NETs can be hormonal (somatostatin), peptide hormone,

which inhibit other hormones, targets for cell proliferations and angiogenesis(Frost et al.,

2018). Such few examples of current biotherapies(Frost et al., 2018) are: Somatostatin



**Figure 1-5.** Treatment for NETs includes surgery, medical and radiological therapies. Figure originally reported in (Frost, Lines, & Thakker, 2018)

analogues, Interferon alpha, mTOR inhibitor (everolimus) etc. Chemotherapy for NETs includes (Frost et al., 2018) alkylating agents (like cisplatin), anti-microtubule agents, topoisomerase inhibitors, cytotoxic antibiotics and antimetabolite (Frost et al., 2018). Radiological treatment mainly includes radiotherapy and interventional radiology.

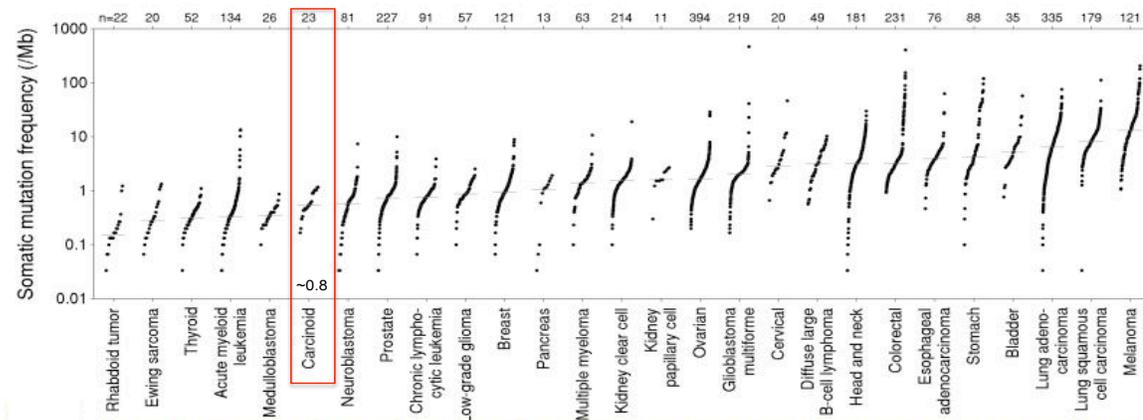
### **1.7 Genomics of NETs**

Recent genomic studies have identified genetic and epigenetic alterations in cancer at a massively unbiased way. Sequencing (DNA, RNA, epigenomic etc.) technological advancement played an important role in finding different hallmark of cancers and its correlation with survival outcome. This emergence of “high-throughput sequencing (HTS)” data fosters an integrative genotype to phenotype association approach with clinical information. This integration approach has led to alter the practice of medicine in clinics to stratify cancers based on molecular aberrations for diagnosis, prognosis, and treatment (Bailey et al., 2016; Cancer Genome Atlas, 2012). According to recent national cancer comprehensive network guidelines, molecular profiling for cancer specific mutational event is recommended to inform decision for mutation-drug targeted therapy (Pal et al., 2016).

The molecular profiling of NETs is recently being appreciated and identified recurrent molecular alterations in the latest years. The genetic landscape of NETs has shed a light on tumorigenesis, oncogenic pathways and prognosis. The majority of NETs are sporadic, however small subset of NETs are heritable groups of neoplasms involving mutations in genes like *MEN1*, *VHL*, *NF1*, *CDKN1B*, *TSC1* and *TSC2*. Recently, many

studies have identified underlying molecular abnormalities for sporadic or evidently sporadic NETs.

The mutational landscape of well differentiated and poorly differentiated NETs revealed distinct altered pathway genes respectively. Poorly differentiated NETs of pancreas and lung are enriched and have recurrent mutations for *TP53* and *RBI* pathways(Simbolo et al., 2017). Well-differentiated NETs of pancreas have inactivation mutations in chromatin remodeling complex genes such as *ATRX*, *DAXX* and *MEN1*(Jiao et al., 2011). It is remarkable that half of well-differentiated NETs have epigenetic related alterations that are not commonly seen in poorly differentiated NETs(Simbolo et al., 2017). These distinct genetically altered pathways highlight different pathogenesis mechanism for NETs and have shown clinical utility for diagnosis and prognosis. Moreover, these altered epigenetic related genes have shown to be sufficient enough to drive tumorigenesis(Fernandez-Cuesta et al., 2014) for NETs but the underlying mechanism is unknown. Well-differentiated NETs have low somatic mutation burden (Fernandez-Cuesta et al., 2014; Lawrence et al., 2013) and mostly diploid genome(Francis et al., 2013).



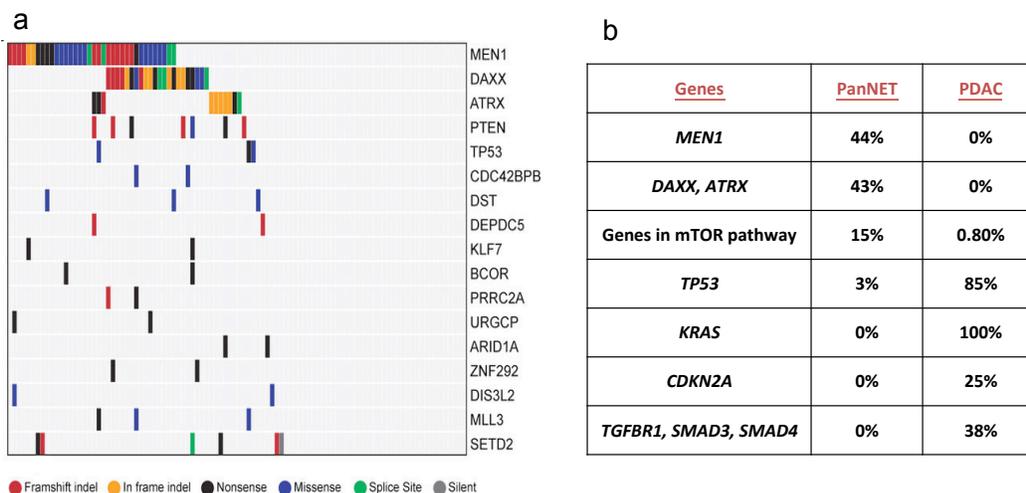
**Figure 1-6.** Somatic mutation burden for major human cancer types from TCGA. Carcinoid tumors are highlighted in red box with  $\sim 0.8$  somatic mutation frequency per MB. Figure adapted from (Lawrence et al., 2013).

### 1.7.1 Pancreatic neuroendocrine tumors (PanNETs)

Pancreas tissue is made of exocrine and endocrine cells. Exocrine (acinar, ductal etc.) cells secrete digestive enzymes and endocrine islet cells (at least five major cell types) secrete hormones (mainly to regulate blood glucose level). Pancreatic islet cells represent  $\sim 1\%$  to  $2\%$  volume of pancreases and have 5 major known NE cell types: alpha (produces Glucagon,  $\sim 20\%$ ), beta (produces Insulin,  $\sim 70\%$ ), delta (produces somatostatin,  $< 10\%$ ), gamma (produces pancreatic polypeptide,  $< 5\%$ ) and epsilon (produces Ghrelin,  $< 1\%$ ) (Da Silva Xavier, 2018). Pancreas has many islets distributed across the tissue (A. Kim et al., 2009). It is known in mouse and human that each islet has differential enrichment of NE cell populations (A. Kim et al., 2009) and number varies between two islets from the same individual (Muraro et al., 2016).

PanNETs introduction is presented in chapter 2 introductions and from our paper (Chan et al., 2018) and here are some details of recent molecular studies of PanNETs.

Molecular studies have identified mutations in *MEN1*, *ATRX*, and *DAXX* to be the most commonly found in PanNETs(Chan et al., 2018; Jiao et al., 2011; Scarpa et al., 2017)(found in approximately 40, 10, and 20% of tumors, respectively). All the three genes play a role in chromatin remodeling and regulate gene expression.



**Figure 1-7.** Mutation profiles of PanNETs, a) recurrently mutated genes in panNETs (Scarpa et al., 2017) b) table represent mutated genes percentage in PanNETs and its counterpart exocrine (PDAC) tumors (Jiao et al., 2011).

Additional mutations in mTOR pathway genes including *TSC2*, *PTEN*, and *PIK3CA* are found in one out of six well-differentiated PanNETs (Jiao et al., 2011). Other reported rare mutations in PanNETs include DNA damage repair (17%) genes (*MUTYH*, *CHEK2*, and *BRCA2*) and chromatin remodeling gene *SETD2* (Scarpa et al., 2017). However, nearly one-third PanNETs do not have mutations in chromatin related genes and also other protein-coding region of the genome.

**The *MEN1* Gene:** An inactivation mutation in tumor suppressor *MEN1* gene leads to multiple endocrine neoplasia type 1 (MEN1), an autosomal dominant disorder(Busygina & Bale, 2006). Menin, a protein product of *MEN1* gene is a histone

methyltransferase and essential component of MLL/SET1 complex(Hughes et al., 2004). *MEN1* is an activator as well as repressor of the gene expression by epigenetic mechanisms and its function depends on the interacting partner of *MEN1*. For example, MEN1 interact with MLL complex leading to methylation of histone H3 (H3K4) and activate gene expression. Several studies have shown that loss of MEN1 does not alter global H3K4me3 level but are responsible for altering H3K4me3 at specific loci(Lin et al., 2015). They found *MEN1* dependent H3K4me3 sites are altered during early stages of PanNETs formation and genes, which are significantly downregulated under *MEN1* loss. However, the human MEN1 studies on proper cell types are missing and need extensive characterization. In majority of *MEN1* PanNETs (>90%), loss of heterozygosity (LOH) is common somatic alterations leading to complete loss of function(Valdes et al., 1998).

***ATRX - DAXX complex:*** This complex is recurrently mutated in PanNETs (44%) and suggested as putative tumor suppressor gene. One of the two subunit of this complex is mutated and exclusive to each other in panNETs cases but can coexist with *MEN1* mutations. *ATRX* (Alpha Thalassemia X-linked intellectual disability) and *DAXX* (Death domain associated protein) form hetero-dimer and is a part of SWI/SNF complex(Lewis, Elsaesser, Noh, Stadler, & Allis, 2010). This complex is known to deposit histone variant H3.3 to telomeric and centromeric region(Lewis et al., 2010) of the genome, maintaining telomeric repeat and tandem repeat elements, DNA methylation and repression of repetitive elements. Intriguingly, inactivation mutations in *ATRX* and *DAXX* are strongly associated with positive altered telomeric length (ALT) in PanNETs(Heaphy et al., 2011).

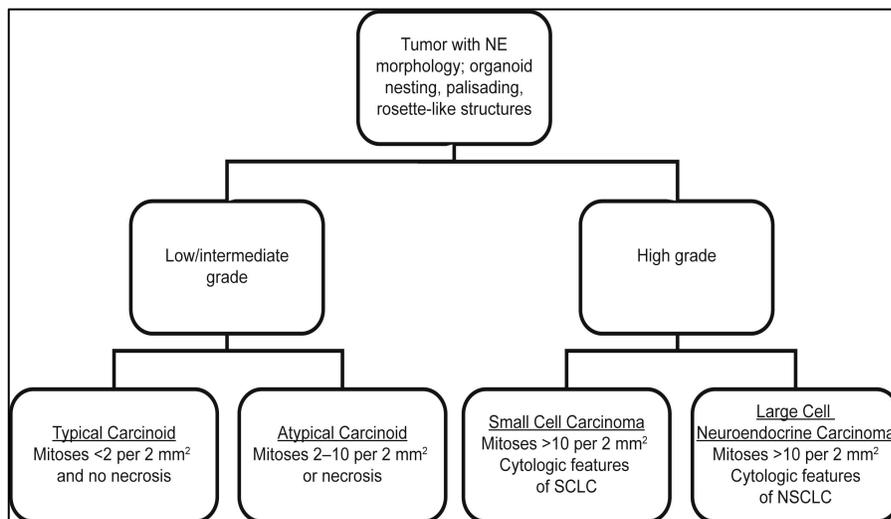
*MEN1* and *ATRX/DAXX* complex are involved in distinct epigenetic pathways and cross talk between them is not well established. Interestingly, little is known about their mutation-specific gene signatures, epigenetic changes, tumorigenesis pathways, oncogenic activation and clinical biomarkers. Also, it is not clear which specific cell type within the Islet of Langerhans is the cell of origin for non-functional well-differentiated NETs. Comprehensive integration of genomics, epigenomic and transcriptomic dataset will elucidate the molecular subtypes of NETs, cell-of-origin, mutations specific transcriptional profile, clinical biomarker for stratifications and choice of targeted therapeutics options.

### **1.7.2 Lung Carcinoids (LCs)**

Tumors originating from lung NE cells are defined as lung neuroendocrine tumors (lung NETs). According to WHO 2015, lung NETs represents ~ 20 to 25% of primary lung neoplasm and are classified as four histological subtypes(Travis et al., 2015) with considerably different biological characteristics: 1) Small Cell Carcinoma (SCLC) (~20% prevalence, high-grade), 2) Large Cell NE Carcinoma (LCNEC) (~3%, high-grade), 3) Atypical Carcinoids (AC) (~0.2%, intermediate grade), 4) Typical Carcinoids (TC) (~2%, low grade) (Figure 1-8). Carcinoids of lung include ACs and TCs and ACs are less frequent, aggressive with adverse clinical outcome.

The key features of this classification rely on morphology, mitotic index per 2mm<sup>2</sup>, and necrosis assessment(Hendifar, Marchevsky, & Tuli, 2017). The reproducibility of this classification and its prognostic efficacy was disputed with high inter-observer variability(Travis et al., 1998; van den Bent, 2010), especially for LCs (TCs Vs

ACs)(Swarts et al., 2014). It has been reported that TCs and ACs tumors are over-diagnosed as SCLC in small crush biopsy specimens(Pelosi, Rodriguez, Viale, & Rosai, 2005).



**Figure 1-8.**Classification of Lung NETs using morphological and proliferation criteria. Figure originally reported in and adapted from (Hendifar et al., 2017)

Recent application of Ki67 proliferation markers for LCs classification revealed promising results yet failed to accurately distinguish ACs from TCs and also between high grade NETs on a large dataset(Pelosi, Rindi, Travis, & Papotti, 2014). Strikingly over the last 30 years, the incidence rate of carcinoids tumors has increased(Gustafsson et al., 2008) in the US, but our molecular understanding of tumorigenesis is incomplete. *MEN1 syndrome* is more prevalent in bronchopulmonary carcinoids(Sachithanandan, Harle, & Burgess, 2005) as compared to high grade Lung NETs. Recent genomics study by Frenandez-Cuesta L et.al(Fernandez-Cuesta et al., 2014) on 74 pulmonary carcinoids (includes TCs and ACs) revealed somatic mutation rate of 0.4 per Mb, which is much lower than respective lung cancers (which is ~9 per Mb). They found genes with somatic mutations enriched for chromatin-remodeling genes with recurrent alteration in *MEN1*,

*PSIP1*, and *ARID1A*. No significant mutations and focal copy alterations were observed for the frequently mutated genes in high-grade lung cancers (such as *KRAS*, *TP53*, *EGFR* etc)(Vollbrecht et al., 2015). This unique mutational profile suggests different cellular and biological mechanisms for LCs pathogenesis than high-grade NETs (LCNEC and SCLC) and lung cancers. In the same study(Fernandez-Cuesta et al., 2014), they also compared gene expression profile of LCs with SCLC and found *RB1*, *EGFR*, *VEGF*, *mTOR*, *PRC2* etc. gene sets are overexpressed in SCLC, which are hallmarks for aggressive growth. It is not known if there are distinct molecular subtypes of LCs or what are their cells of origin. More accurate molecular diagnostic tools and stratification for lung carcinoids will help ensure more appropriate treatment and clinical management.

### **1.7.3 Small Intestinal Neuroendocrine Tumors**

Small intestinal neuroendocrine tumors (SI-NET) are rare, slow growing (produces serotonin), well differentiated, most common type of gastrointestinal endocrines tumors with one case per 100,000 annually, account for 25% of all NETs(Karpathakis et al., 2016). SI-NETs tumor (carcinoids) are small in size (frequently < 2cm) and slow growing with Ki67 index is frequently < 2% but the 5-year survival rate is 50-65%(Yao et al., 2008). Classical NETs of small intestine proposed to originates from enterochromaffin cells (EC) but the origin and development remain undetermined.

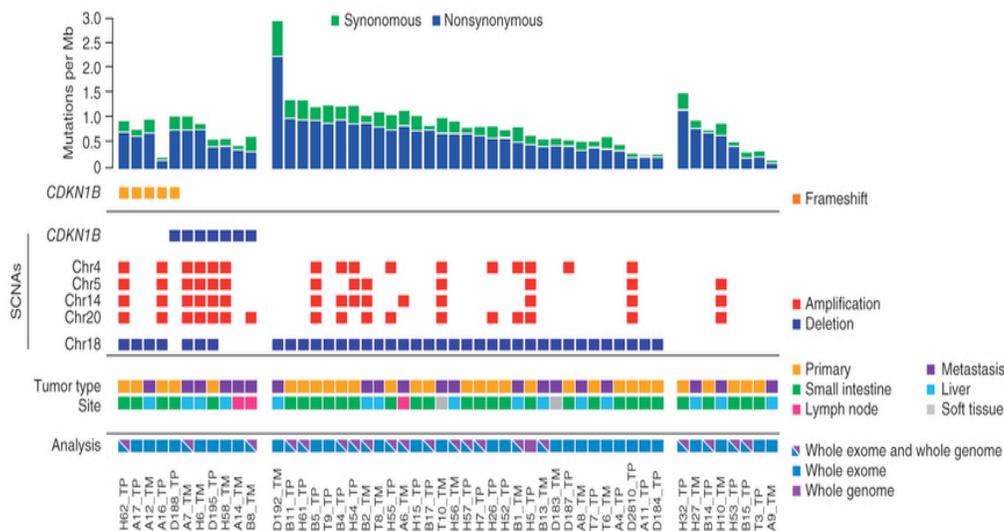
Recent single cell(Haber et al., 2017) RNAseq of mice's small intestine reveal at least 12 distinct neuroendocrine cell populations secreting a variety of hormones such as serotonin, ghrelin, secretin, proglucagon, somatostatin, neurotensin etc. Moreover, four of

the 12 cells have expression of enteroendocrine precursor markers (*SOX4*, *NEUROG3* or *NEUROD1*) and the other eight represent mature enteroendocrine cells (EEC). Interestingly, interaction and plasticity between these cell types have been established (Gribble & Reimann, 2016). The dominant population of EEC in small intestine is enterochromaffin (EC) cells and can be subdivided into EC and EC-Reg4 subgroups based on expression of *REG4* gene (Haber et al., 2017). Moreover, based on *TPHI* and *REG4* gene expression profile, these clusters/cells may be divided into two sub-classes: Major enterochromaffin cells which are *TPHI*<sup>+</sup>/*REG4*<sup>+</sup> and *TPHI*<sup>-</sup>/*REG4*<sup>-</sup> (rare) neuroendocrine cells. The pathways and molecular process regulating these EC cell function and growth remains unknown.

Molecular studies identified handful of recurrent alterations for SI-NETs and found absence of obvious pathogenic genomic alterations. SI-NETs are genetically poorly understood type of NETs. One copy loss of chromosome 18 (chr18) is the most frequent genetic alterations (60 – 80%) of SI-NETs (Figure 1-9) (Francis et al., 2013), although the clinical significance of chr18 loss is unknown. Chr18 host ~400 genes participating in different molecular pathways but the sequencing data did not reveal any recurrent alterations (as a second hit) for chr18 genes. SI-NETs have low somatic mutation burden like other NETs (Francis et al., 2013). Second recurrent (~8%) event in SI-NETs is inactivation mutation in *CDKN1B* (Cyclin dependent kinase inhibitor 1B) gene (Figure 1-9). However, these *CDKN1B* inactivation mutations have no differences in clinical survival or phenotypes (Karpathakis et al., 2016). *CDKN1B* gene has been proposed to acts as a haploinsufficient tumor-suppressor gene in SI-NETs (Francis et al., 2013).

Another study identified three molecular subtypes of SI-NETs based on genetic, epigenetic and gene expression with different clinical outcome (Karpathakis et al., 2016).

However, the clinical understanding of SI-NETs is still limited.



**Figure 1-9.** Genomic landscape of primary and metastatic small intestine NETs. Top panel is for somatic mutation per Mb, bottom panel: *CDKN1B* recurrent somatic mutation and copy number alteration in each sample. Figure originally reported in and adapted from (Francis et al., 2013)

A systematic and comprehensive study is missing to understand the chr18 loss, tumor suppressor genes on chr18 (if any), molecular subtypes, biomarker and targeted therapeutics options.

Taking together, NETs originating from different organs shows different genetic landscape and expression pathways. Interestingly, frequent recurrent mutations in chromatin remodeling genes (like *MEN1*) are commonly seen in panNETs and LCs (part of foregut) but not in SI-NETs (part of midgut). Similarly, Chr18 loss is frequently found with SI-NETs and not in panNETs and LCs. These genetic and expression differences of NETs across GI-track may shed light into their dysregulation and pathogenesis.

## 1.8 Thesis Overview

### 1.8.1 Objectives and Aims

The incidence rate of NETs is increasing but the biological knowledge and clinical biomarkers of NETs diseases is limited. The overall objectives of this thesis were to determine the tumor pathogenesis, clinically relevant molecular subtypes, tumor cell-of-origin, diagnostic biomarkers and clinical prognosis for NETs using integrative multi-layered genomic datasets. My thesis work is mainly focused on **Pancreatic NETs (PanNETs), Lung Carcinoids (LCs) and Small Intestine NETs (SI-NETs)**.

#### The Specific aims of my thesis

**Aim 1:** Conduct comprehensive genomic, transcriptomic, and epigenomic profiles to elucidate molecular mechanisms for pathogenesis and molecular subtyping of NETs.

- a) Identify mutational landscape and driver mutations
- b) Delineate gene expression signatures for diagnostic, prognostic, and biological understanding of subtypes in NETs
- c) Delineate DNA methylation signatures for diagnostic, prognostic, and biological understanding of subtypes in NETs

**Aim 2:** Investigate tumor molecular subtypes by genotype to phenotype associations with clinical data.

### 1.8.2 Multi-Layered NETs dataset and Meta-Analysis for Aim 1 and 2

In collaboration with Prof. Laura Tang's research group at Memorial Sloan-Kettering Cancer Center (MSKCC), we generated genomic dataset for LCs, PanNETs and SI-NETs. Following table outline the dataset, technology and sample size used in this thesis. In addition to omics data generation, immunohistochemistry experiments for chapter 1 and 2 were performed at MSKCC.

	Targeted DNA Seq	Exome Seq	RNAseq	Methylation ChIP	Clinical data
LCs	29	--	30	18	30
PanNETs**	64	--	33	32	60
SI-NETs	--	20	29	21	37

\*\* Genotype for *MEN1*, *ATRX* and *DAXX* gene is probed for all PanNETs samples using Sanger sequencing

**Table 1-3.** Genomic dataset of LCs, PanNETs and SI-NETs. Briefly, RNA sequencing and Methylation 450K array was performed for all three NETs; Targeted DNA sequencing (354 gene panel) was done on LCs samples, Exome sequencing was done on SI-NETs with 10-matched normal. Sanger sequencing was done for *MEN1*, *ATRX* and *DAXX* gene for non-functional well-differentiated PanNETs.

In addition to this omics dataset, we did IHC for biomarkers for respective NETs subtype at MSKCC core facility. More details are presented in each chapter.

### 1.8.3 Chapter Overview

The studies addressing the specific aims outlined are described in chapter 2 through 4.

**Chapter 2** describes the results of Pancreatic Neuroendocrine tumor (PanNETs), where we found two robust subtypes of PanNETs, A-D-M mutant and A-D-M wildtype with distinct clinical phenotypes.

*Published in Nature Communications volume 9, Article number: 4158 (2018) Link:*

<https://www.nature.com/articles/s41467-018-06498-2>

**Chapter 3** describes the results of lung carcinoids (LCs), where we found three (LC1, LC2 and LC3) novel molecular subtypes of LCs with distinct genotype, cell of origin, molecular marker and clinical phenotypes.

*Manuscript is submitted to “Journal of National Cancer Institute” at the time of thesis submission*

**Chapter 4** describes the on-going work and results of small intestine NETs (SI-NETs), where we found two distinct cell-of-origin subtypes of SI-NETs.

*Manuscript in preparation: Brief communication*

**Chapter 5** is a summary of my overall conclusions and covers a description of work-in-progress as well as proposed future research.

## Chapter 2

### **ATRX, DAXX or MEN1 mutant pancreatic neuroendocrine tumors are a distinct alpha-cell signature subgroup**

Results described in this chapter are reproduced/adapted with permission from Chan and Laddha et al, *Nature Communications* volume 9, Article number: 4158 (2018)

Link: <https://www.nature.com/articles/s41467-018-06498-2>

Chang S. Chan\*#, **Saurabh V. Laddha**#, Peter W. Lewis, Matthew S. Koletsky, Kenneth Robzyk, Edaise Da Silva, Paula J. Torres, Brian R. Untch, Janet Li, Promita Bose, Timothy A. Chan, David S. Klimstra, C. David Allis & Laura H. Tang\*

# Equal contribution, \* Corresponding Authors

#### **Contribution:**

SVL performed the data analysis, computational design and manuscript preparation and **contributed equally** with CSC. PL, MK, KR, EDS, PT, BRU, JL, PB, TAC, DSK assisted with experimental study design, implementation and data discussion. CSC, CDA and LHT supervised the project.

**Necessary files and Supplementary data for this chapter available at**  
<https://www.nature.com/articles/s41467-018-06498-2>

## ***ATRX, DAXX* or *MEN1* mutant pancreatic neuroendocrine tumors are a distinct alpha-cell signature subgroup**

### **Abstract**

The most commonly mutated genes in pancreatic neuroendocrine tumors (PanNETs) are *ATRX*, *DAXX*, and *MEN1*. Little is known about the cells-of-origin for non-functional neuroendocrine tumors. Here, we genotyped 64 PanNETs for mutations in *ATRX*, *DAXX*, and *MEN1* and found 37 tumors (58%) carry mutations in these three genes (A-D-M mutant PanNETs) and this correlates with a worse clinical outcome than tumors carrying the wild-type alleles of all three genes (A-D-M WT PanNETs). We performed RNA sequencing and DNA-methylation analysis on 33 randomly selected cases to reveal two distinct subgroups with one group consisting entirely of A-D-M mutant PanNETs. Two biomarkers differentiating A-D-M mutant from A-D-M WT PanNETs were high *ARX* gene expression and low *PDX1* gene expression with *PDX1* promoter hyper-methylation in the A-D-M mutant PanNETs. Moreover, A-D-M mutant PanNETs had a gene expression signature related to that of alpha cells (pval < 0.009) of pancreatic islets including increased expression of *HNF1A* and its transcriptional target genes. This gene expression profile suggests that A-D-M mutant PanNETs originate from or transdifferentiate into a distinct cell type similar to alpha cells.

### **Introduction**

Pancreatic neuroendocrine tumors (PanNETs) or islet cell tumors are a relatively rare neuroendocrine malignancy with an annual incidence of less than 1 per 100,000 per

year(Halldanarson, Rabe, Rubin, & Petersen, 2008) (about 1,000 new cases per year in the United States) but currently represent the second most common epithelial neoplasm after ductal adenocarcinoma of the pancreas and account for 1% to 2% of pancreatic tumors. PanNETs were erroneously considered a benign group of neoplasm because they were initially mostly comprised of benign symptomatic insulin-producing tumors (insulinomas). However, in the past three decades, it has become apparent that at least half of all PanNETs are nonfunctional, and they are a heterogeneous group of tumors with often unpredictable and varying degrees of malignancy. As many as 50% to 80% of PanNETs are associated with synchronous or metachronous metastatic disease(Tang & Klimstra, 2011). Knowledge of functional PanNETs has evolved from insulinoma to almost a dozen other diverse hormone-secreting tumors. These individual lesions may have specific clinical, pathologic, and genetic associations, including multiple endocrine neoplasia type 1 (MEN-1), tuberous sclerosis, and von Hippel-Lindau (VHL) syndromes. Thus, the entity of PanNET represents a diverse group of heterogeneous neoplasms where combined clinical and pathologic assessment is required to further identify their genetic basis for neoplasia and to define their specific clinical behavior. The nonfunctional tumors require further elucidation to characterize their diverse pathogenesis and to predict outcome with potential biomarkers and molecular signatures. Current classification scheme for PanNETs include grade and stage(Halldanarson et al., 2008). The world Health Organization (WHO) classification, which assesses the proliferative index of neoplastic cells, divides PanNETs into low grade (G1), intermediate grade (G2), and high grade (G3). The higher grade PanNETs are generally associated clinically with more aggressive behavior(Tang, Basturk, Sue, & Klimstra,

2016). Poorly differentiated neuroendocrine tumor of the pancreas is extremely rare and clinically aggressive, which represents a different pathogenesis from the well-differentiated counterpart(Tang, Untch, et al., 2016). While well-differentiated PanNETs can be successfully treated with surgery, there are few treatments for metastatic PanNETs, and they do not respond to conventional chemotherapy. A greater understanding of PanNET pathogenesis may guide the development of novel therapeutic options.

Molecular studies have identified mutations in *MEN1*, *ATRX*, and *DAXX* to be the most commonly found in PanNETs(Jiao et al., 2011; Scarpa et al., 2017) (found in approximately 40%, 10%, and 20% of tumors, respectively). All three genes play a role in chromatin remodeling. *MEN1* is a component of a histone methyltransferase complex(Hughes et al., 2004) that specifically methylates Lysine 4 of histone H3 and functions as a transcriptional regulator. *ATRX* and *DAXX* interact to deposit histone H3.3-containing nucleosomes in centromeric and telomeric regions of the genome(Lewis et al., 2010). Additional mutations in mTOR pathway genes including *TSC2*, *PTEN*, and *PIK3CA* are found in one in six well-differentiated PanNETs(Jiao et al., 2011). Other reported rare mutations in PanNETs include DNA damage repair genes (*MUTYH*, *CHEK2*, *BRCA2*) and chromatin remodeling gene *SETD2*(Scarpa et al., 2017).

The neuroendocrine cells in the pancreas include alpha, beta, delta, pancreatic polypeptide (pp)-producing and vasoactive intestinal peptide (VIP)-producing cells. The cell of origin for nonfunctional PanNETs is not well established. Here, we genotyped 64

well differentiated PanNETs for mutations in *ATRX*, *DAXX* and *MEN1* and performed RNA sequencing (n=33) and DNA methylation (n=32) analysis to identify distinct molecular phenotypes of A-D-M mutant PanNETs which potentially reveals their distinct cell of origin or transdifferentiated state.

## **Materials and Methods**

**Patient's information:** Retrospective and prospective reviews of well-differentiated, pancreatic neuroendocrine neoplasms were performed using the pathology files and pancreatic database at MSKCC with IRB approval. All patients were evaluated clinically at our institution with confirmed pathologic diagnoses, appropriate radiological and laboratory studies, and surgical or oncological management. Follow-up information was obtained for all cases.

**Tissue acquisition and nucleotide extraction:** Cases of non-functional well-differentiated pancreatic neuroendocrine tumors were identified. Fresh-frozen tumor and paired normal tissues were obtained from MSKCC's tissue bank under an Institutional Review Board protocol. Histopathology of all tissues was evaluated on hematoxylin and eosin stained sections by an experienced gastrointestinal-hepato-pancreatobiliary pathologist to insure the nature of the tissue, greater than 80% tumor cellularity and absence of necrosis. The relevant tissues were then macro-dissected (20-25 mg) and DNA/RNA extraction using Qiagen DNeasy Blood & Tissue Kit and RNeasy Mini Kit, respectively was carried out according to the manufacturer's protocols (Qiagen, Valencia, CA).

**Sanger sequencing for gene mutation:** All exons of the *DAXX*, *ATRX*, and *MEN1* genes were amplified by PCR and then sequenced using Sanger sequencing. Every mutation detected was validated by bidirectional Sanger sequencing on the tumor-normal pairs. To maintain the correct sample annotation, we used mutation status as sample name with sample ID (For example, A\_mk11 sample is *ATRX* mutant and mk11 is sample ID). Supplementary file 1 a, b and c contains all the clinical information, mutational profile, sample annotation and ESTIMATE tumor purity. Online OncoPrint was used to plot create figure 1a.

**PanNETs Transcriptome Sequencing and data analysis:** RNA Library preparation and RNA sequencing was done by MSKCC Genomics Core Laboratory using Illumina HiSeq with (2 x 75 bp paired end reads) to a minimum depth of ~ 50 million reads were generated for each sample. Raw fastq files were probed for sequencing quality control using FastQC [<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>]. Sequencing reads were mapped to human transcripts corresponding to Genepattern(Reich et al., 2006) genome (hg19 version) GTF annotations using RSEM with default parameters. RSEM package(B. Li & Dewey, 2011) was used to prepare the reference genome with given GTF and calculated expression from mapped BAM files. STAR(Dobin et al., 2013) aligner was used to map reads in RSEM algorithm. Transcripts mapped data were normalized to TPM (Transcript Per Million) from RSEM and log<sub>2</sub> transformed (Supplementary file 6). This log<sub>2</sub>TPM values were used for all downstream analysis. Unsupervised clustering and Principal Component analysis was conducted to elucidate subtypes structure using top 3000 variant genes as input. To query robustness of this subtyping, multiple variant gene sets were used and repeated the same process of

unsupervised clustering. Top 100 variable genes were used to find genes, which were highly expressed in each subtype. Subset of these genes is selected to show in figure 2d for liver and complement system genes. To find differentially expressed genes (DEgenes) between A-D-M mutant PanNETs and A-D-M WT panNETs, we used DeSeq2 R package (Love, Huber, & Anders, 2014) on raw count (values from RSEM). We used significance cutoff with greater than 3 fold change and corrected p-value  $< 0.05$  to call a gene as DEgenes. GSEA Preranked (Subramanian et al., 2005) method was used on DEgenes to find significant KEGG pathways, motif and biological process.

**Clustering and Principal Component Analysis:** For unsupervised clustering on  $\log_2$ TPM, we used Pearson distance metric and ward.D2 hclust method (unless stated otherwise). PCA analysis was done using *prcomp* in R. R (<http://www.r-project.org/>) was used for all the analysis and visualization of data.

**PEEGset from published dataset (Bramswig et al., 2013; Muraro et al., 2016; Y. J. Wang et al., 2016):** The neuroendocrine cells in the pancreas include alpha, beta, delta, pancreatic polypeptide (pp)-producing and vasoactive intestinal peptide (VIP)-producing cells. Gene sets representing different endocrine islet and exocrine pancreatic cells (**PEEGset**) were obtained from three metadata (Bramswig et al., 2013; Muraro et al., 2016; Y. J. Wang et al., 2016) (Supplementary Table 2). We created 13 PEEGset representing all major cells from endocrine and exocrine pancreases. Supplementary Table 2 shows these gene sets with major cell types and number of genes in each set. These gene set were used as prior defined gene set for GSEA analysis.

**Gene Set Enrichment Analysis on Major Islets cell types:** Gene Set Enrichment Analysis (Subramanian et al., 2005) (GSEA) was performed on the  $\log_2$ TPM expression

values of all samples using downloaded version of GSEA software (Broad Institute, Cambridge, MA, USA) to identify the statistically enriched gene sets between A-D-M mutant and A-D-M WT PanNETs. Published pancreatic islet endocrine and exocrine cells signatures were used as prior defined sets as an input. We used all default parameters to perform GSEA on this gene sets to determine the enrichment of specific cell signature enrichment in the PanNET subtypes. We ran GSEA on 1000 permutation mode on phenotypic label to generate FDR and enrichment score (ES) for each gene set. Significant gene set was filtered based on FDR q-values (cutoff of 0.05).

**Bramswig et al., FACs sorted normal alpha and beta cells gene expression:** We extensively used Bramswig et al (Bramswig et al., 2013) FACs sorted RNAseq data to understand normal alpha and beta cells and correlated their gene signature sets with our A-D-M mutant and A-D-M WT panNETs. We downloaded supplement file for total RNA seq normalized expression data for alpha (3 replicate) and beta (3 replicate) and exocrine cells (2 replicate). Bramswig et al (Bramswig et al., 2013) provide strong genes associated with alpha, beta and exocrine cells as supplement file. We used this strong cell specific genes and created gene set for alpha, beta and exocrine and named as Bramswig\_et\_al gene set. *HNF1A* gene expression values were fetched to check whether HNF1A is over expression in normal alpha as compared to beta and exocrine. We applied Student ttest's between three alpha and three beta samples to calculate p-value for *HNF1A* gene expression. Bramswig et al strong alpha cell genes (n=465) were queried to check for *HNF1A* transcription factor motif enriched using online GSEA version (C3 TFs motif database).

**450K DNA methylation array analysis:** DNA extracted from PanNETs samples and interrogated using the Illumina 450K platform (Illumina Inc. San Diego, CA) to access the DNA methylation profiles. All the analysis was performed using ChAMP(Morris et al., 2014) version 2.6.0 open source software implemented in R. Briefly, IDAT file raw data were imported in R and filtered to exclude samples with detection p-value  $<0.01$  and beadcount  $<3$  in at least 5% of samples and normalized using FunctionNormalization(Fortin et al., 2014). This normalization method correct for background; remove dye bias followed by Quantile normalization. Unsupervised clustering and PCA were done on top variants 2000 probes (Var2000) across all samples to find classes of PanNETs. We repeated this clustering using different number (Var10000, Var5000, Var3000, Var1000 and Var500) of probes to check robustness of this subtyping. Differentially methylated CpG sites (DMP) between the A-D-M mutant and A-D-M WT PanNETs were identified using champ.MVP using the all default parameter method (Bonferroni-Hochberg) to adjust the p-value. Significant DMP sites from respective genes were compared to DEgenes to find overlapping dysregulated genes in each subtype.

**A-D-M mutant PanNET Signature and validation:** Significant differentially expressed genes (fold change  $\geq 3$  and BH Corrected Pval  $< 0.05$ ) between A-D-M mutant and WT panNETs were used with log<sub>2</sub> transformation of fold changes to create an A-D-M mutant PanNETs signature for validation. We downloaded gene expression and genotype data from two independent PanNETs cohort a) ICGC Pancreatic Cancer Endocrine Neoplasms(Scarpa et al., 2017) (PAEN) and b) (Sadanandam et al., 2015). ICGC PAEN processed A-D-M mutation status and RNAseq gene expression dataset (FPKM

normalized expression) were downloaded from ICGC website (<http://icgc.org/icgc/cgp/68/304/1003406>) for 29 samples (16 A-D-M Mutant and 13 A-D-M WT). Sadanandam et al., 2015 performed targeted sequencing of *ATRX*, *DAXX*, *MEN1*, *PTEN*, *TSC2* and *ATM* on 75 PanNETs. Genotype information was downloaded from supplementary file of (Sadanandam et al., 2015) and matched microarray gene expression data for 75 PanNETs (28 A-D-M and 47 A-D-M WT) were downloaded from NCBI GEO (Accession Number GSE73338). Normalized gene expression values obtained from GSE73338 were used for gene signature analysis. Pearson correlation was calculated between our A-D-M mutant gene signature and the mean-variance normalized gene expressions from ICGC PAEN and Sadanandam A et al., and pvalue was calculated using Wilcox test (in R). We performed GSEA analysis on A-D-M mutant and WT panNETs from Sadanandam A et al (Sadanandam et al., 2015).

**Immunohistochemistry (IHC):** A representative, formalin-fixed, paraffin-embedded tissue section (4  $\mu$ m thick) of each case was submitted to our institution's core facility to perform immunohistochemistry-using antibodies recognizing the APOH proteins. Briefly, sections were de-paraffinized and pre-treated in Cell Conditioning 1 (CC1 mild; Ventana Medical Systems, AZ, USA) using an automated staining system (Ventana Discovery XT Autostainer; Ventana Medical Systems Inc, Tucson, AZ). Primary antibodies were applied for 60 min at a dilution of 1:100 for *APOH* (anti-APOH, polyclonal antibody; Proteintech). The sections were then incubated for 60min with secondary antibody (1:200) followed by DAB Map detection (DAB visualization; Ventana Medical Systems). Cytoplasmic (*APOH*) labeling in at least 50% of the tumor

cells was considered positive. In the case of *APOH*, normal liver tissue was used as a positive control in each experiment.

**Histone marks IHC:** Serial unstained slides (4 $\mu$ m) were prepared from each block for subsequent immunohistochemistry with the following Histone 3 lysine antibodies clones: H3K4me3 clone C42D8 (Cell Signaling Technologies, 1:1000 dilution), H3K9me3 clone EPR16601 (abCam, 1:1000 dilution), H3K27me3 clone C36B11 (Cell Signaling, 1:100 dilution) and H3K36me3, clone 333 (Active motif, 1:500 dilution). Staining for all clones was performed on the Leica Bond immunohistochemistry platform according to the manufacturer's protocol with the Lieca DAB IHC detection kit. All slides were pretreated with epitope retrieval two solution (Lieca Biostystems) for 30 minutes. Primary antibodies were incubated for 30 minutes. Multi-tissue normal positive control was used. The PanNET cases (n=36, 14 A-D-M mutant and 22 A-D-M WT) were read and interpreted by an independent observer blinded to the clinicopathologic information. Scoring of all histone marks was performed using previously validated scoring systems H3K4me3(Z. J. Wang et al., 2013), H3K9me3(A. Noguchi et al., 2013), H3K27me3(Wei et al., 2008), H3K36me3(Ho et al., 2016). The tumor was considered positive for the histone mark if there was histological evidence of nuclear staining. Every tumor was scored on a scale of 0-3 according to the percentage of cells with nuclear staining: (0, 0%-5% positive cells; 1, 6%-50% positive cells; 2, 51%-75% positive cells; 3, 76%-100% positive cells). Scores of 0-1 were estimated as low expression and scores of 2-3 indicated high expression. Student t-test was used to test significance for histone methylations across *MEN1* panNETs as compared to *MEN1* WT PanNETs.

**Statistical Analysis:** Data are represented as mean  $\pm$  standard deviation. GraphPad Prism 6 (GraphPad Software Inc, La Jolla, Ca) was used for statistical and survival analyses. Survival analysis p-values (2-sided) were based on log-rank tests. Significance was defined as  $P < 0.05$ .

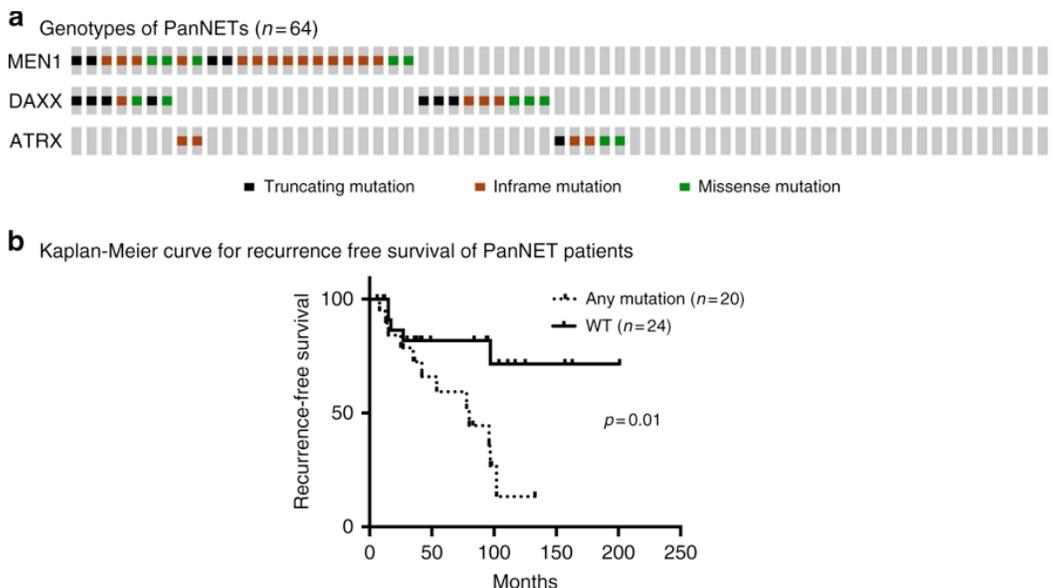
**Supplementary Files, Figures and tables:** All the supplementary contents of this paper is online and can be found at <https://www.nature.com/articles/s41467-018-06498-2#Sec25>

## Results

### **Patient cohort, clinical annotation, and genotyping for *ATRX*, *DAXX* and *MEN1***

We initially performed Sanger sequencing to genotype the *ATRX*, *DAXX* and *MEN1* genes in 64 individual PanNETs. All cases were histologically confirmed to be well-differentiated PanNETs of WHO G1/G2 grade, and cases of poorly differentiated neuroendocrine carcinoma were excluded. The mean patient age was  $52 \pm 1.5$  years (ranging from 26-73) with a 59% male population. The locations of the tumors were 38% proximal/mid body, and 62% distal pancreas. Eighty-one percent of the cases were clinically non-functional and the remaining cases included insulinomas, glucagonomas, gastrinomas, and VIPomas. The median size of tumor was  $3.6 \pm 0.4$  cm (ranging from 1.0 – 14.5 cm). Sixty-eight percent of patients had localized disease without distant metastasis at the time of initial diagnosis (Supplementary file 1).

An A-D-M mutant genotype was identified in 58% (37/64) of cases with *ATRX*, *DAXX*, *MEN1*, *MEN1/ATRX*, and *MEN1/DAXX* mutations in 8%, 16%, 20%, 3%, and 11% cases, respectively (Figure 1a).



**Figure 2-1.** Mutational landscape of *ATRX*, *DAXX* and *MEN1* in PanNETs. **a)** Oncoprint mutational profile for PanNETs samples. *ATRX/DAXX/MEN1* mutations were identified in 37/64 (58 %) of PanNETs using Sanger sequencing. **b)** Among 44 patients who initially presented with localized PanNETs (without distant metastasis), those with A-D-M mutant genotype had a worse recurrence free survival outcome than those A-D-M WT genotype in their primary tumors. A-D-M mutated samples are annotated as any mutation ( $n=20$ ) and A-D-M WT samples annotated as WT ( $n=24$ ).

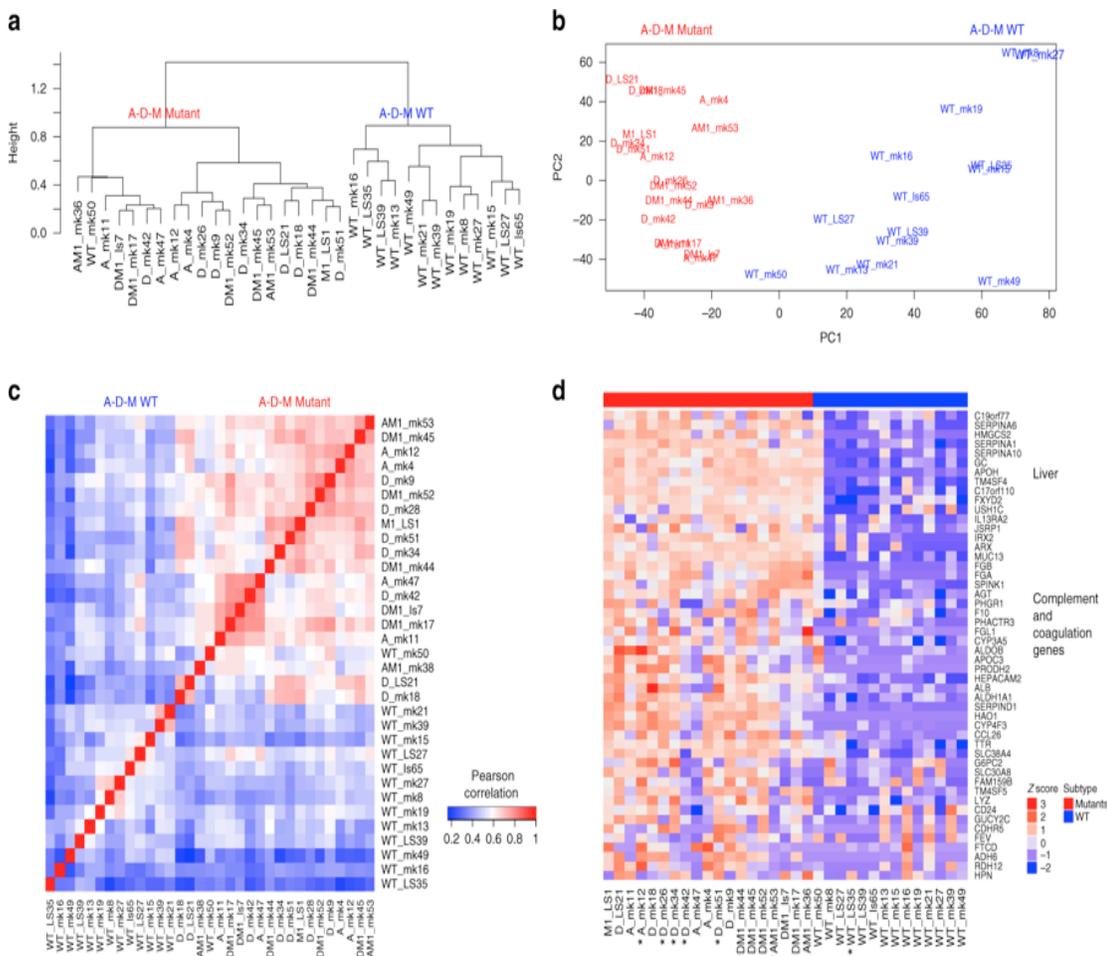
The majority of mutations in *ATRX*, *DAXX* and *MEN1* were truncation mutations (stop-gain or frame-shift) and loss of function consistent with their role as tumor suppressors (Supplementary file 1b). Similar to the observations in our previously published data (Ferrone et al., 2007), the 5-year disease specific survival was associated with tumor stage ( $p$ -value  $< 0.04$ ), tumor grade (G1 vs G2  $p$ -value  $< 0.02$ ), and distant metastasis ( $p$ -value  $< 0.002$ ), respectively. Among 44 patients who initially presented with localized

disease without distant metastasis, those with the A-D-M mutant genotype had a worse recurrence free survival than those of A-D-M WT genotype (Figure 1b). Furthermore, in comparison to A-D-M WT PanNETs, the A-D-M mutant PanNETs were associated with larger tumor size ( $3.6\pm 0.6$  cm vs.  $5.6\pm 0.7$  cm, p-value  $< 0.03$ ) and higher tumor stage (T1 and T2 vs. T3, p-value  $< 0.04$ ). Other demographic and clinical characteristics (including gender, age, tumor functionality, and lymph node metastasis) revealed no statistically significant differences between the two genotypes of PanNETs.

### **Gene expression and DNA methylation reveal two subtypes of PanNETs**

We performed RNA sequencing on 33 randomly selected tumors (19 A-D-M mutant, and 14 A-D-M WT). Unsupervised hierarchical clustering of the top 3000 variable genes across the PanNETs revealed two distinct clusters where almost all A-D-M mutant PanNETs were found in one cluster (Figure 2a). The grouping of A-D-M mutant PanNETs into one distinct cluster by gene expression was robust to the number of most variable genes used for clustering (Supplementary Fig 1). Principal component analysis (PCA) separated the A-D-M mutant PanNETs from the A-D-M WT PanNETs along the first principal component (corresponding to the component comprising the largest variation in gene expression) (Figure 2b). The separation of A-D-M mutant PanNETs from A-D-M WT PanNETs by PCA was robust to the number of top variable genes used (Supplementary Fig 2). These data show that A-D-M mutant tumors have a distinct gene expression pattern from that of A-D-M WT PanNETs. Neither hierarchical clustering nor PCA from gene expression revealed further subgrouping of the tumors with single

mutations in *ATRX*, *DAXX*, or *MEN1* or double mutations in *ATRX/MEN1* or *DAXX/MEN1*.



**Figure 2-2.** A-D-M mutant and WT PanNETs as two distinct gene expression groups. **a)** Unsupervised clustering of PanNETs using top 3000 variant genes across all samples revealed two distinct robust clusters, **b)** Principal component analysis using top 3000 variant genes separated the A-D-M mutant from A-D-M WT PanNETs along the first principal component (PC1), **c)** Heatmap of pair-wise Pearson correlation of panNETs using top 3000 variant genes across all samples revealed a higher correlation among A-D-M mutants as compared to A-D-M WT panNETs. Red color represents higher correlation and blue represents lower correlation, **d)** Heatmap of top variants genes showing liver, complement, and coagulation genes highly expressed in A-D-M mutant panNETs. Star (\*) below sample names represent liver metastatic samples (except for A\_mk12 which is a lymph node).

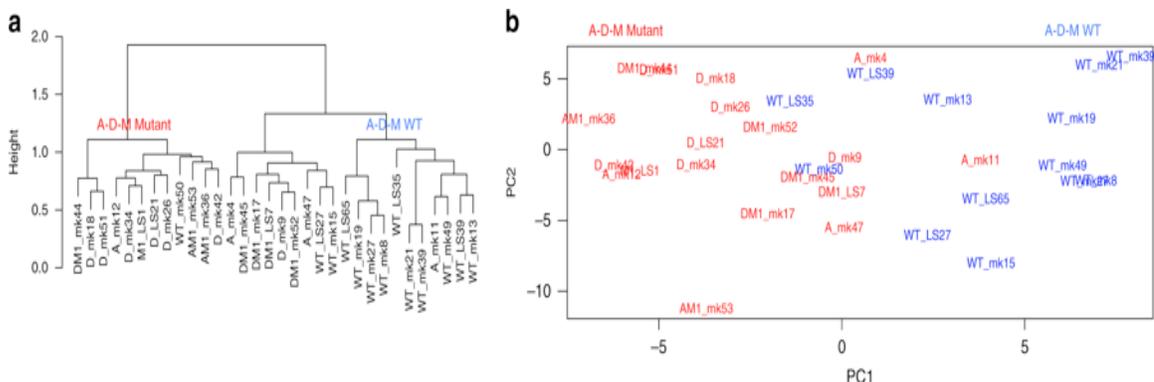
In hierarchical clustering, the A-D-M mutant PanNETs formed a tighter cluster than the A-D-M WT PanNETs. In PCA, the A-D-M mutant PanNETs had smaller variance along

PC1 than A-D-M WT PanNETs. Pair-wise correlation of gene expression between all PanNETs, showed a higher correlation among A-D-M mutant PanNETs as compared to A-D-M WT PanNETs (Figure 2c). Among A-D-M mutant PanNETs, mutants with the same genotype (mutations in *ATRX/DAXX/MEN1*) did not show greater gene expression correlation. These data suggest that A-D-M mutant PanNETs are a more homogeneous group compared to A-D-M WT PanNETs.

Within A-D-M mutant or A-D-M WT PanNETs groups, unsupervised clustering and PCA did not reveal differences between primary and metastatic tumors. Top 100 genes with highest variance across all samples separates mutant from A-D-M WT PanNETs and showed relatively high expression of “liver-specific” genes (*APOH, ALDH1A1, FGB, APOC3* etc.) as well as complement and coagulation pathway genes (*SERPINA1, FGA, F10, CP, MT3* etc.) in A-D-M mutant PanNETs (Figure 2d;Supplementary Fig 3), both in primary (collected in absence of liver tissue) and metastatic tumors. Moreover, the pathological estimate of tumor purity was over 80% for all samples of PanNETs consistent with inference from ESTIMATE(Yoshihara et al., 2013) (median tumor purity of 90%, Supplementary file 1c) showing high tumor purity characteristic of well differentiated PanNETs. In addition, seven A-D-M mutants and one A-D-M WT PanNETs were from the tissue of liver metastases and they had gene expression profile most similar to the genotype group of their primary PanNET counterpart (Figure 2d). We confirmed the distinct gene expression signature of A-D-M mutant PanNETs in a larger tumor set (47 PanNETs including the 33 PanNETs where RNA sequencing was performed) using gene expression microarray technology. The 14 additional samples are

comprised of 3 A-D-M WT PanNETs and 11 A-D-M mutant PanNETs (7 *MEN1* mutant, 2 *DAXX* mutant, and 2 *DAXX/MEN1* mutant)(Supplementary Fig 4).

To investigate epigenetic differences between PanNETs, we used the Illumina 450K chip to assay the DNA methylation at 411,549 CpG sites in 32 PanNETs. Unsupervised hierarchical clustering of the top 2000 variable DNA methylation sites across the PanNETs revealed two distinct clusters where almost all A-D-M mutant PanNETs were found in one cluster (Figure 3a). Principal component analysis (PCA) separated the A-D-M mutant PanNETs from the A-D-M WT PanNETs along the first principal component (corresponding to the component comprising the largest variation in DNA methylation) (Figure 3b).



**Figure 2-3.** Distinct DNA methylation pattern PanNETs subtypes. **a)** Unsupervised clustering of PanNETs using top 2000 variant CpG sites across all samples revealed two clusters, **b)** PCA using top 2000 variant CpG sites separated A-D-M mutant from A-D-M WT PanNETs along PC1.

The separation of A-D-M mutant PanNETs from A-D-M WT PanNETs by PCA was robust to the number of top variable DNA methylation sites used (Supplementary Fig 5).

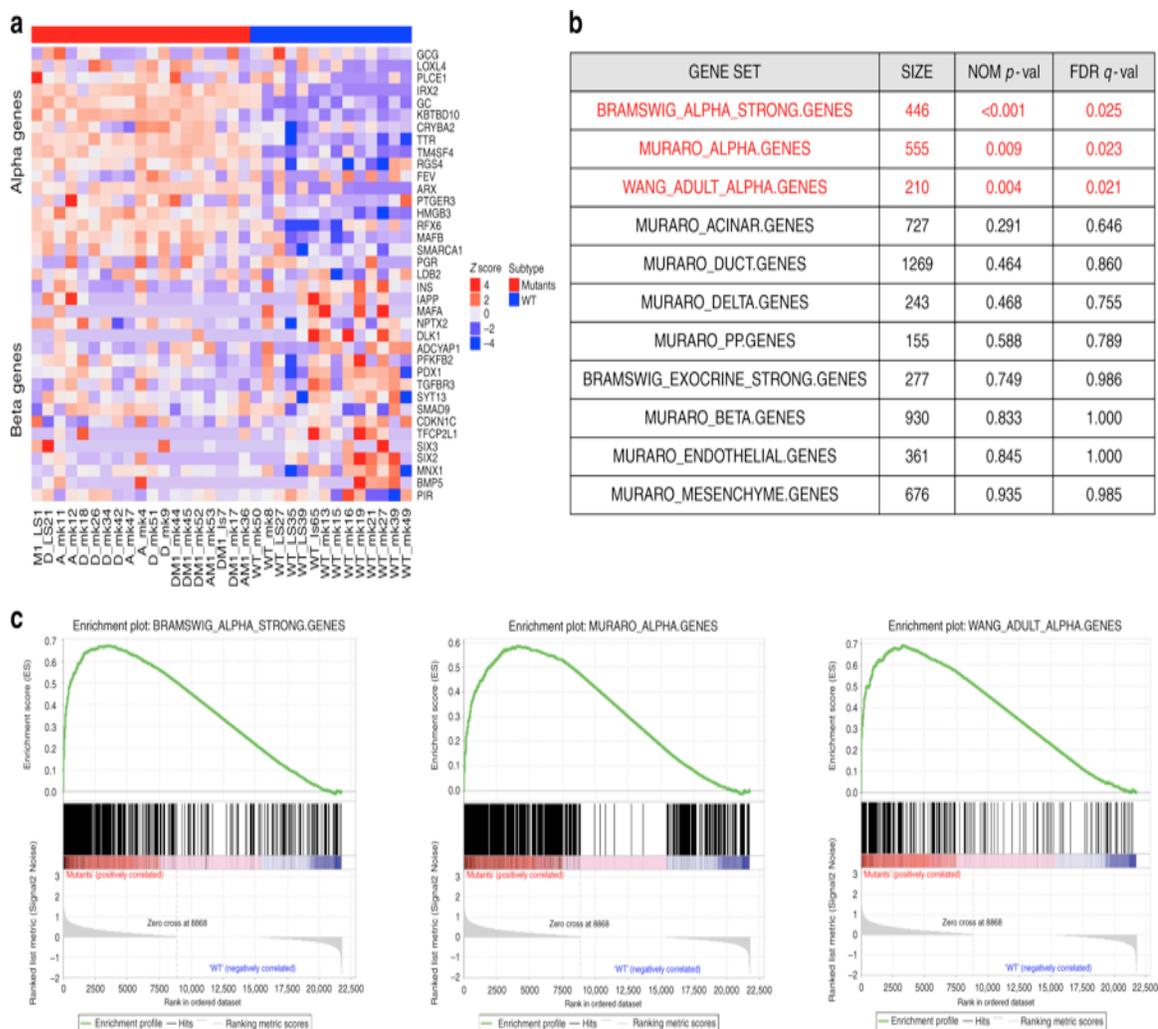
These data reveal that A-D-M mutants PanNETs have a distinct DNA methylation pattern from that of A-D-M WT PanNETs. Neither hierarchical clustering nor PCA revealed differences in DNA methylation sites between the different combinations of genes

mutated among the A-D-M mutant PanNETs. Within A-D-M mutant or A-D-M WT PanNETs groups, unsupervised clustering and PCA of DNA methylation did not reveal differences between primary and metastatic tumors.

To investigate the global histone methylation level in PanNETs with and without A-D-M mutations, we performed immunohistochemistry on H3K4me3, H3K9me3, H3K27me3, and H3K36me3 on 36 PanNETs. There was a general trend of lower histone methylation level for *MEN1* mutated PanNETs when compared to WT PanNETs (Supplementary Fig 6 and Supplementary Table 1).

#### **A-D-M mutant PanNET gene expression resembles that of alpha cells**

There are multiple neuroendocrine cell types in the pancreas including alpha, beta, gamma, delta, and epsilon. We used gene expression data for these various pancreatic neuroendocrine and exocrine cell types from a single cell sequencing study (Muraro et al., 2016) (Supplementary Table 2) to identify gene-set signatures representing highly expressed cell-type-specific genes (Supplementary file 2). The A-D-M mutant PanNETs uniformly exhibited a gene expression signature that was very similar to that of alpha cells (Figure 4a). The A-D-M WT PanNETs were more heterogeneous in their expression of the genes among the gene set signatures for the different pancreatic neuroendocrine cell types. Greater heterogeneity of gene expression signature in A-D-M WT PanNETs was consistent with the greater heterogeneity found in global gene expression.



**Figure 2-4.** A-D-M mutant PanNETs with alpha-cell signature. **a)** Heatmap of gene expression for top 20 alpha and beta cell-specific genes from (Muraro et al. 2016) revealed alpha cell specific genes are highly expressed in A-D-M mutant panNETs. A-D-M WT panNETs are more heterogeneous in gene expression but some show high beta cell specific gene expression. Red color represents higher correlation with alpha cell specific genes, **b)** Gene set enrichment analysis show A-D-M mutant PanNETs to be enriched for expression of alpha cell specific genes. Pancreas cell type (alpha, beta, delta, PP, acinar, ductal) gene signatures were obtained from three different published dataset to access enrichment of cell type signatures in A-D-M mutant vs A-D-M WT PanNETs. Table represents GSEA results where size is the number of genes in gene set. All alpha cell gene sets (from three different sources) are significantly enriched in A-D-M mutant panNETs (highlighted in red). No other cell types were enriched in A-D-M mutant or A-D-M WT panNETs, **c)** GSEA plots of significant alpha cell signatures (from Bramswig et al., 2013; Wang et al., 2016; Muraro et al., 2016)

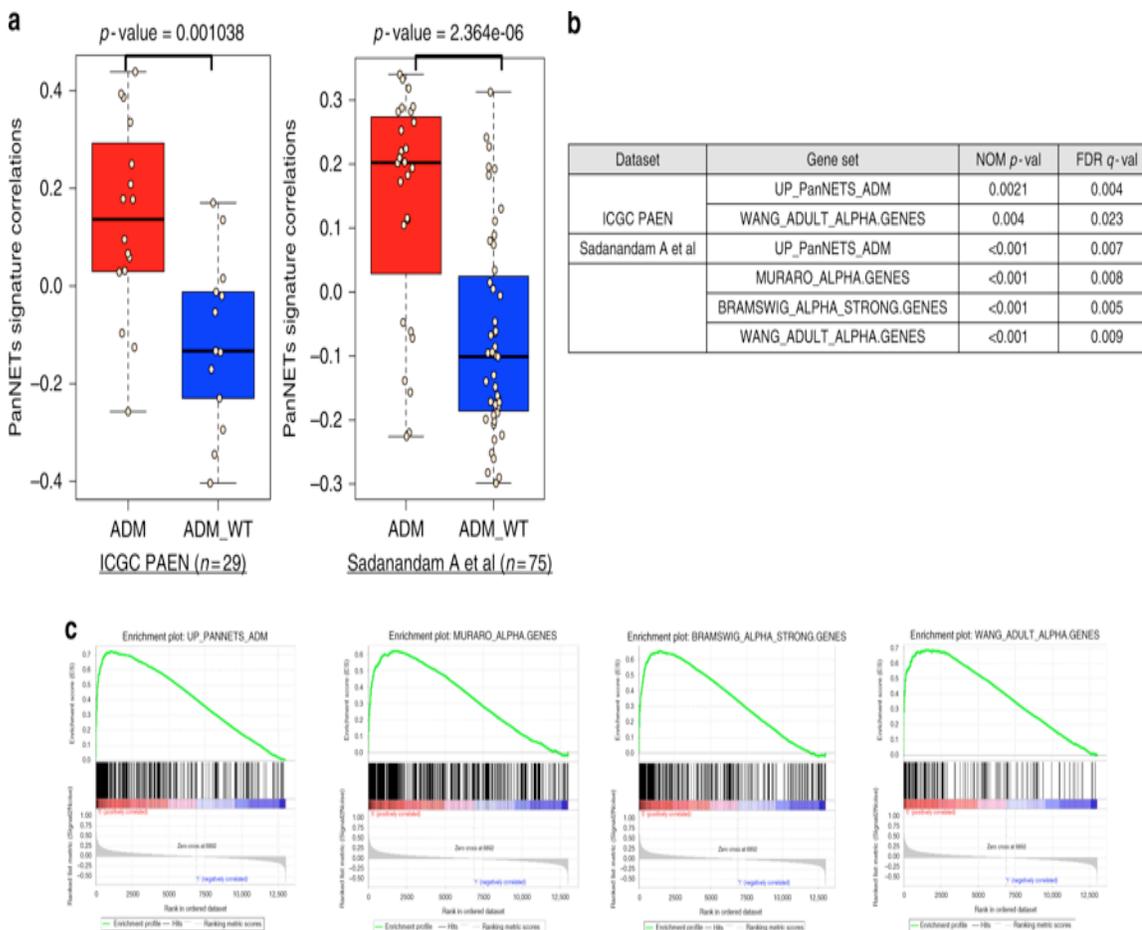
To further investigate the gene expression signature of A-D-M mutant PanNETs we performed gene set enrichment analysis (Subramanian et al., 2005) (GSEA) on the thirteen manually curated gene sets for pancreatic endocrine and exocrine cells from a previous study. This study assessed gene expression of individual pancreatic cell types (alpha, beta, delta, PP, acinar, ductal, mesenchyme and endothelial) enriched by flow cytometry and using single cell RNAseq (Supplementary Table 2). Our analysis indicates that only the alpha cell gene signature was significantly enriched in A-D-M mutant PanNETs (FDR q-value < 0.009) (Figure 4b and c) (Supplementary table 3).

Alpha and beta cell lineage specific genes were examined for the A-D-M mutant and WT PanNETs. *ARX*, *IRX2*, and *TM4SF4* were all highly expressed in A-D-M mutant PanNETs compared to A-D-M WT PanNETs (Supplementary Fig 7). Surprisingly, *GCG* (glucagon) expression was lower in A-D-M mutants as compared A-D-M WT PanNETs. For beta cell specific genes, *PDX1*, *MAFA*, *INS*, and *DLKI*, all had lower expression in A-D-M mutant PanNETs than A-D-M WT PanNETs (Supplementary Fig 7). However, these genes had much greater expression heterogeneity in A-D-M WT PanNETs suggesting that some A-D-M WT PanNETs resemble beta cells and others did not (Supplementary Fig 7).

### **Validation of distinct subtype and alpha cell signature in A-D-M mutant PanNETs**

We derived an A-D-M mutant gene expression signature from significant differentially expressed genes between the A-D-M mutant and WT PanNETs from our data set (n=33). We used two independent panNET (Sadanandam et al., 2015; Scarpa et al., 2017) data

sets to validate the gene expression signature of A-D-M mutant panNETs. We obtained A-D-M mutation status and gene expression dataset from International Cancer Genome Consortium Pancreatic Endocrine Neoplasm (ICGC PAEN) (n=29)(Scarpa et al., 2017) and Sadanandam, A. et al (n=75)(Sadanandam et al., 2015).

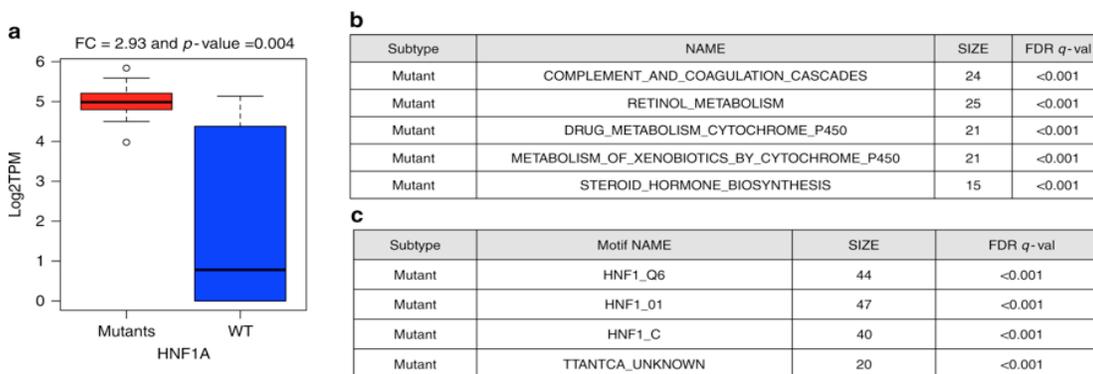


**Figure 2-5.** Validation of A-D-M mutant PanNET and alpha cell signatures. **a)** Pearson correlation boxplot for two independent PanNET datasets show significant positive and negative correlations of A-D-M mutant and WT PanNETs with our A-D-M mutant PanNETs signature respectively (red represent A-D-M mutant and blue represent A-D-M WT with Wilcox p-value; center line is median, bounds of box are first and third quartile, and whiskers are min and max). **b)** GSEA analysis shows A-D-M mutant PanNETs from ICGC PAEN (Scarpa A et al., 2017) and Sadanandam et al., 2015 are enriched for A-D-M mutant and alpha cell gene signatures. **c)** GSEA enrichment plot for significant gene set for A-D-M mutant and alpha cell gene signatures from Sadanandam et al., 2015 dataset.

The A-D-M mutant and WT PanNETs from both data sets have significant positive and negative correlations with our A-D-M mutant PanNET signature respectively (Figure 5a) (Supplementary file 3). Additionally, we found alpha cell signatures to be significantly enriched (FDR  $q < 0.001$ ) only in the A-D-M mutant PanNETs from the two validation data sets using GSEA (Figure 5b and 5c).

### **HNF1A pathway is transcriptionally upregulated in A-D-M mutant PanNETs and alpha cells**

HNF1A is one of the most significantly differentially expressed genes between A-D-M mutant and WT PanNETs. HNF1A is a homeobox family transcription factor that is highly expressed in the liver and is involved in the regulation of several liver-specific genes. The expression of HNF1A was 2.93 fold higher in A-D-M mutant PanNETs than A-D-M WT PanNETs (corrected  $p$ -value  $< 0.004$ ) (Figure 6a). Differentially expressed genes (DEgenes) between the A-D-M mutant and A-D-M WT PanNETs were found in 1478 genes (with greater than 3 fold change and corrected  $p$ -value  $< 0.05$ , see Methods section)(Supplementary file 4).



**Figure 2-6.** *HNF1A* motif and pathways with transcriptionally up-regulation in A-D-M mutant panNETs and alpha cells. **a)** Boxplot of *HNF1A* gene expression for A-D-M mutant and A-D-M WT PanNETs. *HNF1A* was homogeneously expressed 2.93 fold higher in A-D-M mutants

PanNETs (corrected  $p$ -val  $< 0.004$ , Benjamini–Hochberg). **b)** Table represents significant KEGG pathways where genes were differentially expressed between A-D-M mutants and A-D-M WT panNETs. **c)** Table represents transcription factor motifs significantly enriched in promoters of genes differentially expressed in A-D-M mutants and A-D-M WT panNETs. Three HNF1 related motif gene sets from GSEA showed significant enrichment in genes over-expressed in A-D-M mutant panNETs. GSEA was used to find pathway enrichment from genes differentially expressed between A-D-M mutant and A-D-M WT PanNETs

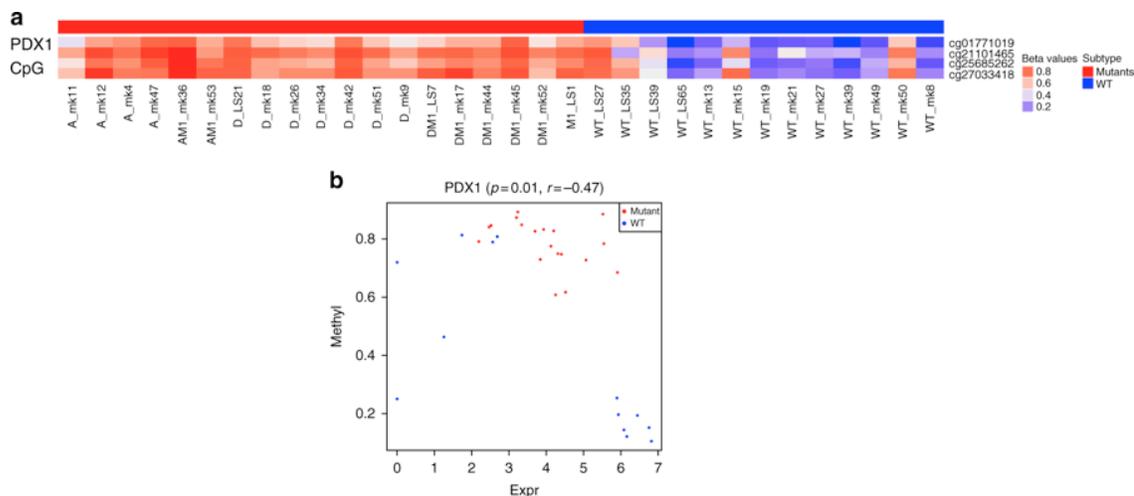
Functional pathway enrichment for DEgenes using preranked GSEA revealed the complement and coagulation cascades, retinol metabolism, and drug metabolism to be upregulated in A-D-M mutant PanNETs (see Methods section) (Figure 6b; Supplementary file 4). The differentially expressed genes were also enriched for HNF1A transcription factor motifs in their promoters (FDR  $< 0.001$ , Figure 6c). The complete list of significant TF motifs is presented in Supplementary file 4. Taken together, the A-D-M mutant PanNETs had higher expression of *HNF1A* along with many of its transcriptional target genes associated with liver function. In addition, the transcriptional regulator of *HNF1A*, *HNF4A* (J. Li, Ning, & Duncan, 2000) was expressed 3.02 fold higher in A-D-M mutant PanNETs (p-value  $< 0.009$ ). We used gene expression data from Bramswig et al. (Bramswig et al., 2013) to show *HNF1A* expression was increased in alpha cells compared to beta cells (p-value  $< 0.008$ ), and the 465 alpha cell specific genes in the pancreas were enriched for transcriptional targets of *HNF1A* and for having *HNF1A* TF motif in their promoters (Methods section; Supplementary Table 4).

Many of the most differentially expressed genes and highly expressed in A-D-M mutant PanNETs are targets of *HNF1A* and are involved in protein secretion, transport and metabolism (*APOH*, *ALB*, *AFM*, *HAO1*, *UGT1A3*, *UGT1A1*, *GC*, *G6PC*, *TM4SF4*, *PKLR* etc). *APOH* is expressed 8.46 fold higher in A-D-M mutant PanNETs (p-value  $< 10^{-5}$ )

and is potentially a good diagnostic biomarker for A-D-M mutant PanNETs. Moreover, *APOH* has been shown to have high expression in only the alpha cells of the pancreatic islet in a single cell RNA sequencing study (Baron et al., 2016). We perform IHC staining for *APOH* and show positive staining in  $70\pm 2.5\%$  of A-D-M mutant and only  $18\pm 2.0\%$  of A-D-M WT PanNETs (Supplementary Fig 8).

### **Integrative analysis reveals *PDX1* gene is hypermethylated with low expression in A-D-M mutant PanNETs**

There is no genome wide hypo or hypermethylation of DNA in A-D-M mutant or WT panNETs. DNA methylation differences between the A-D-M mutant and A-D-M WT PanNETs were found at 378 CpG sites (corrected p-value < 0.05 and difference in beta value > 0.2, see Methods section), 287 of which were found in genes and 91 in intergenic regions (Supplementary file 5). Of the 287 differentially methylated genic CpG sites, 70 (associated with 59 genes) were found at promoter (transcriptional start site, TSS1500 and TSS200) or within first exon, a region where DNA methylation is associated with transcriptional repression (Brenet et al., 2011). Thirteen of the 59 genes were also found to be differentially expressed (with fold change greater than 3 and corrected p-value < 0.05, see Methods section) and seven genes that were hypomethylated in A-D-M mutant and over-expressed are *APOH*, *CCL15*, *EMID2*, *PDZK1*, *HAO1*, *BAIAP2L2*, and *NPC1L1*. One gene, *TACR3*, was hypomethylated in A-D-M WT and over-expressed (Supplementary file 5). Four of the 70 CpG sites were found in the gene *PDX1* (pancreatic and duodenal homeobox 1), a transcription factor necessary for pancreatic development and beta cell maturation.



**Figure 2-7.** PDX1 has promoter hypermethylated and lower gene expression in A-D-M mutant panNETs. **a)** Four PDX1 promoter CpG sites show strong hypermethylation in A-D-M mutant PanNETs (corrected  $p$ -val < 0.05, Benjamini–Hochberg). The range of beta values is from 0 to 1 and represented as blue (hypo-methylation) to red (hyper-methylation). **b)** PDX1 expression and promoter methylation (TSS1500 cg27033418 CpG site) across all samples showing separation of A-D-M mutant and A-D-M WT PanNETs.

*PDX1* functions in the cell fating of endocrine cells, favoring the production of insulin positive beta cells and somatostatin positive delta cells while repressing glucagon positive alpha cells (Mansouri, 2012). These four CpG sites were all hypermethylated in A-D-M mutant PanNETs (Figure 7a) and the expression of *PDX1* was 2.92 fold higher in A-D-M WT PanNETs ( $p$ -value < 0.005) (Figure 7a and b). In contrast, while *ARX* was highly expressed in A-D-M mutant PanNETs compared to A-D-M WT PanNETs, the promoter and first exon of *ARX* are not differentially methylated.

## Discussion and Conclusions

Similar to a number of recent studies (Marinoni et al., 2014; Park et al., 2017), we have demonstrated in this cohort of PanNETs that, in addition to pathologic stage and grade of the tumor, mutations in *DAXX*, *ATRX*, and *MEN1* are associated with adverse clinical

outcome in comparison to those without these mutations. Our results seem to be in contradiction to the findings initially reported by Jiao et al.(Jiao et al., 2011) in that 15 patients with PanNETs carrying mutations in *DAXX* or *ATRX* genes had better survival than did 12 patients with wild-type PanNET. This discrepancy between our data and their data could be attributed to a different composition of the tumors. Indeed, all the tumors analyzed in Jiao's study(Jiao et al., 2011) were liver metastases from PanNETs as opposed to only 19 % (12 out of 64) in our study. Other factors including sample size and length of follow up time may also contribute to discrepancies between different studies.

Here, we found that A-D-M mutant PanNETs form a distinct subgroup on the basis of their gene expression profile and DNA methylation pattern. Moreover, this subgroup is more homogeneous based on gene expression profile than the A-D-M WT PanNETs. The gene signature of the A-D-M mutant PanNETs strongly corresponds to the genes that are specifically expressed in alpha cells including genes known to define alpha cells such as *ARX* and "liver-specific" genes such as *HNF1A* and its transcriptional targets. Conversely, *PDX1*, a gene critical to the beta cell lineage is transcriptionally repressed in A-D-M mutant PanNETs and the *PDX1* promoter is hypermethylated. On the other hand, WT PanNETs have heterogeneous gene expression profiles and their gene mutational landscape is less understood.

The pancreas is comprised of many different cell types including acinar, ductal, and at least five neuroendocrine cell types including alpha, beta, gamma, delta, and epsilon cells. There are two plausible explanations for the "alpha cell-like" expression pattern of

A-D-M mutant PanNETs. Either an alpha cell or an uncharacterized cell type with an alpha-cell like gene expression profile is the cell-of-origin for PanNETs with mutations in *ATRX*, *DAXX* and *MEN1*, or loss of *ATRX*, *DAXX*, or *MEN1* genes may promote pancreatic neuroendocrine (or progenitor) cell types to reprogram their gene expression profiles to resemble alpha cells. It remains unclear whether there are pancreatic stem or progenitor cells in adult pancreas.

*ATRX-DAXX* and *MEN1* are involved in distinct biochemical pathways to regulate gene expression. Therefore, we would expect that loss of these proteins during transformation of A-D-M mutant PanNETs would result in a more heterogeneous gene expression profile. Due to the high degree of homogeneity of the A-D-M mutant PanNETs at the level of gene expression and the strong expression of genes that are known to be alpha cell specific, we hypothesize that alpha cells are the cell-of-origin for this group of tumors. In addition, *MEN1* and *ATRX/DAXX* mutations occur alone or in a combined pattern suggest that they have independent oncogenic activities in A-D-M mutant PanNETs, making the idea of reprogramming to a homogeneous alpha-like cell state less probable. Some of the A-D-M WT PanNETs have a strong beta cell signature and these may have arisen from beta cells (Figure 4a). However, other A-D-M WT PanNETs have neither alpha nor beta cell signatures, which may arise from other cell types in the pancreas.

Conditional knockouts of *MEN1* in mice support the model of an alpha cell origin for A-D-M PanNETs (Shen et al., 2009). The restricted deletion of *MEN1* to alpha cells

surprisingly led to the development of insulinomas(Lu et al., 2010; Shen et al., 2010). Most of our PanNETs were nonfunctional (26 of 33 PanNETs) but the functional tumors were insulinoma and VIPoma, even though their gene expressions have alpha cell signature. Moreover, some PanNETs express combinations of neuroendocrine hormones (*GCG*, *INS*, *SST*, *PPY*, *GHRL*, *VIP*, and *GAST*), suggesting that regulation of cell type specific hormone may be disrupted. To create robust gene signatures that are not sensitive to changes in expression of a few genes, we use a large number of genes to create the A-D-M mutant and alpha cell signatures. In other mouse models of PanNETs(Bertolino et al., 2003; F. Li et al., 2015; Shen et al., 2009), *MEN1* deletion using the insulin or *PDX1* promoter driven Cre construct, insulinomas, glucagon-expressing tumors and well differentiated PanNETs were also observed. However, Cre expression may be leaky in these models and further study is needed to understand the heterogeneity of the cells in the tumors that develop and trace the cell of origin or transdifferentiated state of the cancer cells.

In our gene expression analysis, we have not identified the oncogenic pathways activated in A-D-M mutant PanNETs. *MEN1* has been shown to upregulate expression of long noncoding RNA *MEG3* in MIN6 mouse insulinoma cell line(Modali, Parekh, Kebebew, & Agarwal, 2015). In the same study, they show *MEG3* represses expression of the oncogene *MET* leading to delayed cell cycle progression and reduced cell proliferation. In a different study, *MEN1* and *DAXX* were shown to repress the expression of the membrane metalloendopeptidase (*MME*) and mutations in *MEN1* or *DAXX* result in loss of this repression leading to neuroendocrine tumor proliferation(Feng et al., 2017). Our data is consistent with these studies when comparing A-D-M mutant to WT PanNETs,

showing that A-D-M mutant PanNETs have lower expression of *MEG3* (7.3 fold lower, p-value < 4.3E-07), higher expression of *MET* (3 fold higher, p-value < 0.003), and higher expression of *MME* (4 fold higher, p-value < 0.001). Among A-D-M mutant PanNETs, we do not see expression differences of *MEG3*, *MET*, and *MME* depending on mutation status of *ATRX*, *DAXX*, and *MEN1*.

While PanNETs may seemingly represent as a single clinical disease, they can be further characterized into different subtypes based upon their cell lineage and the associated molecular genotype. Understanding the epigenetic and transcriptional dysregulation of PanNETs will require comparison to their proper cells of origin and may explain the unpredictable outcome of the disease and facilitate the development of unique and targeted therapeutic strategies.

### Chapter 3

## **Integrative Genomic Characterization Identifies Molecular Subtypes of Lung Carcinoids**

Results described in this chapter are reproduced/adapted with permission from Laddha S.V et al,

*Manuscript submitted to “Journal of National Cancer Institute” at the time of thesis submission*

**Saurabh V. Laddha**<sup>1</sup>, Edaise M. da Silva<sup>2</sup>, Kenneth Robyzk<sup>2</sup>, Brian R. Untch<sup>3</sup>, Hua Ke<sup>1</sup>, Natasha Rekhtman<sup>2</sup>, John T. Poirier<sup>4</sup>, William D. Travis<sup>2</sup>, Laura H. Tang<sup>2\*</sup>, Chang S. Chan<sup>1, 5\*</sup>

<sup>1</sup>Rutgers Cancer Institute of New Jersey, Rutgers University, New Brunswick, NJ, USA

<sup>2</sup>Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York NY, USA

<sup>3</sup>Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York NY, USA

<sup>4</sup>Thoracic Oncology Service, Memorial Sloan-Kettering Cancer Center, New York NY, USA

<sup>5</sup>Rutgers Cancer Institute of New Jersey, Rutgers University and Department of Pediatrics, Robert Wood Johnson Medical School, New Brunswick, NJ, USA

\*Corresponding Authors

#### **Contributions:**

**S.V.L:** Computational design, computational data analysis and manuscript preparation.

**E.M.S and K.R:** specimen preparation, experimental design (mainly IHC experiments) and manuscript review. **B.R.U:** data discussion and manuscript review. **H.K:** analysis.

**J.T.P, N.R and W.D.T:** Tissue Microarray IHC experimental implementation. **L.H.T and**

**C.S.C:** study design, data analysis and manuscript preparation.

Supplementary information will be available online after the acceptance of this study.

## **Integrative Genomic Characterization Identifies Molecular Subtypes of Lung Carcinoids**

### **Abstract**

Lung carcinoids (LCs) are rare and slow growing primary lung neuroendocrine tumors. We performed targeted exome sequencing using a 354-cancer gene panel (n=29), mRNA sequencing (n=30) and DNA methylation array (450K, n=18) on macro-dissected lung carcinoids. The recurrent mutations we identified were enriched for genes involved in covalent histone modification/chromatin remodeling (34.5%) (*MEN1*, *ARID1A*, *KMT2C* and *KMT2A*) as well as DNA repair (17.2%) pathways. Unsupervised clustering and principle component analysis on gene expression and DNA methylation profiles showed three robust molecular subtypes (LC1, LC2, LC3) with distinct clinical features. *MEN1* gene mutations were found to be exclusively enriched in the LC2 subtype (p-value < 0.001). Subtype LC1 and LC3 is predominately found at peripheral and endobronchial lung respectively. Subtype LC3 is diagnosed at younger age than LC1 and LC2. Immunohistochemical staining of two biomarkers, *ASCL1* and *SI00*, was found to be sufficient to stratify the three subtypes. This molecular classification of lung carcinoids into three subtypes may facilitate the understanding of their molecular mechanisms and improve treatment decision and clinical management.

### **Introduction**

Lung carcinoids are an indolent and rare type of primary lung neoplasms that are, in general, understudied. The 2015 World Health Organization(Travis et al., 2015) (WHO)

classification of Lung Carcinoids (LCs) includes Atypical Carcinoids (AC) (~0.2% prevalence) and Typical Carcinoids (TC) (~2% prevalence). TCs are slow growing tumors that rarely spread beyond the lungs while ACs are faster growing tumors and have a greater chance of metastasizing to other tissues(Caplin et al., 2015). The WHO classification relies mainly on morphology, proliferation rate (mitotic index) and necrosis assessment(Travis et al., 1998). This current method of classification has its drawbacks as studies have shown that the reproducibility of cancer classification and its prognostic efficacy have high inter-observer variability(Travis et al., 1998; van den Bent, 2010), especially for differentiating between TC and AC(Swarts et al., 2014). Recent WHO classifications highlight use of the Ki67 cell proliferation marker to distinguish ACs from TCs(Travis et al., 2015). However, overlapping distribution of Ki67 between ACs and TCs does not enable reliable stratification between well-differentiated lung carcinoids<sup>1</sup>(Pelosi, Papotti, Rindi, & Scarpa, 2014; Volante, Gatti, & Papotti, 2015). It has also been reported that TCs and ACs are over-diagnosed as small cell lung carcinomas (SCLC) in small crush biopsy specimens(Pelosi et al., 2005), a situation where artifacts in specimens appear as bluish clusters in which cellular details are not recognizable. As SCLCs are highly malignant, incorrect diagnosis of TC and AC tumors as SCLC can subject patients to unnecessary stress and treatment(Pelosi et al., 2005). More accurate molecular diagnostic tools and stratification for lung carcinoids will help ensure more appropriate treatment and clinical management.

Previous genomic analysis of lung carcinoid tumors has identified recurrent mutations in *MEN1*, *PSIP1*, and *ARID1A*(Fernandez-Cuesta et al., 2014). No significant mutations or

focal copy alterations were observed in genes that are frequently mutated in non-small cell lung cancer (NSCLC), large cell neuroendocrine carcinoma (LCNEC) and SCLC (e.g. *KRAS*, *TP53*, *EGFR* and *RB1*)(Vollbrecht et al., 2015). The different mutation spectrum and low mutation burden(Fernandez-Cuesta et al., 2014) of lung carcinoids indicate they are distinct from NSCLC and high-grade lung NETs. It is not known if there are distinct molecular subtypes of lung carcinoids or what are their cells of origin.

In this study, we performed genotyping on 29 LCs to detect mutations in a 354-cancer gene panel, mRNA sequencing (n=30) and DNA methylation 450K-array analysis (n=18) and thus found three molecular subtypes with distinct clinical features. We also identified two key biomarkers (*ASCL1* and *SI00*) to stratify these subtypes. Integration of genetic and epigenetic hallmarks distinguishes each subtype of carcinoid (irrespective of their TCs or ACs WHO classification), providing deeper insight into their distinctive molecular and biological mechanisms of tumorigenesis as well as cell of origin.

## **Materials and Methods**

**Patient Data:** Retrospective and prospective reviews of 30 lung carcinoids neoplasms were performed using the pathology files and institutional database at MSKCC with IRB approval. All patients were evaluated clinically at MSKCC institution with confirmed pathologic diagnoses, appropriate radiological and laboratory studies, and surgical or oncological management. Relevant clinical and pathologic information is presented in Supplementary File 1. Tissue microdissection and nucleic acid extraction process were

followed as described in chapter 2 under “Tissue microdissection and nucleic acid extraction” section.

**DNA sequencing:** We performed next-generation targeted sequencing on 29 LCs using MSK-IMPACT(Cheng et al., 2015) hybrid capture based array enclosed cancer gene panel (n=341). Single nucleotide variants and short Indel (<30bp) were annotated using MSK-IMPACT pipeline, as previously described(Cheng et al., 2015). Briefly, reads were filtered based on quality, mapped to NCBI b37 genome using BWA–MEM, coordinate sorted, duplicates marked, recalibration, IndelRealigner using GATK and finally variant discovery using MuTect. Variants were annotated based on its entry in NCBI-dbSNPs (<http://www.ncbi.nlm.nih.gov/snp>), 1000G project (<http://www.1000genomes.org/>) and COSMIC (<http://cancer.sanger.ac.uk/cosmic>). Filtered variants were manually reviewed on IGV. We created mutational OncoPrint plot using online cBioPortal website (<http://www.cbioportal.org/oncoprinter.jsp>) on our 29 LCs dataset. The MSK-IMPACT mutational dataset is available on MSKCC cBioportal (under Pulmonary NET, Tang CMO5837).

**RNA sequencing and analysis:** RNA Library preparation and RNA sequencing was done by MSKCC Genomics Core Laboratory using Illumina HiSeq with (2 x 100 bp paired end reads) to a minimum depth of ~ 50 million reads for each sample. We performed standard RNAseq data analysis as previously described in chapter 2 under “RNA sequencing and data analysis” and (Chan et al., 2018)(Conesa et al., 2016). The R package DeSeq2(Love et al., 2014) was used to identify find differentially expressed genes between three subtypes, between ACs vs TCs , between subtype LC1 ACs vs TCs

and between LC2 ACs vs TCs. We used Benjamini & Hochberg (BH) corrected p-value ( $< 0.05$ ) and  $\log_2$ foldchange of  $\geq 2$  to filter differentially expressed genes.

**Subtype validation using independent lung carcinoid dataset:** To validate our novel molecular subtyping, we used Fernandez-Cuesta L et al (Fernandez-Cuesta et al., 2014) gene expression and mutational dataset of lung carcinoids (n=65). The gene expression and mutational data was downloaded from supplement data files (<https://www.nature.com/articles/ncomms4518#supplementary-information>) reported in ref.9. This gene expression data was reported as transcript expression instead of gene expression. We used collapseRows (Miller et al., 2011) on transcript expression to convert to respective gene expression using MaxVariance option. Gene expression for LCs signature (top 100 variant genes from our 30 LCs dataset) was fetched from this RNAseq dataset. Unsupervised clustering and PCA analysis on this dataset was performed and *MEN1* mutation data was overlaid in respective samples. Boxplot was created for LCs biomarker (*ASCL1* and *S100*) and heatmap for *HNF1A*, *HNF4A* and *FEV* using gene expression.

**DNA Methylation Analyses:** DNA extracted from LC samples and interrogated for DNA CpG methylation using the Illumina 450K array platform (Illumina Inc. San Diego, CA). All the analysis were performed using ChAMP (Morris et al., 2014) version 2.6.0 open source software implemented in R/Bioconductor and followed same steps as mentioned in chapter 2 under “Genome wide 450k analysis” section. Subtype specific differentially methylated CpG probes (DMP) and CpG island were identified using COHCAP (Warden et al., 2013) using all default parameter.

**COHCAP analysis on matched LCs methylation and expression:** We performed an integrative analysis on methylation beta values (CpG island) and gene expression values for matched methylation and expression (n=18) samples using R/Bioconductor package COHCAP(Warden et al., 2013) with default parameter. We pursued COHCAP protocol for 450K array to identify differentially methylated CpG probes followed by differentially methylated CpG island using delta beta value  $>0.2$  and corrected FDR P-value  $< 0.05$ . We focused on probes present at TSS1500/200 and 1<sup>st</sup> Exon for subsequent analysis. Differentially methylated CpG sites were annotated to respective CpG Island (UCSC 450k) and average of this island were used for comparisons. Next, we integrated differentially methylated events with gene expression and found subtype specific genes anti-correlated at methylation and gene expression level (default parameter for COHCAP.integrate.avg.by.island with FDR p-value  $<0.01$ ).

**Immunohistochemical staining:** Immunohistochemical staining was performed using commercially available antibodies at optimal dilutions as follows: ASCL1 (a-MASH1) [monoclonal, 1:300, BD] and S100 [monoclonal, 1:4000, BG]. Tissue Microarray (TMA) constructed from 173 independent lung carcinoid tumors to check for clinical correlates.

## **Results**

### **Patient cohort, clinical annotations and mutational profile of lung carcinoids**

We analyzed 30 randomly selected and histologically confirmed, well-differentiated LCs (17 TCs and 13 ACs). Most specimens were from pulmonary lobectomy with lymph node detection. Tumor locations, i.e. peripheral versus central (endobronchial), were assessed by the combination of radiographic reveal and the pathologic observations. Fifty-four

percent (7/13) of ACs had either lymph node or distant metastasis while 6% (1/17) of TCs had local lymph node metastasis. The 5-year disease specific survival was 89% and 55% for TC and AC, respectively. Clinical information and features are presented in Supplementary File 1. In addition, a tissue microarray (TMA) containing 173 cases of lung carcinoid had been prepared previously (Rekhtman et al., 2018) and used for biomarker validation in this study.

We performed targeted sequencing of a 354-cancer gene panel (MSK-IMPACT (Cheng et al., 2015)) on 29 LCs. The mutated genes were enriched for those implicated in covalent histone modification/chromatin remodeling and found in 10 samples (*MEN1* (13.8%), *ARID1A* (10%), *KMT2A* (3%), *KMT2C* (7%), *KMT2D* (3%) and *SMARCA4* (3%)) recapitulating the results from a previous study where they also show mutations in genes involved in histone methylation leading to trend of global decrease in H3K9me3 and H3K27me3 methylation (Fernandez-Cuesta et al., 2014). We also found mutations in DNA repair (17.2%) pathways (Figure 1a) (Supplementary File 2). Mutations were not detected in the 354-cancer gene panel for 13 LCs samples. Mutations in *MEN1*, the most frequently mutated gene, were found in four samples (4 ACs) and four of these mutations had variant allele frequencies higher than 70% indicating loss of heterozygosity (LOH) (Supplementary Figure 1). One sample (Lu-Aty9) has two *MEN1* mutations (an in-frame deletion and a missense substitution) a few bases apart on the same copy of *MEN1* (Supplementary File 2). The *ARID1A* gene is mutated in three samples with LOH occurring in one of the three samples. Using variant allele frequencies and LOH status of *MEN1* and *ARID1A*, we found median tumor purity to be 91% (Supplemental File 1),



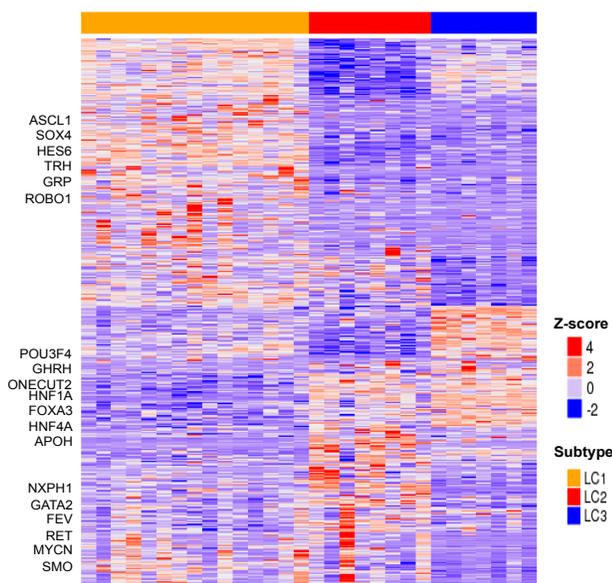
according to their gene expression and DNA methylation pattern. Orange = subtype 1 (LC1), Red = subtype 2 (LC2), Blue = subtype 3 (LC3). Column represents sample and row represents gene name with mutation frequency. Gene expression (n=30) and DNA CpG methylation (n=18) analysis revealed novel three Lung Carcinoid subtypes using unsupervised clustering and PCA analysis: b) Heatmap of unsupervised clustering of top 100 variably expressed gene across all samples. c) PCA of top 3000 variably expressed genes. d) Heatmap of unsupervised clustering of top 500 variable methylated CpG probes. e) PCA on top 3000 variable methylated CpG probes.

The most recurrent CNV are single copy deletions in *FANCA* (17%), *FAT1* (10%), *MEN1* (7%), *ATM* (17%), *SDHD* (17%), and *CHEK1* (17%), many of which reside on chr11q. We did not observe changes in the transcription levels of these genes with hemizygous deletions in comparison to wild type samples. There are 18 samples (4 ACs and 14 TCs) with normal karyotype, 6 samples (4 ACs and 2 TCs) with nearly normal karyotype (aneuploid for only one or two different chromosomes), and 6 samples (5 ACs and 1 TCs) with aneuploidy in more than two different chromosomes in our dataset (Supplementary file 2). We did not find any known pathogenic germline mutations in the panel of cancer-associated genes in our samples. *TP53* and *RBI* genes were not mutated in this cohort, unlike high-grade lung NETs and SCLC.

### **Transcriptome and methylome profiles reveal three distinct subtypes**

We performed RNA sequencing on 30 LCs (including 13 atypical and 17 typical samples) and DNA methylation analysis on 18 LCs (12 of the 30 samples did not have sufficient material for analysis). Unsupervised clustering and principal component analysis (PCA) on the top 3000 variable (Var3000) genes showed three distinct clusters (Figure 1b and 1c). These clusters are robust when different number of top variable genes was used for clustering (Supplementary Figure 2). We named these subtypes LC1, LC2 and LC3 (Lung Carcinoid 1, 2 and 3). Pearson correlation heatmap on Var3000 genes

shows three blocks representing the three evident subtypes (Supplementary Figure 3). The top 100 variable genes (Supplementary Figure 4) across all LCs show enrichment for gene ontologies related to hormonal secretions, endogenous stimulus, wound healing and developmental processes (Supplementary Table 2). Genome wide expression analysis revealed greater similarity between LC2 and LC3 as compared to LC1 (Figure 2, heatmap of differentially expressed genes between three subtypes, Supplementary File 3).



**Figure 3-2.** Heatmap of differentially expressed genes between LC subtypes. Supervised analysis on LC subtypes reveals differential expression of transcription factor and neuropeptide (some are highlighted on the left side of the heatmap). Heatmap expression level is in Z-score.

To investigate the epigenetic profiles of LCs (n=18), we used Illumina 450K chip array to assay DNA methylation. Unsupervised clustering and principal component analysis of top 3000 variable CpG sites revealed three subtypes in complete agreement with the gene expression based subtypes (Figure 1c and 1d). These data revealed distinct methylation sites between the three subtypes. Consistent with gene expression, we also observed greater similarity of DNA methylation levels for LC2 and LC3 subtype when compared to LC1. The three grouping of subtypes was robust and reproducible using different

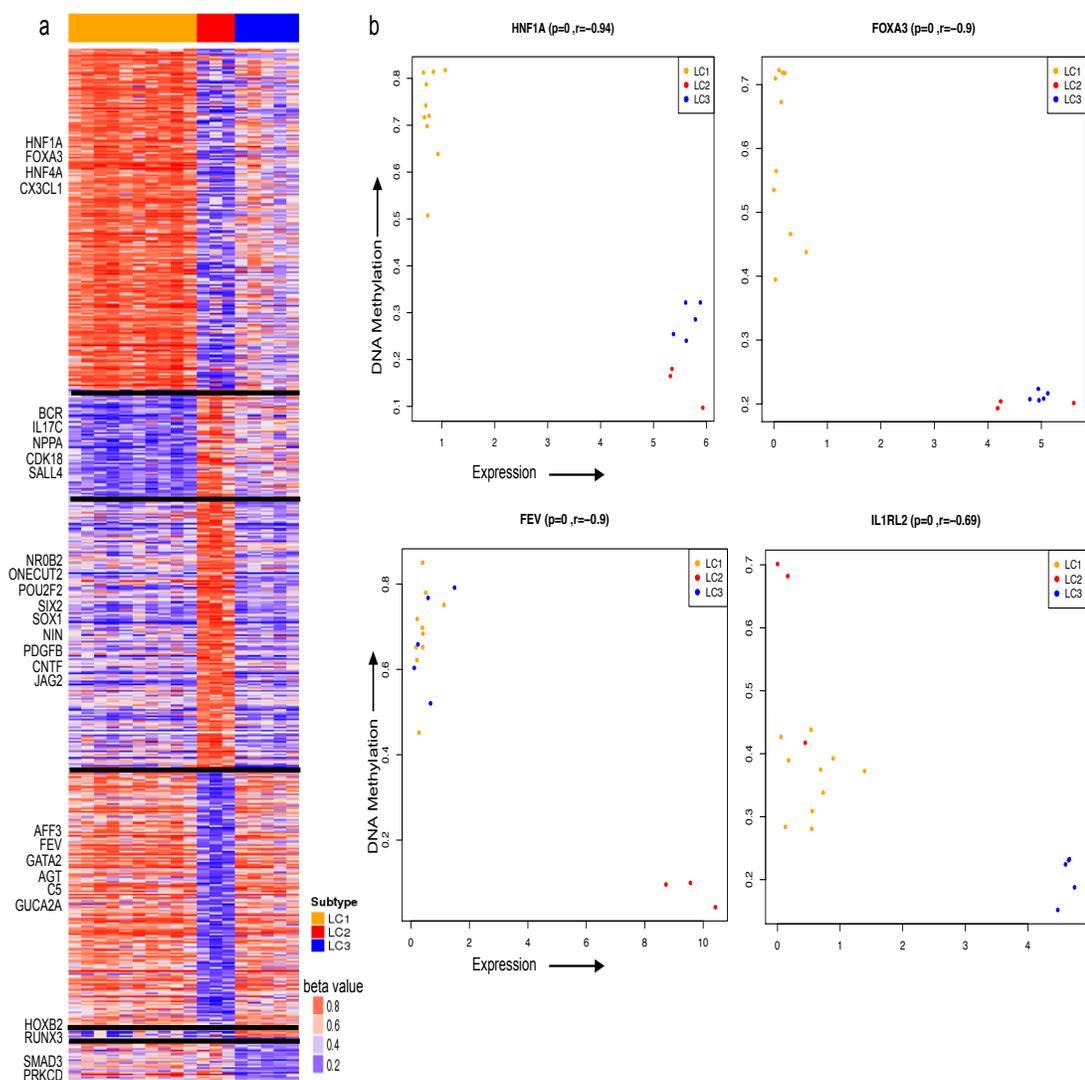
numbers of top variable CpG sites (Supplementary Figure 5). Total genome-wide DNA methylation level is not different in the three subtypes (Supplementary Table 3).

### **Subtype-specific molecular characterization of lung carcinoids**

We investigated gene expression and DNA methylation (CpG site and CpG island) profiles to determine subtype-specific molecular alterations (see Materials and Methods section). Genes up-regulated in LC2 and LC3 as compared to LC1 are enriched for having the transcription factor motifs for *HNF1* (FDR q-value < 0.001) and *HNF4* (FDR q-value < 0.001)(Supplementary File 3). This is in agreement with the observed high gene expression and DNA hypomethylation of *HNF1A*, *FOXA3* and *HNF4A* in LC2 and LC3 as compared to LC1 (Figure 3a and 3b, Supplementary Figure 6). In fact, many of the most highly expressed genes (*APOH*, *GC*, *HAO1*, *G6PC*, *TM4SF4*, *PKLR*, *UGT2B17*, *CDH1*, and *SERPINA1/2/6*) in LC2 and LC3 are targets of these hepatocyte nuclear factors. Cancer hallmark gene set enrichment analysis shows complement and coagulation, xenobiotic, retinol and bile acid metabolism to be significantly up-regulated in LC2 and LC3 as compared to LC1 (Supplementary File 3). However, we also found many subtype specific TFs that are differentially expressed between LC2 and LC3 (*FEV* is more highly expressed in LC2 whereas *POU3F4* has higher expression *in LC3*)(Figure 3-2).

*MEN1* gene is shown to regulate several members of the HOX gene family(Yokoyama et al., 2005). Indeed, the LC2 subtype, which included all of the *MEN1* mutant samples, has low expression of *HOXB2/3/4/5/6* genes as compared to LC1 and LC3

(Supplementary Figure 7). We found several key neuro-peptides that are highly and significantly differentially expressed in the three subtypes. *TRH* (10.7 log<sub>2</sub>FC and BH corrected p-value 4.16E-49), *GRP* (7.98 log<sub>2</sub>FC and BH corrected p-value 7.83E-31) and *NPPA* (3.39 log<sub>2</sub>FC and BH corrected p-value 1.44E-14) are highly expressed in LC1 only. *NXPPI* (3.90 log<sub>2</sub>FC and BH corrected p-value 2.33E-05) and *GHRH* (5.29 log<sub>2</sub>FC and BH corrected p-value 8.62E-09) are only highly expressed in LC2 and LC3 respectively.



**Figure 3-3.** Subtype specific molecular characterization of gene expression and DNA methylation profiles. a) Heatmap of differentially methylated CpG sites (probes from TSS1500,

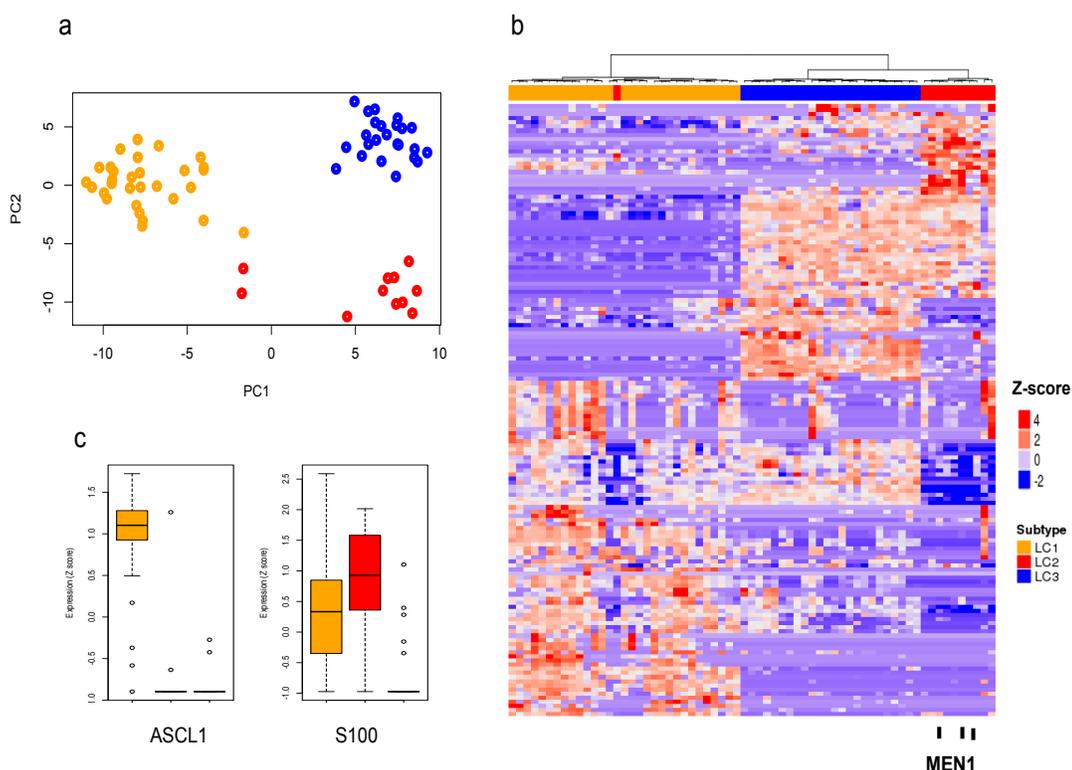
TSS200 and first exon) of genes among the three LC subtypes. Some genes with altered gene expression and CpG sites are highlighted on the left of heatmap. Dark black line represent subtype specific blocks b) Anti-correlation of gene expression and respective CpG island methylation (18 matched samples) for *HNF1A*, *FOXA3*, *FEV* and *ILRL2* across three subtypes. Each plot represents gene expression on x-axis and average CpG island beta value on y-axis along with Pearson correlation ( $r$ ) and p-value ( $p$ ) on top of the plot.

In addition to subtype-specific, differentially expressed genes, we integrated subtype-specific DNA methylation at CpG sites and CpG islands (see method section). We focused on CpG sites between 1500 bps and 200 bps upstream to the transcription start site (TSS) and in the first exon, which have been shown to inversely correlate with gene expression (Brenet et al., 2011). Figure 3a shows subtype-specific differentially methylated CpG probes (DMP) and these subtypes-specific DMPs are inversely correlated with neighboring gene expression. We found 75 genes with expression to be significantly anti-correlated with respective CpG island methylation level (FDR P-value < 0.01) (Supplementary File 4). *HNF1A* and *FOXA3* are hypermethylated and low-expressed in LC1. *FEV*, *GATA2* and *PROCR* are hypomethylated and highly expressed in LC1. *SOX1* is hypermethylated and low-expressed in LC2. *SIX2*, *ONECUT2* and *ILRL2* are hypomethylated and highly expressed in LC3 (Figure 3b). Many of these observations suggest further mechanistic studies but there are currently no appropriate lung carcinoids cell lines or animal models that can be used.

### **Independent validation of lung carcinoid classification**

We validated our novel classification and gene expression of biomarkers using published lung carcinoid genomic data from (Fernandez-Cuesta et al., 2014) which include genome/exome and RNA sequencing of 65 samples (56 TCs, 6 ACs and 3 carcinoids).

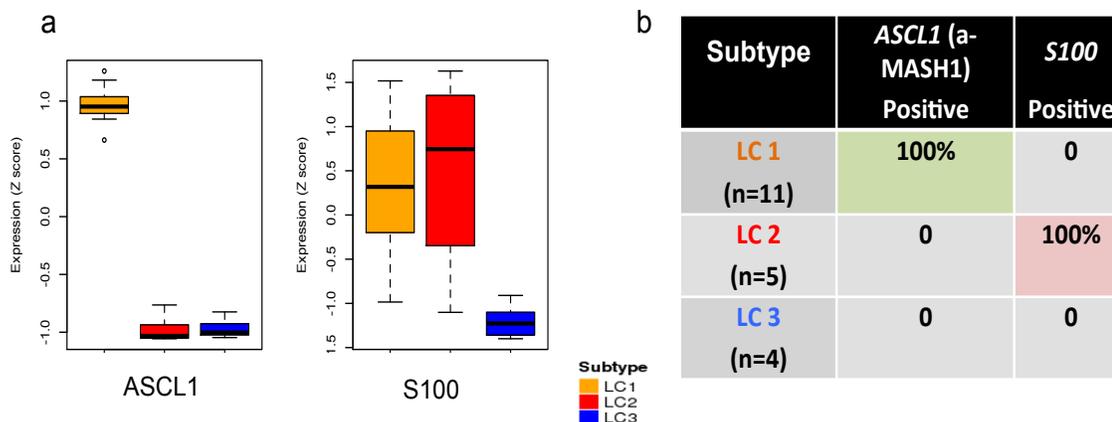
Using our gene signatures derived from the top 100 most variable genes (Supplementary File 5) across lung carcinoids, we found three distinct subtypes using unsupervised clustering and PCA that are consistent with the subtypes identified from our data (Figure 4a,b) (Supplementary Figure 8). Moreover, all *MEN1* mutated lung carcinoids in their study are found exclusively in LC2 (Figure 4b). In addition, we found *HNF1A* and *FOXA3* are more highly expressed in LC2 and LC3 whereas *FEV* is more highly expressed in LC2 consistent with our data (Supplementary Figure 9).



**Figure 4-4.** Validation of novel classification of LC on an independent collection of LCs (Fernandez-Cuesta et al., 2014) a) and b) PCA and Heatmap of hierarchical clustering of gene expression of LCs from (Fernandez-Cuesta et al., 2014) using our top 100 gene set signature show three distinct subtypes LC1 (orange), LC2 (red) and LC3 (blue). Black sticks represent samples with *MEN1* mutations and they are all found in subtype LC2. c) Boxplot of *ASCL1* and *S100* gene expression from Fernandez-Cuesta et al. is consistent with LC subtypes.

***ASCL1* and *S100* are novel biomarkers for lung carcinoid subtypes**

We selected genes with distinct subtype-specific expression to test for use as biomarkers. *ASCL1* encodes a transcription factor that plays a role in neuronal differentiation and proliferation (Castro et al., 2011), neuroepithelial bodies formation (Guha et al., 2012) and is a lineage specific oncogene for high-grade neuroendocrine lung cancer (Borromeo et al., 2016). *ASCL1* is significantly highly expressed in LC1 along with its transcriptional targets (Figure 5a and 4c) (Supplementary Figure 10). *S100*, a family of proteins containing two EF-hand calcium-binding motifs, is implicated in tumor progression and poor prognosis (Chen, Xu, Jin, & Liu, 2014). Its gene expression levels are significantly higher in subtype LC2 (Figure 5a and 4c). We performed IHC staining of *ASCL1* and *S100* to use as biomarkers. *ASCL1* stained positively only for LC1 samples (n=11) and *S100* stained positively only for LC2 samples (n=5). Both of these genes stained negatively for LC3 samples (n=4) (Supplementary Table 4).



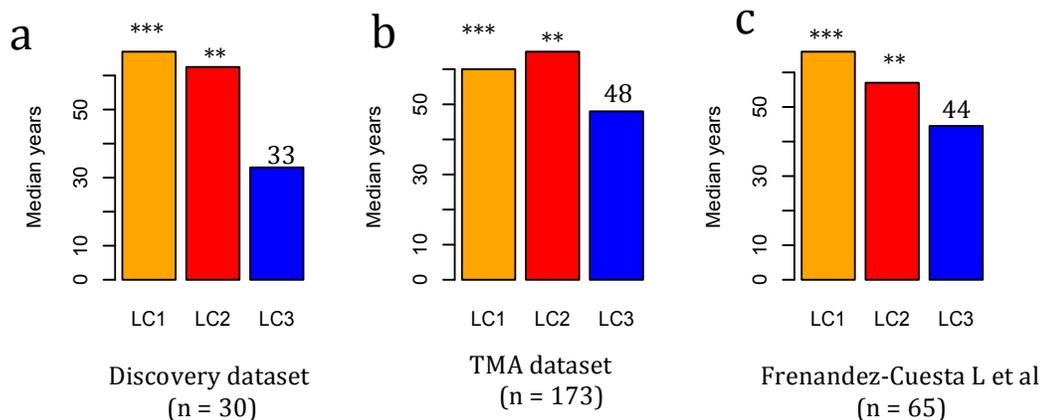
**Figure 3-5.** Gene expression and immunohistochemistry for *ASCL1* and *S100* biomarker genes. a) Boxplot of *ASCL1* and *S100* Gene expression. b) IHC staining results for *ASCL1* and *S100* in LC samples: LC1 (n=11), LC2 (n=5) and LC3 (n=4). Supplementary table 4 has IHC results for all samples.

Additionally, we performed *ASCL1* and *S100* IHC staining on a panel of 173 independent lung carcinoids tissue microarray (TMA) (Supplementary File 6). *ASCL1* positive and

*S100* negative samples (n=54) were designated TMA-LC1. ASCL1 negative and S100 positive samples (n=15) were designated TMA-LC2. ASCL1 negative and S100 negative samples (n=71) were designated TMA-LC3. Fifteen percent of the TMA samples stained positive for ASCL1 and S100, which is not represented in our 30 LC samples for which we did gene expression analysis.

### **Lung carcinoid subtypes have distinct clinical phenotypes**

The three subtypes of lung carcinoids have distinct clinical phenotypes. Subtype LC1 is enriched for peripheral lung (p-value <0.003 in 30 LC dataset; p-value < 0.002 in TMA-LC1) while subtype LC3 is found predominately at endobronchial lung (p-value < 0.054 in 30 LC dataset; p-value < 3.8e-5 in TMA) (Figure 1a box)(Supplementary file 1). Subtype LC3 has significantly younger age of diagnosis (median age of 33, 44.5, and 48 years in 30 LC dataset, Fernandez-Cuesta et al, and TMA respectively) than LC1 (median age of 67, 66, and 60 years respectively) and LC2 (median age of 62.5, 57, and 65 years respectively)(Supplementary Figure 11a and b). LC1 subtype was enriched for female patients (p-value < 0.007 in 30 LC dataset, p-value < 1.4e-5 in TMA) but not for LC2 or LC3.



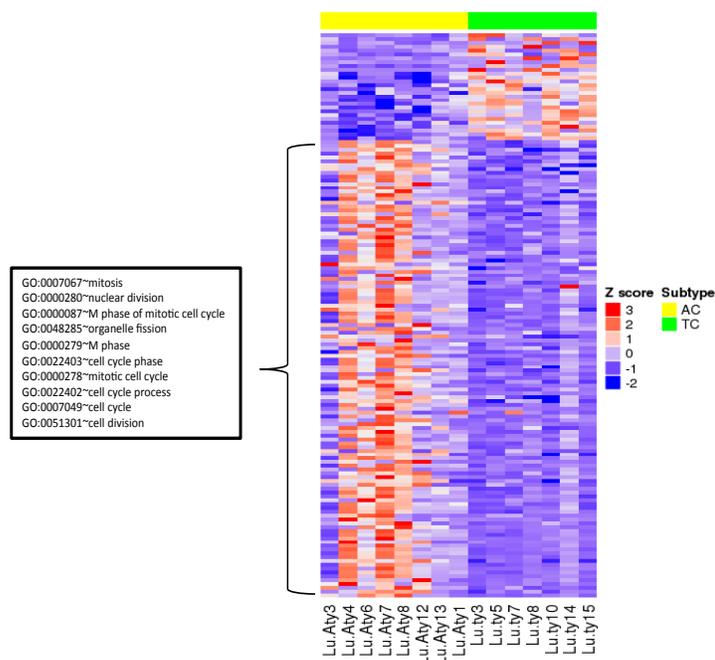
Dataset	LC1 median age	LC2 median age	LC3 median age
a) Discovery dataset	67	62.5	33
b) TMA dataset	60	65	48
c) Frenandez-Cuesta L et al	66	57	44

**Supplementary Figure 11:** Clinical features of lung carcinoid subtypes with median age of diagnosis. a) Our dataset , b) Tissue Microarray dataset from MSKCC and c) Frenandez-Cuesta L et al., 2014 dataset for median age across all three subtypes. \*\*\* Represent pval < 0.0001 and \*\* represents < 0.001 significance level.

### Cell cycle and mitotic genes are highly expressed in atypical carcinoids

Pathologically, atypical carcinoids are more aggressive and have a higher mitotic index in comparison to TCs. To find the gene signature responsible for these features of ACs, we compared ACs (n=13) and TCs (n=17) from our 30 LC cohort. Surprisingly, we did not find cell cycle or mitosis-related genes to be differentially expressed. We then controlled for LC subtypes and compared ACs (n=8) and TCs (n=7) from subtype LC1 and found differentially expressed genes (Figure 6) were then enriched for mitotic and cell cycle related pathways with high expression in the ACs (Supplementary File 7). Of the eight AC tumors, three with highest gene expression signature for mitotic/cell-cycle pathway

have metastases or recurrences while only one of the five with low gene expression signature have recurrence or metastases (Figure 6). We also observed high aneuploidy in the ACs with high gene expression signature of mitotic/cell-cycle pathway. We did the same analysis for ACs and TCs from LC2 and did not find any significant gene signatures, which could be due to the small sample size of LC2.

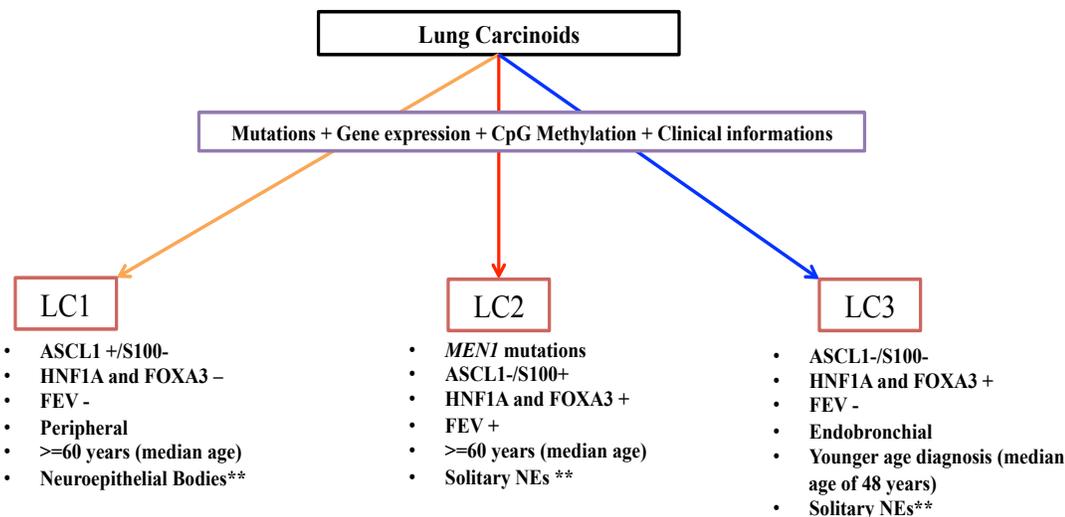


**Figure 3-6.** Heatmap of differentially expressed genes between ACs and TCs within LC1 subtypes. Upregulated genes in ACs are significantly enriched for cell cycle and mitotic gene ontologies.

## Discussion and Conclusions

We identified three novel molecular subtypes of lung carcinoids with distinct clinical features on the basis of gene expression, DNA methylation and mutational profile (Figure 7). Integrative analysis of gene expression and DNA methylation identified subtype

specific transcriptional profiles of key differentiation transcription factors (*ASCL1*, *HNF1A*, *FOXA1*) and their downstream target genes.



**Figure 3-7.** Subtype-specific molecular events for LCs based on mutation, gene expression, DNA methylation and clinical informations. \*\* Represents nature of neuroendocrine cells based on gene expression of *ASCL1/ROBO1/SLIT* (Supplementary Figure 12). The age mentioned in this schematic figure is median age.

Mutational analysis revealed recurrent mutations in chromatin remodeling genes found in all subtypes with exception of *MEN1* mutations occurring only in subtype LC2 tumors. Importantly, we found mutations in DNA repair genes in 17% of our LC samples. Interestingly, germline mutations in DNA repair genes *MUTYH*, *CHEK2*, and *BRCA2* were found at larger than expected proportions in clinically sporadic pancreatic neuroendocrine tumors which is thought to share embryologic lineage with the pulmonary counterpart(Scarpa et al., 2017). Subtype LC3 have younger age of diagnosis and are predominantly endobronchial, whereas subtypes LC1 are predominantly found in peripheral regions of the lung. The younger age of diagnosis for LC3 by 15-20 years as compared to LC1 and LC2 may be due to earlier diagnosis from the clinically

symptomatic tumors located in the central lung as compared to asymptomatic tumors at the peripheral lung or suggest a possible germline mutation predisposing to LC3. However, we were not able to detect any pathogenic germline mutations in the panel of cancer associated genes used in the MSK IMPACT testing for any of the lung carcinoids.

Our classification and gene expression biomarkers were validated in 65 additional lung carcinoid samples from (Fernandez-Cuesta et al., 2014). Using our subtype classification, we found gene signature for cell cycle and mitotic processes activated in ACs as compared to TCs of the LC1 subtype and those ACs with the high gene signature have worst outcome. This gene signature may potentially serve as a diagnostic and prognostic biomarker to differentiate malignant from more benign ACs from subtype LC1. This gene signature is specific to LC1 and would not have been found from comparing ACs to TCs from all lung carcinoids.

The three distinct subtypes of LCs identified may arise from distinct population of cells or the genetic or epigenetic alterations acquired in tumorigenesis may give rise to the distinct gene expression profiles. Neuroendocrine cells (NECs) in the lung occur as randomly scattered solitary cells known as Kultchitsky (K) cells and as clusters known as neuroepithelial bodies (NEB). NECs have many functions in lung development and as chemoreceptors in the airway (Van Lommel, 2001). The development and cell lineage of lung NECs are not fully understood. Solitary NECs migrate distally and form intermodal NEBs along the airway and nodal NEBs at bifurcation points in the airway (M. Noguchi, Sumiyama, & Morimoto, 2015). Gene expression studies identified selective expression

and protein localization of *DLL3*(Verckist et al., 2017) and *GAD1/2*(Schnorbusch et al., 2013) only in NEBs. *ROBO1/2* is required for the ability of NECs to cluster into NEBs and for maintenance of NEBs after clustering(Branchfield et al., 2016). *SLIT1/2* are ligands to ROBO and the SLIT-ROBO signaling complex is required to drive NEC clustering through cellular attraction(Branchfield et al., 2016). Genetic ablation of *ASCL1* prevents the formation of NEB(Guha et al., 2012). We found higher expression of *ASCL1*, *DLL3*, *GAD1/2*, *ROBO1/2* and *SLIT1* in LC1 compared to LC2 and LC3 (Supplementary Figure 12), suggesting that LC1 carcinoids originate from NEB cells and LC2 and LC3 from K cells from the peripheral and endobronchial lung respectively. Alternatively, genetic and epigenetic alterations producing three distinct subtypes from a homogeneous cell population is less likely because besides the *MEN1* mutations found exclusively in LC2, there were no recurrent genes mutated that were subtype specific.

Our molecular classification introduces three subtypes of lung carcinoids with distinct clinical phenotypes. This can refine and complement the WHO classification of lung carcinoids into typical or atypical carcinoids and help diagnose ambiguous cases of lung carcinoids from the more malignant LCNEC and SCLC. The stratification of LCs into distinct molecular subtypes will help with future study of prognosis and treatment options.

## Chapter 4

### Genomic Analyses Identify Two Distinct Cell-Of-Origin Subtypes Of Small Intestine

#### Neuroendocrine Tumors

#### *Manuscript under Preparation*

**Saurabh V. Laddha**<sup>1</sup>, Kenneth Robyzk<sup>2</sup>, Laura H. Tang<sup>2\*</sup>, Chang S. Chan<sup>1,3\*</sup>

<sup>1</sup>Rutgers Cancer Institute of New Jersey, Rutgers University, New Brunswick, NJ, USA

<sup>2</sup>Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York NY, USA

<sup>3</sup>Rutgers Cancer Institute of New Jersey, Rutgers University and Department of Pediatrics, Robert Wood Johnson Medical School, New Brunswick, NJ, USA

#### **Contributions:**

S V. L: computational design, computational data analysis and manuscript preparation, K

R: experimental design and specimen preparations, L H. T and C S.C: study design and

data analysis.

## Genomic Analyses Identify Two Distinct Cell-Of-Origin Subtypes Of Small Intestine Neuroendocrine Tumors

### Abstract

Small intestine neuroendocrine tumors (SI-NETs) are rare, slow growing, well differentiated and the most common neoplasms of small bowel. SI-NETs are proposed to originate from enterochromaffin (EC) cells but their cell-of-origin and development remains undetermined. We performed exome sequencing (n=20) and mRNA sequencing (n=29) on well-differentiated SI-NETs. We identified loss of chromosome 18 (chr18) in 75% (22 of 29) of samples. We did not find any mutation on chr18 genes. Unsupervised clustering and principle component analysis on gene expression profile showed two robust molecular subgroups (SINET-A and SINET-B) with distinct cell-of-origin signature. SINET-B subtype uniformly expresses *TPHI* and *REG4* gene (*TPHI*+/*REG4*+), which are the key markers for major EC cells, whereas SINET-A exhibit gene signature for rare neuroendocrine cells that are *TPHI*-/*REG4*-. SINET-A (n=3) has chromosome 18 wildtype copy while SINET-B mostly lose one copy of chromosome 18 (22 of 26 samples). Gene expression profile of two potential biomarkers (*LMX1A* and *ONECUT2*) was found to stratify the two subtypes. This molecular classification of SI-NETs will enable comparison to the proper cell of origin to improve the study of their molecular mechanisms and potentially improve clinical management.

### Introduction

SI-NETs are slow growing and well differentiated (serotonin producing). They are the most common type of gastrointestinal endocrine tumors with one case per 100,000

annually and account for 25% of all NETs(Karpathakis et al., 2016). SI-NETs are generally small in size (frequently < 2cm) and slow growing (Ki67 index frequently < 2%) but for metastatic tumors, the 5-year survival rate is 50-65%(Walter et al., 2018; Yao et al., 2008). SI-NETs are proposed to originate from enterochromaffin cells(Lundqvist & Wilander, 1987) but their origin and development remains undetermined.

Molecular studies identified only a few recurrent alterations for SI-NETs. One copy loss of chromosome 18 (chr18) is the most frequent (60% to 80%) genetic alterations in SI-NETs (Francis et al., 2013), however, the clinical significance of chr18 loss is unknown. SI-NETs have low somatic mutation burden like other NETs and most of the tumors are diploid with no genome instability except for one copy loss of chr18. Second most frequent (~8%) alteration in SI-NETs is inactivation mutations in *CDKN1B* (Cyclin dependent kinase inhibitor 1B) gene. However, tumors with *CDKN1B* inactivation mutations have no difference in clinical survival or phenotypes(Karpathakis et al., 2016) as compared to wild type and *CDKN1B* is proposed to act as a haploinsufficient tumor-suppressor gene in SI-NETs(Francis et al., 2013).

Recent single cell(Haber et al., 2017) surveys of mice small intestine revealed at least 12 distinct neuroendocrine cell populations secreting a variety of hormones such as serotonin, ghrelin, secretin, proglucagon, somatostatin, neurotensin etc. Moreover, four of the 12 cells have expression of enteroendocrine precursor markers (*SOX4*, *NEUROG3* or *NEUROD1*) and the other eight represent mature enteroendocrine cells (EEC). Interestingly, interaction and plasticity between these cell types have been

established (Gribble & Reimann, 2016). The dominant population of EEC in small intestine is enterochromaffin (EC) cells and can be subdivided into EC and EC-Reg4 subgroups based on expression of *REG4* gene (Haber et al., 2017). Moreover, based on *TPHI* and *REG4* gene expression profile, these clusters/cells may be divided into two sub-classes: Major enterochromaffin cells which are *TPHI*<sup>+</sup>/*REG4*<sup>+</sup> and *TPHI*<sup>-</sup>/*REG4*<sup>-</sup> (rare) neuroendocrine cells. The pathways and molecular process regulating these EC cell function and growth remains unknown.

In this study, we performed exome sequencing (n=20) and mRNA sequencing (n=29) and found two distinct cell-of-origin subtypes of SI-NETs. We also identified two potential biomarkers (*LMX1A* and *ONECUT2*) to stratify these subtypes based on gene expression. This molecular subtyping of SI-NETs provides deeper insights into the cell-of-origin and biomarkers for clinical research.

## **Materials and Methods**

**Patient Data:** Retrospective and prospective reviews of 29 SI-NETs neoplasms were performed using the pathology files and institutional database at MSKCC with IRB approval. All patients were evaluated clinically at MSK institution with confirmed pathologic diagnoses, appropriate radiological and laboratory studies, and surgical or oncological management. Follow-up information was obtained for all cases. DNA extraction from microdissected tumor samples and normal adjacent tissues (if available) was performed using the same protocol as mentioned in chapter 2 under “Tissue Acquisition and nucleic acid extraction” section.

**Single Nucleotide variants, Indel and Copy Number variation analysis:** We followed the same protocol as mentioned chapter 3 method section for MSK-IMPACT data analysis for calling variants at single nucleotide and copy number variations. Briefly, exome seq fastq files were mapped to hg19 genome using BWA-MEM(H. Li & Durbin, 2009) and followed by post-alignment processing (mark duplicates, .bam file conversion etc). *Mutect* (Cibulskis et al., 2013) was used to call single nucleotide variants and ExomeCNVs(Sathirapongsasuti et al., 2011) algorithm was used to find focal and large CNVs.

**RNA sequencing and data analysis:** We followed the same protocol as mentioned in chapter 3 method section for RNAseq analysis for gene expression quantifications, unsupervised clustering, PCA, differentially expressed gene test, pathway analysis and visualization in R.

To investigate chr18 status for samples with no exome seq data, we did qPCR and used gene expression pattern. For gene expression, we summed TPM values of chr18 genes and ranked with known chr18 copy status from exome seq and qPCR. We normalized using median sumTPM for each sample and assigned chr18 copy based on positive (two copy) and negative (loss). This analysis is not binary but certainly helped to validate the known chr18 pattern and predicted for samples where exome seq data is not available.

## Results

### Loss of Chromosome 18 is frequent genetic alterations in SI-NETs

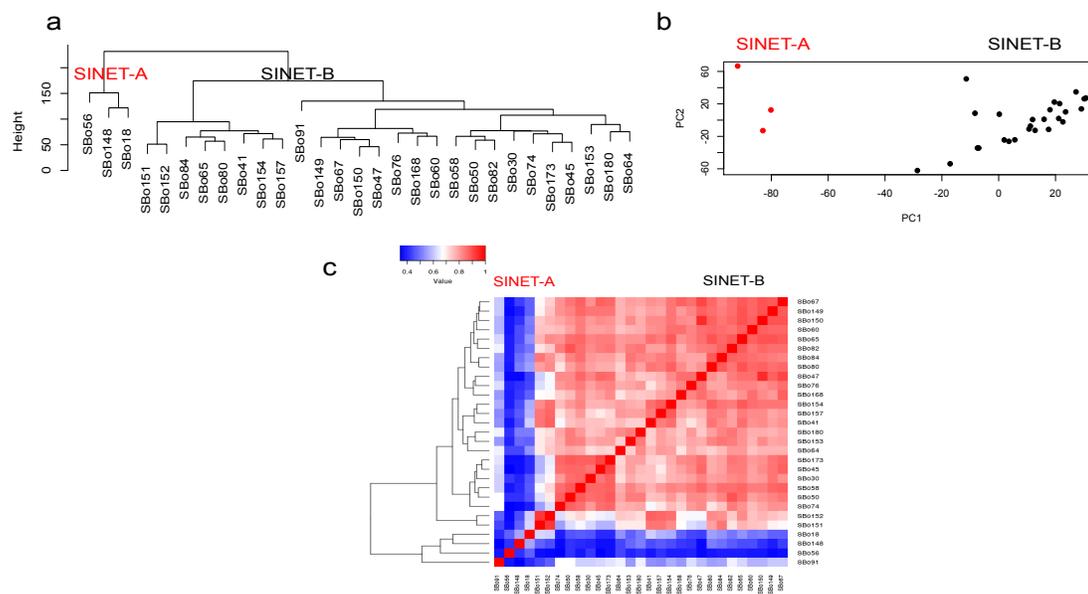
Copy number variations analysis identified 75% of SI-NETs with hemizygous loss of chr18 from exome sequencing and RNAseq dataset (see method section)(table 4-1), which is consistent with previous studies(Francis et al., 2013). However, we did not find any mutation for chr18 genes. Most chromosomes of SI-NETs from our dataset have two copies (except for loss of chr18). Other reported recurrent alteration we found was gain of chr5 (in 3 samples) and chr10 (in 3 sample). We did not find any mutations in *CDKN1B* gene, which reported to be altered in 8% of SI-NETs(Francis et al., 2013).

Subtype	Sample ID	chr18 Status
SINET-B	SB0149	LOSS
SINET-B	SB0151	LOSS
SINET-B	SB0152	LOSS
SINET-B	SB0153	LOSS
SINET-B	SB0154	LOSS
SINET-B	SB0157	LOSS
SINET-B	SB0168	LOSS
SINET-B	SB0173	LOSS
SINET-B	SB0180	LOSS
SINET-B	SB030	LOSS
SINET-B	SB045	LOSS
SINET-B	SB050	LOSS
SINET-B	SB058	LOSS
SINET-B	SB060	LOSS
SINET-B	SB064	LOSS
SINET-B	SB065	LOSS
SINET-B	SB067	LOSS
SINET-B	SB074	LOSS
SINET-B	SB076	LOSS
SINET-B	SB080	LOSS
SINET-B	SB082	LOSS
SINET-B	SB084	LOSS
SINET-B	SB0150	WT
SINET-B	SB041	WT
SINET-B	SB047	WT
SINET-B	SB091	WT
SINET-A	SB0148	WT
SINET-A	SB018	WT
SINET-A	SB056	WT

**Table 4-1.** Chromosome 18 status for SI-NET subtypes. Samples are arranged based on the subtype from gene expression and grouped as SINET-A and SINET-B class.

### Gene expressions reveal two distinct subtypes of SI-NET

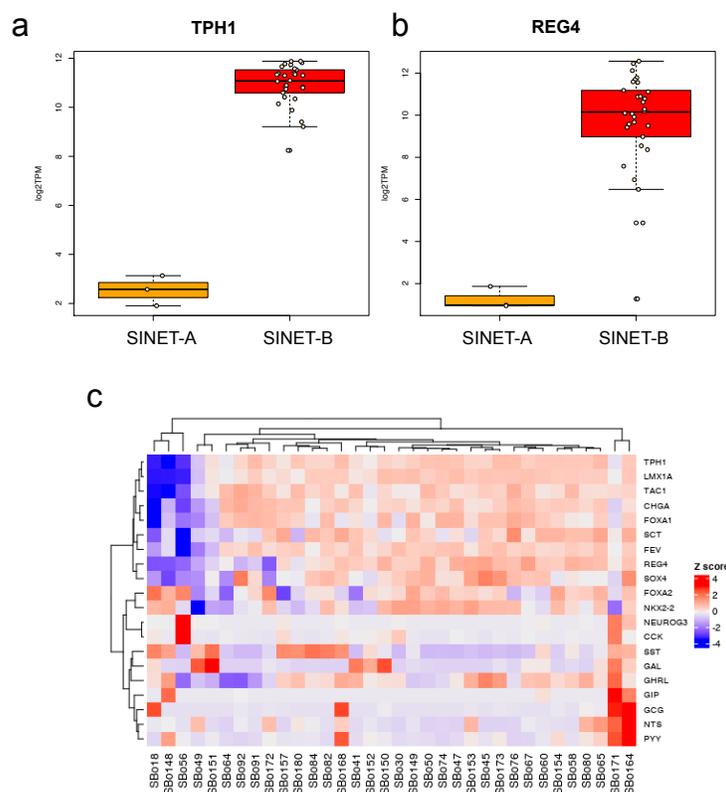
We performed RNA sequencing on 29 well-differentiated SI-NETs. Unsupervised clustering and principal component analysis (PCA) on the top 3000 variable (Var3000) genes showed two distinct clusters (Figure 1a and 1b). These clusters are robust when different number of top variable genes was used for clustering and PCA. We named these subtypes SINET-A and SINET-B. Heatmap of top 100 variant genes distinctly revealed genes with subtype specific expression patterns. To understand the gene expression differences between subtypes, we performed supervised differential expression test (DeSeq2) and identified pathways related to metabolism, cAMP and hypoxia processes. Interestingly SINET-A subtypes (n=3) are wildtype for chr18 copy number while most subtype B has loss of one copy of chr18 (22 of 26 samples). We also performed comparison within SINET-B subtypes for samples with and without chr18 loss and did not find any chr18 genes to be significantly differentially expressed.



**Figure 4-1.** Molecular subtypes of SI-NET using gene expression profile. a) Unsupervised clustering, b) PCA and c) correlation coefficient heatmap on top 3000 variants genes across all samples.

### Subtypes of SI-NET reveal distinct cell-of-origin signature

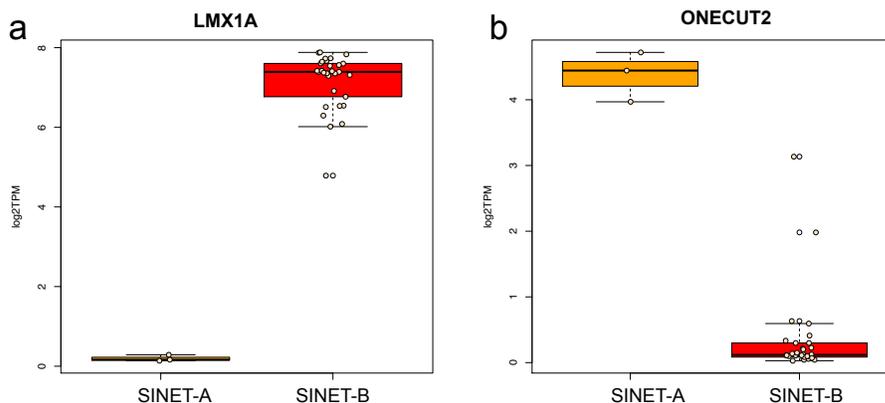
Recent single cell (Haber et al., 2017) analysis of small intestine from mice reveals at least 12 distinct neuroendocrine cell populations. Chromogranin A (CHGA) gene expression is distinctly high in all 29 SI-NETs confirming the neuroendocrine origin for these tumors. Interestingly, the SINET-B subtype uniformly expresses *TPHI* and *REG4* gene (*TPHI*<sup>+</sup>/*REG4*<sup>+</sup>), which is a key marker for major EC cells, whereas SINET-A exhibit gene signature for rare neuroendocrine cells, which is *TPHI*<sup>-</sup>/*REG4*<sup>-</sup> (Figure 4-2a and b). Based on this observation, we found that SINET-A may originate from *TPHI*<sup>-</sup>/*REG4*<sup>-</sup> rare neuroendocrine cells and SINET-B from EC cells.



**Figure 4-2.** Cell type specific marker for SI-NETs subtypes. Boxplot of *TPHI* (a) and *REG4* (b) gene for two subtypes, c) Heatmap of small intestine neuroendocrine cell-specific marker (obtained from (Haber et al., 2017)) on SI-NETs gene expression.

### ***LMX1A* and *ONECUT2* are potential biomarkers to distinguish SI-NET subtypes**

To further investigate the cell-of-origin for these subtypes, we selected genes with distinct cell-type-specific expression. *LMX1A* (LIM Homeobox Transcription Factor 1, Alpha) a transcription factor that positively regulates the insulin gene expression and one of the important downstream targets of *NKX2.2* transcription factor in EC cells of small intestine for production of serotonin (Gross et al., 2016). *LMX1A* gene is highly expressed in normal EC cells and SINET-B subtype (with log<sub>2</sub>FC of 9.82 and corrected pval < E<sup>-94</sup>) (Figure 4-3a). *ONECUT2* (One cut homeobox 2) is a transcription factor known that regulate genes involved in differentiation (Klimova, Antosova, Kuzelova, Strnad, & Kozmik, 2015). *ONECUT2* gene is not expressed in normal EC cells (Haber et al., 2017) but highly expressed in *TPHI-/REG4-* normal (rare) neuroendocrine cells of small intestine. We found *ONECUT2* gene expression level is significantly high in SINET-A subtype (with log<sub>2</sub>FC of 5.60 and corrected pval < E<sup>-05</sup>) (Figure 4-3b).



**Figure 4-3.** Gene expression boxplot of potential biomarkers (*LMX1A* and *ONECUT2*) for SI-NET subtypes.

## Discussion and Conclusion

We identified two distinct cell-of-origin signature molecular subtypes of SI-NETs. Genetic analysis revealed chr18 loss in ~75% of SI-NETs. We did not find any mutation in chr18 genes. Interestingly, we found SINET-A subtype members (n=3) have chr18 wildtype copy and mostly loss of chr18 (loss in 22 samples and wildtype in 4 samples) in SINET-B subtype members. We found metabolism, cAMP and hypoxia hallmark processes to be differentially expressed between these subtypes. Using gene expression signature from normal neuroendocrine cells of small intestine, we found that SINET-A may originate from *TPHI-/REG4-* neuroendocrine cells of small intestine and SINET-B from EC cells that are *TPHI+/REG4+*. Gene expression profile of two potential biomarkers, *LMX1A* and *ONECUT2*, was found to be sufficient to stratify the two subtypes.

Our molecular classification introduces two subtypes of SI-NETs with distinct cell-of-origin signature and novel biomarkers. This molecular classification and knowledge of cell-of-origin of SI-NETs subtypes may facilitate the understanding of their molecular mechanisms by comparing within appropriate cell population and improve biological and clinical research of SI-NETs.

## Chapter 5: Conclusion and Future Directions

### 5.1 Thesis Conclusions

The overall goal of this thesis was to provide greater insights into neuroendocrine tumor biology, molecular subtypes, pathogenesis, and cell-of-origins and to translate this knowledge to the clinic for the advancement of NETs treatment. The project discussed in chapter 2, 3 and 4 involved the integration of genomics dataset (DNaseq, RNAseq and DNA CpG methylation) to uncover the clinically relevant molecular subtypes and their cell-of-origin. Chapter 2 covers the two molecular subtypes of panNETs with distinct clinical phenotypes. Non-functional well-differentiated panNETs with mutations in *ATRX*, *DAXX* or *MEN1* have worst clinical outcome and resemble the gene expression profile of pancreatic alpha cells. We identified novel gene signature and biomarkers that differentiate panNETs genotype and gained an enhanced understanding of biology of panNETs from the cell lineage viewpoint. Chapter 3 demonstrates the molecular subtypes of lung carcinoids with distinct clinical phenotypes. We identified three novel molecular subtypes (LC1, LC2, and LC3) using genomics dataset. We found two biomarkers, *ASCL1* and *S100* that can stratify the three subtypes and performed IHC on 173 lung carcinoid tissue microarray with detailed clinical information. *MEN1* mutations were found to be enriched and exclusively in subtype LC2. Subtype LC1 and LC3 is predominately found at peripheral and endobronchial lung respectively. Subtype LC3 is diagnosed on average 10 years earlier than LC1 and LC2. While frequently mutated chromatin remodeling complex genes were previously found in lung carcinoids, we found that DNA repair genes are mutated in 17% of our samples. PanNETs and LCs are part of foregut and *MEN1* mutations is one of the common genetic alterations found to be

present in foregut NETs. In addition to this, we found common gene signature between PanNETs (with *ATRX*, *DAXX* and *MEN1* mutations subtype) and LCs (LC2 (*MEN1*) and LC3 subtypes) including high expression of the hepatocyte nuclear factors (*HNF1A*, *HNF4A*) and their transcriptional targets as well as the transcription factor *FEV*. This suggest there are commonalities between the subtypes of lung and pancreatic neuroendocrine tumors harboring *MEN1* mutations that may shed some light into some shared gene dysregulation and pathogenesis. Chapter 4 discusses the SI-NETs subtypes with distinct cell-of-origin signature. We identified two (SINET-A and B) subtypes of SI-NETs using gene expression and genetic dataset with distinct cell-of-origin signature. We identified one copy loss of chromosome 18 (chr18) in 85% (22 of 26) of subtype SINET-B while subtype SINET-A is diploid for chr18. We found that SINET-A subtype may originate from *TPHI*<sup>-</sup>/*REG4*<sup>-</sup> neuroendocrine cells of the small intestine and SINET-B from enterochromaffin (EC) cells, which are *TPHI*<sup>+</sup>/*REG4*<sup>+</sup>. Gene expression profile of two potential biomarkers (*LMX1A* and *ONECUT2*) was found to stratify the two subtypes.

While well-differentiated NETs of an organ may seemingly represent as a single clinical disease, they can be further characterized into different molecular subtypes based upon their cell lineage and the associated molecular genotype. Understanding the epigenetic and transcriptional dysregulation of NETs will require comparison to their proper cells of origin which may explain the unpredictable outcome of the disease and facilitate the development of unique and targeted therapeutic strategies.

## 5.2 Work in Progress and Future Directions

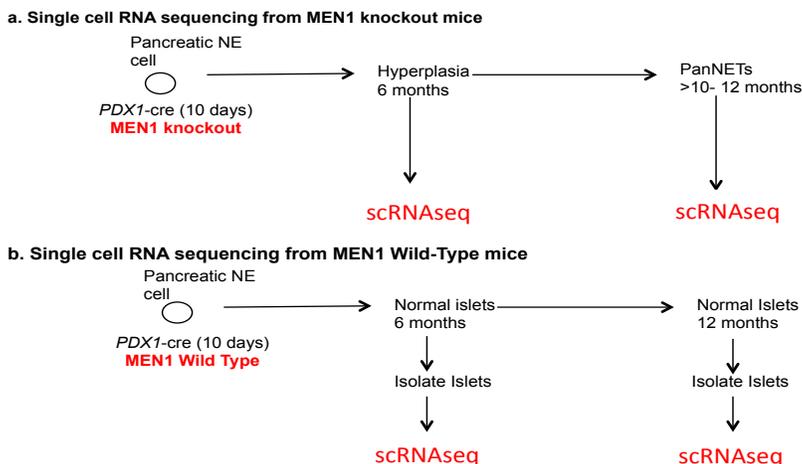
### 5.2.1 Investigate Single Cell RNA sequencing data of *MEN1* knockout PanNETs

We found two distinct molecular subtypes of human PanNETs using mutational, gene expression and DNA methylation profile (Chapter 2). PanNETs subtype with *ATRX*, *DAXX* and *MEN1* are associated with adverse clinical outcome as compared to PanNETs without these three mutations. Moreover, their gene expression profile resembles that of pancreatic alpha cells. This result is surprising, as pancreatic beta cell has been postulated to be the cell-of-origin for PanNETs. Further work is needed to characterize the cellular origin for PanNETs at single cell resolution. Particularly, 1) Identification of cell-of-origin and tumor heterogeneity for PanNETs, 2) *MEN1* PanNET tumorigenesis should be evaluated in comparison to appropriate cell type.

Unlike *MEN1* knockout mice models, *DAXX* and *ATRX* knockout mice models do not promote PanNETs tumorigenesis. Hence we selected *MEN1* knockout mice model to study PanNETs. We plan to do single cell RNA sequencing of *MEN1* knockout mice to:

- 1:** To characterize gene expression profiles of normal pancreatic neuroendocrine cells, pancreatic hyperplasia (6 months of age) and PanNETs (12 months of age) in *MEN1* knockout mouse model
- 2:** Identify the cell of origin and cellular heterogeneity of PanNETs

To conduct the studies at single cell resolution, we will use Drop-seq RNA sequencing technology, which is a droplet based high throughput method for RNA profiling. The following schematic diagram shows the model and single cell data generation strategies for our further study.



**Figure 5-1.** Schematic representation of *MEN1* knockout mice model for single cell RNA sequencing from normal islets, hyperplasia and PanNETs

### 5.2.2 Continued Characterization of *MEN1* PanNETs and identifications of genome wide binding sites of *MEN1*

The work presented in this thesis has uncovered genetic alterations in *MEN1* gene that are strongly enriched for NET subtypes. However, the functional mechanistic understanding of *MEN1* tumorigenesis in NETs is limited and complicated by the different cell-of-origin signatures between subtypes. *MEN1* encodes a histone methyltransferase and regulates gene expression. Further work should characterize the genome wide binding of *MEN1* in wildtype neuroendocrine cells as compared to *MEN1* mutated NETs. Particularly, which binding sites of endogenous *MEN1* are lost in *MEN1* mutated samples and which oncogenic genes are activated or which tumor suppressor genes are repressed during tumorigenesis.

### 5.3 Concluding Remarks

NETs are a rare, slow growing and benign neoplasm with malignant potentials. The classification of NETs (specifically for GI-NETs) has evolved over the last two decades and revisited many times. With greater awareness of NETs in clinic and improvement in diagnostic imaging techniques, there has been steady increase in incidence rate of NETs. Despite the increased incidence rate, the biological knowledge, molecular subtypes, mutational landscape, clinical diagnostic and prognostic markers are limited. Another rate limiting step in NET research is lack of proper cell of origin to compare tumors with. Recently, genomic studies decoded the mutational landscape of NETs and found most NETs have low mutation burden with few recurrent alterations. However, there are some key pathways frequently targeted in tumorigenic process and result in loss of *MEN1* function, inactivation of chromatin remodeling complex, activation of the PI3K/mTOR pathways, loss of chromosome 18 and altered telomere length. The work presented in this thesis was focused on deciphering ways to understand the underlying biology, molecular subtypes and cell-of-origin of NETs using integrative genomics dataset, with an objective to bridge the gap between basic genomic research and the clinic. Collectively, the primary research presented here should provide a stepping-stone to advance knowledge of NET biology and the molecular subtypes of NETs to improve clinical decision-making.

## References

- Anlauf, M., Enosawa, T., Henopp, T., Schmitt, A., Gimm, O., Brauckhoff, M., . . . Kloppel, G. (2008). Primary lymph node gastrinoma or occult duodenal microgastrinoma with lymph node metastases in a MEN1 patient: the need for a systematic search for the primary tumor. *Am J Surg Pathol*, *32*(7), 1101-1105. doi:10.1097/PAS.0b013e3181655811
- Arnold, R., Wilke, A., Rinke, A., Mayer, C., Kann, P. H., Klose, K. J., . . . Barth, P. (2008). Plasma chromogranin A as marker for survival in patients with metastatic endocrine gastroenteropancreatic tumors. *Clin Gastroenterol Hepatol*, *6*(7), 820-827. doi:10.1016/j.cgh.2008.02.052
- Bailey, P., Chang, D. K., Nones, K., Johns, A. L., Patch, A. M., Gingras, M. C., . . . Grimmond, S. M. (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature*, *531*(7592), 47-52. doi:10.1038/nature16965
- Baron, M., Veres, A., Wolock, S. L., Faust, A. L., Gaujoux, R., Vetere, A., . . . Yanai, I. (2016). A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst*, *3*(4), 346-360 e344. doi:10.1016/j.cels.2016.08.011
- Basu, B., Sirohi, B., & Corrie, P. (2010). Systemic therapy for neuroendocrine tumours of gastroenteropancreatic origin. *Endocr Relat Cancer*, *17*(1), R75-90. doi:10.1677/ERC-09-0108
- Bertolino, P., Tong, W. M., Herrera, P. L., Casse, H., Zhang, C. X., & Wang, Z. Q. (2003). Pancreatic beta-cell-specific ablation of the multiple endocrine neoplasia type 1 (MEN1) gene causes full penetrance of insulinoma development in mice. *Cancer Res*, *63*(16), 4836-4841.
- Borromeo, M. D., Savage, T. K., Kollipara, R. K., He, M., Augustyn, A., Osborne, J. K., . . . Johnson, J. E. (2016). ASCL1 and NEUROD1 Reveal Heterogeneity in Pulmonary Neuroendocrine Tumors and Regulate Distinct Genetic Programs. *Cell Rep*, *16*(5), 1259-1272. doi:10.1016/j.celrep.2016.06.081
- Bramswig, N. C., Everett, L. J., Schug, J., Dorrell, C., Liu, C., Luo, Y., . . . Kaestner, K. H. (2013). Epigenomic plasticity enables human pancreatic alpha to beta cell reprogramming. *J Clin Invest*, *123*(3), 1275-1284. doi:10.1172/JCI66514
- Branchfield, K., Nantie, L., Verheyden, J. M., Sui, P., Wienhold, M. D., & Sun, X. (2016). Pulmonary neuroendocrine cells function as airway sensors to control lung immune response. *Science*, *351*(6274), 707-710. doi:10.1126/science.aad7969
- Brenet, F., Moh, M., Funk, P., Feierstein, E., Viale, A. J., Socci, N. D., & Scandura, J. M. (2011). DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS One*, *6*(1), e14524. doi:10.1371/journal.pone.0014524
- Busygina, V., & Bale, A. E. (2006). Multiple endocrine neoplasia type 1 (MEN1) as a cancer predisposition syndrome: clues into the mechanisms of MEN1-related carcinogenesis. *Yale J Biol Med*, *79*(3-4), 105-114.
- Cancer Genome Atlas, N. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, *490*(7418), 61-70. doi:10.1038/nature11412

- Capdevila, J., Casanovas, O., Salazar, R., Castellano, D., Segura, A., Fuster, P., . . . Castano, J. P. (2017). Translational research in neuroendocrine tumors: pitfalls and opportunities. *Oncogene*, *36*(14), 1899-1907. doi:10.1038/onc.2016.316
- Caplin, M. E., Baudin, E., Ferolla, P., Filosso, P., Garcia-Yuste, M., Lim, E., . . . participants, E. c. c. (2015). Pulmonary neuroendocrine (carcinoid) tumors: European Neuroendocrine Tumor Society expert consensus and recommendations for best practice for typical and atypical pulmonary carcinoids. *Ann Oncol*, *26*(8), 1604-1620. doi:10.1093/annonc/mdv041
- Castro, D. S., Martynoga, B., Parras, C., Ramesh, V., Pacary, E., Johnston, C., . . . Guillemot, F. (2011). A novel function of the proneural factor *Ascl1* in progenitor proliferation identified by genome-wide characterization of its targets. *Genes Dev*, *25*(9), 930-945. doi:10.1101/gad.627811
- Chan, C. S., Laddha, S. V., Lewis, P. W., Koletsky, M. S., Robzyk, K., Da Silva, E., . . . Tang, L. H. (2018). *ATRX*, *DAXX* or *MEN1* mutant pancreatic neuroendocrine tumors are a distinct alpha-cell signature subgroup. *Nat Commun*, *9*(1), 4158. doi:10.1038/s41467-018-06498-2
- Chen, H., Xu, C., Jin, Q., & Liu, Z. (2014). S100 protein family in human cancer. *Am J Cancer Res*, *4*(2), 89-115.
- Cheng, D. T., Mitchell, T. N., Zehir, A., Shah, R. H., Benayed, R., Syed, A., . . . Berger, M. F. (2015). Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. *J Mol Diagn*, *17*(3), 251-264. doi:10.1016/j.jmoldx.2014.12.006
- Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., . . . Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*, *31*(3), 213-219. doi:10.1038/nbt.2514
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., . . . Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol*, *17*, 13. doi:10.1186/s13059-016-0881-8
- Da Silva Xavier, G. (2018). The Cells of the Islets of Langerhans. *J Clin Med*, *7*(3). doi:10.3390/jcm7030054
- Dasari, A., Shen, C., Halperin, D., Zhao, B., Zhou, S., Xu, Y., . . . Yao, J. C. (2017). Trends in the Incidence, Prevalence, and Survival Outcomes in Patients With Neuroendocrine Tumors in the United States. *JAMA Oncol*, *3*(10), 1335-1342. doi:10.1001/jamaoncol.2017.0589
- de Mestier, L., Cros, J., Neuzillet, C., Hentic, O., Egal, A., Muller, N., . . . Hammel, P. (2017). Digestive System Mixed Neuroendocrine-Non-Neuroendocrine Neoplasms. *Neuroendocrinology*, *105*(4), 412-425. doi:10.1159/000475527
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15-21. doi:10.1093/bioinformatics/bts635
- Feng, Z., Wang, L., Sun, Y., Jiang, Z., Domsic, J., An, C., . . . Hua, X. (2017). *Menin* and *Daxx* Interact to Suppress Neuroendocrine Tumors through Epigenetic

- Control of the Membrane Metallo-Endopeptidase. *Cancer Res*, 77(2), 401-411. doi:10.1158/0008-5472.CAN-16-1567
- Fernandez-Cuesta, L., Peifer, M., Lu, X., Sun, R., Ozretic, L., Seidal, D., . . . Thomas, R. K. (2014). Frequent mutations in chromatin-remodelling genes in pulmonary carcinoids. *Nat Commun*, 5, 3518. doi:10.1038/ncomms4518
- Ferrone, C. R., Tang, L. H., Tomlinson, J., Gonen, M., Hochwald, S. N., Brennan, M. F., . . . Allen, P. J. (2007). Determining prognosis in patients with pancreatic endocrine neoplasms: can the WHO classification system be simplified? *J Clin Oncol*, 25(35), 5609-5615. doi:10.1200/JCO.2007.12.9809
- Fortin, J. P., Labbe, A., Lemire, M., Zanke, B. W., Hudson, T. J., Fertig, E. J., . . . Hansen, K. D. (2014). Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol*, 15(12), 503. doi:10.1186/s13059-014-0503-2
- Francis, J. M., Kiezun, A., Ramos, A. H., Serra, S., Pedamallu, C. S., Qian, Z. R., . . . Meyerson, M. (2013). Somatic mutation of CDKN1B in small intestine neuroendocrine tumors. *Nat Genet*, 45(12), 1483-1486. doi:10.1038/ng.2821
- Frost, M., Lines, K. E., & Thakker, R. V. (2018). Current and emerging therapies for PNETs in patients with or without MEN1. *Nat Rev Endocrinol*, 14(4), 216-227. doi:10.1038/nrendo.2018.3
- Gribble, F. M., & Reimann, F. (2016). Enteroendocrine Cells: Chemosensors in the Intestinal Epithelium. *Annu Rev Physiol*, 78, 277-299. doi:10.1146/annurev-physiol-021115-105439
- Gross, S., Garofalo, D. C., Balderes, D. A., Mastracci, T. L., Dias, J. M., Perlmann, T., . . . Sussel, L. (2016). The novel enterochromaffin marker Lmx1a regulates serotonin biosynthesis in enteroendocrine cell lineages downstream of Nkx2.2. *Development*, 143(14), 2616-2628. doi:10.1242/dev.130682
- Guha, A., Vasconcelos, M., Cai, Y., Yoneda, M., Hinds, A., Qian, J., . . . Cardoso, W. V. (2012). Neuroepithelial body microenvironment is a niche for a distinct subset of Clara-like precursors in the developing airways. *Proc Natl Acad Sci USA*, 109(31), 12592-12597. doi:10.1073/pnas.1204710109
- Gustafsson, B. I., Siddique, L., Chan, A., Dong, M., Drozdov, I., Kidd, M., & Modlin, I. M. (2008). Uncommon cancers of the small intestine, appendix and colon: an analysis of SEER 1973-2004, and current diagnosis and therapy. *Int J Oncol*, 33(6), 1121-1131.
- Haber, A. L., Biton, M., Rogel, N., Herbst, R. H., Shekhar, K., Smillie, C., . . . Regev, A. (2017). A single-cell survey of the small intestinal epithelium. *Nature*, 551(7680), 333-339. doi:10.1038/nature24489
- Halfdanarson, T. R., Rabe, K. G., Rubin, J., & Petersen, G. M. (2008). Pancreatic neuroendocrine tumors (PNETs): incidence, prognosis and recent trend toward improved survival. *Ann Oncol*, 19(10), 1727-1733. doi:10.1093/annonc/mdn351
- Hashemi, J., Fotouhi, O., Sulaiman, L., Kjellman, M., Hoog, A., Zedenius, J., & Larsson, C. (2013). Copy number alterations in small intestinal neuroendocrine tumors determined by array comparative genomic hybridization. *BMC Cancer*, 13, 505. doi:10.1186/1471-2407-13-505

- Heaphy, C. M., de Wilde, R. F., Jiao, Y., Klein, A. P., Edil, B. H., Shi, C., . . . Meeker, A. K. (2011). Altered telomeres in tumors with ATRX and DAXX mutations. *Science*, 333(6041), 425. doi:10.1126/science.1207313
- Hendifar, A. E., Marchevsky, A. M., & Tuli, R. (2017). Neuroendocrine Tumors of the Lung: Current Challenges and Advances in the Diagnosis and Management of Well-Differentiated Disease. *J Thorac Oncol*, 12(3), 425-436. doi:10.1016/j.jtho.2016.11.2222
- Ho, T. H., Kapur, P., Joseph, R. W., Serie, D. J., Eckel-Passow, J. E., Tong, P., . . . Parker, A. S. (2016). Loss of histone H3 lysine 36 trimethylation is associated with an increased risk of renal cell carcinoma-specific death. *Mod Pathol*, 29(1), 34-42. doi:10.1038/modpathol.2015.123
- Hughes, C. M., Rozenblatt-Rosen, O., Milne, T. A., Copeland, T. D., Levine, S. S., Lee, J. C., . . . Meyerson, M. (2004). Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. *Mol Cell*, 13(4), 587-597.
- Jann, H., Roll, S., Couvelard, A., Hentic, O., Pavel, M., Muller-Nordhorn, J., . . . Pape, U. F. (2011). Neuroendocrine tumors of midgut and hindgut origin: tumor-node-metastasis classification determines clinical outcome. *Cancer*, 117(15), 3332-3341. doi:10.1002/cncr.25855
- Jiao, Y., Shi, C., Edil, B. H., de Wilde, R. F., Klimstra, D. S., Maitra, A., . . . Papadopoulos, N. (2011). DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science*, 331(6021), 1199-1203. doi:10.1126/science.1200609
- Karpathakis, A., Dibra, H., Pipinikas, C., Feber, A., Morris, T., Francis, J., . . . Thirlwell, C. (2016). Prognostic Impact of Novel Molecular Subtypes of Small Intestinal Neuroendocrine Tumor. *Clin Cancer Res*, 22(1), 250-258. doi:10.1158/1078-0432.CCR-15-0373
- Kim, A., Miller, K., Jo, J., Kilimnik, G., Wojcik, P., & Hara, M. (2009). Islet architecture: A comparative study. *Islets*, 1(2), 129-136. doi:10.4161/isl.1.2.9480
- Kim, J. Y., & Hong, S. M. (2016). Recent Updates on Neuroendocrine Tumors From the Gastrointestinal and Pancreatobiliary Tracts. *Arch Pathol Lab Med*, 140(5), 437-448. doi:10.5858/arpa.2015-0314-RA
- Klimova, L., Antosova, B., Kuzelova, A., Strnad, H., & Kozmik, Z. (2015). Onecut1 and Onecut2 transcription factors operate downstream of Pax6 to regulate horizontal cell development. *Dev Biol*, 402(1), 48-60. doi:10.1016/j.ydbio.2015.02.023
- Kloppel, G. (2017). Neuroendocrine Neoplasms: Dichotomy, Origin and Classifications. *Visc Med*, 33(5), 324-330. doi:10.1159/000481390
- Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., . . . Getz, G. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, 499(7457), 214-218. doi:10.1038/nature12213
- Lewis, P. W., Elsaesser, S. J., Noh, K. M., Stadler, S. C., & Allis, C. D. (2010). Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc Natl Acad Sci U S A*, 107(32), 14075-14080. doi:10.1073/pnas.1008850107

- Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*, 323. doi:10.1186/1471-2105-12-323
- Li, F., Su, Y., Cheng, Y., Jiang, X., Peng, Y., Li, Y., . . . Ning, G. (2015). Conditional deletion of Men1 in the pancreatic beta-cell leads to glucagon-expressing tumor development. *Endocrinology*, *156*(1), 48-57. doi:10.1210/en.2014-1433
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324
- Li, J., Ning, G., & Duncan, S. A. (2000). Mammalian hepatocyte differentiation requires the transcription factor HNF-4alpha. *Genes Dev*, *14*(4), 464-474.
- Lin, W., Watanabe, H., Peng, S., Francis, J. M., Kaplan, N., Pedamallu, C. S., . . . Meyerson, M. (2015). Dynamic epigenetic regulation by menin during pancreatic islet tumor formation. *Mol Cancer Res*, *13*(4), 689-698. doi:10.1158/1541-7786.MCR-14-0457
- Linnoila, R. I. (2006). Functional facets of the pulmonary neuroendocrine system. *Lab Invest*, *86*(5), 425-444. doi:10.1038/labinvest.3700412
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, *15*(12), 550. doi:10.1186/s13059-014-0550-8
- Lu, J., Herrera, P. L., Carreira, C., Bonnavion, R., Seigne, C., Calender, A., . . . Zhang, C. X. (2010). Alpha cell-specific Men1 ablation triggers the transdifferentiation of glucagon-expressing cells and insulinoma development. *Gastroenterology*, *138*(5), 1954-1965. doi:10.1053/j.gastro.2010.01.046
- Lundqvist, M., & Wilander, E. (1987). A study of the histopathogenesis of carcinoid tumors of the small intestine and appendix. *Cancer*, *60*(2), 201-206.
- Mafficini, A., & Scarpa, A. (2018). Genomic landscape of pancreatic neuroendocrine tumours: the International Cancer Genome Consortium. *J Endocrinol*, *236*(3), R161-R167. doi:10.1530/JOE-17-0560
- Mansouri, A. (2012). Development and regeneration in the endocrine pancreas. *ISRN Endocrinol*, *2012*, 640956. doi:10.5402/2012/640956
- Marinoni, I., Kurrer, A. S., Vassella, E., Dettmer, M., Rudolph, T., Banz, V., . . . Perren, A. (2014). Loss of DAXX and ATRX are associated with chromosome instability and reduced survival of patients with pancreatic neuroendocrine tumors. *Gastroenterology*, *146*(2), 453-460 e455. doi:10.1053/j.gastro.2013.10.020
- Miller, J. A., Cai, C., Langfelder, P., Geschwind, D. H., Kurian, S. M., Salomon, D. R., & Horvath, S. (2011). Strategies for aggregating gene expression data: the collapseRows R function. *BMC Bioinformatics*, *12*, 322. doi:10.1186/1471-2105-12-322
- Modali, S. D., Parekh, V. I., Kebebew, E., & Agarwal, S. K. (2015). Epigenetic regulation of the lncRNA MEG3 and its target c-MET in pancreatic neuroendocrine tumors. *Mol Endocrinol*, *29*(2), 224-237. doi:10.1210/me.2014-1304
- Modlin, I. M., Shapiro, M. D., & Kidd, M. (2004). Siegfried Oberndorfer: origins and perspectives of carcinoid tumors. *Hum Pathol*, *35*(12), 1440-1451.

- Morris, T. J., Butcher, L. M., Feber, A., Teschendorff, A. E., Chakravarthy, A. R., Wojdacz, T. K., & Beck, S. (2014). ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics*, *30*(3), 428-430. doi:10.1093/bioinformatics/btt684
- Muraro, M. J., Dharmadhikari, G., Grun, D., Groen, N., Dielen, T., Jansen, E., . . . van Oudenaarden, A. (2016). A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst*, *3*(4), 385-394 e383. doi:10.1016/j.cels.2016.09.002
- Noguchi, A., Li, X., Kubota, A., Kikuchi, K., Kameda, Y., Zheng, H., . . . Takano, Y. (2013). SIRT1 expression is associated with good prognosis for head and neck squamous cell carcinoma patients. *Oral Surg Oral Med Oral Pathol Oral Radiol*, *115*(3), 385-392. doi:10.1016/j.oooo.2012.12.013
- Noguchi, M., Sumiyama, K., & Morimoto, M. (2015). Directed Migration of Pulmonary Neuroendocrine Cells toward Airway Branches Organizes the Stereotypic Location of Neuroepithelial Bodies. *Cell Rep*, *13*(12), 2679-2686. doi:10.1016/j.celrep.2015.11.058
- Oronsky, B., Ma, P. C., Morgensztern, D., & Carter, C. A. (2017). Nothing But NET: A Review of Neuroendocrine Tumors and Carcinomas. *Neoplasia*, *19*(12), 991-1002. doi:10.1016/j.neo.2017.09.002
- Pal, S. K., Agarwal, N., Boorjian, S. A., Hahn, N. M., Siefker-Radtke, A. O., Clark, P. E., & Plimack, E. R. (2016). National Comprehensive Cancer Network Recommendations on Molecular Profiling of Advanced Bladder Cancer. *J Clin Oncol*, *34*(27), 3346-3348. doi:10.1200/JCO.2016.68.1429
- Park, J. K., Paik, W. H., Lee, K., Ryu, J. K., Lee, S. H., & Kim, Y. T. (2017). DAXX/ATRX and MEN1 genes are strong prognostic markers in pancreatic neuroendocrine tumors. *Oncotarget*, *8*(30), 49796-49806. doi:10.18632/oncotarget.17964
- Pelosi, G., Papotti, M., Rindi, G., & Scarpa, A. (2014). Unraveling tumor grading and genomic landscape in lung neuroendocrine tumors. *Endocr Pathol*, *25*(2), 151-164. doi:10.1007/s12022-014-9320-0
- Pelosi, G., Rindi, G., Travis, W. D., & Papotti, M. (2014). Ki-67 antigen in lung neuroendocrine tumors: unraveling a role in clinical practice. *J Thorac Oncol*, *9*(3), 273-284. doi:10.1097/JTO.0000000000000092
- Pelosi, G., Rodriguez, J., Viale, G., & Rosai, J. (2005). Typical and atypical pulmonary carcinoid tumor overdiagnosed as small-cell carcinoma on biopsy specimens: a major pitfall in the management of lung cancer patients. *Am J Surg Pathol*, *29*(2), 179-187.
- Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., & Mesirov, J. P. (2006). GenePattern 2.0. *Nat Genet*, *38*(5), 500-501. doi:10.1038/ng0506-500
- Rekhtman, N., Pietanza, C. M., Sabari, J., Montecalvo, J., Wang, H., Habeeb, O., . . . Joubert, P. (2018). Pulmonary large cell neuroendocrine carcinoma with adenocarcinoma-like features: napsin A expression and genomic alterations. *Mod Pathol*, *31*(1), 111-121. doi:10.1038/modpathol.2017.110
- Sachithanandan, N., Harle, R. A., & Burgess, J. R. (2005). Bronchopulmonary carcinoid in multiple endocrine neoplasia type 1. *Cancer*, *103*(3), 509-515. doi:10.1002/cncr.20825
- Sadanandam, A., Wullschleger, S., Lyssiotis, C. A., Grotzinger, C., Barbi, S., Bersani, S., . . . Hanahan, D. (2015). A Cross-Species Analysis in Pancreatic

- Neuroendocrine Tumors Reveals Molecular Subtypes with Distinctive Clinical, Metastatic, Developmental, and Metabolic Characteristics. *Cancer Discov*, 5(12), 1296-1313. doi:10.1158/2159-8290.CD-15-0068
- Sathirapongsasuti, J. F., Lee, H., Horst, B. A., Brunner, G., Cochran, A. J., Binder, S., . . . Nelson, S. F. (2011). Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics*, 27(19), 2648-2654. doi:10.1093/bioinformatics/btr462
- Scarpa, A., Chang, D. K., Nones, K., Corbo, V., Patch, A. M., Bailey, P., . . . Grimmond, S. M. (2017). Whole-genome landscape of pancreatic neuroendocrine tumours. *Nature*, 543(7643), 65-71. doi:10.1038/nature21063
- Scherubl, H., Streller, B., Stabenow, R., Herbst, H., Hopfner, M., Schwertner, C., . . . Zappe, S. M. (2013). Clinically detected gastroenteropancreatic neuroendocrine tumors are on the rise: epidemiological changes in Germany. *World J Gastroenterol*, 19(47), 9012-9019. doi:10.3748/wjg.v19.i47.9012
- Schimmack, S., Svejda, B., Lawrence, B., Kidd, M., & Modlin, I. M. (2011). The diversity and commonalities of gastroenteropancreatic neuroendocrine tumors. *Langenbecks Arch Surg*, 396(3), 273-298. doi:10.1007/s00423-011-0739-1
- Schnorbusch, K., Lembrechts, R., Pintelon, I., Timmermans, J. P., Brouns, I., & Adriaensen, D. (2013). GABAergic signaling in the pulmonary neuroepithelial body microenvironment: functional imaging in GAD67-GFP mice. *Histochem Cell Biol*, 140(5), 549-566. doi:10.1007/s00418-013-1093-x
- Shen, H. C., He, M., Powell, A., Adem, A., Lorang, D., Heller, C., . . . Libutti, S. K. (2009). Recapitulation of pancreatic neuroendocrine tumors in human multiple endocrine neoplasia type I syndrome via Pdx1-directed inactivation of Men1. *Cancer Res*, 69(5), 1858-1866. doi:10.1158/0008-5472.CAN-08-3662
- Shen, H. C., Ylaya, K., Pechhold, K., Wilson, A., Adem, A., Hewitt, S. M., & Libutti, S. K. (2010). Multiple endocrine neoplasia type 1 deletion in pancreatic alpha-cells leads to development of insulinomas in mice. *Endocrinology*, 151(8), 4024-4030. doi:10.1210/en.2009-1251
- Simbolo, M., Mafficini, A., Sikora, K. O., Fassan, M., Barbi, S., Corbo, V., . . . Scarpa, A. (2017). Lung neuroendocrine tumours: deep sequencing of the four World Health Organization histotypes reveals chromatin-remodelling genes as major players and a prognostic role for TERT, RB1, MEN1 and KMT2D. *J Pathol*, 241(4), 488-500. doi:10.1002/path.4853
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., . . . Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*, 102(43), 15545-15550. doi:10.1073/pnas.0506580102
- Swarts, D. R., van Suylen, R. J., den Bakker, M. A., van Oosterhout, M. F., Thunnissen, F. B., Volante, M., . . . Speel, E. J. (2014). Interobserver variability for the WHO classification of pulmonary carcinoids. *Am J Surg Pathol*, 38(10), 1429-1436. doi:10.1097/PAS.0000000000000300
- Taal, B. G., & Visser, O. (2004). Epidemiology of neuroendocrine tumours. *Neuroendocrinology*, 80 Suppl 1, 3-7. doi:10.1159/000080731

- Tang, L. H., Basturk, O., Sue, J. J., & Klimstra, D. S. (2016). A Practical Approach to the Classification of WHO Grade 3 (G3) Well-differentiated Neuroendocrine Tumor (WD-NET) and Poorly Differentiated Neuroendocrine Carcinoma (PD-NEC) of the Pancreas. *Am J Surg Pathol*, 40(9), 1192-1202. doi:10.1097/PAS.0000000000000662
- Tang, L. H., & Klimstra, D. S. (2011). Conundrums and Caveats in Neuroendocrine Tumors of the Pancreas. *Surg Pathol Clin*, 4(2), 589-624. doi:10.1016/j.path.2011.03.003
- Tang, L. H., Untch, B. R., Reidy, D. L., O'Reilly, E., Dhall, D., Jih, L., . . . Klimstra, D. S. (2016). Well-Differentiated Neuroendocrine Tumors with a Morphologically Apparent High-Grade Component: A Pathway Distinct from Poorly Differentiated Neuroendocrine Carcinomas. *Clin Cancer Res*, 22(4), 1011-1017. doi:10.1158/1078-0432.CCR-15-0548
- Travis, W. D., Brambilla, E., Nicholson, A. G., Yatabe, Y., Austin, J. H. M., Beasley, M. B., . . . Panel, W. H. O. (2015). The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol*, 10(9), 1243-1260. doi:10.1097/JTO.0000000000000630
- Travis, W. D., Rush, W., Flieder, D. B., Falk, R., Fleming, M. V., Gal, A. A., & Koss, M. N. (1998). Survival analysis of 200 pulmonary neuroendocrine tumors with clarification of criteria for atypical carcinoid and its separation from typical carcinoid. *Am J Surg Pathol*, 22(8), 934-944.
- Valdes, N., Alvarez, V., Diaz-Cadorniga, F., Aller, J., Villazon, F., Garcia, I., . . . Coto, E. (1998). Multiple endocrine neoplasia type 1 (MEN1): LOH studies in a affected family and in sporadic cases. *Anticancer Res*, 18(4A), 2685-2689.
- van den Bent, M. J. (2010). Interobserver variation of the histopathological diagnosis in clinical trials on glioma: a clinician's perspective. *Acta Neuropathol*, 120(3), 297-304. doi:10.1007/s00401-010-0725-7
- Van Lommel, A. (2001). Pulmonary neuroendocrine cells (PNEC) and neuroepithelial bodies (NEB): chemoreceptors and regulators of lung development. *Paediatr Respir Rev*, 2(2), 171-176.
- Verckist, L., Lembrechts, R., Thys, S., Pintelon, I., Timmermans, J. P., Brouns, I., & Adriaensen, D. (2017). Selective gene expression analysis of the neuroepithelial body microenvironment in postnatal lungs with special interest for potential stem cell characteristics. *Respir Res*, 18(1), 87. doi:10.1186/s12931-017-0571-4
- Vinik, A. I., Silva, M. P., Woltering, E. A., Go, V. L., Warner, R., & Caplin, M. (2009). Biochemical testing for neuroendocrine tumors. *Pancreas*, 38(8), 876-889. doi:10.1097/MPA.0b013e3181bc0e77
- Vinik, A. I., Woltering, E. A., Warner, R. R., Caplin, M., O'Dorisio, T. M., Wiseman, G. A., . . . North American Neuroendocrine Tumor, S. (2010). NANETS consensus guidelines for the diagnosis of neuroendocrine tumor. *Pancreas*, 39(6), 713-734. doi:10.1097/MPA.0b013e3181ebaffd
- Volante, M., Gatti, G., & Papotti, M. (2015). Classification of lung neuroendocrine tumors: lights and shadows. *Endocrine*, 50(2), 315-319. doi:10.1007/s12020-015-0578-x

- Vollbrecht, C., Werner, R., Walter, R. F., Christoph, D. C., Heukamp, L. C., Peifer, M., . . . Mairinger, F. D. (2015). Mutational analysis of pulmonary tumours with neuroendocrine features using targeted massive parallel sequencing: a comparison of a neglected tumour group. *Br J Cancer*, *113*(12), 1704-1711. doi:10.1038/bjc.2015.397
- Waldum, H. L., Oberg, K., Sordal, O. F., Sandvik, A. K., Gustafsson, B. I., Mjones, P., & Fossmark, R. (2018). Not only stem cells, but also mature cells, particularly neuroendocrine cells, may develop into tumours: time for a paradigm shift. *Therap Adv Gastroenterol*, *11*, 1756284818775054. doi:10.1177/1756284818775054
- Walter, D., Harter, P. N., Battke, F., Winkelmann, R., Schneider, M., Holzer, K., . . . Waidmann, O. (2018). Genetic heterogeneity of primary lesion and metastasis in small intestine neuroendocrine tumors. *Sci Rep*, *8*(1), 3811. doi:10.1038/s41598-018-22115-0
- Wang, Y. J., Schug, J., Won, K. J., Liu, C., Naji, A., Avrahami, D., . . . Kaestner, K. H. (2016). Single-Cell Transcriptomics of the Human Endocrine Pancreas. *Diabetes*, *65*(10), 3028-3038. doi:10.2337/db16-0405
- Wang, Z. J., Yang, J. L., Wang, Y. P., Lou, J. Y., Chen, J., Liu, C., & Guo, L. D. (2013). Decreased histone H2B monoubiquitination in malignant gastric carcinoma. *World J Gastroenterol*, *19*(44), 8099-8107. doi:10.3748/wjg.v19.i44.8099
- Warden, C. D., Lee, H., Tompkins, J. D., Li, X., Wang, C., Riggs, A. D., . . . Yuan, Y. C. (2013). COHCAP: an integrative genomic pipeline for single-nucleotide resolution DNA methylation analysis. *Nucleic Acids Res*, *41*(11), e117. doi:10.1093/nar/gkt242
- Wei, Y., Xia, W., Zhang, Z., Liu, J., Wang, H., Adsay, N. V., . . . Hung, M. C. (2008). Loss of trimethylation at lysine 27 of histone H3 is a predictor of poor outcome in breast, ovarian, and pancreatic cancers. *Mol Carcinog*, *47*(9), 701-706. doi:10.1002/mc.20413
- Welin, S., Stridsberg, M., Cunningham, J., Granberg, D., Skogseid, B., Oberg, K., . . . Janson, E. T. (2009). Elevated plasma chromogranin A is the first indication of recurrence in radically operated midgut carcinoid tumors. *Neuroendocrinology*, *89*(3), 302-307. doi:10.1159/000179900
- Wiedenmann, B., & Huttner, W. B. (1989). Synaptophysin and chromogranins/secretogranins--widespread constituents of distinct types of neuroendocrine vesicles and new tools in tumor diagnosis. *Virchows Arch B Cell Pathol Incl Mol Pathol*, *58*(2), 95-121.
- Williams, E. D., & Sandler, M. (1963). The classification of carcinoid tumours. *Lancet*, *1*(7275), 238-239.
- Yao, J. C., Hassan, M., Phan, A., Dagohoy, C., Leary, C., Mares, J. E., . . . Evans, D. B. (2008). One hundred years after "carcinoid": epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *J Clin Oncol*, *26*(18), 3063-3072. doi:10.1200/JCO.2007.15.4377
- Yokoyama, A., Somerville, T. C., Smith, K. S., Rozenblatt-Rosen, O., Meyerson, M., & Cleary, M. L. (2005). The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell*, *123*(2), 207-218. doi:10.1016/j.cell.2005.09.025

Yoshihara, K., Shahmoradgoli, M., Martinez, E., Vegesna, R., Kim, H., Torres-Garcia, W., . . . Verhaak, R. G. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*, 4, 2612. doi:10.1038/ncomms3612