Independent somatic evolution underlies multifocal small intestine neuroendocrine tumors

Erik Elias[2,3]†, Arman Ardalan[1]†, Markus Lindberg[1], Susanne Reinsbach[1], Andreas Muth[2,3], Ola Nilsson[4,5], Yvonne Arvidsson[4] and Erik Larsson[1]*

[†]These authors contributed equally

[1]Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden.

[2]Endocrine and Sarcoma Surgery, Department of Surgery, Institute of Clinical Sciences, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden.

[3]Department of Endocrine and Sarcoma Surgery, Surgical Clinic SU/S, Sahlgrenska University Hospital, Gothenburg, Sweden.

[4]Department of Laboratory medicine, Institute of Biomedicine, Sahlgrenska Cancer Center, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden.

[5]Department of Pathology, Sahlgrenska University Hospital, Gothenburg, Sweden.

*Correspondence to Erik Larsson, Institute of Biomedicine, Box 440, SE-405 30, Gothenburg, Sweden. Email: erik.larsson@gu.se. Phone: +46 31 786 6942

Abstract

Small intestine neuroendocrine tumor (SI-NET), the most common cancer of the small bowel, often displays a curious multifocal phenotype with several intestinal tumors centered around a regional lymph node metastasis, yet the typical path of evolution of these lesions remains unclear. Here, we determined the complete genome sequences of 37 tumors from 5 patients with multifocal SI-NET, allowing elucidation of phylogenetic relationships between multiple intestinal tumors and metastases in individual patients. Notably, lack of shared somatic mutational events strongly supported that the intestinal tumors were of independent clonal origin. Furthermore, adjacent lymph node metastases arose from a single, typically centrally located, intestinal tumor. Thus, we propose that multifocal SI-NETs form primarily by independent somatic evolution rather than local metastatic spread, suggesting a contribution from of a cancer-priming local environmental factor.

Significance statement

In multifocal cancer, several tumors arise in the same tissue area. Typically, this is due local metastatic spread, eventually leading to multiple tumors arising from a unique precursor clone, or due to a common heritable oncogenic genetic variant shared among all cells in an individual. Here, we find tumor development is independently initiated in multiple separate cells in a cancer of the small intestine, leading to a development of a polyclonal group of tumors that are genetically unrelated in terms of somatic evolution. This implies a contribution from a yet unknown non-genetic factor, thus providing an important future research direction for an enigmatic form of cancer in which genetic driver events are generally lacking.

Introduction

Cancer is considered to be an evolutionary process involving cycles of random oncogenic mutations, selection and clonal expansion (1). In most malignant epithelial tumors (carcinomas), a single primary tumor will typically form, followed by metastases arising as cells disperse from the tumor. More rarely, carcinomas instead display *multifocality*, i.e. multiple separate tumors at the primary site, which is typically due to predisposing germline variants or local spread of cancer-primed mutated clones (2). Of particular interest are small intestine neuroendocrine tumors (SI-NETs), the most common cancer of the small intestine, where ~50% of cases display a striking multifocal phenotype that often involves 10 or more morphologically identical tumors clustered within a limited intestinal segment, commonly centered around a regional lymph node metastasis (3, 4). SI-NET often presents with distant metastases at diagnosis, yet the somatic mutational burden of SI-NET is low and there is an unusual absence of somatic driver mutations (5-7).

The evolutionary relationships that govern multiple intestinal tumors and metastases in SI-NET can give insight into the underlying biology, but earlier efforts to determine this have yielded conflicting results (8, 9). A study of copy number alterations (CNA) in multifocal SI-NET primary tumors found that loss of chromosome 18 (chr18), the most frequent CNA, can affect different chromosome homologs in different samples within a patient (10), compatible with an independent clonal origin or late loss of chr18. Alternatively, multifocal SI-NETs have been proposed to represent "drop metastases" originating from regional lymph nodes, consistent with the limited spatial distribution of the intestinal tumors (11). Adding to the challenge is the low mutational burden, which reduces the number of exonic genetic markers.

Results

To conclusively determine the evolutionary trajectory of multifocal SI-NET in a single individual, we performed whole genome sequencing (WGS) on 6 intestinal tumors, 3 adjacent lymph node metastases, 2 peritoneal metastases and a blood normal from a patient (Patient 1) that underwent surgery with curative intent (**Fig. 1a**). The sampled peritoneal metastases were located approx. 50 cm from the tumors, which were all found within a limited section of the small intestine. All lesions showed typical SI-NET morphology and stained positively for established diagnostic SI-NET markers (**Supplementary Fig. 1**). The samples were sequenced at an average coverage of 35.7-43.1× (**Supplementary Table 1**). This was followed by somatic mutation calling using strict filters including rigorous population variant filtering to avoid false positives in phylogenetic analyses (see **Methods**). Resulting genome-wide single nucleotide variant (SNV) burdens varied from 363 to 1,175 (0.13-0.42 per Mb; **Supplementary Table 1**).

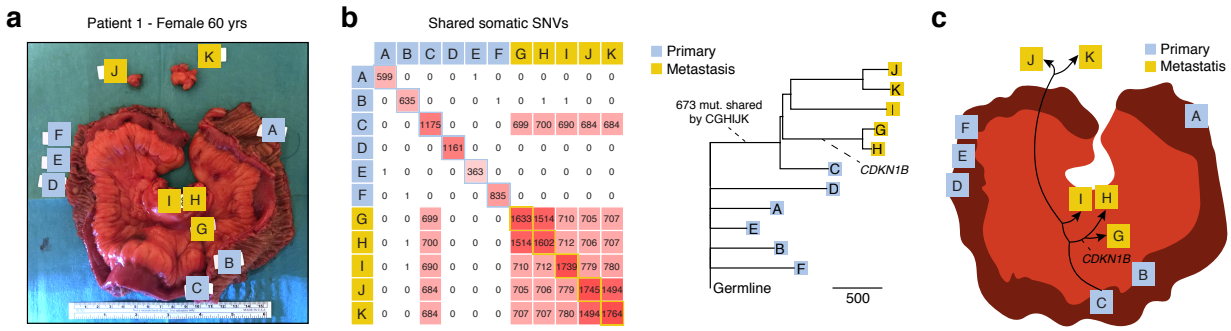


Figure 1. Whole genome sequencing of 11 primary tumors and metastases from a single SI-NET patient supports independent clonal evolution. (**a**) Section of the small intestine from a 60 year-old female patient harboring 6 primary tumors (A-F), 3 lymph node metastases (G-I) and 2 peritoneal metastases (J-K). (**b**) Pairwise analysis of shared somatic SNVs based on whole genome sequencing at 34.5-43.1× coverage. Primary tumor C and the 5 metastases shared a common set of 673 mutations. A maximum parsimony phylogenetic tree is shown, with the number of SNVs in each branch given by the scale marker. An indel in *CDKN1B*, a known driver events, is indicated. (**c**) Proposed model, where all metastases originate from a single primary tumor, and where all primaries are unrelated in terms of somatic evolution.

We next determined pairwise shared somatic SNVs between samples. Surprisingly, a common set of 673 SNVs was shared between a single primary tumor (denoted C) and all 5 metastases (G-K), while overlapping mutations were essentially lacking between the other samples (**Fig. 1b**). The results were thus not compatible with a common clonal origin for the primary tumors, expected to result in hundreds to thousands of shared mutations genome-wide in the case of a late-onset cancer. Instead, phylogenetic analysis supported that the primary tumors had developed independently, with a single, centrally located, tumor metastasizing first to the lymph nodes and then further to the peritoneum (**Fig. 1b-c**).

To validate the initial findings, we sampled multiple tumors from an additional SI-NET patient (Patient 2). Patient 2 had 30 macroscopically detectable tumors clustered within a limited intestinal segment that were isolated along with a single adjacent lymph node metastasis, all of which stained positively for SI-NET immunohistochemical markers. A subset of 11 samples with sufficient tumor material and purity, including the metastasis, were subjected to WGS at a 35.9-43.0× coverage (**Fig. 2a**, **Supplementary Table 1**, **Supplementary Fig. 2**). One sample had lower-than-expected mutational burden (144 SNVs), likely due to low sample purity, while others varied between 776 and 2,170 SNVs (0.28-0.78 per Mb; **Supplementary Table 1**). A single primary tumor (G) showed a striking relatedness to the metastasis (L; 1,219 shared somatic SNVs), while shared mutations were otherwise largely absent (**Fig. 2b**). Similar to Patient 1, the data therefore supported independent somatic evolution of multiple primary tumors, with a single, centrally located, tumor being the source of metastatic spread (**Fig. 1b-c**).

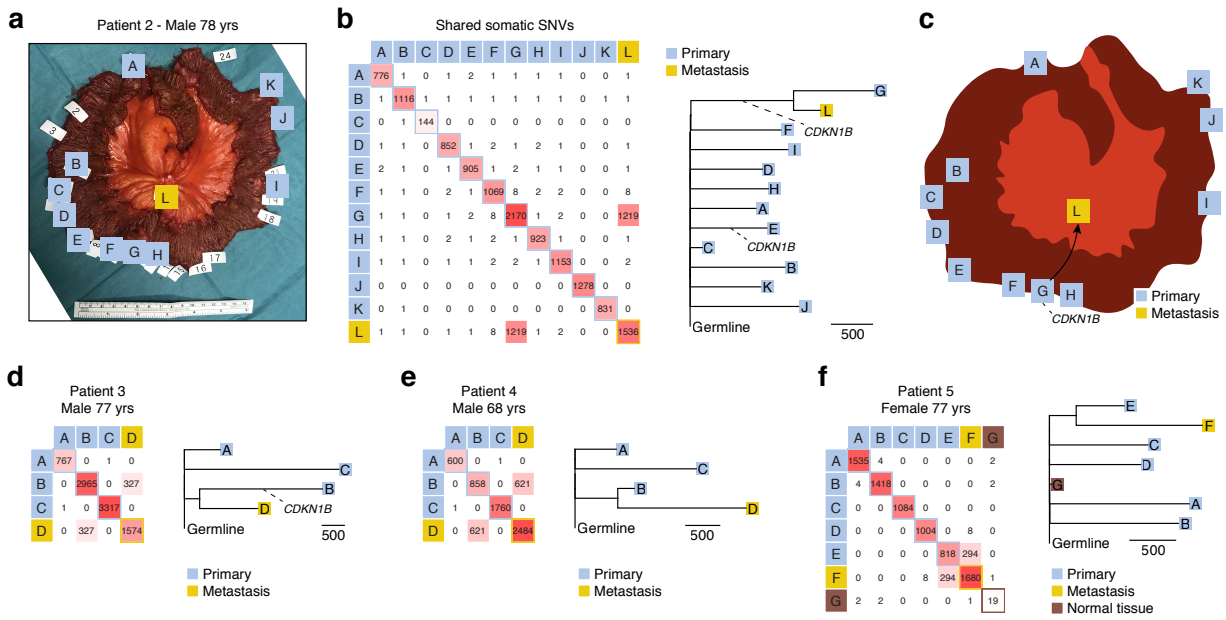


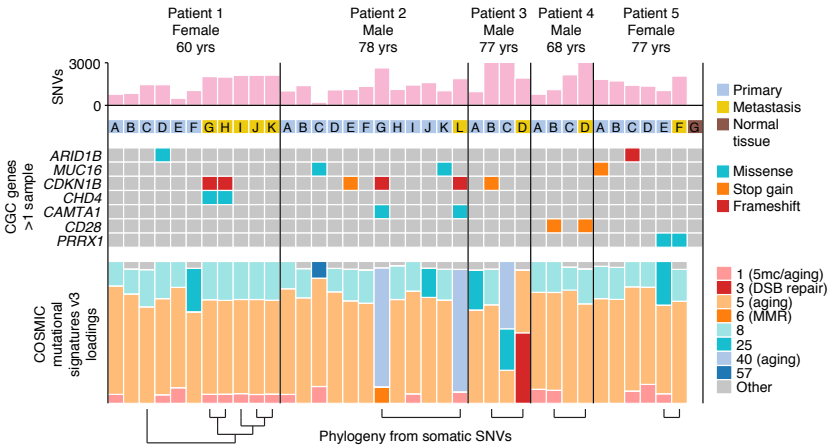
Figure 2. Whole genome sequencing of 26 primary tumors and metastases from 4 additional SI-NET patients. (**a**) Section of the small intestine from a 78 year-old male patient harboring 10 primary tumors (A-K) and one lymph node metastasis (L). White labels indicate all identified tumors while letters indicate sequenced samples with sufficient purity and tumor material. (**b**) Pairwise analysis of shared somatic SNVs based on whole genome sequencing at 36.9-43.0× coverage. Primary tumor G and the metastasis shared 1219 mutations. The known *CDKN1B* driver event is indicated in the phylogenetic tree (the number of SNVs in each branch is given by the scale marker). (**c**) Proposed model. (**d-f**) Similar to panel b, based on archival material from 3 additional patients, ages 68-77. Whole genome sequencing was performed at 36.7-45.0x coverage. A tissue normal sample was included for Patient 5 (G).

We finally retrieved previously sampled tumor tissue from 3 additional SI-NET patients (Patients 3-5) stored at a local biobank, each having between 3 and 5 primary tumors and one adjacent lymph node metastasis. A macroscopically normal small intestinal mucosa sample was also included in one case (Patient 5). All tumor samples stained positively for SI-NET markers (**Supplementary Fig. 3**). WGS was performed at 36.7-45.0× coverage, with between 600 and 3317 SNVs detected in each tumor (0.21-1.18 per Mb). In comparison, 19 SNVs were called genome-wide in the tissue normal, where widespread clonal somatic mutations are not expected. Mirroring the earlier results, all 3 patients displayed strong overlaps in the somatic mutational profile of the metastasis and a single intestinal tumor, ranging from 294 to 621 SNVs, while shared mutations were otherwise largely absent (**Fig. 2d-f**).

False shared somatic SNVs can arise due to the fact that all samples from a given patient use the same blood normal sequencing data, in particular in whole genome analyses. However, only a small number of additional common SNVs were observed beyond the main genetic relationships (**Fig. 1b** and **Fig. 2b,d-f**), many of which could be dismissed as false positives by manual inspection (**Supplementary Table 2**). In Patient 5, one additional tumor (D) shared 8 credible SNVs with the metastasis (F) (**Fig. 2f**). This overlap grew to 70 when high-confidence variants were whitelisted for relaxed calling in other samples, allowing more sensitive detection of subclonal shared SNVs (**Supplementary Fig. 4**). Notably, these variants were present as low variant allele frequency (VAF) traces in the metastasis while having normal VAFs in the tumor, compatible with contamination during sample handling or, less likely, metastatic spread to the same lymph node from two tumors with a minor contribution from D (**Supplementary Fig. 5**). In Patient 2, 8 SNVs were shared between the main metastatic primary (G),

129 an adjacent (5 mm apart) primary (F) and the metastasis (L) (**Fig. 2b**; **Supplementary Fig. 6**). All had
130 normal VAF distributions in all 3 samples arguing against contamination **(Supplementary Fig. 7**) and
131 all but one passed manual inspection (**Supplementary Table 2**). While too small to represent common
132 clonal origin in a late-onset cancer, we speculate that overlaps of this size may arise due to lineage-
133 specific somatic mutations established early during organismal development (12).

134 With the ages of the patients ranging from 60 to 78 years, the observed burdens and numbers of
135 overlapping variants (294-1,219 pairwise shared SNVs) are roughly on par with a mutation rate similar
136 to healthy human neurons, estimated to accumulate ~23-40 SNVs/year (13) (**Fig. 3**). Analysis of
137 mutational signatures indeed supported the major contributor was COSMIC Signature SBS5, an
138 ubiquitous aging-associated mutational process, or SBS40 which is closely related to SBS5 (14) (**Fig.
139 3**). Within-patient burden variability was to a large degree explained by variable sample purity
140 (**Supplementary Fig. 8**). These results confirm the mutationally quiet nature of SI-NET.



141

142 **Figure 3. Overview of potential driver mutations and mutational signatures for all 5 patients.** All CGC genes
143 with non-synonymous mutations in more than one sample are shown. Mutational signature loadings, i.e. relative
144 contributions, refer to COSMIC v3 definitions (14) and were dominated by the closely related signatures SBS5
145 and SBS40. Phylogenetic relationships previously inferred from somatic SNVs (**Fig. 1-2**) are indicated at the
146 bottom. Somatic SNV burdens, potential driver mutations (Cancer Gene Census genes mutated in >1 sample) and
147 mutational signature loadings in this figure were determined using less rigorous population variant filtering
148 compared to the phylogenetic analyses (see **Methods**). CGC, Cancer Gene Census.

149 Analysis of potential somatic driver mutations in coding genes mirrored earlier published results, with
150 infrequent variants in *CDKN1B* emerging as the main recurrent event (6 samples; **Fig. 3;
151 Supplementary Table 3**) (5). In Patient 1, the same *CDKN1B* frameshift indel occurred in two of the
152 metastases (G, H) in a pattern consistent with the inferred phylogeny (**Fig. 1b-c**). In Patient 2, the same
153 *CDKN1B* frameshift indel was found in the metastatic primary (G) and the metastasis (L), thus again
154 consistent with the phylogeny, while one non-metastatic primary carried a *CDKN1B* stop gain variant
155 (**Fig. 2b-c**). Patient 3 carried a stop gain variant in *CDKN1B* in the metastatic primary (B) that was
156 missing in the corresponding metastasis, possibly due to allelic dropout during sequencing (**Fig. 2d**).
157 Other Cancer Gene Census (15) genes mutated in >1 sample were *MUC16* (n = 3), a common false
158 positive gene encoding the second largest human protein (16), and *ARID1B*, *CHD4*, *CAMTA1*, *CD28*
159 and *PRRX1* (n = 2; **Fig. 3**). *TERT* promoter mutations, a frequent non-coding driver event in several
160 cancer types (17, 18), were absent. Consistent with other reports, there was thus a lack of obvious driver
161 mutations beyond *CDKN1B*, which was not essential for metastasization.

Analysis of somatic copy number alterations (CNAs) gave further support for the phylogenetic relationships inferred from somatic SNVs (**Fig. 4a**). The metastatic tumor in Patient 1 (C) carried a distinct segmental loss on chr11 found in all metastases (G-K) but not the other intestinal tumors. Similarly, the metastatic tumor in Patient 2 (G) exhibited chromothripsis (abundant clustered CNAs) on chr13, and this exact complex CNA pattern was mirrored in the corresponding lymph node metastasis (L) (**Fig. 4b**). A deletion on chr11 in this primary tumor that was absent in the metastasis presumably occurred after metastasization. Approximately 70-75% of SI-NETs have been shown to harbor hemizygous loss of chr18 (19), and we accordingly observed chr18 loss in all patients and in 22 of the 37 tumor samples. However, where chr18 loss could be phased based on germline single nucleotide polymorphisms (SNPs), we found that tumors deemed unrelated based on SNVs often had undergone loss of different chromosome homologs, while related tumors always showed loss of the same homolog (**Fig. 4a,c** and **Supplementary Fig. 9**). Among other recurrent events were segmental deletions on chr11 and chr13, as shown previously (5, 20, 21). No distinctive CNAs were shared in ways that contradicted the SNV-based evolutionary trees (**Fig. 4a**).

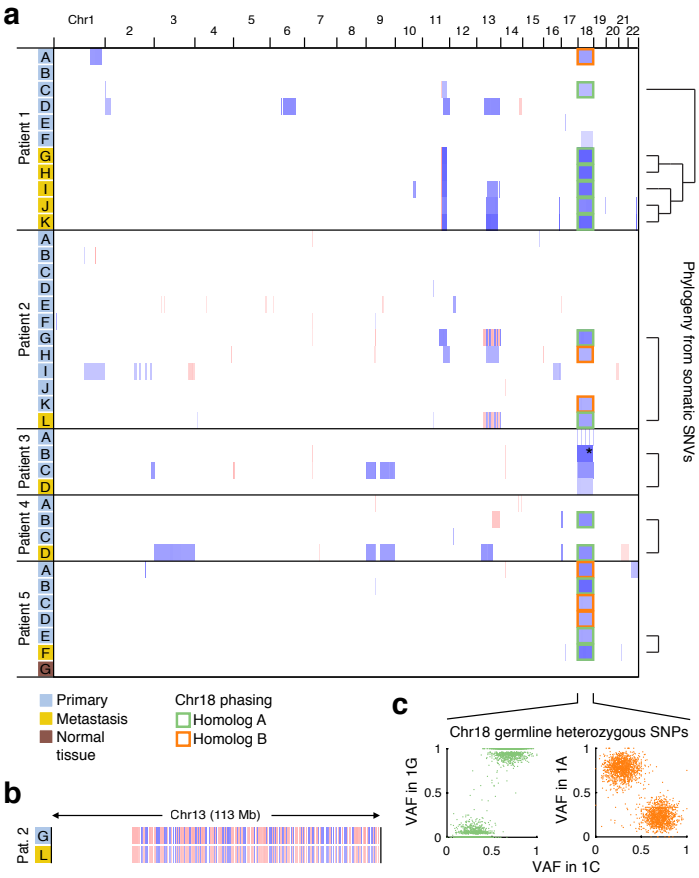


Figure 4. Somatic copy number alterations agree with phylogenies inferred from somatic SNVs. (**a**) Somatic CNA profiles for all samples (blue = deletion, red = amplification; segments <25 kb were excluded for visualization purposes). Phylogenetic relationships previously inferred from somatic SNVs (**Fig. 1-2**) are indicated to the right. Chr18 loss was phased based on germline SNPs, with the two chromosome homologs indicated in green and orange. *, phase could not be determined in Patient 3 (see **Supplementary Fig. 9**). (**b**) Detailed view of chromothripsis on chr13 in Patient 2, present in the metastatic primary (G) and the metastasis (L). (**c**) Example of phasing of chr18 deletions in Patient 1.

Chromothripsis, not previously reported in SI-NET to our knowledge, was further confirmed by analysis of somatic structural alterations. Of 17 in-frame gene fusion events, none of which were obvious drivers, 8 were intrachromosomal events on chr13, all found in both the primary and the metastasis shown by CNA to harbor chromothripsis on this chromosome (Patient 2, G and L; **Supplementary Fig. 10**).

Discussion

Based on our findings, we propose that multifocal SI-NETs evolve by independent somatic evolution. Furthermore, despite similar morphology for all primary tumors, metastases exclusively originated from a single lesion in all 5 patients, and where positional data was available (Patients 1 and 2) it can be noted that this tumor was located near the center of the affected intestinal segment. Previous exome-based analyses of paired single intestinal and metastatic samples have yielded puzzling results with a highly varying degree of genetic overlap and in some cases no overlap at all (5, 21), and our results thus suggest that a non-metastatic intestinal tumor was sampled when overlap was absent.

Multifocal gastrointestinal carcinomas are typically seen in hereditary cancer syndromes such as MEN-1 and familial adenomatous polyposis (FAP). Cohort studies support that some SI-NETs may have a heritable component (22, 23), and multifocality has been suggested to occur preferentially in familial cases (24, 25). However, multifocality is common in all SI-NET (3, 26), and key clinical parameters such as survival or age of diagnosis, which is typically lower for hereditary cancer, are similar in patients with or without multifocality (3, 26). The observation that the most recurrent somatic genetic event in SI-NET, chr18 loss, can affect both the maternal and the paternal chromosome homologs in the same patient argues against contribution from a predisposing chr18 germline variant by loss of heterozygosity (10). Furthermore, in stark contrast to germline-induced gastrointestinal tumors, SI-NET only affects a limited intestinal segment. Given the general lack of genetic drivers, including somatic events as confirmed by our study, and in the light of our results showing clonal independence, we suggest that future studies could focus on cancer-priming local environmental factors that may drive the emergence of multifocal SI-NET. While such factors remain elusive, it can be noted that SI-NET originates from epithelial enterochrommafin cells (EC-cells) that intrinsically interact with their surroundings by paracrine and endocrine serotonin secretion, synaptic-like connections with the enteric nervous system, and receptor-mediated nutrient sensing of luminal content (27).

Materials and methods

Patients

5 patients who underwent surgery for SI-NET at Sahlgrenska University Hospital, Gothenburg, Sweden, were included in the study. The clinical characteristics of the patients are summarized in **Supplementary Table 4**. All patients were postoperatively diagnosed with having well-differentiated neuroendocrine tumour grade 1-2 of the ileum (WHO 2019) and had metastatic disease at the time of diagnosis (TNM stage IIIB or IV). The diagnosis was based on one primary intestinal tumor and extent of lymph node metastases was evaluated as a standard clinical routine. The patients lacked family history of SI-NET and all had late onset (60-78 yrs). Tumor samples were collected at surgery and were immediately snap-frozen in liquid nitrogen. A piece of each tumor was formalin-fixed and paraffin-embedded for studies by immunohistochemistry. The tumors had typical SI-NET morphology and expression of neuroendocrine markers. Blood was collected and used as normal tissue in sequencing. We obtained consent from the patients and approval from the Regional Ethical Review Board in Gothenburg for the use of clinical materials for research purposes (Dnr:558-07, 833-18).

228 **Immunohistochemistry**
229 Sections of all sampled tumors from 5 patients with SI-NET were subjected to antigen retrieval using
230 EnVision FLEX Target Retrieval Solution (high pH) in a Dako PT-Link. Immunohistochemical staining
231 was performed in a Dako Autostainer Link using EnVision FLEX according to the manufacturer's
232 instructions (DakoCytomation). The histopathological evaluation and assessment of tumor cell content
233 was performed on hematoxylin and eosin-stained sections. Immunohistochemical staining was also
234 performed for chromogranin A, synaptophysin, serotonin and somatostatin receptor 2. The following
235 primary antibodies were used: anti-chromogranin A (MAB319; Chemicon; diluted 1:1000), anti-
236 synaptophysin (SY38, M0776; Dako; diluted 1:100), anti-serotonin (H209; Dako; diluted 1:10) and
237 anti-SSTR2 (UMB-1; Abcam; diluted 1:50). All collected tumor samples were reviewed by a board
238 certified surgical pathologist (ON).

239 **DNA extraction**
240 For DNA extraction, biopsies with purity less than 30%, as assessed by immunohistochemistry, were
241 excluded, as were samples that were deemed to small for DNA extraction. DNA from fresh-frozen
242 biopsies as well as blood from each patient was isolated using the allprep DNA/RNA Mini Kit (Qiagen)
243 according to the manufacturer's protocol.

244 **Whole genome sequencing**
245 WGS libraries were constructed for 37 primary and metastatic tumors from the 5 patients, selected
246 based on estimated tumor purity from immunohistochemistry staining. 5 blood normals and one
247 adjacent tissue normal was additionally sequenced, for a total of 43 samples. The prepared libraries
248 were sequenced on a Illumina Novaseq 6000 using 150 bp paired-end reads to an average depth of
249 38.85× (34.5-45.0). The reads were mapped to the human reference genome hg19 using BWA as part
250 of Sentieon Genomics Tools (v201808), which further also performs deduplication, realignment and
251 sorting. The data has been deposited in European Genome-Phenome Archive (EGA) under the
252 accession number EGAS00001004435.

253 **Somatic variant calling and annotation**
254 Somatic variant calling of matched tumor-normal specimens was performed using combined outputs
255 from Mutect2 (GATK v4.1.4.0) and VarScan (v2.3.9) (28). Mutect2 was run using default parameters,
256 and the output was fed into the filterMutectCalls tool with a mean median mapping quality score
257 requirement of 10. With VarScan, the somatic tool was used to call variants by default cutoffs and a
258 minimum variant allele frequency of 0.01 from SAMtools (v1.9) pileups of reads with Phred-scaled
259 mapping quality >= 15 and variant base quality >= 20. The VarScan processSomatic tool was then used
260 to get only the somatic SNVs and indels, where a minimum support of two reads was required for a
261 variant to be called. The somaticFilter tool was also used to discard SNVs called within SNP clusters
262 or within 3 bp of somatic and germline indels. Variants were further filtered to have coverage >=20
263 reads in the normal, to be present on both strands, and to be absent in the normal. The final set of
264 variants was yielded by intersecting outputs from the two callers, followed by annotation using
265 ANNOVAR (v2019Oct24) (29) and Variant Effect Predictor (VEP) (30). To minimize false positive
266 calls in phylogenetic analyses, population variants in dbSNP150, provided by ANNOVAR, and the
267 ENSEMBL variant database (corresponding to dbSNP153), provided by VEP, were removed. Analysis
268 of potential driver mutations and mutational signatures was based on less stringent population variant
269 filtering (dbSNP138), to avoid false negatives and since extensive SNP filtering can skew results from
270 mutational signatures analyses. For high-sensitivity analysis of shared mutations, mutations detected
271 by the standard pipeline as described above in at least one sample were whitelisted for detection using

272 the unfiltered VarScan somaticFilter output in other samples, allowing more sensitive detection of
273 overlapping subclonal mutations.

274 **Phylogenetic analyses**

275 Phylogenetic analyses were based on SNVs only as indel calls contained a higher fraction problematic
276 calls. Maximum parsimony phylogenies were reconstructed and visualized using MEGAX with default
277 settings (31).

278 **Mutational signatures**

279 The trinucleotide profile for each tumor was matched against known mutational signatures from
280 COSMIC (14) (v3 release) using the R package deconstructSigs (32) (version 1.9.0). We used default
281 parameters and a maximum of three signatures for each sample.

282 **Copy number analysis and homolog phasing**

283 CNAs were called using XCAVATOR (33) using the RC mode and a window size of 2000 bp. The
284 segmented copy number results were filtered to remove small segments (<25 kb) prior to visualization
285 using IGV. Deletions on chr18 were phased to determine what chromosome homolog was deleted based
286 on SNP calls reported by VarScan somatic ("germline" and "LOH" sets with a required coverage of
287 20). Only heterozygous SNPs were considered (variant allele frequency ranging from 0.25 to 0.75 in
288 the blood normal). Variant allele frequencies for individual SNPs were compared between a reference
289 sample, which was always the metastatic primary tumor in each patient, and another sample of interest
290 by means of a scatter plot.

291 **Structural variant analyses**

292 Structural variants (SVs) were called based on discordant read pairs and split reads using Manta
293 (v.1.5.0) (34). Genomic coordinates of high confidence SV calls were annotated using AnnotSV (v.2.1)
294 (35). For each patient, all identified translocations were visualised in a circos plot (36). A custom script
295 was used to evaluate whether gene-gene fusions were in-frame or out-of-frame. To calculate the frames,
296 exon and coding sequence coordinates were taken from the annotation file provided by AnnotSV.

297 Acknowledgements

298 The work described here was supported by the Swedish Research Council (E.L.), the Swedish Cancer
299 Society (E.L. and O.N.), the Knut and Alice Wallenberg Foundation (E.L.), the BioCARE National
300 Strategic Research Programme (O.N.), and grants from the Swedish state under the agreement between
301 the Swedish government and the county councils, the ALF agreement (O.N.). We wish to thank
302 laboratory assistant Gulay Altiparmak, research nurse Maria Nilsson and surgical coordinator Jenny
303 Oliver for skilled technical assistance, and Kerryn Elliot for critical reading of the manuscript. We
304 further want thank professor emeritus Bo Wängberg as well as previous and current surgical staff at the
305 Section of Endocrine and Sarcoma Surgery, Department of Surgery, Sahlgrenska University Hospital.

306 Author contributions

307 E.E. and E.L. designed the study; E.E., A.A., M.L., S.R, O.N., Y.A. and E.L. analyzed data; E.E., A.M.
308 and O.N acquired clinical materials; E.L. and E.E. drafted the manuscript with contributions from all
309 other authors.

References

1. L. R. Yates, P. J. Campbell, Evolution of the cancer genome. *Nat Rev Genet* **13**, 795-806 (2012).
2. K. Curtius, N. A. Wright, T. A. Graham, An evolutionary perspective on field cancerization. *Nat Rev Cancer* **18**, 19-32 (2018).
3. A. Gangi *et al.*, Multifocality in Small Bowel Neuroendocrine Tumors. *J Gastrointest Surg* **22**, 303-309 (2018).
4. A. Dasari *et al.*, Trends in the Incidence, Prevalence, and Survival Outcomes in Patients With Neuroendocrine Tumors in the United States. *JAMA Oncol* **3**, 1335-1342 (2017).
5. J. M. Francis *et al.*, Somatic mutation of CDKN1B in small intestine neuroendocrine tumors. *Nat Genet* **45**, 1483-1486 (2013).
6. J. Crona *et al.*, Somatic Mutations and Genetic Heterogeneity at the CDKN1B Locus in Small Intestinal Neuroendocrine Tumors. *Ann Surg Oncol* **22 Suppl 3**, S1428-1435 (2015).
7. P. Priestley *et al.*, Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* **575**, 210-216 (2019).
8. Z. Guo, Q. Li, E. Wilander, J. Ponten, Clonality analysis of multifocal carcinoid tumours of the small intestine by X-chromosome inactivation analysis. *J Pathol* **190**, 76-79 (2000).
9. T. M. Katona *et al.*, Molecular evidence for independent origin of multifocal neuroendocrine tumors of the enteropancreatic axis. *Cancer Res* **66**, 4936-4942 (2006).
10. Z. Zhang *et al.*, Patterns of chromosome 18 loss of heterozygosity in multifocal ileal neuroendocrine tumors. *Genes Chromosomes Cancer* 10.1002/gcc.22850 (2020).
11. Y. Z. Wang *et al.*, Reappraisal of lymphatic mapping for midgut neuroendocrine patients undergoing cytoreductive surgery. *Surgery* **156**, 1498-1502; discussion 1502-1493 (2014).
12. M. A. Lodato *et al.*, Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94-98 (2015).
13. M. A. Lodato *et al.*, Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**, 555-559 (2018).
14. L. B. Alexandrov *et al.*, The repertoire of mutational signatures in human cancer. *Nature* **578**, 94-101 (2020).
15. J. G. Tate *et al.*, COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **47**, D941-D947 (2019).
16. M. S. Lawrence *et al.*, Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214-218 (2013).
17. F. W. Huang *et al.*, Highly recurrent TERT promoter mutations in human melanoma. *Science* **339**, 957-959 (2013).
18. S. Horn *et al.*, TERT promoter mutations in familial and sporadic melanoma. *Science* **339**, 959-961 (2013).
19. E. Andersson, C. Sward, G. Stenman, H. Ahlman, O. Nilsson, High-resolution genomic profiling reveals gain of chromosome 14 as a predictor of poor outcome in ileal carcinoids. *Endocr Relat Cancer* **16**, 953-966 (2009).
20. M. S. Banck *et al.*, The genomic landscape of small intestine neuroendocrine tumors. *J Clin Invest* **123**, 2502-2508 (2013).
21. D. Walter *et al.*, Genetic heterogeneity of primary lesion and metastasis in small intestine neuroendocrine tumors. *Sci Rep* **8**, 3811 (2018).
22. J. P. Dumanski *et al.*, A MUTYH germline mutation is associated with small intestinal neuroendocrine tumors. *Endocr Relat Cancer* **24**, 427-443 (2017).
23. K. M. Walsh *et al.*, A pilot genome-wide association study shows genomic variants enriched in the non-tumor cells of patients with well-differentiated neuroendocrine tumors of the ileum. *Endocr Relat Cancer* **18**, 171-180 (2011).
24. Y. Sei *et al.*, Polyclonal Crypt Genesis and Development of Familial Small Intestinal Neuroendocrine Tumors. *Gastroenterology* **151**, 140-151 (2016).
25. Y. Sei *et al.*, A Hereditary Form of Small Intestinal Carcinoid Associated With a Germline Mutation in Inositol Polyphosphate Multikinase. *Gastroenterology* **149**, 67-78 (2015).

363 26. A. B. Choi *et al.*, Is Multifocality an Indicator of Aggressive Behavior in Small Bowel
364 Neuroendocrine Tumors? *Pancreas* **46**, 1115-1120 (2017).
365 27. N. W. Bellono *et al.*, Enterochromaffin Cells Are Gut Chemosensors that Couple to Sensory
366 Neural Pathways. *Cell* **170**, 185-198 e116 (2017).
367 28. D. C. Koboldt *et al.*, VarScan 2: somatic mutation and copy number alteration discovery in
368 cancer by exome sequencing. *Genome Res* **22**, 568-576 (2012).
369 29. K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from
370 high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
371 30. W. McLaren *et al.*, Deriving the consequences of genomic variants with the Ensembl API and
372 SNP Effect Predictor. *Bioinformatics* **26**, 2069-2070 (2010).
373 31. G. Stecher, K. Tamura, S. Kumar, Molecular Evolutionary Genetics Analysis (MEGA) for
374 macOS. *Mol Biol Evol* **37**, 1237-1239 (2020).
375 32. R. Rosenthal, N. McGranahan, J. Herrero, B. S. Taylor, C. Swanton, DeconstructSigs:
376 delineating mutational processes in single tumors distinguishes DNA repair deficiencies and
377 patterns of carcinoma evolution. *Genome Biol* **17**, 31 (2016).
378 33. A. Magi, T. Pippucci, C. Sidore, XCAVATOR: accurate detection and genotyping of copy
379 number variants from second and third generation whole-genome sequencing experiments.
380 *BMC Genomics* **18**, 747 (2017).
381 34. X. Chen *et al.*, Manta: rapid detection of structural variants and indels for germline and cancer
382 sequencing applications. *Bioinformatics* **32**, 1220-1222 (2016).
383 35. V. Geoffroy *et al.*, AnnotSV: an integrated tool for structural variations annotation.
384 *Bioinformatics* **34**, 3572-3574 (2018).
385 36. M. Krzywinski *et al.*, Circos: an information aesthetic for comparative genomics. *Genome Res*
386 **19**, 1639-1645 (2009).

387

Supplementary Figures

Pat. 1

Tumor	H&E	SYP	CHGA	5-HT	SSTR2	Tumor content
A						70%
B						40%
C						50%
D						70%
E						70%
F						70%
G						50%
H						60%
I						70%
J						50%
K						60%

Supplementary Figure 1. Immunohistochemical analysis of all sequenced samples from Patient 1. All tumors display similar morphology, and all express the neuroendocrine markers, SYP (synaptophysin), CHGA (chromogranin A) and 5-HT (serotonin) as well as the clinically relevant SSTR2 (somatostatin receptor 2). The assessment of tumor cell content was performed visually on hematoxylin and eosin-stained (H&E) sections of the tumors. The letter in each row indicates the sample ID.

Pat. 2

Tumor	H&E	SYP	CHGA	5-HT	SSTR2	Tumor content
A						50%
B						50%
C						60%
D						50%
E						30%
F						30%
G						50%
H						50%
I						60%
J						60%
K						60%
L						50%

Supplementary Figure 2. Immunohistochemical analysis of all sequenced samples from Patient 2. All tumors display similar morphology, and all express the neuroendocrine markers, SYP (synaptophysin), CHGA (chromogranin A) and 5-HT (serotonin) as well as the clinically relevant SSTR2 (somatostatin receptor 2). The assessment of tumor cell content was performed visually on hematoxylin and eosin-stained (H&E) sections of the tumors. The letter in each row indicates the sample ID.

Pat. 3

Tumor	H&E	SYP	CHGA	5-HT	SSTR2	Tumor content
A						50%
B						50%
C						70%
D						90%

Pat. 4

Tumor	H&E	SYP	CHGA	5-HT	SSTR2	Tumor content
A						70%
B						70%
C						70%
D						80%

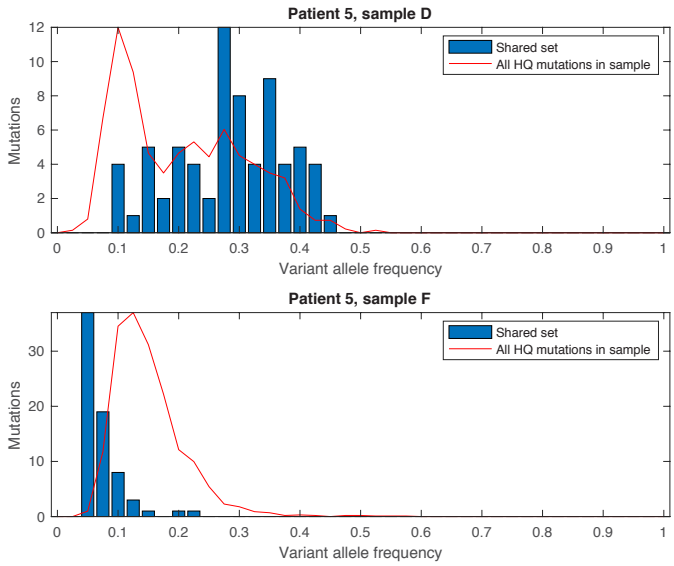
Pat. 5

Tumor	H&E	SYP	CHGA	5-HT	SSTR2	Tumor content
A						40%
B						50%
C						70%
D						70%
E						70%
F						40%
G						0%

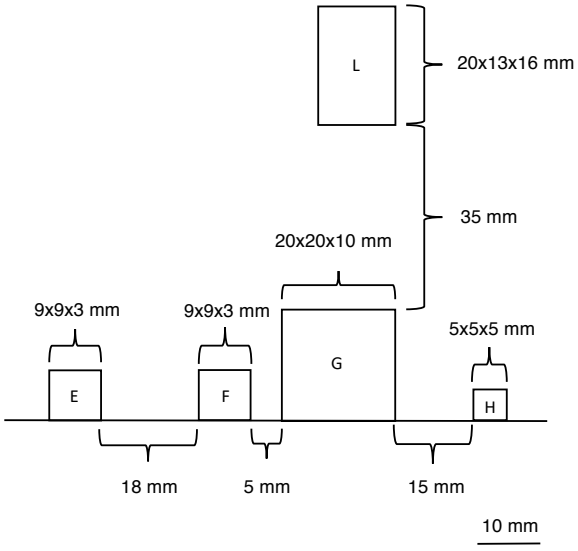
Supplementary Figure 3. Immunohistochemical analysis of all sequenced samples from Patient 3-5. All tumors display similar morphology, and all express the neuroendocrine markers, SYP (synaptophysin), CHGA (chromogranin A) and 5-HT (serotonin) as well as the clinically relevant SSTR2 (somatostatin receptor 2). The assessment of tumor cell content was performed visually on hematoxylin and eosin-stained (H&E) sections of the tumors. The letter in each row indicates the sample ID. Sample G is from normal intestinal mucosa.

	1A	1B	1C	1D	1E	1F	1G	1H	1I	1J	1K	2A	2B	2C	2D	2E	2F	2G	2H	2I	2J	2K	2L	3A	3B	3C	3D	4A	4B	4C	4D	5A	5B	5C	5D	5E	5F	5G
1A	599	1	2	2	2	1	3	4	3	5	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0
1B	1	635	3	1	2	3	2	1	2	2	1	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
1C	2	3	1175	0	1	1	742	736	724	715	716	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
1D	2	1	0	1161	1	0	0	1	2	1	0	0	0	0	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
1E	2	2	1	1	363	2	3	2	2	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	1	0	0	0
1F	1	3	1	0	2	835	2	1	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
1G	3	2	742	0	3	2	1633	1580	728	718	715	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	2	1	1	0
1H	4	1	736	1	2	1	1580	1602	726	715	714	2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0
1I	3	2	724	2	2	1	728	726	1739	868	869	1	0	0	2	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
1J	5	2	715	1	2	1	718	715	868	1745	1559	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0
1K	1	1	716	0	1	0	715	714	869	1559	1764	0	0	0	2	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1
2A	1	1	1	0	1	1	0	2	1	1	0	776	4	4	3	5	2	4	4	3	3	2	2	0	0	0	0	0	0	0	0	0	0	1	0	0	1	2
2B	0	0	0	0	0	0	0	0	0	0	0	4	1116	6	4	3	2	4	4	2	3	3	3	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
2C	0	0	0	0	0	0	0	0	0	0	0	4	6	144	5	2	4	6	4	4	2	2	5	0	1	0	0	0	0	0	0	0	0	0	1	1	1	1
2D	0	1	0	1	0	0	0	0	2	0	2	3	4	5	852	6	3	3	3	4	3	1	1	0	0	0	0	1	1	0	1	0	0	1	0	0	0	0
2E	0	0	0	0	0	0	0	0	0	0	0	5	3	2	6	905	4	6	5	5	4	2	2	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
2F	0	1	0	1	0	0	0	0	1	0	1	2	2	4	3	4	1069	10	3	3	4	3	8	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2G	0	0	0	0	0	0	0	0	0	1	0	4	4	6	3	6	10	2170	4	5	4	3	1587	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
2H	0	1	1	1	0	0	0	0	1	0	1	4	4	4	3	5	3	4	923	6	2	1	2	0	1	0	0	0	0	0	0	0	0	0	0	0	1	2
2I	0	0	0	0	0	0	0	0	0	0	0	3	2	4	4	5	3	5	6	1153	4	3	3	0	0	0	0	0	0	0	0	0	0	0	2	0	1	2
2J	0	0	0	1	0	1	0	0	0	1	0	3	3	2	3	4	4	4	2	4	1278	0	3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
2K	0	0	0	0	0	0	0	0	0	0	0	2	3	2	1	2	3	3	1	3	0	831	2	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
2L	0	0	0	0	0	0	0	0	0	0	0	2	3	5	1	2	8	1587	2	3	3	2	1536	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3A	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	767	6	5	3	0	0	0	1	1	1	1	2	0	0	0
3B	1	0	0	0	1	0	0	0	0	0	0	0	1	1	0	1	1	1	1	0	0	1	0	6	2965	4	698	0	0	0	0	0	0	0	0	0	0	0
3C	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	4	3317	3	1	1	0	2	0	0	1	1	1	0	0
3D	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3	698	3	1574	0	1	0	2	0	0	0	2	0	0	0
4A	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	600	3	2	5	1	0	0	0	0	0	0
4B	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	3	858	1	719	0	1	0	0	0	0	0
4C	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1760	3	0	0	0	0	0	0	0
4D	0	0	1	1	0	0	1	1	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	2	2	5	719	3	2484	1	0	0	1	1	1	1
5A	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	1535	8	6	4	3	5	10
5B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	8	1418	7	4	3	4	7
5C	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	6	7	1084	3	1	4	5
5D	0	0	0	1	1	0	2	0	1	0	0	0	1	1	0	1	0	1	0	2	1	1	0	2	0	1	2	0	0	0	1	4	4	3	1004	3	70	5
5E	0	0	0	0	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	3	3	1	3	818	483	5
5F	0	1	0	0	0	0	1	0	0	0	1	1	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	5	4	4	70	483	1680	5
5G	0	2	1	0	0	0	0	0	0	0	1	2	0	1	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	1	10	7	5	5	5	5	19

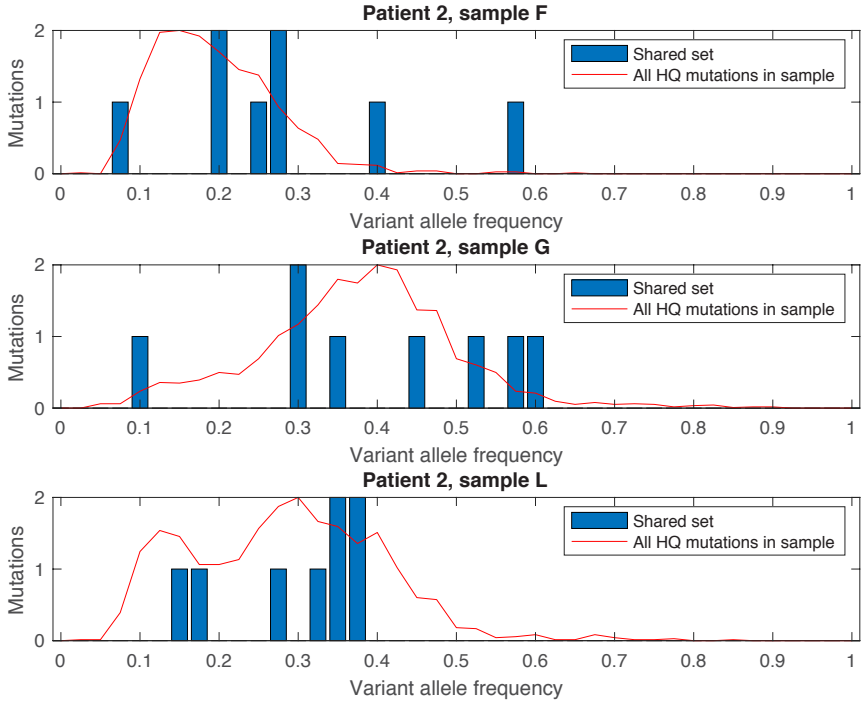
Supplementary Figure 4. High-sensitivity search for shared variants by whitelisting of SNVs called at high confidence in at least one sample. A two-step procedure was used, where any somatic SNVs called at high confidence (detected by both Mutect and VarScan using various filters including a strand filter as described in **Methods**) in at least one sample were "whitelisted". This was followed by a more sensitive search for presence of the whitelisted variants using VarScan without stringent filters (e.g. exclusion of strand filter, see **Methods**). The number and letter in each row/column label indicates the patient number and sample ID, respectively.



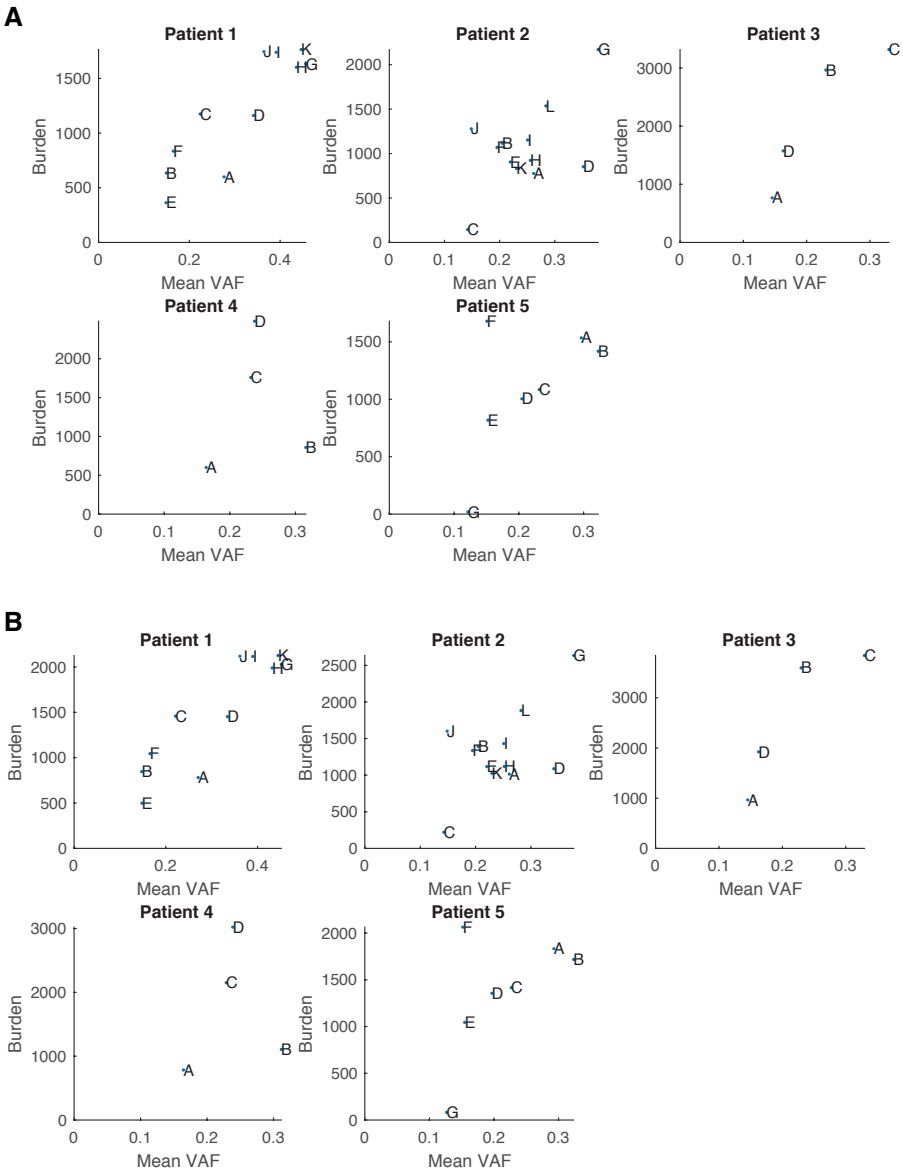
Supplementary Figure 5. Variant allele frequency (VAF) distributions support contamination between samples D and F from Patient 5. Samples D (primary tumor) and F (lymph node metastasis) in Patient 5 shared 8 SNVs in the main phylogenetic analysis (**main Fig. 2**). A high-sensitivity search for shared variants (**Supplementary Fig. 4**) revealed 62 additional shared variants called at high confidence in at least one of the samples while being detectable using relaxed filters (e.g. exclusion of strand filter, see **Methods**) in the other sample. These 70 shared SNVs were present at normal VAFs in sample D (top histogram) while being highly subclonal in sample F (bottom histogram). This supports that the metastasis sample (F) was contaminated with DNA from D, either during sample handling or, less likely, through a biological mechanism. Overall VAF distributions for all high-quality mutations in each sample are shown for comparison (red).



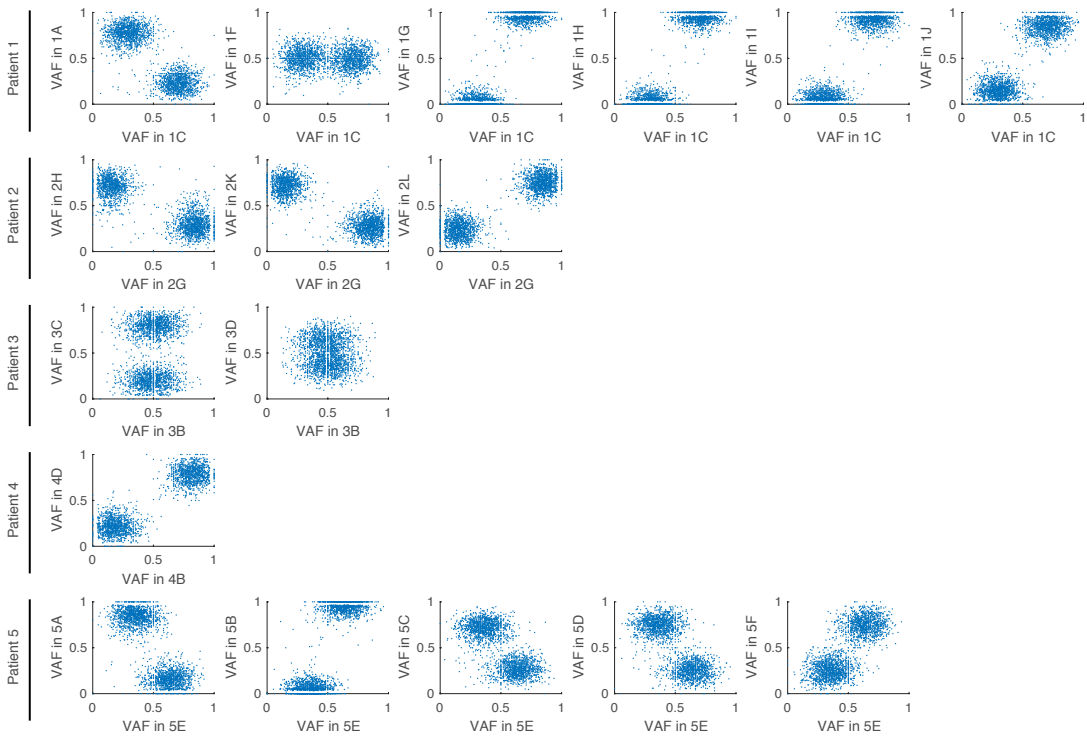
Supplementary Figure 6. Physical positioning of tumors F, G and L in Patient 2.



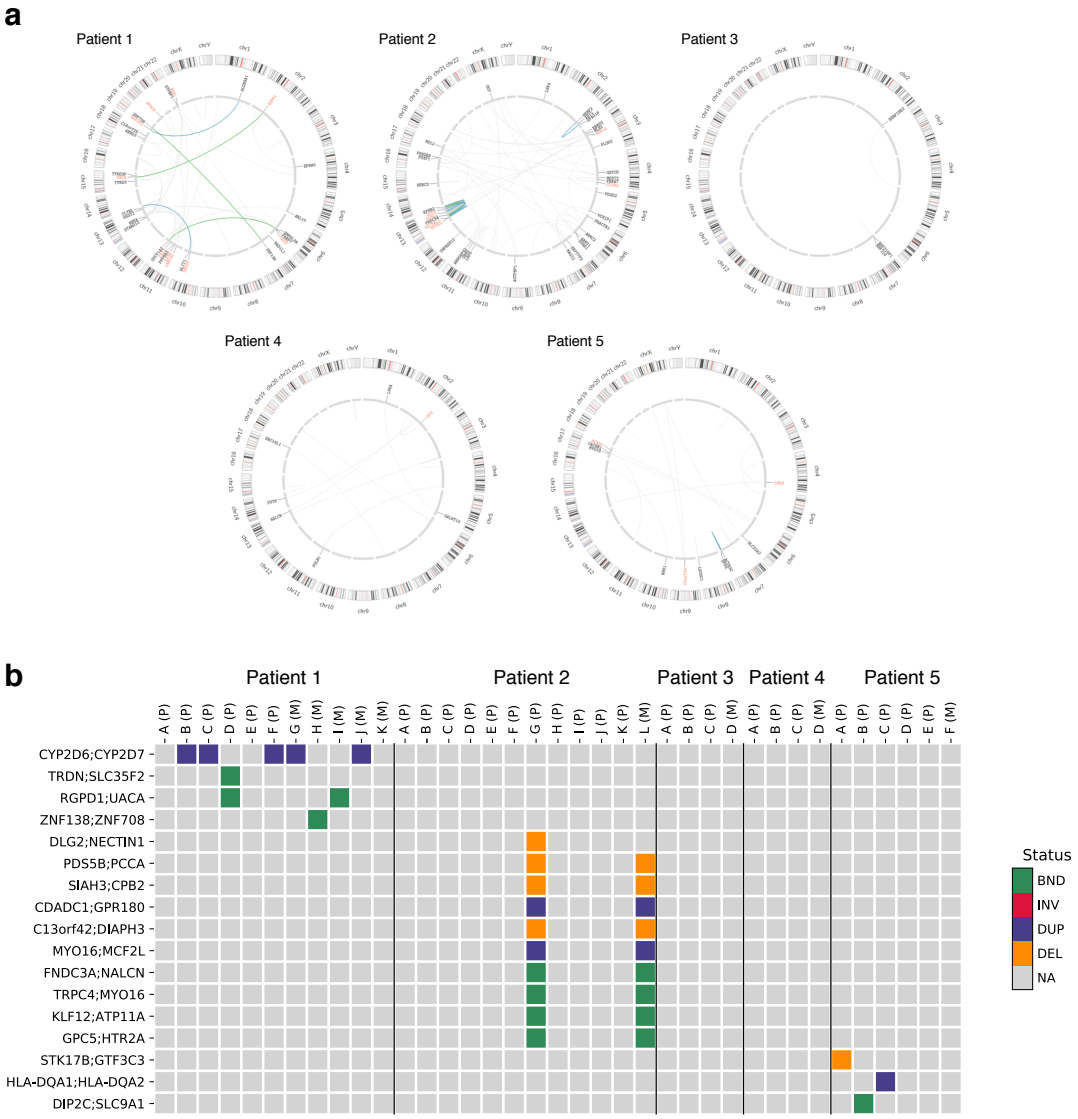
Supplementary Figure 7. Variant allele frequency (VAF) distributions for 8 variants found in samples F, G and L from Patient 2. A set of 8 SNVs was shared between samples F, G and L (lymph node metastasis) in Patient 2 (**main Fig. 2**). This number stayed constant when using a high-sensitivity approach to uncover subclonal shared SNVs (**Supplementary Fig. 4**). Manual inspection in IGV supported that the shared calls were mostly of high quality, with only one having characteristics of being a false positive call (**Supplementary Table 2**). The shared SNVs were present at normal VAFs in all samples. Overall VAF distributions for all high-quality mutations in each sample are shown for comparison (red).



Supplementary Figure 8. Positive correlations between per-sample SNV burden and average variant allele frequency (VAF). Mean VAF, which is indicative of sample purity, correlated positively with the number of called SNVs (burden) in each patient. (**a**) Results using SNV calls used for phylogenetic analyses, which were stringently filtered for known population variants (dbSNP150 and dbSNP153). (b) Results using relaxed population variant filtering (dbSNP138), as used for driver mutations and mutational signature analyses, were nearly identical.



Supplementary Figure 9. Chromosomal phasing of chr18 deletions based on germline heterozygous SNPs. Heterozygous germline SNPs on chr18 (read coverage 20 and VAF ranging from 0.25 to 0.75 in the blood normal) were used to determine what chromosome homolog was deleted, using the metastatic primary tumor in each patient as reference. The scatter plots compare VAFs for individual SNPs in the reference sample compared to other relevant samples. The data was down-sampled to 20% to reduce dot density in the plots. Phasing was inconclusive in Patient 3 due to lack of bimodality in the VAF distribution for the reference sample (3B), which could theoretically be due to a biallelic deletion in combination with whole genome duplication or due the deletion being homozygous.



Supplementary Figure 10. Analysis of somatic genomic structural alterations. (**a**) Circos plots showing all identified translocations per patient. Genes marked in red were involved in a translocation in more than one sample. Links involving one or no gene are shown in grey, links involving two in-frame genes are shown in green, and out-of-frame in blue. (**b**) In-frame fusion gene pairs colored by alteration type, with translocations (BND) shown in green, inversions in red, duplications in purple, and deletions in orange.