Whole Genome Sequencing of Single Circulating Tumor Cells from Neuroendocrine Neoplasms

Alexa Childs¹, Christopher D. Steele¹, Clare Vesely¹, Francesca M. Rizzo¹, Leah Ensell¹,
Helen Lowe¹, Pawan Dhami¹, Heli Vaikkinen¹, Tu Vinh Luong², Lucia Conde¹, Javier
Herrero¹, Martyn Caplin³, Christos Toumpanakis³, Christina Thirlwell¹,⁴, John A.
Hartley¹, Nischalan Pillay⁵,⁶, Tim Meyer¹,⁴

¹UCL Cancer Institute, University College London, London WC1E 6DD, UK
²Department of Histopathology, Royal Free London NHS Foundation Trust, London
NW3 2QG, UK
³Department of Gastroenterology, Royal Free London NHS Foundation Trust, London
NW3 2QG, UK
⁴Department of Oncology, Royal Free London NHS Foundation Trust, London NW3
2QG, UK
⁵Research Department of Pathology, Cancer Institute, University College London,
London WC1E 6BT, UK
⁶Department of Cellular and Molecular Pathology, Royal National Orthopaedic
Hospital NHS Trust, Stanmore, Middlesex HA7 4LP, UK

Corresponding author: Professor Tim Meyer, UCL Cancer Institute, University
College London. 72 Huntley Street, London WC1E 6DD, email; t.meyer@ucl.ac.uk
Tel; +44 207 679 6731.

Running title: WGS of single circulating tumor cells from NEN

Keywords: Neuroendocrine tumors, circulating tumor cells, single cell, copy number
variation, whole genome sequencing

Word Count: 4998
Abstract

Single-cell profiling of circulating tumor cells (CTCs) as part of a minimally invasive liquid biopsy presents an opportunity to characterize and monitor tumor heterogeneity and evolution in individual patients. In this study, we aimed to compare single-cell copy number variation (CNV) data with tissue, and define the degree of intra- and inter-patient genomic heterogeneity. We performed next generation sequencing (NGS) whole genome CNV analysis of 125 single CTCs derived from seven patients with neuroendocrine neoplasms (NEN) alongside matched white blood cells (WBC), formalin fixed paraffin embedded (FFPE) and fresh frozen (FF) samples. CTC CNV profiling demonstrated recurrent chromosomal alterations in previously reported NEN copy number hotspots, including the prognostically relevant loss of chromosome 18. Unsupervised hierarchical clustering revealed CTCs with distinct clonal lineages as well as significant intra- and inter-patient genomic heterogeneity, including subclonal alterations not detectable by bulk analysis and previously unreported in NEN. Notably, we also demonstrate the presence of genomically distinct CTCs according to the enrichment strategy utilized (EpCAM-dependent versus size-based). This work has significant implications for the identification of therapeutic targets, tracking of evolutionary change and the implementation of CTC-biomarkers in cancer.

Background

The molecular characterization of tumors has advanced our understanding of the major somatic driver mutations and informed the development of targeted therapies, which have transformed outcomes in selected patient populations (Vogel
et al., 2002, Sharma et al., 2007, Sosman et al., 2012). Whilst tissue biopsy remains central to diagnostic work-up, it is invasive, limited by the overall percentage of tumor cells and subject to heterogeneity exhibited in primary and metastatic tumors (Gerlinger et al., 2012, Walter et al., 2018, Navin et al., 2010). Furthermore, bulk genomic analysis cannot provide resolution at the single-cell level, which is required to fully define the extent of tumor heterogeneity.

Technological advances in whole genome amplification (WGA) and next generation sequencing (NGS) methods now permit genomic analysis at the single-cell level and are uniquely placed to unravel complex clonal structures and track tumor evolution over time. Furthermore, characterization of single circulating tumor cells (CTCs) as part of a minimally invasive “liquid biopsy” provides an opportunity to explore tumor biology and identify therapeutic targets.

The first clinical applications of CTCs focused on enumeration using the EpCAM-dependent CellSearch® platform, which has been shown to be both prognostic and predictive across a wide range of epithelial malignancies (Cristofanilli et al., 2004, Cohen et al., 2008, de Bono et al., 2008, Krebs et al., 2011, Poveda et al., 2011), including neuroendocrine neoplasms (Khan et al., 2013, Khan et al., 2016, Khan et al., 2011, Mandair et al., 2021). More recently, molecular analysis of single CTCs has been used to identify predictive biomarkers, such as the T790M resistance allele in NSCLC (Maheswaran et al., 2008). In SCLC, a pretreatment CTC-based biomarker has been shown to predict sensitivity to first line chemotherapy (Carter et al., 2017).
Neuroendocrine neoplasms (NEN) represent a heterogeneous disease entity with diverse histology, clinical features and prognosis (Dasari et al., 2017). They are characterized by a low mutational burden (Banck et al., 2013), but recurrent patterns of copy number variation (CNV) have been observed (Capurso et al., 2012, Cunningham et al., 2011, Kulke et al., 2008). CNVs affect a greater portion of the cancer genome than any other somatic genetic alteration (Heitzer et al., 2016) and CNV burden is prognostic for cancer-free and overall survival in multiple tumor types (Hieronymus et al., 2018) including NEN, where aneuploidy can be used to define distinct molecular subgroups of prognostic relevance (Karpathakis et al., 2016).

In this study, we perform CNV analysis of single NEN CTCs, aiming to define the extent of genomic heterogeneity both within and between patients, and to compare single-cell CTC data with bulk tissue analysis. CTC enrichment in NEN patients has to date been confined to EpCAM-dependent methodologies, which may fail to capture the full diversity of CTCs seen in this disease (Gorges et al., 2012). Here, we utilize both the EpCAM-based CellSearch® and epitope-independent Parsortix® systems in order to interrogate the full diversity of cells at the CNV level and investigate whether single cell CTCs may differ at a genomic level, according to EpCAM expression.

Methods
Patients

NEN patients were recruited at the Royal Free Hospital, London, between September 2014 and February 2018. The study was approved by the Local Ethics Committee (NRES Committee London – Bromley, IRAS ref 13/LO/0376) and all participants were required to provide written informed consent. Eligible patients had a histologically confirmed diagnosis of metastatic NEN in the absence of any other active malignancy. Tumors were graded according to the European Neuroendocrine Tumor Society (ENETS) guidelines (Bosman et al., 2010).

CTC enrichment using CellSearch®

Peripheral blood samples (7.5mls) were collected into CellSave tubes (Veridex LLC) stored at room temperature, and processed within 96 hours using the Celltracks® Autoprep® and Analyzer II® platform for the semi-automated staining, enrichment and enumeration of CTCs as previously described (Cristofanilli et al., 2005, Riethdorf et al., 2007). CTCs were defined as cells with a DAPI positive nucleus and positive EpCAM and cytokeratin expression in the absence of CD45 staining. All evaluations regarding enumeration of CTCs were made by two independent operators without knowledge of patient pathology. Enriched samples were re-suspended, aspirated from the CellSearch® cartridge and stored at -20°C in 50% glycerol.
CTC enrichment using Parsortix®

Blood was collected in Streck tubes (10ml) and incubated for 24-48 hours prior to size-based enrichment with the Parsortix® platform (ANGLE) using software protocols provided by the manufacturer. Following enrichment, samples were harvested in a total volume of 1.2mL of HBS by applying a reverse flow to the separation cassette. Enriched samples were resuspended in 200μl of autoMACS® Running Buffer and fixed and stained for further processing on a sterile Transwell polycarbonate membrane insert placed within a 50ml Falcon tube. BSA 3% (200μl) was pipetted to entirely cover its surface for a 10minute incubation. The 50ml tube was centrifuged at 500g for 2 minutes to elute the BSA solution from the filter prior to transferring the enriched patient sample onto the insert surface. 100μl of a 10% CD45 staining solution (10μl anti CD45-APC [Miltenyi Biotec] and 90μl of Running Buffer) and 100μl of a 10% CK staining solution solution (10μl anti CK-PE [Abcam] and 90μl Inside Perm [Miltenyi Biotec]) were used to sequentially stain samples for CD45 and cytokeratin prior to staining for nuclear content using 100μl of a 0.001mg/ml solution of Hoechst 33342 (Sigma Aldrich). After washing with SB115 buffer, the cell suspension was transferred into a sterile 1.5ml tube prior to volume reduction and loading into the DEPArray™ cartridge.

Cell Isolation from FFPE

FFPE tissue sections of 40-60μm thickness were dissociated into single cell suspensions and stained as previously described (Bolognesi et al., 2016). To enable visualization and identification of cells using the DEPArray™, cytokeratin and
vimentin were used as tumor and stromal cell markers respectively. Cell suspensions were stained with anti-cytokeratin MNF116 (IgG1) (DAKO), anti-cytokeratin AE1/AE3 (IgG1) (Millipore-Chemicon) and anti-Vimentin 3B4 (IgG2A) (DAKO).

Dissociated FFPE samples were subjected to a DNA quality control assay using the DEPArray™ FFPE QC kit (Silicon Biosystems). Each sample was given a QC score between 0-1 based on a qPCR-based assay. Samples with a sufficiently high DNA quality as determined by a QC score ≥0.4 according to manufacturer’s guidelines were processed on the DEPArray™ platform for retrieval of single tumor cells.

**Cell Isolation from fresh tissue**

Fresh tissue samples were collected in RPMI 1640 medium (Gibco) and processed within 3 hours of collection. The tumor sample was placed in 1ml of dissociation solution (240μl collagenase, 150μl DNAse and 13.85ml of RPMI media) and processed in a gentleMacs™ dissociater for one cycle, followed by two consecutive 30 minute incubations at 37°C. Single cell suspensions were created using a 50μl cell strainer and centrifuged and re-suspended in 5ml of RPMI prior to re-suspending in 1ml of freezing medium (10% DMSO in FBS) for storage at -80°C. Samples were fixed with 2% paraformaldehyde (Fluka) for 20 minutes at room temperature prior to staining for cytokeratin, vimentin and DAPI performed as per FFPE samples.

**DEPArray™ Sorting and Recovery**

Both CellSearch® and Parsortix enriched samples were imaged and sorted using the DEPArray™ system (Silicon Biosystems) as per manufacturer’s instructions (Abonnenc et al., 2013). Image-based selection was used to identify and recover individual cells.
of interest as either single-cells or pools of cells, based on their morphological features, DNA content and fluorescence labelling; CTCs (CK-PE<sup>+</sup>/CD45-APC<sup>-</sup>/DAPI<sup>+</sup>) and WBC (CK-PE<sup>-</sup>/CD45-APC<sup>+</sup>/DAPI<sup>+</sup>).

For analysis of FFPE samples with the DEPArray™, between 5000 to 10000 stained cells were loaded into the cartridge and cell sorting executed according to DEPArray™ User’s Manual rev 1.1_sw 2.1.1. The Cytokeratin<sup>+</sup> Vimentin<sup>-</sup> tumor cell population and Cytokeratin-Vimentin<sup>+</sup> stromal cell population were gated separately to evaluate morphology and staining characteristics prior to selecting cells for recovery.

Whole genome amplification of single-cell DNA and quality-control assay

WGA was performed on all recovered single-cells using the Ampli1™ WGA kit version 02 (Silicon Biosystems) as per manufacturer’s instructions to generate a 50μl WGA product. For single cells derived from blood (CTCs and WBC) and fresh tissue (tumor and stromal cells), the quality of the WGA product was determined using the Ampli1™ QC Kit (Silicon Biosystems). A genomic integrity index (GII) was allocated for each sample, scored from 0-4. Only single-cells with sufficiently good quality DNA as determined by a GII≥2 were selected for downstream analysis.

Nucleic acid extraction

For bulk sequencing, DNA was extracted from 5 -10 sections of 10μm thickness from three FFPE blocks using the DNAstorm FFPE DNA Isolation Kit (CELLDATA) following manufacturers instructions. DNA was eluted into 75μl of nuclease free water and
concentrations measured using the NanoDrop-1000 Spectrophotometer (NanoDrop) and Qubit 2.0 Fluorometer (Invitrogen). Haematoxylin and eosin-stained sections were evaluated to ensure >80% purity of tumor specimens prior to processing.

LowPass Whole Genome Sequencing and Bioinformatics

Ampli1™ LowPass kit for Illumina (Menarini Silicon Biosystems) was used for preparing low-pass Whole Genome Sequencing (WGS) libraries from single cells. For high-throughput processing, the manufacturer’s procedure was implemented in a fully automated workflow on a STARlet Liquid Handling Robot (Hamilton®). Ampli1™ LowPass libraries were normalized and sequenced by HiSeq 2500 instrument using 150 SR rapid-run mode. The obtained FASTQ files were aligned to the hg19 human reference sequence using Burrows-Wheeler Aligner version 0.7.12 (BWA). Copy-number alterations in the data were identified using Control-FREEC software (version 11.0).

For bulk analysis of FFPE samples, genomic DNA was quantified using Qubit 3 fluorometer with dsDNA BR kit according to manufacturer’s instructions. One microgram of genomic DNA was used to prepare whole genome sequencing libraries using Nonacus Cell 3 Target: Library Preparation kit. Library preparation was done according to manufacturer’s instructions. Enzymatic fragmentation was performed at 32°C for 14 minutes to obtain library fragments with an average size of 250bp followed by ligation of UMI Adapters on both ends of the 5’-phosphorylated / 3’-dA-tailed DNA fragments. Libraries were purified using Target Pure NGS clean-up beads and minimal PCR amplification was carried out using 4 cycles of amplification.
Libraries were quantified using Qubit 3 fluorometer with dsDNA BR kit and run on an Agilent Bioanalyzer DNA 1000 chip according to manufacturer’s instructions. Average library fragment length was determined from the bioanalyzer trace. Library molar concentration was determined based on the average fragment size and the Qubit concentration. All libraries were normalised to 10 nM working concentration and pooled. The dual-indexed library pool was sequenced on Illumina Nextseq 500/550 platform to generate paired end reads. The Nonacus Cell 3 Target: Library preparation protocol adds unique molecular identifiers (UMIs) to the sequencing libraries which were sequenced by additional 9 cycles of sequencing added on to the i7 index read.

Bulk sequencing data was processed with the nextflow Sarek v2.3.FIX1 pipeline (https://github.com/UCL-BLIC/Sarek_v2.3.FIX1) following GATK best practices. Specifically, reads were aligned against hg38 with BWA v0.7.17, duplicated reads were marked and reads were recalibrated with GATK v4.1.1.0. CNV profiles were obtained by running Control-FREEC v11.5 with WGS recommended parameters.

Statistics
All statistical analyzes were performed in R. Pairwise Manhattan distances were calculated for all samples, using only copy number bins that were not NA for each pair. Hierarchical clustering of copy number profiles using these distances was performed with Ward’s minimum variance method.
When comparing bulk and CTC copy number profiles, the mean copy number across CTC copy number bins that overlapped a bulk bin was taken. Any bulk bin without an overlapping CTC bin was not given a copy number designation.

t-distributed stochastic neighbour embedding (TSNE) analysis was performed using the R package Rtsne, using only the genomic bins that were non-missing for all samples analyzed, with a perplexity of 30.

Correlations between copy number profiles were calculated with respect to a base copy number of 2, as described in Zhang et al 2017 (Gao et al., 2017):

$$\rho_{mn} = \frac{\sum_i(Cm_i - 2)(Cn_i - 2)}{\sqrt{\sum_j(Cm_j - 2)^2 \sum_k(Cn_k - 2)^2}}$$

Where $\rho_{mn}$ is the correlation between samples $m$ and $n$, while $Cm_i$ is the copy number for sample $m$ at bin $i$.

To account for differences in ploidy, correlations were also calculated with respect to the average copy number across all bins for each sample:

$$\rho_{mn} = \frac{\sum_i(Cm_i - P_m)(Cn_i - P_n)}{\sqrt{\sum_j(Cm_j - P_m)^2 \sum_k(Cn_k - P_n)^2}}$$

where $P_m$ is the mean copy number for sample $m$ across all bins.

Metrics chosen to investigate copy number dynamics within a sample were the proportion of genome altered (number of CN!=2 bins divided by the total number of bins) and Shannon’s diversity index, $-\sum_i p_i \ln p_i$, where $p_i$ is the proportion of copy number bins with copy number state $i$ i.e. CN=2. Tests for statistical differences...
between distributions for these metrics were performed using the Kolgomorov-
Smirnov test.

Copy number gains and losses were defined in relation to ploidy. Gains were defined
as log2(CN/ploidy)>0.9, while losses were defined as log2(CN/ploidy)<-0.9. The
proportion of cells with a loss at a given genomic bin was used as a metric for a
single patient. When combining multiple patients the mean proportion of cells
across all patients considered was used. A threshold for statistically significant
recurrent gain or loss was determined by bootstrapping the original copy number
data; for each patient copy number states were sampled with replacement from
every copy number state seen in the original data for that patient, this was
performed for the same number of cells as were originally profiled for that patient.
Gains and losses were defined as previously, and the proportion of simulated cells
with a gain/loss at each genomic bin was calculated. This was repeated 1000 times
per patient, and the threshold for determining recurrent gains/losses was set as
99.9\textsuperscript{th} percentile value across all genomic bins for gains or losses separately. For a
threshold where multiple patients are being considered, the same bootstrapping was
performed for each patient, but the threshold was determined as the 99.9\textsuperscript{th}
percentile of the mean proportion of cells with gain/loss across the patients being
evaluated.

Results
Patient characteristics and sample collection

Seven NEN patients were included with primary tumor sites comprising the small intestine (SINET) (n=4), pancreas (n=1), gastro-oesophageal junction (GOJ) (n=1) and kidney (n=1). All patients had peripheral blood samples taken for CTC enrichment using the EpCAM-dependent CellSearch® platform and three patients had concomitant samples enriched using the size-based Parsortix® device (Figure 1).

Blood samples were taken from new patients at time of first presentation to our clinic (patients 1, 3, 5, 6) or at time of disease progression prior to commencing systemic therapy (patients 2, 4, 7). Matched WBC were analyzed as negative controls. A total of 7 tissue samples (6 FFPE, 1 FF) from six patients were analyzed. Of the seven samples, four were primary tumor samples (3 small intestine, 1 GOJ) and three were metastatic sites (2 liver, 1 brain). One patient (patient 1) had no available tissue for analysis. The clinical and treatment characteristics as well as the samples analyzed per patient are summarized in Table 1.

CTC sequencing

In total, 125 single CTCs were isolated from seven patients and successfully subjected to whole genome amplification (WGA), quality-control PCR and low pass whole genome sequencing (LPWGS). Single CTCs displayed high quality metrics, with only 3.5% failing to pass the quality checks for single-cell CNV. As a control, 17 single WBC (CD45 positive cells) were isolated and subjected to the same procedures. CD45 positive cells showed balanced copy number profiles (Supplementary Figure 1) whereas CTCs showed multiple gains and losses (Figures 2 and 3), confirming the aberrant nature of these tumor cells and the uniformity of single-cell WGA with the
Ampli1™ kit. The sensitivity and specificity of CTC identification and recovery by the DEPArray™ was assessed across all single cells subjected to LPWGS. Cells with CNV profiles demonstrating an overabundance of substantial chromosomal gains and losses were considered CTCs, whilst cells demonstrating flat profiles were classified as WBC (Ferrarini et al., 2018, Mangano et al., 2019). Using CNV profiles as the ultimate classifier of cell status, DEPArray™ selection had a positive predictive value of 95% and negative predictive value of 100% (P<0.0001).

Single tumor cells derived from FFPE surgical specimens/biopsies were also subjected to the same procedures as CTCs. DNA quality of single cell suspensions was assessed using the Ampli1™ QC Kit (Silicon Biosystems) prior to cell sorting. Four of the seven samples had QC values ≥0.4 indicating a sufficient DNA quality for single-cell CNV analysis and 8-10 single cells from each sample were processed for CNV analysis. The majority of single tumor cells had high derivative log ratio spread values in keeping with low library quality and only 15% of recovered single cells yielded sufficient quality results for CNV analysis.

**CTC versus tumor tissue CNV profiles**

For the three patients with sufficient matched FFPE tissue available for bulk analysis, whole-genome CNV profiles were compared between CTCs and bulk FFPE samples (Figure 2). The CNVs demonstrated in bulk tissue analysis were predominantly losses and these were also detectable in most CTCs. For example, in patient 2, losses in chromosomes 6, 9 and 18 are seen in bulk tissue and in 25, 80 and 65% of CTCs respectively, while patient 3, losses in chromosome 16 were observed in bulk tissue.
and 100% of CTCs (Figure 2). The majority of these concordant genomic losses are located in regions of the genome previously described as altered in NENs, with loss of chromosome 9 and 18 reported in 20% and 60-78% of SINETS respectively. However, single CTC data demonstrated the presence of clones enriched in additional somatic copy number alterations not detectable at bulk level, including the presence of a subclone of cells with evidence of whole genome doubling, observed in patients 2 (10% of CTCs) and 4 (6%). These reproducible CNV patterns were not evident in bulk sequencing analysis, and only detectable due to the resolution afforded by single cell sequencing. Such subclonal copy number alterations were most pronounced in patient 4, where appreciable CNV gains or losses were only detectable at the single cell level and not in the bulk tissue.

In patient 3, single tumor cells derived from a fresh frozen (FF) liver biopsy exhibited identical copy number profiles as CTCs and unsupervised hierarchical clustering of CTC and tumor copy number profiles demonstrated clustering of these cells together.

**CTC analysis reveals significant inter- and intrapatient heterogeneity**

To fully explore inter- and intrapatient CNV heterogeneity in NEN patients, the full set of 125 single CTCs from seven patient samples were further interrogated (Figure 3). Copy number losses were seen more frequently than amplifications, however, whole genome doubling was detected in all CTCs derived from two patients (Patients 1 and 6). Despite the preponderance of losses, the CNV patterns of individual patients are dissimilar and this remains the case when considering only those
patients of small intestinal primary site (patients 1-4). These patient-specific patterns of CNV were confirmed using t-distributed stochastic neighbour embedding (TSNE; Figure 4), which demonstrated clear clustering of individual patients, with no segregation according to primary site. Conversely, all WBC clustered together regardless of patient of origin in keeping with their flat CNV profiles (Figure 4b).

Within individual patients there were observations of clonal CN alterations seen in 100% of CTCs, but also clear evidence of subclonal changes and of individual cells with unique CNV profiles indicative of divergent evolution (Figure 3). This intra-patient heterogeneity was only detectable at the single cell level. The degree of intra-patient heterogeneity varied according to patient, with patients 3 and 6 demonstrating the highest average pairwise correlation of CTC CNV profiles, and hence the most homogenous copy number landscape across CTCs (Supplementary Figure 2). However, the correlation of CNV profiles within patients remains higher than that observed between patients, underscoring the independent nature of CNV profiles originating in different patients, and the shared evolutionary history of CTCs, and thus CTC CNV profiles, within individual patients.

**CNV profiles vary according to enrichment strategy**

In patient 7, hierarchical clustering of CNV profiles demonstrated distinct clustering of CTCs enriched by the EpCAM-dependent CellSearch® as compared to the epitope-independent, size-based Parsortix® platform (Figure 3). This is also demonstrated in Figure 4 where Parsortix® and CellSearch® CTCs from patient 7 form largely separate groups. To investigate this further, we summarized single CTC profiles via two
metrics; the proportion of the genome that is aberrant (copy number other than 2), and copy number diversity as enumerated by Shannon’s diversity index, and compared these metrics across cells according to the enrichment strategy utilized. There was a statistically significant difference in the distribution of both metrics between different enrichment strategies within patient 7 (Kolgomorov-Smirnov test, p<0.01, Figure 5A), where Parsortix® CTCs demonstrate a larger range in both metrics as compared to CellSearch® CTCs, indicating greater cell-to-cell variation. Interestingly, the difference seen in patient 7 were not found to be statistically significant across all patients (Figure 5B), indicating that these differences may vary on a patient-to-patient basis. This data suggests that restricting analysis of CTCs to only those that express EpCAM may exclude subsets of tumor cells that could be clinically relevant.

**CTC Molecular Characterization**

In order to evaluate the clinical application of CTC CNV profiling as a surrogate for tissue biopsy, we interrogated CTC CNV profiles for prognostic or actionable copy number changes described in the NEN literature. Evaluation of the frequency of copy number amplifications and deletions within CTC CNV profiles from SINET patients revealed recurrent losses of Chromosomes 9, 13q, 16q and 18 (Figure 6A). These have previously been described in SINETs supporting the technical reliability of our data and the potential use of CTCs as a tissue surrogate(Di Domenico et al., 2017, Karpathakis et al., 2016, Hashemi et al., 2013, Banck et al., 2013, Kulke et al., 2008). Of particular note is Chromosome 18, loss of which is the most frequently reported
genomic event in SINET, occurring in 60-78% of tumors, and is of prognostic relevance (Karpathakis et al., 2016). Previously unreported alterations, including loss of Chromosome 2p and 7q22 were also identified. Although not reported in SINET, allelic losses in Chromosome 2p are reported in colorectal, lung and endometrial malignancies. The tumor suppressor gene \textit{CUX1} is located at Chromosome 7q22, knockdown of which causes increased PI3K signaling and AKT phosphorylation (Ramdzan and Nepveu, 2014). This may be relevant in this patient population as deregulation of the PI3K/Akt/mTOR pathway is well-established in NEN, supported by the clinical efficacy of the mTOR inhibitor everolimus (Yao et al., 2011, Pavel et al., 2011).

Whole chromosome and arm gains at Chromosome 4 have previously been described in SINET. We did not observe such large-scale gains, instead, we observed focal gains in the TEC gene on Chromosome 4p12, which encodes a protein belonging to the Tec family of non-receptor protein-tyrosine kinases involved in the T-lymphocyte activation pathway and implicated in myelodysplastic syndrome.

CTCs from patient 7 (renal NET) demonstrated recurrent chromosomal alterations of likely clinical significance. Loss of Chromosome 3p was observed in a high proportion of CTCs, and harbors several tumor suppressor genes including the VHL gene at 3p25. Loss of heterozygosity (LOH) of 3p has been reported in the limited renal NET sequencing data available and is also found in over 90% of clear cell renal carcinoma (Alimov et al., 2000, el-Naggar et al., 1995). Loss of Chromosomes 10q and 13q were also observed, the former of which encodes the tumor suppressor gene
PTEN and is of prognostic relevance in renal cell carcinoma (Velickovic et al., 2002).

Finally, as with SINET, Chromosome 16q loss was frequently identified across patient
7 CTCs. Deletion of 16q is demonstrated across multiple malignancies, and LOH has
been indicated as an early event in the development of breast and hepatocellular

Discussion

Copy number analysis of NEN CTCs confirmed a wide range of genomic aberrations
making them readily distinguishable from WBC. All cells classified as WBC using the
pre-determined DEPArray™ criteria demonstrated balanced copy number profiles,
confirming the specificity and reproducibility of these criteria and accuracy of
DEPArray™ sorting.

In this study we show for the first time that somatic CNVs of NEN CTCs mirror those
seen in FFPE tissue, validating these CTC enrichment and isolation technologies in
NEN and confirming their potential use as a surrogate for tissue biopsy. The clinical
applications of this finding have been demonstrated in other tumor types such as
NSCLC, where good concordance between ALK-rearranged CTCs and ALK-positive
tumor biopsies has been demonstrated (Pailler et al., 2013). This finding is
particularly relevant in tumor types where tissue biopsy is not readily available or as
in NEN, where the relatively good prognosis of patients with low grade disease
means surgical specimens or biopsies may have been taken several years previously
and therefore not be representative of the current genomic landscape of the disease
after multiple lines of systemic therapy. CTCs have the additional benefit of being
non-invasive and therefore easily repeatable, thus allowing the monitoring of
genomic change in real-time. Beyond this, serial CTC monitoring may also enable the
detection of mechanisms of resistance (Pailler et al., 2015). Importantly, subclonal
CNVs not discernible in bulk tissue analysis were detectable in single CTC samples
thus allowing the identification of intra-patient genomic heterogeneity.

Unsupervised hierarchical clustering identified intrapatient genomic heterogeneity in
NEN patients, with diverse single CTC CNV traces observed in some patients. The
intrapatient CNV heterogeneity demonstrated in this study has also been observed
in other tumor types such as prostate (Lambros et al., 2018, Dago et al., 2014) and
colorectal (Heitzer et al., 2013) cancer. This is in contrast to lung adenocarcinoma,
SCLC, breast and gastric cancer, where more homogeneous CNV have been observed
in CTCs from individual patients (Heidary et al., 2014, Ni et al., 2013, Gao et al., 2017).

Intrapatient heterogeneity is of clinical relevance as it may impact upon prognosis,
response to treatment and biomarker development. High intratumoral
heterogeneity in tissue samples has been associated with a worse overall survival
across different tumor types (Mroz et al., 2013, Seol et al., 2012). This relationship
has not yet been examined with regards to the genomic profiling of CTCs, but low
phenotypic diversity of prostate cancer CTCs has been shown to correlate with
improved OS in patients treated with androgen receptor signalling inhibitors (ARSI),
whereas high heterogeneity was associated with increased risk of death on ARSI
relative to taxanes. Considerable heterogeneity was also demonstrated in CNV
patterns between patients. This appears to be cancer-type specific. Ni et al observed
almost identical global CNV patterns in 5 different patients with lung
adenocarcinoma with 78% of the gain and loss regions shared between any two
patients (Ni et al., 2013) and similar findings have been reported in gastric
cancer (Gao et al., 2017). However, increased inter-patient heterogeneity is seen in
other tumor types, such as SCLC and breast cancer (Gao et al., 2017, Ni et al., 2013).

The inter-patient heterogeneity in CNV profiles demonstrated in this study persists
even when analysis is confined to those patients with small-intestinal primaries.

Epitope-dependent enrichment technologies such as the CellSearch® platform limit
recovery of CTCs to an EpCAM-positive subpopulation. In this study we performed
the first direct comparison of CTC CNV profiles using identical blood draws between
the epitope-independent size-based Parsortix® and EpCAM-based CellSearch®. In
patient 7, CTCs enriched using the CellSearch® platform demonstrate reproducible
CNV with high inter-cell concordance. However, CTCs enriched using Parsortix®
appear genomically distinct, lacking the conserved CNV demonstrated in CellSearch®
CTCs and displaying a wider range of inter-cell heterogeneity. Different methods of
enrichment may therefore impact on the results of single-cell genomic analysis and
have implications for serial monitoring of CNV profiles. This finding is clinically
significant as it may impact on biomarker development. For example, a CNV-based
classifier of CTCs has been shown to predict chemosensitivity in SCLC patients (Carter
et al., 2017). In that study, all CTCs were enriched using CellSearch® and the classifier
was less effective in those patients demonstrating intra-patient heterogeneity. The
data presented in our study suggests that the efficacy of CNV-based classifiers such
as this may be affected by the form of enrichment used and could not be directly
extrapolated to CTCs enriched using alternative technologies. Furthermore, it
suggests combining epitope-independent enrichment strategies with CellSearch® may allow sampling of a wider population of CTCs with greater potential to fully capture CTC diversity.

SINET are characterized by a low mutational burden, with the most frequent mutation occurring in the cell cycle regulator CDKN1B (cyclin-dependent kinase inhibitor 1B) in only 8% of tumors (Francis et al., 2013, Crona et al., 2015). In this study, we identify recurrent loss of Chromosome 18, the most common genomic event in SINET and predictive of PFS in SINET. Karpathakis et al have previously demonstrated that CNV analysis of SINET primary tissue can be used to divide patients into three molecular subtypes with significant impact on PFS (Karpathakis et al., 2016). We also demonstrate novel and potentially targetable alterations such as focal gains in Chromosome 4p12, which encodes the TEC gene (Yu and Smith, 2011).

Further work is required to validate this finding in a larger cohort of patients.

Despite the novel findings reported, we acknowledge some limitations; namely the relatively low number of patients involved, as well as their heterogeneity in terms of grade and primary site. However, limiting analysis to a smaller patient cohort allowed assessment of multiple CTCs per patient in order to better characterize intra-patient heterogeneity, whilst the overall large number of single cells analysed allowed comparison with bulk tissue data and of cell enrichment techniques at the molecular level.
In conclusion, this is the first study to demonstrate that CNV analysis of single CTCs in NEN patients is feasible. We have demonstrated significant intra- and inter-patient genomic heterogeneity undetected by bulk tissue analysis. Additionally, we demonstrate for the first time, the presence of genomically distinct CTCs according to the enrichment strategy utilized, which has implications for the study of CTCs across all tumour types.

**Conflict of interest:** there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported

**Funding:** This work was supported by the CRUK & EPSRC Comprehensive Cancer Imaging Centre, European Neuroendocrine Tumor Society Fellowship, University College London (UCL) CRUK and NIHR Experimental Cancer Medicine Centre Grant No. C12125/A15576, MRC award MR/M009033/1 and the UCL Hospitals NIHR Biomedical Research Centre.

**References**


BOSMAN, F. T., CARNEIRO, F., HRUBAN, R. H. & THEISE, N. D. 2010. WHO Classification of Tumours, IARC.


bowel carcinoid tumors reveals areas of recurrent amplification and loss. Genes
Chromosomes Cancer, 47, 591-603.

LAMPROS, M. B., SEED, G., SUMANASURIYA, S., GIL, V., CRESPO, M., FONTES, M., CHANDLER,

MAHESWARAN, S., SEQUIST, L. V., NAGRATH, S., ULKUS, L., BRANNIGAN, B., COLLURA, C. V.,

MANDAIR, D., KHAN, M. S., LOPES, A., FURTADO O’MAHONY, L., ENSELL, L., LOWE, H.,
Threshold for Circulating Tumor Cells in Patients With Pancreatic and Midget

MANGANO, C., FERRARINI, A., FORCATO, C., GARONZI, M., TONONI, P., LANZELLOTTO, R.,
detection of genomic imbalances at single-cell resolution reveals intra-patient

2013. High intratumor genetic heterogeneity is related to worse outcome in patients

NAVIN, N., KRASNITZ, A., RODGERS, L., COOK, K., METH, J., KENDALL, J., RIGGS, M.,
EBERLING, Y., TROGE, J., GRUBOR, V., et al. 2010. Inferring tumor progression from

NI, X., ZHUO, M., SU, Z., DUAN, J., GAO, Y., WANG, Z., ZONG, C., BAI, H., CHAPMAN, A. R.,
ZHAO, J., et al. 2013. Reproducible copy number variation patterns among single

PAILLER, E., ADAM, J., BARTHELEMY, A., OULHEN, M., AUGER, N., VALENT, A., BORGET, I.,
cells harboring a unique ALK rearrangement in ALK-positive non-small-cell lung

PAILLER, E., AUGER, N., LINDSAY, C. R., VIELH, P., ISLAS-MORRIS-HERNANDEZ, A., BORGET, I.,
NGO-CAMUS, M., PLANCHARD, D., SORIA, J. C., BESSE, B., et al. 2015. High level of
chromosomal instability in circulating tumor cells of ROS1-rearranged non-small-cell

PAVEL, M. E., HAINSWORTH, J. D., BAUDIN, E., PEETERS, M., HORSCR, D., WINKLER, R. E.,
octreotide long-acting repeatable for the treatment of advanced neuroendocrine
tumours associated with carcinoid syndrome (RADIANT-2): a randomised, placebo-

POVEDA, A., KAYE, S. B., MCCORMACK, R., WANG, S., PAREKH, T., RICCI, D., LEBEDINSKY, C.
progression free survival and overall survival in patients with relapsed/recurrent

RAMDZAN, Z. M. & NEPVEU, A. 2014. CUX1, a haploinsufficient tumour suppressor gene

RIETHDORF, S., FRITSCHE, H., MULLER, V., RAU, T., SCHINDLBECK, C., RACK, B., JANNI, W.,
COITH, C., BECK, K., JANICKE, F., et al. 2007. Detection of circulating tumor cells in
peripheral blood of patients with metastatic breast cancer: a validation study of the


Figure Legends:

Figure 1. Experimental design of the study. Workflow used in the study to enrich for CTCs and CNV profiling using Ampl1 WGA and LowPass kit for Illumina. Following enrichment (EpCAM-dependent versus size-based platforms), single NEN CTCs and matched WBC are selectively recovered in dynamically controlled dielectrophoretic cages using the DEPArray Image-Assisted Digital Cell Sorter. CTC samples undergo WGA and QC prior to low-resolution whole-genome sequencing for CNV profiling. Where surgical resection or biopsy specimens are available, samples are processed for bulk LPWGS and single cell LPWGS as per CTCs.

Figure 2. Comparison of low-resolution whole-genome copy number profiles for CTCs and bulk tissue reveals reproduction of the majority of the CNV from the formalin fixed paraffin embedded (FFPE) and fresh frozen (FF) tissue in CTC samples. Unsupervised hierarchical clustering heat map of each analyzed individual CTC and tissue sample based on CNV from 3 SINET patients. Each patient is depicted with one color as shown on the phenobar at the bottom of the heat map. Individual CTCs are categorized according to enrichment method and tissue into bulk versus single cell FFPE (see key). Chromosomal CNV are shown from top to bottom for each individual cell or sample; copy-number gains are depicted in blue, losses in orange.

Figure 3. Individual CTC CNV data depicting complex intrapatient and interpatient genomic diversity. Unsupervised hierarchical clustering heat map of all analysed CTCs based on CNV across seven patients. Each patient is depicted with one color as shown on the phenobar at the bottom of the heat map along with cell sorting method and primary NET site.
Figure 4. Relationship between CTCs from all 7 NEN patients is revealed through TSNE analysis. (A) Single CTCs from all seven patients are visualised and can be identified by color in the phenobar at the top of the figure. Cells are also depicted according to enrichment strategy (see key). (B) TSNE of all analyzed CTCs and WBC.

Figure 5. Distribution plot describing the impact of enrichment strategy in patient 7 (A) and all patients (B) on the proportion of the genome that is aberrant and CNV diversity as quantified by Shannon Index. Each small line represents the described value for a single CTC. Large bars represent mean values.

Figure 6. Frequency of genomic amplifications and deletions across all CTCs. Profiles demonstrated for SINET patients (A) Patient 5; pancreatic NEN (B) Patient 6; GOJ NEN (C) and patient 7; renal NEN (D).

Supplementary Figure 1. Cluster analysis of copy number profiles for CD45 positive cells reveals balanced copy number profiles. Each patient is depicted with one color as shown on the phenobar at the bottom of the heat map. Profiles are distinct from CTCs and in keeping with WBC populations.

Supplementary Figure 2. Average pairwise correlation for CNV profiles within (diagonal) and between (off-diagonal) patients. After adjusting for ploidy, there was low correlation between individual patients. The degree of heterogeneity varied on a per patient basis, with LPWGS demonstrating more homogenous CNV profiles and therefore lower intra-patient heterogeneity in patients 3 and 6.
<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Sex</th>
<th>Age</th>
<th>Primary Site</th>
<th>Grade</th>
<th>Treatment</th>
<th>CellSearch CTCs</th>
<th>Parsortix CTCs</th>
<th>WBC</th>
<th>Fresh tissue single cells</th>
<th>FFPE single cells</th>
<th>FFPE bulk samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>74</td>
<td>Small intestine</td>
<td>2</td>
<td>na</td>
<td>4</td>
<td>na</td>
<td>4</td>
<td>na</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>45</td>
<td>Small intestine</td>
<td>3</td>
<td>PEN-221</td>
<td>15</td>
<td>5</td>
<td>na</td>
<td>2</td>
<td>Pituitary metastasis</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>69</td>
<td>Small intestine</td>
<td>2</td>
<td>na</td>
<td>7</td>
<td>8</td>
<td>na</td>
<td>8</td>
<td>na</td>
<td>Small bowel</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>65</td>
<td>Small intestine</td>
<td>1</td>
<td>SSA</td>
<td>18</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>Small bowel</td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>47</td>
<td>Pancreas</td>
<td>2</td>
<td>na</td>
<td>12</td>
<td>na</td>
<td>4</td>
<td>na</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>64</td>
<td>GOJ</td>
<td>3</td>
<td>na</td>
<td>11</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>1</td>
<td>na</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>33</td>
<td>Renal</td>
<td>2</td>
<td>PEN-221</td>
<td>21</td>
<td>11</td>
<td>10</td>
<td>na</td>
<td>2</td>
<td>na</td>
</tr>
</tbody>
</table>

**Table 1. Summary of clinical characteristics**

M: male, F: female, na: not applicable, GOJ: Gastro-oesophageal junction, SSA: Somatostatin analogues, PEN-221: novel antibody-drug conjugate, FFPE: formalin fixed paraffin embedded. All tissue samples are FFPE unless specifically indicated otherwise.
Figure 1. Experimental design of the study. Workflow used in the study to enrich for CTCs and CNV profiling using Ampli1 WGA and LowPass kit for Illumina. Following enrichment (EpCAM-dependent versus size-based platforms), single NEN CTCs and matched WBC are selectively recovered in dynamically controlled dielectrophoretic cages using the DEPArray Image-Assisted Digital Cell Sorter. CTC samples undergo WGA and QC prior to low-resolution whole-genome sequencing for CNV profiling. Where surgical resection or biopsy specimens are available, samples are processed for bulk LPWGS and single cell LPWGS as per CTCs.
<table>
<thead>
<tr>
<th>Patient</th>
<th>Sorting</th>
<th>Primary NET</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CellSearch</td>
<td>Tumour cell (FFPE)</td>
</tr>
<tr>
<td>2</td>
<td>Parsortix</td>
<td>Tumour cell (FF)</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**CN**
- 0.0
- 1.0
- 2.0
- 3.0
- 4.0
- 5.0
- 6.0
- 7.0
- 8.0
- 9.0+
- na

**Primary NET**
- GOJ
- Midgut
- Pancreas
- Renal
a

Proportion genome aberrant

CellSearch

Parsortix

Tumour cell (FFPE)

n.s.

Proportion genome aberrant

CellSearch

Parsortix

Tumour cell (FF)

Tumour cell (FFPE)

b

CN diversity (Shannon's)

n.s.

CN diversity (Shannon's)