

# UCSF

## UC San Francisco Previously Published Works

### Title

Whole genome sequencing reveals the independent clonal origin of multifocal ileal neuroendocrine tumors

### Permalink

<https://escholarship.org/uc/item/1xp247ph>

### Journal

Genome Medicine, 14(1)

### ISSN

1756-994X

### Authors

Mäkinen, Netta  
Zhou, Meng  
Zhang, Zhouwei  
[et al.](#)

### Publication Date

2022-12-01

### DOI

10.1186/s13073-022-01083-1

Peer reviewed

RESEARCH

Open Access



# Whole genome sequencing reveals the independent clonal origin of multifocal ileal neuroendocrine tumors

Netta Mäkinen<sup>1,2</sup>, Meng Zhou<sup>1,2</sup>, Zhouwei Zhang<sup>1,2</sup>, Yosuke Kasai<sup>3</sup>, Elizabeth Perez<sup>1</sup>, Grace E. Kim<sup>4</sup>,  
Chrissie Thirlwell<sup>5,6</sup>, Eric Nakakura<sup>3</sup> and Matthew Meyerson<sup>1,2,7\*</sup> 

## Abstract

**Background:** Small intestinal neuroendocrine tumors (SI-NETs) are the most common neoplasms of the small bowel. The majority of tumors are located in the distal ileum with a high incidence of multiple synchronous primary tumors. Even though up to 50% of SI-NET patients are diagnosed with multifocal disease, the mechanisms underlying multiple synchronous lesions remain elusive.

**Methods:** We performed whole genome sequencing of 75 de-identified synchronous primary tumors, 15 metastases, and corresponding normal samples from 13 patients with multifocal ileal NETs to identify recurrent somatic genomic alterations, frequently affected signaling pathways, and shared mutation signatures among multifocal SI-NETs. Additionally, we carried out chromosome mapping of the most recurrent copy-number alterations identified to determine which parental allele had been affected in each tumor and assessed the clonal relationships of the tumors within each patient.

**Results:** Absence of shared somatic variation between the synchronous primary tumors within each patient was observed, indicating that these tumors develop independently. Although recurrent copy-number alterations were identified, additional chromosome mapping revealed that tumors from the same patient can gain or lose different parental alleles. In addition to the previously reported *CDKN1B* loss-of-function mutations, we observed potential loss-of-function gene alterations in *TNRC6B*, a candidate tumor suppressor gene in a small subset of ileal NETs. Furthermore, we show that multiple metastases in the same patient can originate from either one or several primary tumors.

**Conclusions:** Our study demonstrates major genomic diversity among multifocal ileal NETs, highlighting the need to identify and remove all primary tumors, which have the potential to metastasize, and the need for optimized targeted treatments.

**Keywords:** Small bowel, Small intestinal neuroendocrine tumors, Multifocality, Whole genome sequencing, Independent clonal origin

## Background

Small intestinal neuroendocrine tumors (SI-NETs) are the most common neoplasms of the small bowel with an estimated annual age-adjusted incidence ranging from 0.67 to 1.12 per 100,000 persons [1–3]. SI-NETs originate from enterochromaffin cells of the digestive tract, and most tumors arise in the terminal ileum [4, 5]. SI-NETs

\*Correspondence: matthew\_meyerson@dfci.harvard.edu

<sup>1</sup> Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215, USA  
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

are usually well-differentiated tumors characterized by a low proliferation rate, but also by a high percentage of distant metastases at diagnosis [6]. The 5-year survival rate is less than 50% in patients with metastatic disease [1, 6]. The only curative treatment of SI-NETs is complete surgical resection. Development of targeted therapies has been impeded by the lack of apparent driver genes in SI-NETs.

Previous high-throughput sequencing studies, which have primarily focused on targeted gene panel and whole exome sequencing, have reported low somatic mutation rates in SI-NETs [7–11]. The most frequent genomic alteration identified to date is loss of heterozygosity (LOH) at chromosome (chr) 18, occurring in 70% of tumors [12–14]. Other recurrent whole chromosome and whole chromosome arm copy-number alterations (CNAs) have been observed in 10–30% of SI-NETs, including gains of chromosomes 4, 5, 7, 14, and 20 [7–10]. The only recurrent mutations identified in SI-NETs are loss-of-function mutations in cyclin-dependent kinase inhibitor 1B (*CDKN1B*) in approximately 8–10% of tumors [8, 15]. Furthermore, a recent whole genome sequencing (WGS) study of 2520 metastatic solid tumors reported SI-NETs to rarely harbor a candidate driver mutation or germline predisposition variant. A total of 34 samples in the study had no identified drivers, and 18 of these samples were SI-NETs [16].

The majority of large-scale sequencing studies to date have concentrated on sequencing single primary tumors from each patient. Multiple synchronous lesions, however, have been observed in up to 50% of SI-NET patients [17, 18]. The molecular mechanisms underlying these multifocal lesions are not yet understood. Recently, two high-throughput sequencing studies of multifocal SI-NETs by us and others have suggested these tumors to develop independently [19, 20]. Our high-throughput sequencing-based copy-number profiling of 40 multifocal ileal NETs revealed distinct patterns of chr18 allelic loss in individual tumors from the same patient, suggesting these tumors originate independently and that no specific germline allele on chr18 is targeted by somatic LOH [19]. Additionally, Elias et al. performed WGS of 61 tumor samples (42 primary tumors and 19 metastases) from 11 patients with multifocal SI-NET to study the evolutionary trajectory of multifocal SI-NETs within single patients [20]. They observed lack of shared somatic variation among the primary tumors within the patients, supporting the independent clonal origin of multifocal SI-NETs.

To obtain a comprehensive molecular genomic characterization of multifocal SI-NETs and to confirm previous findings, we performed WGS of 90 tumor samples (75 primary tumors and 15 metastases) from 13 patients with

multifocal ileal NETs. Our analysis of somatic single-nucleotide variants (SNVs), insertion-deletion mutations (indels), CNAs, and structural variants (SVs) revealed substantial genomic diversity among the multifocal ileal NETs, indicating that these tumors develop independently. We also provide evidence showing that metastases in a multifocal ileal NET patient can occasionally originate from different primary tumors.

## Methods

### Samples

Each patient provided informed consent in accordance with the protocols approved by the Institutional Review Boards of the Dana-Farber Cancer Institute and the University of California San Francisco. The initial sample material consisted of 87 de-identified synchronous primary tumors, 20 metastases and matched adjacent normal ileal mucosa and/or whole blood specimens from 13 multifocal ileal NET patients, who underwent surgery at the University of California San Francisco. The clinical characteristics of the patients are summarized in Table 1. All patients had been diagnosed with well-differentiated ileal NETs (grades 1–2) [21], which was confirmed post-operatively, and developed metastatic disease. The tissue specimens were snap-frozen in liquid nitrogen after surgical excision and stored at  $-80^{\circ}\text{C}$ . Tumor purity was assessed for each specimen using hematoxylin and eosin (H&E) staining before sequencing. Samples with tumor purity  $\geq 20\%$  were selected for DNA extraction, resulting in a total of 78 synchronous primary tumors and 16 metastases.

### DNA extraction

Genomic DNA was extracted from the whole blood specimens using the QIAamp<sup>®</sup> DNA Blood kit (Qiagen, Germantown, MD) according to the manufacturer's instructions, whereas gDNA extraction of fresh-frozen primary ileal NETs, metastases, and normal ileal mucosa specimens was performed at the Genomics Platform (GP), Broad Institute of MIT, and Harvard, Cambridge, MA. For three primary tumors, the amount of extracted DNA was too low to proceed with WGS.

### Whole genome sequencing

A total of 75 synchronous primary tumors, 16 metastases, and 18 normal tissue specimens entered WGS. Library construction, WGS, and preprocessing of raw sequencing reads were carried out at the GP [22]. Briefly, gDNA libraries were sequenced on HiSeq X Ten (Illumina, San Diego, CA) to generate 151-bp paired-end reads with a mean target depth of  $60\times$  coverage for tumor specimens, and  $30\times$  coverage for the normal tissue specimens. Sequenced reads were aligned to GRCh38/hg38 reference

**Table 1** Patient cohort characteristics

| Patient | Gender | Average age [range] | Average number of primary tumors [range] | T <sup>a</sup> | N <sup>a</sup> | M <sup>a</sup> | Stage <sup>a</sup> | Grade <sup>b</sup> |
|---------|--------|---------------------|--|----------------|----------------|----------------|--------------------|--------------------|
| 744     | Male   |                     |  | T3             | N1             | M1a            | IV                 | G2                 |
| 760     | Male   |                     |  | T2             | N0             | M1a            | IV                 | G2                 |
| 772     | Male   |                     |  | T3             | N2             | M1a            | IV                 | G1                 |
| 825     | Male   |                     |  | T4             | N1             | M1a            | IV                 | G1                 |
| 848     | Female |                     |  | T3             | N1             | M1a            | IV                 | G1                 |
| 850     | Male   |                     |  | T3             | N2             | M0             | III                | G2                 |
| 852     | Male   | 64 [53–75]          | 7 [2–18]                                 | T3             | N2             | M1c            | IV                 | G1                 |
| 876     | Female |                     |  | T3             | N2             | M0             | III                | G1                 |
| 947     | Female |                     |  | T4             | N2             | M1b            | IV                 | G1                 |
| 952     | Female |                     |  | T3             | N1             | M0             | III                | G1                 |
| 1060    | Male   |                     |  | T3             | N2             | M1a            | IV                 | G2                 |
| 1076    | Male   |                     |  | T4             | N2             | M1c            | IV                 | G1                 |
| 1089    | Male   |                     |  | T2             | N2             | M0             | III                | G1                 |

<sup>a</sup> TNM staging was based on American Joint Committee on Cancer 8th edition

<sup>b</sup> Grade was based on WHO Classification 5th edition

assembly using BWA–MEM [23] and duplicate-marked with Picard tools [24]. GATK was utilized for base score recalibration and local indel re-alignments [25]. One metastasis specimen failed quality control of WGS data and was excluded from the study. The mean coverage achieved was  $80.5 \times [60.0–171.3 \times]$  for tumor specimens and  $75.3 \times [30.5–102.3 \times]$  for normal tissue samples (Additional file 1: Table S1).

### Somatic variant calling

SNVs and indels were called with Mutect2 [26] from GATK v4.1.9.0 [27] and Strelka2 v2.9.10 [28] and functionally annotated with GATK's Funcotator using default parameters. Normal ileal mucosa samples were used as matched normal tissue, apart from patient 952, from which only whole blood sample was available. Additionally, a panel of sequences derived from normal samples was applied for Mutect2, including both normal ileal mucosa and whole blood specimens from ileal NET patients. Further filtering for SNVs and indels included removal of variants with coverage  $\leq 6$  reads and population variants with minor allele frequency (MAF)  $> 0.001$  (gnomAD genome data). All the coding variants were individually visualized with Integrative Genomics Viewer (IGV) v2.7.2 to exclude those present only in the same direction reads or repetitive genomic regions as likely artifacts. Only noncoding variants called by both Mutect2 and Strelka2 were included in the study. CNAs were called using GATK's somatic copy-number variant discovery workflow. Non-overlapping genomic intervals were used for collecting read counts. CNAs with median logR-ratio  $> 0.15$  and  $< -0.15$  were included in the study with an aim to

capture major CNA events while filtering out noisy signals. It is possible that using these thresholds some low-abundance subclonal events may have been filtered out. CNA breakpoints located in chromosomal centromeres or within 1 Mb at the end of a chromosome, and CNAs  $< 10$  kb in size were excluded from the study. SVs were called using SvABA v1.1.3 with default parameters [29]. SVs  $< 10$  kb in size, defined by the distance of two breakpoints if they were located at the same chromosome, and with breakpoints locating in a centromeric region were disregarded. All SVs were individually visualized with IGV to exclude likely artifacts.

### Mutation significance analysis

MutSig2CV [30] was used to identify genes that were mutated more often in the tumors than expected by chance given the inferred background mutation processes. The analysis was performed for somatic coding region SNVs and indels using both patient- and tumor-level information. Due to the low somatic mutation rate in SI-NETs, we increased the sample size for this analysis by combining our data with previously published high-throughput sequencing data of SI-NETs [8, 16], resulting in 176 tumors from 99 SI-NET patients. For the patient-level analysis, all samples from the same patient were collapsed together and the analysis included a union of unique mutations from each patient. For the tumor-level analysis, tumors with highly similar somatic coding region SNV and indel profiles were removed by MutSig2CV from the analysis. Default parameters were applied, and FDR-adjusted  $P < 0.1$  was considered statistically significant.

### Clustering analysis

We used fishhook [31] to identify statistical enrichment or depletion of SNVs and indels in the genomes of all primary ileal NETs and metastases. The Gamma-Poisson regression model was corrected by incorporating covariates of genomic features. The covariates used included common fragile sites from HGNC BioMart [32], known retrotransposons annotated by RepeatMasker [33], and nucleotide context, including GC, CpG, and TpC contents. A non-overlapping bin size of 10 kb was used for SNVs and 100 kb for indels. False discovery rate (FDR; Benjamini-Hochberg) < 0.1 was considered statistically significant.

### Signaling pathway analysis

Signaling pathway analysis was performed for all 90 tumor samples. We used a statistical test (hypergeometric distribution) to determine whether certain Reactome pathways were enriched among the mutated genes in the primary ileal NETs and metastases [34]. Synonymous variation was excluded from the analysis. Overall, 560 out of 880 genes were found in Reactome v75, where 1378 pathways were hit by at least one of them. The probability score for each pathway was FDR corrected using the Benjamini-Hochberg method. FDR-adjusted  $P < 0.1$  was considered statistically significant. As described in the results, no pathways were statistically significant after FDR correction.

### Mutation signature analysis

Mutation signature analysis was performed for all 90 tumor samples. Mutation signature analysis focused on single base substitutions (SBS) in 96 trinucleotide contexts. Mutational matrices were created using SigProfilerMatrixGenerator [35] with default parameters. SigProfilerExtractor [36] was used to perform de novo extraction of a maximum of five mutational signatures per sample and decomposition of the signatures into COSMIC mutational signatures v3.2 [37].

### Chromosome mapping

Germline single-nucleotide polymorphisms (SNPs) were called with HaplotypeCaller [38] from GATK v4.1.9.0 using allele-specific filtering workflow with a truth sensitivity filter level of 99%. Chromosome mapping was performed for all recurrent CNAs that were present in multiple tumors from at least two ileal NET patients as described previously [19]. First, heterozygous germline SNPs were identified from the normal tissue samples using the following filters: read depth > 10 and variant allele frequency between 0.4 and 0.6. The allelic depths of these SNPs were retrieved from the corresponding tumor samples if the total read depth of a given SNP

was > 10 in the tumor samples. Next, a binomial test was applied to the read counts of the reference and alternative alleles of each SNP with the null hypothesis of 0.5, meaning that both alleles were expected to occur in half of the reads. FDR correction was performed for all SNPs across the segment in question. SNPs with FDR-adjusted  $P < 0.05$  were considered as informative SNPs. For tumors with < 1000 shared informative SNPs, FDR-adjusted  $P < 0.1$  was applied. The deleted and amplified chromosome alleles were assigned for each informative SNP by comparing the read counts of reference and alternative alleles. We acknowledge that this approach may not be optimal for capturing subclonal CNAs or allelic imbalance in samples with low tumor purity.

## Results

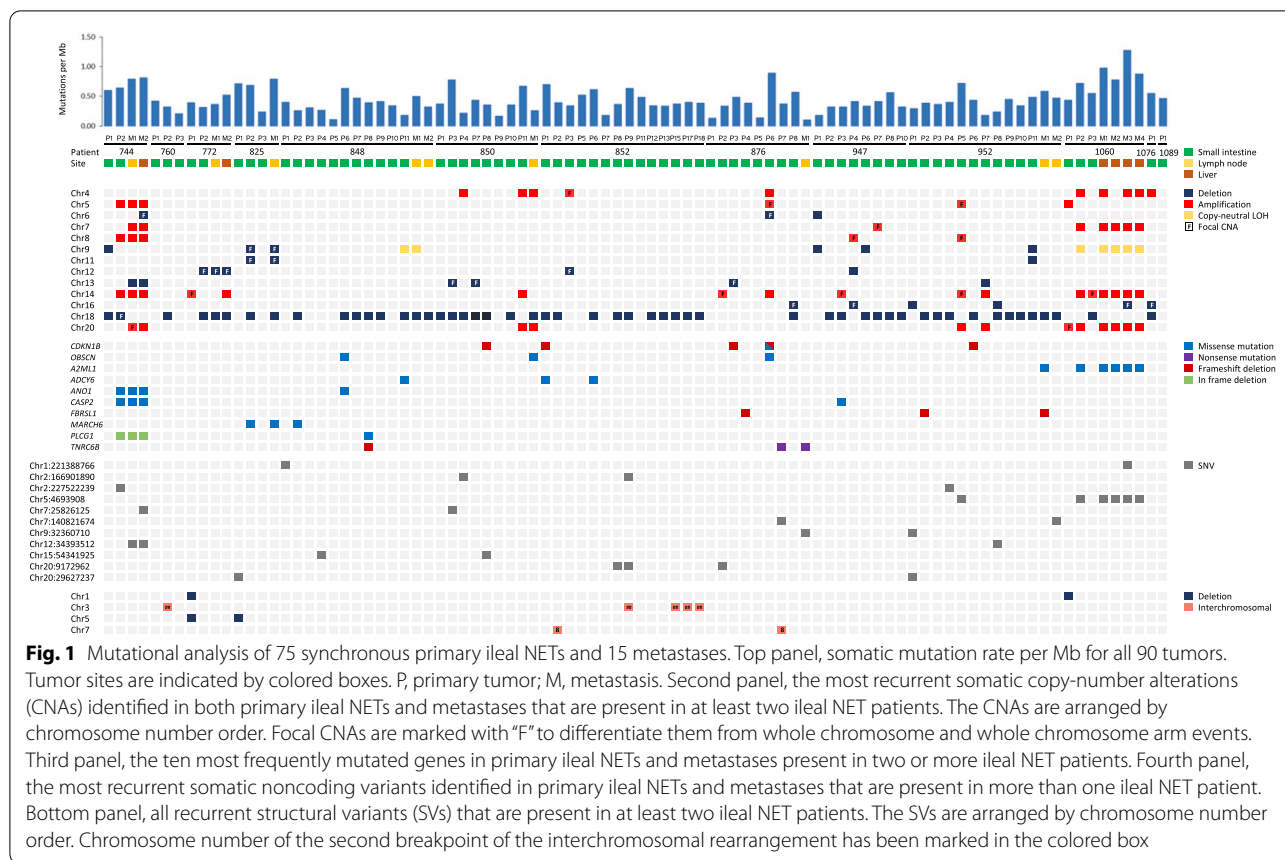
### Patient cohort

Patient cohort characteristics are detailed in Table 1. Most were white males (8/13, 62%). The median age at surgery was 64 years (range 53–75), and all patients had developed metastatic disease. The number of synchronous primary tumors varied from two to 18 among the multifocal ileal NET patients. The total number of metastases per patient was unknown. WGS was successfully performed for 75/91 (82%) primary ileal NETs and 15 metastases, including nine lymph node and six liver metastases (Additional file 1: Table S1).

### Mutational landscape of primary ileal NETs and metastases

WGS data analysis identified 124,550 somatic SNVs and indels across all 90 sequenced tumor samples, consisting of 1447 coding region variants and 123,103 noncoding region variants (Additional file 1: Table S1). The average somatic mutation burden was 0.41 mutations per megabase (mut/Mb) per primary ileal NET (range 0.11–0.89) and 0.63 mut/Mb per metastasis (range 0.10–1.27) (Fig. 1, Additional file 1: Table S1). The mean number of coding variants per sample was 15 in primary ileal NETs (range 2–71) and 20 in metastases (range 7–35), whereas the mean number of noncoding variants per sample was 1256 (range 346–2736) and 1926 (range 288–3914), respectively (Additional file 1: Table S1). Most of the coding variants were either missense (65%) or silent mutations (27%) (Additional file 2: Fig. S1, Additional file 3: Table S2).

Copy-number analysis identified 107 regions of somatic deletion, 101 regions of somatic amplification and nine copy-neutral LOH events across all 90 tumor samples (Additional file 1: Tables S1 and S3). The majority of the CNAs were either whole chromosome or chromosome arm events (58%) and present in both primary ileal NETs and metastases (Additional file 2: Fig. S2). The mean number of CNAs per sample was 2.15 in primary



ileal NETs (range 0–15) and 3.73 in metastases (0–10) (Additional file 1: Table S1). We also identified 147 intra-chromosomal (47%) and 168 interchromosomal somatic SVs (53%) across all 90 tumor samples (Additional file 1: Tables S1 and S4). Intrachromosomal SVs included deletions (33%), duplications (31%), translocations (24%), and inversions (13%). The mean number of SVs per sample was 3.68 in primary ileal NETs (range 0–23) and 2.60 in metastases (0–12) (Additional file 1: Table S1). The higher mean value in primary ileal NETs can be explained by a few individual tumors that harbored markedly more SVs than the rest of the tumors.

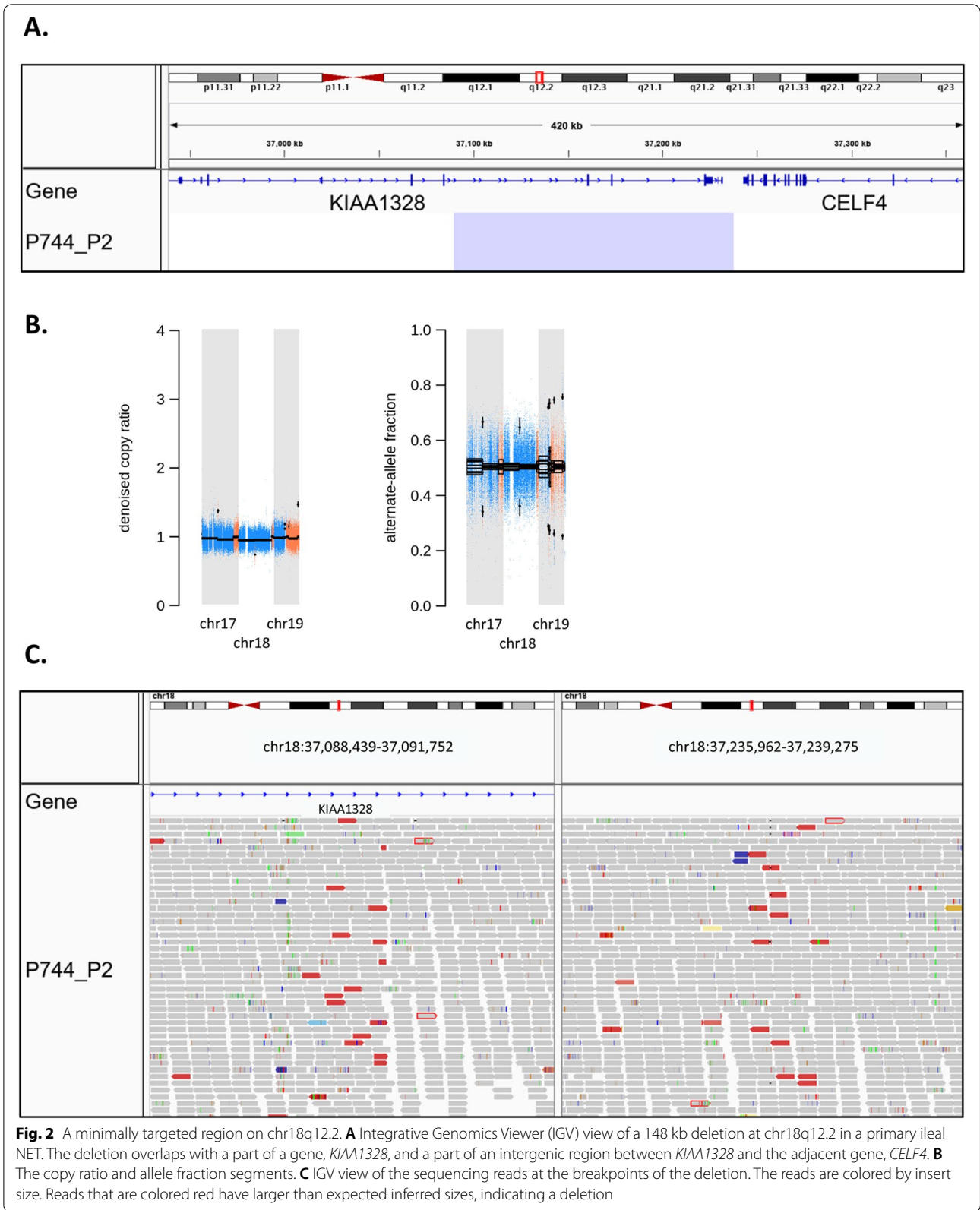
**CNAs are the most recurrent somatic alterations in multifocal ileal NET patients**

The most recurrent somatic genomic alterations in primary ileal NETs and metastases were whole chromosome and chromosome arm events. Chr18 LOH (51/90, 57%) was clearly the most frequent CNA followed by amplifications of chr14 (12/90, 13%), 20 (10/90, 11%), 4 (9/90, 10%), and 7 (7/90, 8%) (Fig. 1, Additional file 1: Table S3). Additionally, six minimally targeted regions of size <5 Mb were identified, 4p15.2, 10q11.21, 14q32.2–32, 18q12.2, 19p13.11, and 20p13-12.3, each present in tumors of at

least two multifocal ileal NET patients (Additional file 2: Fig. S3). One of the regions, 18q12.2, contained only a part of a single gene, *KIAA1328*, resulting in loss of the last three exons of the gene, as well as a part of an intergenic region between *KIAA1328* and the adjacent gene, *CELF4*, which harbors a lncRNA (*AC015961.1*) and a candidate cis-regulatory element (cCRE) of CTCF-only group as reported by the ENCODE Encyclopedia [39] (Fig. 2). All the other regions were comprised of multiple genes, including a few known cancer genes according to COSMIC Cancer Gene Census v92: *SLC34A2* on 4p15.2, *RET* on 10q11.21, *HSP90AA1* on 14q32.31, and *JAK3* on 19p13.11. No additional coding region variants, however, were observed in these genes.

**TNRC6B is a candidate tumor suppressor gene in SI-NETs**

We also identified 96 recurrently mutated genes harboring nonsynonymous mutations across all tumor samples (Additional file 2: Fig. S4, Additional file 3: Table S2). Many of the genes (66/96, 69%) were mutated in both primary ileal NETs and metastases, though, typically only in the tumors of one ileal NET patient at a time (Additional file 2: Fig. S2 and S4). Only two genes among all recurrently mutated genes, *CDKN1B* (5/90, 6%) and



*OBSCN* (3/90, 3%), displayed mutations in the tumors of more than two ileal NET patients (Fig. 1). Furthermore, we identified only one recurrently mutated gene, *DROSHA*, that harbored somatic alterations affecting the same exact genomic location in tumors from more than one ileal NET patient. Two tumors from two different patients (P852\_P15 and P952\_P10) displayed different missense mutations in this location: c.1282G>A, p.D428N and c.1282G>C, p.D428H (Additional file 2: Fig. S4). This location is not a known mutation site in *DROSHA*. A clear majority of the somatic alterations in the recurrently mutated genes were missense mutations (85%) (Additional file 2: Fig. S4). We observed three genes that harbored frameshift deletions and/or nonsense mutations in tumors from multiple ileal NET patients: *CDKN1B*, *FBRSL1*, and *TNRC6B* (Fig. 1). Additional larger deletions affecting *CDKN1B* and *TNRC6B* were identified in five primary ileal NETs in both cases (Additional file 2: Fig. S5). *CDKN1B* was also affected by copy-neutral LOH in one primary ileal NET. Lastly, ten out of 96 (10%) recurrently mutated genes were known cancer genes according to COSMIC Cancer Gene Census v92: *ARID2*, *BCL11B*, *BCOR*, *CDKN1B*, *DROSHA*, *FAT4*, *HIP1*, *NUTM1*, *PLCG1*, and *ROBO2* (Additional file 2: Fig. S4). Nonsynonymous mutations were also identified in 51 additional cancer genes in single primary ileal NETs and metastases, including *ALK*, *DAXX*, *KRAS*, and *MEN1* (Additional file 2: Fig. S6).

To identify genes showing statistical evidence of positive selection for mutations in SI-NETs, we performed a mutation significance analysis for somatic coding region SNVs and indels. Due to the low somatic mutation rate in SI-NETs, we combined our data with previously published high-throughput sequencing data of SI-NETs [8, 16], resulting in 176 tumors from 99 SI-NET patients. The analysis was performed for both patient- and tumor-level data separately. *CDKN1B* was the most significant gene in both analyses ( $P_{adj}=6.3 \times 10^{-12}$ , patient-level data;  $P_{adj}=1.5 \times 10^{-11}$ , tumor-level data) (Additional file 1: Tables S5 and S6). The patient-based analysis identified one additional significant gene, *ZNF845* ( $P_{adj}=2.6 \times 10^{-2}$ ), whereas the tumor-based analysis identified four, *TNRC6B* ( $P_{adj}=2.9 \times 10^{-4}$ ), *ZNF780B* ( $P_{adj}=4.2 \times 10^{-2}$ ), *RPP30* ( $P_{adj}=6.4 \times 10^{-2}$ ), and *CASQ1* ( $P_{adj}=8.7 \times 10^{-2}$ ). Only two significant genes, *CDKN1B* and *TNRC6B*, were mutated across the studies. A closer look at previous high-throughput sequencing data of SI-NETs revealed three somatic nonsense mutations and two frameshift deletions in *TNRC6B* among 195 sequenced sample pairs, as well as ten deletions of chr22 [7, 8, 16, 20]. Together with our data, *TNRC6B* is affected in approximately 8% of SI-NETs and metastases.

### The majority of recurrent noncoding region and structural variants are patient specific

In terms of the noncoding genome, recurrent noncoding variants affecting the same exact genomic location were observed in both primary ileal NETs and metastases, the latter harboring most of the variants (Fig. 1, Additional file 2: Fig. S2). As in the case of recurrently mutated genes, the majority of the recurrent noncoding variants (6505/6533, 99.6%) were present in the tumors of only one ileal NET patient at a time (Additional file 3: Table S2). Further visualization of the remaining 28 variants resulted in the identification of 18 somatic recurrent noncoding variants (0.3%) that were displayed concurrently in the tumors of more than one ileal NET patient (Additional file 1: Table S7). We also detected 27 recurrent SVs, four of which were present in tumors from more than one ileal NET patient (Fig. 1, Additional file 1: Table S4). Two of the recurrent SVs were deletions affecting genes *PEX14* and *CASZ1* on chr1 and *SGCD* on chr5, respectively. The breakpoints of the remaining two recurrent interchromosomal rearrangements located in introns of the following genes: *AC034195.1* on chr3, *WWOX* on chr16, *HOXA3* on chr7, and *SGCZ* on chr8.

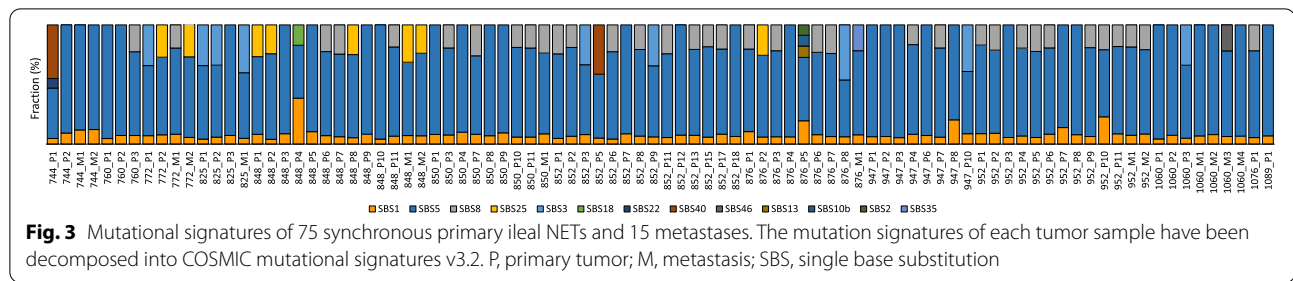
### CDKN1B frameshift deletions cause a statistically significant cluster of indels on chr12

In the absence of apparent recurrent genomic driver alterations among the tumors of multifocal ileal NET patients, we next looked for potential enrichment of SNVs and indels in the genomes of primary ileal NETs and metastases. We identified one statistically significant cluster of SNVs on chr2 ( $P_{adj}=7.2 \times 10^{-2}$ ) caused by a somatic intronic SNV in *FOXN2* (g.chr2:48340593 T>C), which was present in eight tumors from one ileal NET patient (Additional file 2: Fig. S7, Additional file 3: Table S2). Also, one statistically significant cluster of indels was observed on chr12 ( $P_{adj}=2.4 \times 10^{-2}$ ), consisting of four *CDKN1B* frameshift deletions from three different ileal NET patients and one intronic deletion in *APOLD1* (Additional file 2: Fig. S7, Additional file 3: Table S2).

### No evidence for significantly mutated signaling pathways in multifocal ileal NET patients

Due to the lack of recurrently mutated genes among the multifocal ileal NET patients, we examined if the mutated genes fell into the same signaling pathways. We identified 13 pathways with nominally significant  $P$ -value ( $<0.05$ ); however, none of the pathways remained significant after multiple testing correction (Additional file 1: Table S8).





### Mutational signatures SBS1 and SBS5 occur in all primary ileal NETs and metastases

The most common substitution types were transitions C > T and T > C that accounted for 30% and 24% of all the detected mutations, respectively. SBS signatures 1 and 5 were observed in all tumor samples and the only signatures present in 36% of the tumors, the latter contributing the most to the mutational profiles of primary ileal NETs and metastases (Fig. 3). Both signatures are considered clock-like signatures, indicating that the number of mutations correlates with the age of the individual. Additionally, SBS signatures 3 (10%), 8 (39%), and 25 (9%) were identified in multiple tumors from various ileal NET patients and the signatures were mutually exclusive. SBS8 and SBS25 are unknown signatures, whereas SBS3 has been associated with defective homologous recombination-based DNA damage repair. No clear differences were observed between the mutational signatures of primary ileal NETs and metastases. However, the patterns and fractions of the signatures varied between the tumors from the same patient.

### Synchronous primary tumors arise independently from the normal ileal mucosa

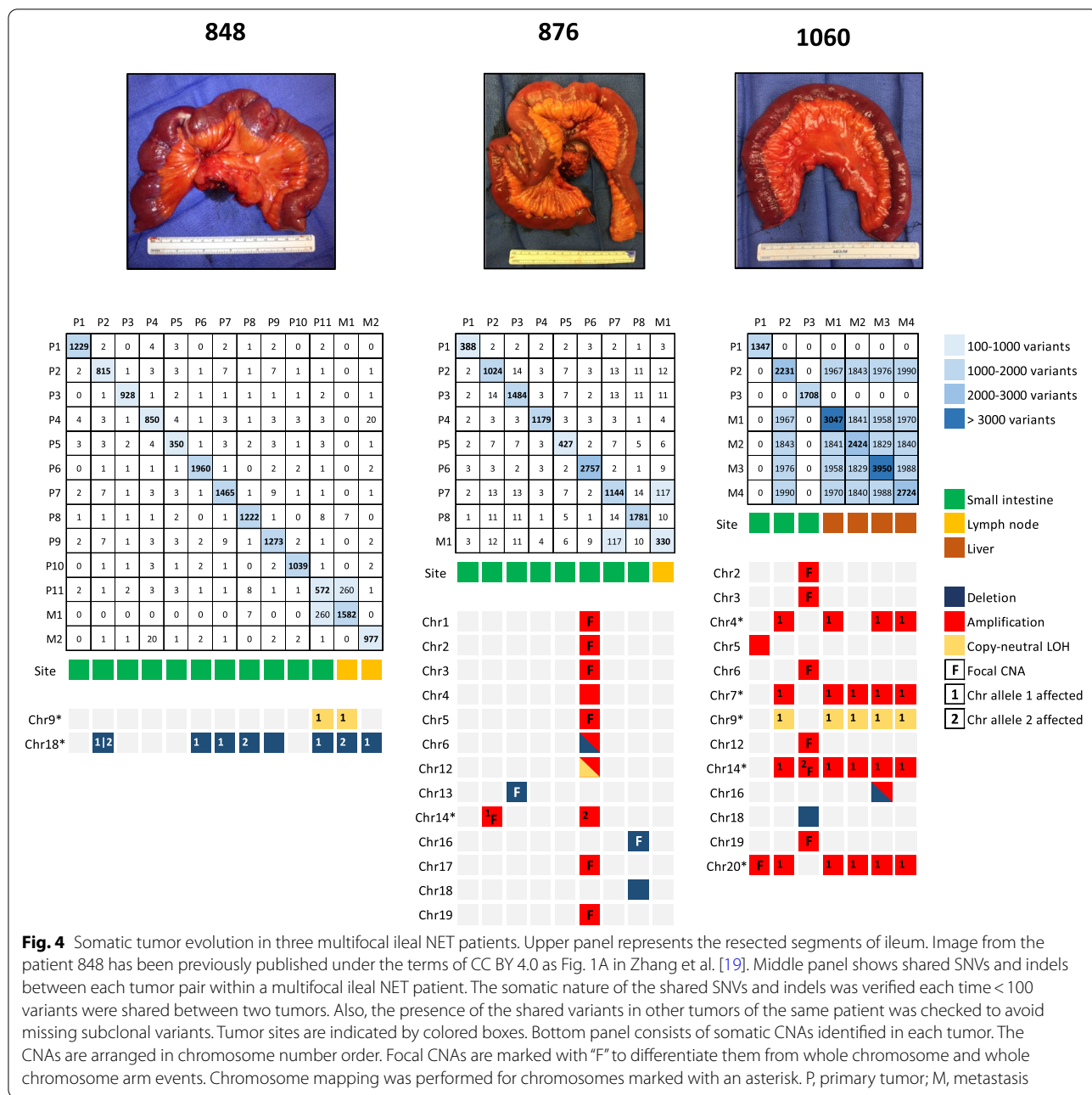
Next, we studied the clonal relationship of the tumors within each multifocal ileal NET patient. WGS data of multiple tumor samples were available from 11 patients. Pairwise comparison of the numbers of shared SNVs and indels indicated that multifocal primary tumors within each patient arise independently from the normal ileal mucosa (Fig. 4, Additional file 2: Fig. S8). On average, only 1.9 (0.08%) somatic SNVs and/or indels (range 0–14; 0–0.7%) were shared between the primary tumors. All shared variants, except one in *JADE2* (c.1465C > T, p.R489C) observed in two out of eight primary tumors in patient 947, were noncoding variants. WGS data of metastases were available from eight of the patients. Based on the numbers of shared SNVs and indels, we identified the putative primary tumors of origin for the sequenced metastatic tumors in seven patients as follows (Fig. 4, Additional file 2: Fig. S8): P744\_P2 for P744\_M1 and P744\_M2, P772\_P2 for P772\_M1 and P772\_M2, P825\_P2 for P825\_M1, P848\_P11 for P848\_M1, P876\_P7

for P876\_M1, P952\_P2 for P952\_M1 and P952\_P10 for P952\_M2, and P1060\_P2 for P1060\_M1–4. None of the primary tumors in patient 850 shared SNVs or indels with the metastasis (Additional file 2: Fig. S8). However, WGS data were not available from three of the patient's primary tumors, suggesting that one of those tumors could be the origin for the metastasis. Intriguingly, two different dissemination patterns were identified in patients with multiple metastases. Metastases were either clonal, originating from a single primary tumor, or independent, originating from two separate primary tumors within a patient (Fig. 4, Additional file 2: Fig. S8). We did not observe any common somatic alterations or mutational signatures between the metastasized primary tumors that would separate them from the other primary tumors.

Although CNAs were the most recurrent somatic genomic alterations identified in primary ileal NETs and metastases, none of them were present in all tumors of the same patient. Additional chromosome mapping of the most frequent CNAs among the tumor samples revealed that primary tumors and metastases from the same patient can present different CNA patterns, including gains or losses of the same parental allele, a different parental allele consistent across the whole chromosomal arms, or different parental alleles in the short (p) and long (q) arms (Additional file 1: Table S9). Gains or losses of different parental alleles were identified in chromosomes 4, 14, and 20, as well as chromosomes 13 and 18 in the multifocal ileal NET patients, respectively (Fig. 4, Additional file 2: Fig. S8). The clear majority of the observed CNA patterns were concordant with the results from the pairwise comparisons of the amounts of shared SNVs and indels between the tumors within each patient. However, there were two occasions, one in patient 744 and the other in patient 848, where clonal tumors had gained or lost different parental alleles in a subset of CNAs (Fig. 4, Additional file 2: Fig. S8).

### Discussion

To better understand the molecular mechanisms underlying the growth and development of multifocal SI-NETs, we whole genome sequenced 75 primary tumors, 15



**Fig. 4** Somatic tumor evolution in three multifocal ileal NET patients. Upper panel represents the resected segments of ileum. Image from the patient 848 has been previously published under the terms of CC BY 4.0 as Fig. 1A in Zhang et al. [19]. Middle panel shows shared SNVs and indels between each tumor pair within a multifocal ileal NET patient. The somatic nature of the shared SNVs and indels was verified each time < 100 variants were shared between two tumors. Also, the presence of the shared variants in other tumors of the same patient was checked to avoid missing subclonal variants. Tumor sites are indicated by colored boxes. Bottom panel consists of somatic CNAs identified in each tumor. The CNAs are arranged in chromosome number order. Focal CNAs are marked with "F" to differentiate them from whole chromosome and whole chromosome arm events. Chromosome mapping was performed for chromosomes marked with an asterisk. P, primary tumor; M, metastasis

metastases, and the corresponding normal tissue samples from 13 multifocal ileal NET patients. Low somatic mutation rate was observed across all tumor samples, which is in line with the previous literature [7–11, 16]. CNAs were the most recurrent somatic alterations identified. Chr18 LOH was present in 57% of primary tumors, representing the most recurrent somatic alteration in multifocal ileal NETs, along with amplifications of chromosomes 4, 7, 14, and 20 in 9–13% of tumors. These results are consistent with previous high-throughput sequencing studies

of both uni- and multifocal ileal NETs [7, 8, 10, 19, 20]. In addition to the previous literature, we observed copy-neutral LOH of chr9 in 8% of tumors. Interestingly, we identified a small deletion on 18q12.2 in one out of 44 primary ileal NETs displaying chr18q LOH. The deletion was supported by both the CNA and SV data, and it primarily affected a part of *KIAA1328*, but also an intergenic region between *KIAA1328* and the adjacent gene, *CELF4*. *KIAA1328* encodes a protein called hinderin, which has been shown to bind to SMC3, a subunit of the cohesin

complex, which plays an important role in mediating sister chromatid cohesion, homologous recombination, and DNA looping [40]. Currently, there are no functional data available for the lncRNA (AC015961.1) or cCRE that are located in the intergenic region affected by the deletion. One of the previous high-throughput sequencing studies of SI-NETs has reported another small deletion on 18q12.2 in one out of 22 primary NETs displaying chr18q LOH in the region, as well as a deletion breakpoint at 18q12.2 in a second primary NET [7]. Both the deletion and the breakpoint affected *CELF4*, which encodes an RNA-binding protein (RBP) implicated in the regulation of pre-mRNA alternative splicing [41]. The protein is used as a part of RBP-based models to predict the survival of colorectal cancer (CRC) patients [42]. Also, a rare intronic germline variant in *CELF4* has been associated with CRC risk [43]. Neither *KIAA1328* nor *CELF4* have previously been implicated in small bowel cancer.

Along with the recurrent CNAs, we observed recurrent frameshift and/or nonsense mutations, as well as larger deletions affecting two genes, *CDKN1B* and *TNRC6B*, in 11% and 9% of tumors across multiple ileal NET patients, respectively. *CDKN1B* has previously been implicated as a potential tumor suppressor gene in SI-NETs, linking cell cycle dysregulation in their tumorigenesis [8, 15]. In addition to the previous studies, we demonstrated that *CDKN1B* frameshift deletions cause a statistically significant enrichment of indels on chr12 in multifocal ileal NETs, strengthening the role of *CDKN1B* as a driver gene in SI-NETs and suggesting that the gene displays a specific mutational pattern in the tumors. Like *CDKN1B*, *TNRC6B* was identified as one of the genes showing statistical evidence of positive selection for mutations in SI-NETs, suggesting a candidate tumor suppressor role for this gene in a subset of tumors. Trinucleotide repeat containing 6 (TNRC6) proteins, including TNRC6A, TNRC6B, and TNRC6C, are important for miRNA- and siRNA-mediated gene silencing through their functions within RNA-induced silencing complex [44]. Downregulation of TNRC6B has been suggested to contribute to tumorigenesis of different cancers, such as prostate cancer and lung adenocarcinoma [45, 46]. Inhibition of TNRC6B has been shown to lead to acceleration of cell proliferation and deceleration of cell adhesion in hepatoma cell lines [47]. We did not observe other apparent candidate driver genes among the multifocal ileal NET patients.

Further comparison of the tumors within each multifocal ileal NET patient confirmed that multifocal primary tumors arise independently from the normal ileal mucosa. The primary tumors rarely shared any SNVs, indels, or SVs, and the observed recurrent CNAs were not present in all tumors of the same patient. Additional

chromosome mapping revealed that tumors from the same patient can display different CNA patterns, including gains or losses of either parental allele. In the majority of the cases, where the tumors had gained or lost a different parental allele, the change was consistent across the whole chromosomal arms. We observed, however, one patient with a tumor that had lost different parental alleles in the short (p) and long (q) arms of chr18. This case has been discussed in our previous paper, where we speculate that there are two different mechanisms that may lead to this event; either homologous recombination after which one of the recombined copies is lost during tumorigenesis, or two independent genomic events [19]. We also detected variation in the patterns and fractions of mutational signatures between tumors from the same patient. Intriguingly, we show that multiple metastases in the same patient can originate from one or several primary tumors. Our results are corroborated by recent findings of Elias et al. [20] and have an important clinical implication, supporting the concept that identification of all multifocal primary tumors by careful palpation of the entire jejunum and ileum is essential at the time of surgery [48]. We could not detect common somatic alterations among the metastasized primary tumors in our data.

Together with previous high-throughput sequencing studies on multifocal SI-NETs [19, 20], our data confirm the lack of shared driver genes among these tumors, suggesting that multifocal ileal NETs are not driven by genomic alterations that are detectable with our current genome sequencing and analysis approaches. Additionally, our previous discovery of distinct chr18 LOH patterns in multifocal ileal NETs from the same patient excludes the possibility of a germline loss-of-function mutation in a tumor suppressor gene on chr18 as a cause for SI-NETs [19]. These findings suggest that SI-NETs could be mainly driven by epigenetic mechanisms, or alternatively arise from morphologically normal small intestine as a result of a field cancerization. Also, the tumor microenvironment in the ileum may play a role in the growth and development of SI-NETs.

## Conclusions

Our study indicates notable genomic diversity among multifocal ileal NETs, suggesting that these tumors develop independently. We identified potential loss-of-function gene alterations in *TNRC6B*, a candidate tumor suppressor gene in a small subset of ileal NETs, as well as a minimally deleted region on chr18q12.2, providing new candidate genes to study to better understand the molecular mechanisms of SI-NETs. Additionally, we observed that multiple metastases in the same patient can originate from either one or several primary tumors, which

highlights the need to identify and remove all primary tumors, which have the potential to metastasize. Altogether, our results suggest the tumorigenesis of SI-NETs is unlikely to be driven exclusively by genomic alterations and underscore the need of a deeper understanding of the molecular mechanisms that underlie SI-NETs and to apply that knowledge toward development of new and effective treatments.

### Abbreviations

CCRE: Candidate cis-regulatory element; CDKN1B: Cyclin-dependent kinase inhibitor 1B; CNA: Copy-number alteration; FDR: False discovery rate; Indel: Insertion-deletion mutation; LOH: Loss of heterozygosity; SBS: Single base substitution; SI-NET: Small intestinal neuroendocrine tumor; SNP: Single-nucleotide polymorphism; SNV: Single-nucleotide variant; SV: Structural variant; TNRC6B: Trinucleotide repeat containing 6 B; WGS: Whole genome sequencing.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-022-01083-1>.

**Additional file 1: Table S1.** WGS sample information and somatic genomic alteration counts. **Table S3.** Filtered copy-number alteration calls. **Table S4.** Filtered structural variant calls. **Table S5.** MutSig2CV results for patient-level information. **Table S6.** MutSig2CV results for tumor-level information. **Table S7.** Recurrent non-coding variants. **Table S8.** Results of the signaling pathway analysis. **Table S9.** Results of chromosome mapping analysis

**Additional file 2: Figure S1.** Classification of coding variants in 75 primary ileal NETs and 15 metastases. **Figure S2.** Overlap of somatic variation between primary ileal NETs and metastases. **Figure S3.** Minimally targeted regions of size < 5Mb. **Figure S4.** Recurrently mutated genes in 75 primary ileal NETs and 15 metastases. **Figure S5.** Deletions affecting *CDKN1B* and *TNRC6B*. **Figure S6.** Known cancer genes mutated in single primary ileal NETs and metastases. **Figure S7.** Statistically enriched regions of SNVs and indels in 75 primary ileal NETs and 15 metastases. **Figure S8.** Somatic tumor evolution in multifocal ileal NET patients.

**Additional file 3: Table S2.** Filtered SNV and indel calls.

### Acknowledgements

The authors would like to acknowledge the Broad Institute's Genomics Platform for whole genome sequencing and preprocessing of the raw sequencing data.

### Authors' contributions

Conceptualization, M.M., E.N., C.T., and N.M.; Formal Analysis, N.M. and M.Z.; Investigation, N.M., M.Z., Z.Z., E.P., and G.E.K.; Writing—Original Draft, N.M.; Writing—Review & Editing, M.M., E.N., C.T., and M.Z.; Funding Acquisition, M.M., E.N., C.T., and N.M.; Resources, E.N. and Y.K.; Supervision, M.M. and N.M. All authors read and approved the final manuscript.

### Funding

This study was supported by the Neuroendocrine Tumor Research Foundation.

### Availability of data and materials

The WGS data generated and analyzed in this study are accessible at the European Genome-phenome Archive (EGA) website under the accession number EGAS00001006294 (<https://ega-archive.org/studies/EGAS00001006294>) [22].

### Declarations

#### Ethics approval and consent to participate

Each patient provided informed consent to participate in this study in accordance with the protocols approved by the Institutional Review Boards of the Dana-Farber Cancer Institute (IRB#: 17–349) and the University of California San Francisco (IRB#: 13–12574). Research performed conformed to the principles of the Helsinki Declaration.

#### Consent for publication

Not applicable.

#### Competing interests

Dr. Meyerson declares the following general conflicts of interest: research support from Bayer, Ono, and Janssen; patent licensing royalties from Bayer and LabCorp; and serving as scientific advisory board member and consultant for Interline and Isabl. The remaining authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215, USA. <sup>2</sup>Cancer Program, Broad Institute of Harvard and MIT, Cambridge, MA, USA. <sup>3</sup>Department of Surgery, University of California, San Francisco, CA, USA. <sup>4</sup>Department of Pathology, University of California, San Francisco, CA, USA. <sup>5</sup>Research Department of Oncology, UCL Cancer Institute, London, UK. <sup>6</sup>School of Medicine and Health, University of Exeter, RILD Building, Exeter, UK. <sup>7</sup>Departments of Genetics and Medicine, Harvard Medical School, Boston, MA, USA.

Received: 18 October 2021 Accepted: 6 July 2022

Published online: 03 August 2022

### References

1. Yao JC, Hassan M, Phan A, Dagohoy C, Leary C, Mares JE, et al. One hundred years after "carcinoid": epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *J Clin Oncol*. 2008;26:3063–72. <https://doi.org/10.1200/JCO.2007.15.4377>.
2. Bilimoria KY, Bentrem DJ, Wayne JD, Ko CY, Bennett CL, Talamonti MS. Small bowel cancer in the United States: changes in epidemiology, treatment, and survival over the last 20 years. *Ann Surg*. 2009;249:63–71. <https://doi.org/10.1097/SLA.0b013e31818e4641>.
3. Landerholm K, Falkmer S, Järhult J. Epidemiology of small bowel carcinoids in a defined population. *World J Surg*. 2010;34:1500–5. <https://doi.org/10.1007/s00268-010-0519-z>.
4. Maggard MA, O'Connell JB, Ko CY. Updated population-based review of carcinoid tumors. *Ann Surg*. 2004;240:117–22. <https://doi.org/10.1097/01.sla.0000129342.67174.67>.
5. Modlin IM, Champaneria MC, Chan AK, Kidd M. A three-decade analysis of 3,911 small intestinal neuroendocrine tumors: the rapid pace of no progress. *Am J Gastroenterol*. 2007;102:1464–73. <https://doi.org/10.1111/j.1572-0241.2007.01185.x>.
6. Modlin IM, Lye KD, Kidd M. A 5-decade analysis of 13,715 carcinoid tumors. *Cancer*. 2003;97:934–59. <https://doi.org/10.1002/cncr.11105>.
7. Banck MS, Kanwar R, Kulkarni AA, Boora GK, Metge F, Kipp BR, et al. The genomic landscape of small intestine neuroendocrine tumors. *J Clin Invest*. 2013;123:2502–8. <https://doi.org/10.1172/JCI67963>.
8. Francis JM, Kiezun A, Ramos AH, Serra S, Pedamallu CS, Qian ZR, et al. Somatic mutation of CDKN1B in small intestine neuroendocrine tumors. *Nat Genet*. 2013;45:1483–6. <https://doi.org/10.1038/ng.2821>.
9. Karpathakis A, Dibra H, Pipinikas C, Feber A, Morris T, Francis J, et al. Prognostic impact of novel molecular subtypes of small intestinal neuroendocrine tumor. *Clin Cancer Res*. 2016;22:250–8. <https://doi.org/10.1158/1078-0432.CCR-15-0373>.
10. Walter D, Harter PN, Battke F, Winkelmann R, Schneider M, Holzer K, et al. Genetic heterogeneity of primary lesion and metastasis in small intestine neuroendocrine tumors. *Sci Rep*. 2018;8:3811. <https://doi.org/10.1038/s41598-018-22115-0>.

11. Simbolo M, Vicentini C, Mafficini A, Fassan M, Pedron S, Corbo V, et al. Mutational and copy number asset of primary sporadic neuroendocrine tumors of the small intestine. *Virchows Arch.* 2018;473:709–17. <https://doi.org/10.1007/s00428-018-2450-x>.
12. Kytölä S, Höög A, Nord B, Cedermarck B, Frisk T, Larsson C, et al. Comparative genomic hybridization identifies loss of 18q22-qter as an early and specific event in tumorigenesis of midgut carcinoids. *Am J Pathol.* 2001;158:1803–8. <https://doi.org/10.1016/S00428-018-2450-x>.
13. Löllgen RM, Hessman O, Szabo E, Westin G, Akerström G. Chromosome 18 deletions are common events in classical midgut carcinoid tumors. *Int J Cancer.* 2001;92:812–5. <https://doi.org/10.1002/ijc.1276> (PMID: 11351300).
14. Cunningham JL, Díaz de Ståhl T, Sjöblom T, Westin G, Dumanski JP, Janson ET. Common pathogenetic mechanism involving human chromosome 18 in familial and sporadic ileal carcinoid tumors. *Genes Chromosomes Cancer.* 2011;50:82–94. <https://doi.org/10.1002/gcc.20834>.
15. Crona J, Gustavsson T, Norlén O, Edfeldt K, Åkerström T, Westin G, et al. Somatic mutations and genetic heterogeneity at the CDKN1B locus in small intestinal neuroendocrine tumors. *Ann Surg Oncol.* 2015;22(Suppl 3):S1428–35. <https://doi.org/10.1245/s10434-014-4351-9>.
16. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analysis of metastatic solid tumours. *Nature.* 2019;575:210–6. <https://doi.org/10.1038/s41586-019-1689-y>.
17. Kim JY, Hong SM. Recent updates on neuroendocrine tumors from the gastrointestinal and pancreatobiliary tracts. *Arch Pathol Lab Med.* 2016;140:437–48. <https://doi.org/10.5858/arpa.2015-0314-RA>.
18. Gangi A, Siegel E, Barmparas G, Lo S, Jamil LH, Hendifar A, et al. Multifocality in small bowel neuroendocrine tumors. *J Gastrointest Surg.* 2018;22:303–9. <https://doi.org/10.1007/s11605-017-3586-8>.
19. Zhang Z, Mäkinen N, Kasai Y, Kim GE, Diosdado B, Nakakura E, et al. Patterns of chromosome 18 loss of heterozygosity in multifocal ileal neuroendocrine tumors. *Genes Chromosomes Cancer.* 2020;59:535–9. <https://doi.org/10.1002/gcc.22850>.
20. Elias E, Ardalán A, Lindberg M, Reinsbach S, Muth A, Nilsson O, et al. Independent somatic evolution underlies clustered neuroendocrine tumors in the human small intestine. *bioRxiv.* 2020 <https://doi.org/10.1101/2020.05.06.080499> version 3
21. Digestive System Tumours WHO Classification of Tumours. 5th ed. Edited by the WHO Classification of Tumours Editorial Board. Lyon: International Agency for Research on Cancer; 2019.
22. Mäkinen N, Zhou M, Zhang Z, Kasai Y, Perez E, Kim GE, et al. Whole genome sequencing reveals the independent clonal origin of multifocal ileal neuroendocrine tumors. EGAS00001006294, European Genome-phenome Archive (EGA). 2022. <https://ega-archive.org/studies/EGAS00001006294>. Accessed 25 Jun 2022.
23. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25:1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
24. Picard. <https://broadinstitute.github.io/picard/>
25. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43:491–8. <https://doi.org/10.1038/ng.806>.
26. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol.* 2013;31:213–9. <https://doi.org/10.1038/nbt.2514>.
27. Van der Auwera, G.A. and O'Connor, B.D. Genomics in the cloud: using Docker, GATK, and WDL in Terra. 1st ed. Sebastopol, MA: O'Reilly Media; 2020.
28. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics.* 2012;28:1811–7. <https://doi.org/10.1093/bioinformatics/bts271>.
29. Wala JA, Bandopadhyay P, Greenwald NF, O'Rourke R, Sharpe T, Stewart C, et al. SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Res.* 2018;28:581–91. <https://doi.org/10.1101/gr.221028.117>.
30. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature.* 2014;505:495–501. <https://doi.org/10.1038/nature12912>.
31. Imielinski M, Guo G, Meyerson M. Insertions and deletions target lineage-defining genes in human cancers. *Cell.* 2017;168:460–472.e14. <https://doi.org/10.1016/j.cell.2016.12.025>.
32. Tweedie S, Braschi B, Gray K, Jones TEM, Seal RL, Yates B, et al. GeneNames.org: the HGNC and VGNC resources in 2021. *Nucleic Acids Res.* 2021;49:D939–46. <https://doi.org/10.1093/nar/gkaa980>.
33. RepeatMasker. <http://repeatmasker.org/>
34. Fabregat A, Sidiropoulos K, Viteri G, Forner O, Marin-Garcia P, Arnau V, et al. Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics.* 2017;18:142. <https://doi.org/10.1186/s12859-017-1559-2>.
35. Bergstrom EN, Huang MN, Mahto U, Barnes M, Stratton MR, Rozen SG, et al. SigProfilerMatrixGenerator: a tool for visualizing and exploring patterns of small mutational events. *BMC Genomics.* 2019;20:685. <https://doi.org/10.1186/s12864-019-6041-2>.
36. AlexandrovLab/SigProfilerExtractor. <https://github.com/AlexandrovLab/SigProfilerExtractor>
37. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. *Nature.* 2020;578:94–101. <https://doi.org/10.1038/s41586-020-1943-3>.
38. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv.* 2017. <https://doi.org/10.1101/201178>.
39. ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shores N, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature.* 2020;583:699–710. <https://doi.org/10.1038/s41586-020-2493-4>.
40. Patel CA, Ghiselli G, Hinderin, a five-domains protein including coiled-coil motifs that binds to SMC3. *BMC Cell Biol.* 2005;6:3. <https://doi.org/10.1186/1471-2121-6-3>.
41. Ladd AN, Charlet N, Cooper TA. The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. *Mol Cell Biol.* 2001;21:1285–96. <https://doi.org/10.1128/MCB.21.4.1285-1296.2001>.
42. Miao Y, Zhang H, Su B, Wang J, Quan W, Li Q, et al. Construction and validation of an RNA-binding protein-associated prognostic model for colorectal cancer. *PeerJ.* 2021;9:e11219. <https://doi.org/10.7717/peerj.11219>.
43. Teerlink CC, Stevens J, Hernandez R, Facelli JC, Cannon-Albright LA. An intronic variant in the CELF4 gene is associated with risk for colorectal cancer. *Cancer Epidemiol.* 2021;72: 101941. <https://doi.org/10.1016/j.canep.2021.101941>.
44. Gebert LFR, MacRae IJ. Regulation of microRNA function in animals. *Nat Rev Mol Cell Biol.* 2019;20:21–37. <https://doi.org/10.1038/s41580-018-0045-7>.
45. Sun J, Zheng SL, Wiklund F, Isaacs SD, Li G, Wiley KE, et al. Sequence variants at 22q13 are associated with prostate cancer risk. *Cancer Res.* 2009;69:10–5. <https://doi.org/10.1158/0008-5472.CAN-08-3464>.
46. Chiosea S, Jelezcova E, Chandran U, Luo J, Mantha G, Sobol RW, et al. Overexpression of Dicer in precursor lesions of lung adenocarcinoma. *Cancer Res.* 2007;67:2345–50. <https://doi.org/10.1158/0008-5472.CAN-06-3533>.
47. Murakami Y, Tamori A, Itami S, Tanahashi T, Toyoda H, Tanaka M, et al. The expression level of miR-18b in hepatocellular carcinoma is associated with the grade of malignancy and prognosis. *BMC Cancer.* 2013;13:99. <https://doi.org/10.1186/1471-2407-13-99>.
48. Nakakura EK. Challenges staging neuroendocrine tumors of the pancreas, jejunum and ileum, and appendix. *Ann Surg Oncol.* 2018;25:591–3. <https://doi.org/10.1245/s10434-017-6026-9>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.