Genomic characteristics and clinical significance of CD56+ Circulating Tumor Cells in Small Cell Lung Carcinoma

Charles RICORDEL*[1,2], Laura CHAILLOT*[1], Efstathios-Iason VLACHAVAS*[3,4], Marianthi LOGOTHETI[3], Anna JOUANNIC[1], Thomas DESVALLEES[5,6], Gwendoline LECUYER[1], Marc AUBRY[1], Georgia KONTOGIANNI[6], Chara MASTROKALOU[3], Florence JOUAN[1], Ulrich JARRY[5,6], Romain CORRE[2], Yannick LE GUEN[2], Thierry GUILLAUDEUX[1,5], Hervé LENA[1, 2], Aristotelis CHATZIIOANNOU[3,7], Rémy PEDEUX[1,5]

[1] Univ Rennes 1, INSERM, OSS (Oncogenesis Stress Signaling), UMR_S 1242, CLCC Eugene Marquis, F-35000, Rennes, France

[2] Service de Pneumologie, CHU Rennes, Rennes, France

[3] e-NIOS PC, Kallithea-Athens, Greece

[4] Division of Molecular Genome Analysis (B050), German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany.

[5] Univ Rennes, CNRS, INSERM, BIOSIT UAR 3480, US_S 018, Oncotrial, F-35000 Rennes, France

[6] Biotrial Pharmacology, Unité De Pharmacologie Préclinique, Rennes, France.

[7] Centre of Systems Biology, Biomedical Research Foundation of the Academy of Athens, 4 Soranou Ephessiou Street, 11527, Athens, Greece.

[*] These authors contributed equally to this work.

Corresponding author:
Charles Ricordel
Service de Pneumologie, Hôpital Pontchaillou
2 Rue Henri Le Guilloux
35033 RENNES
France
Phone: +33 (0)2 99 28 37 38
Fax: +33 (0)2 99 28 24 80
Mail: charles.ricordel@chu-rennes.fr
ORCID ID: 0000-0001-6160-2732

Co-corresponding author :
Rémy Pedeux
Université Rennes 1
CLCC Eugène Marquis, INSERM U1242-COSS
Rue Bataille Flandres Dunkerque
35042 RENNES, FRANCE
Tel : 33 (0)2 23 23 47 02
E-mail : remy.pedeux@univ-rennes1.fr

Abstract

Introduction: Circulating Tumor Cells (CTC) have been studied in various solid tumors. However, clinical utility of CTC in Small Cell Lung Cancer (SCLC) remains to be demonstrated. Most published studies are based on Cellsearch® isolation methods that did not capture CTC undergoing epithelial-mesenchymal transition. The aim of the CTC-CPC study was to develop an EpCAM-independent CTC isolation method allowing isolation of a broader range of living CTC from SCLC and decipher their genomic and biological characteristics.

Method: We initiated a monocentric prospective non-interventional study (CTC-CPC) from April 2016 including only treatment-naïve newly diagnosed SCLC. CD56+CTC were isolated from whole blood samples using an immunofluorescence-based method, at diagnosis and relapse after first-line treatment. Whole-exome-sequencing (WES) was performed on isolated CTC, matched PBMCs and tumor biopsies for genomic characterization, mutational signature classification and biological altered pathways identification in four patients. In addition, CTC Derived Xenografts (CDXs) were generated and submitted to RNA-sequencing.

Results: Phenotypic study confirms tumor lineage and tumorigenic properties of isolated cells for the 4 patients analyzed with WES. Notably, CDX model derived from CD56+CTC was successfully classified with an ASCL1 transcriptomic profile. WES of CD56+CTC and matched tumor biopsy reveal common genomic alteration frequently impaired in SCLC. Moreover, CD56+CTC were characterized by a high mutation load and a distinct mutational profile compared to match tumors biopsies. The BioInfoMiner (BIM) web platform allowed the identification of a genomic signature for the CD56+CTC at the time of diagnosis. In addition to classical pathways altered in SCLC related to neural cell adhesion, extracellular matrix organization, hypoxia, signaling kinase cascaded and immune response, we found new biological processes specifically affected in CD56+CTC at diagnosis, sumoylation and transcriptional regulation by AP-2 and RUNX2. High numeration of CD56+CTC (>7/mL) at

diagnosis was associated with ES-SCLC in a cohort of 33 patients. Finally, comparing genomic features of diagnosis and relapse CD56+CTC, we identify reminiscent altered oncogenic pathways (e.g. DLL3 or MAPK pathway) that are currently investigated in targeted therapy trials.

Conclusions: We report a versatile and easy-to-use method of CD56+CTC detection for newly diagnosed SCLC. Numeration of CD56+CTC at diagnosis is correlated with disease extension. Isolated CD56+CTC conserve tumorigenic properties and show a distinct mutational profile. We report a minimal gene set as a unique signature of CD56+CTC and identify new affected biological pathways enriched in EpCAM-independent isolated CTC in SCLC. Finally, our results suggest that studying CD56+CTC at relapse might help to identify the oncologic pathways that seeds tumor resistance to first-line treatment.

Introduction

Lung cancer is the leading cause of cancer-related death worldwide, accounting for 1.6 million death per year (1). Small cell lung cancer (SCLC) is an aggressive subset of lung cancer, representing 10 to 15% of newly diagnosed lung cancers per year and characterized by rapid-onset symptoms due to high tumor growth rate (2). At diagnosis, most patients present an extensive stage disease (ES-SCLC) for whom the recommended treatment has been for the last decades a palliative platinum-based chemotherapy regimen. Even if recently, the combination of chemotherapy with immune checkpoint inhibitors showed substantial outcomes improvement in two phase III trials (3,4), the prognosis remains poor with a median overall survival of about one year. For patients diagnosed with limited stage disease (LS-SCLC), concomitant chemo-radiotherapy is the recommended treatment option showing a median overall survival of over 20 months (5).

SCLC shows a high propensity for metastatic spreading and therefore generates a higher number of circulating tumor cells (CTC) compared to other tumor types (6). Recent studies have demonstrated that CTC isolated from patients diagnosed with SCLC can be deeply characterized at the genomic level and have a potential prognostic significance (7–9). Moreover, an in-depth analysis of single-cell CTC copy number variation (CNV) by whole-genome sequencing in a cohort of 31 SCLC patients was used to create a CNV classifier, that could reliably predict chemosensitivity (10). However, these genomic analyses, especially on single cell CTC are not sustainable in routine clinical practice. Moreover, it should be noted that in most studies, CTC isolation was based on the expression of epithelial markers (EpCAM). It is well documented that tumor cells migrating through the bloodstream undergo epithelio-mesenchymal transition (EMT), characterized by the loss of expression of epithelial

markers expression (11,12). Coherently, studies using a EpCAM-independent isolation method show a higher detection rate of CTC (13). A recent immunofluorescence study on a large cohort of 108 SCLC patients demonstrate that EpCAM-based methodology (Cellsearch®) was unable to detect CD56+ and/or TTF1+-CTC in more than 20% of patients (14). These results suggest that CTC isolation method relying on EpCAM markers does not circumvent the issue of EMT of CTC in SCLC. Consequently, there is a clear unmet need for comprehensive genomic studies performed on CTC isolated independently of EpCAM to achieve a better understanding of the biology of SCLC.

Here we describe a versatile and easy-to-use workflow for the detection, count and isolation of CD56+CTC from whole blood samples in a prospective cohort of 33 newly diagnosed patients with SCLC at our institution. Phenotypic analysis of isolated cells confirmed their tumor lineage. Notably, generation of CTC-Derived Xenografts (CDX) in immunocompromised mice support the tumorigenic properties of CD56+CTC. Our findings indicate that while there are genes altered between biopsy and CD56+CTC samples, there is noticeable mutational diversity in the liquid biopsies, highlighting a putative distinct molecular signature. Finally, CD56+CTC count at diagnosis was associated with the stage of the disease but not with clinical outcome.

Materials and methods

Patient and samples

We conducted a prospective non-interventional study in our institution (Rennes University Hospital), including only patients with a histologically or cytologically confirmed chemotherapy-naïve SCLC and eligible to a systemic treatment, starting from April 2016. The data lock was planned in December 2021 for all patients that achieve a minimal follow-up of at least 6 months. The recruitment is still active beyond this date. Any active neoplasia other than SCLC or known HIV, HCV or HBV infection, were considered as exclusion criterion. CD56+CTC were numerated and isolated from a whole blood sample, before the first administration of any cytotoxic agent and when disease relapsed (Fig 1A). Paired tumor tissue samples were obtained for four patients in order to compare CD56+CTC and biopsies genomic characteristics. Patients underwent conventional management of SCLC during all the course of the study and all clinical data was prospectively collected. LS-SCLC was defined as disease confined to one hemi-thorax and that could be safely encompassed in a single radiation field, otherwise patients were considered ES-SCLC. The International Association for the Study of Lung Cancer (IASCL) eighth edition of the TNM Classification for small cell lung cancer was used to stage patient's SCLC (15). Chemosensitive status was defined as patient without progressive disease within 3 months after the first-line of treatment, as opposed to chemorefractory status. Concerning the smoking exposure, patients were classified as active smoker, former smoker (no exposition during the last 12 months at diagnosis), never-smoker, or unknown. Progression-free survival (PFS) was defined as the time from the start of chemotherapy until disease progression or death from any cause and assessed using RECIST version 1.1. Overall survival (OS) was defined as the time from the start of chemotherapy until death from any cause. Patients who had not progressed at the time of the statistical analysis or who were lost to follow-up before progression or death were censored at

their last evaluation. This study was conducted in accordance with the Declaration of Helsinki and was approved by our institutional ethic committee (n° 15.120). All participants provided written non-opposition form.

Workflow Isolation of CD56+CTC from whole blood samples

The first step consists in a CTC enrichment from whole blood sample by immunomagnetic negative selection using the EasySep™ Direct Human CTC Enrichment Kit (StemCell Technologies®, Vancouver, Canada) according to manufacturer instruction. Briefly, blood samples were collected in EDTA-K2 tube (1x5mL tube) and in a heparin tube (2x10mL tubes) before first chemotherapy administration. Approximately 10 ml of the heparin blood samples were gently mixed and incubated with a CTC enrichiment coktail (StemCell Technologies®) at room temperature for 5 min in constant mixing. After dilution of 1:2 in PBS 2%-BSA the blood was incubated with buffer EasySep™ and RapidSpheres™ in "Big Easy" StemCell magnet for 10 minutes. The cells suspension was collected and the same immunomagnetic depletion step was repeated once. Then cells suspension was centrifuged and the pellet resuspended in 70µL of PBS 2% BSA. Because we expected to isolate a very small number of cells and to accurately sort the CTCs, the remaining cells were mixed with $1x10^5$ PBMCs. Cell suspension was washed in PBS 2% BSA and incubated with saturating concentrations of fluorescent labelled antibodies against human CD56, CD45 and Cd235a (Miltenyi, Bergisch Gladbach, Germany) for 15 min at 4°C. Cells were then washed with PBS 2%-BSA and analyzed and sorted by flow cytometry using a FACSAria II (Becton Dickinson, Franklin Lakes, NJ). The population of interest was gated according to its CD56/CD45 criteria to avoid NK lymphocytes contamination (exclusion of CD56+/CD45+ cells). Cells CD56+/CD45- were sorted, centrifuged, and immediately freeze at -80°C.

CD45+ peripheral blood mononuclear cells (PBMC) were used as a negative control. Data were analyzed with FlowJo™ Software (Ashland, OR: Becton, Dickinson and Company).

PBMC isolation from whole blood samples

The blood sample collected in EDTA-K2 tube (1x5mL tubes) was diluted in an equal volume of PBS 2% BSA. Five milliters of Ficoll-Paque™ (Miltenyi) was carefully added on the suspension, and then centrifuged 20 minutes at 1200 x g at room temperature without brake. The mononuclear cell layer was carefully aspirated and diluted in an equal volume of PBS 2% BSA. The suspension was centrifuged, washed in PBS 2% BSA and immediately freeze at -80°C.

Extraction of DNA

Genomic DNA (DNAg) of CTC and PBMC were prepared using the REPLI-g Single Cell Kit and QIAamp DNA Mini Kit respectively (both from Quiagen, Hilden, Germany), according to manufacturer's recommendations. Tumor biopsies were obtained from primary site for 4 patients. Biopsy sections was deparaffinized with Deparaffinization Solution (Qiagen) at 56°C for 3 minutes. Biopsy DNA was extracted using QIAamp DNA FFPE Tissue Kit (Quiagen) according to manufacturer's recommendations.

Whole exome sequencing

For each sample, total genomic DNA library was prepared and control quality of process (Fragment Analyzer and qPCR assays) was performed at each step. Library preparation, exome capture and sequencing were performed by Integragen SA (Evry, France) according to manufacturer’s instruction and relative protocols (SureSelect, Agilent) without modification except for library preparation performed with NEBNext® Ultra II kit (New England

Biolabs®). Genomic DNA was captured using Agilent in-solution enrichment methodology with their biotinylated oligonucleotides probes library (SureSelect Clinical Research Exome V2, Agilent Technologies). Sequencing was performed on a Illumina® HiSeq4000™ platform in paired-end mode (2x75 bases). The average depth of constitutional, CTC and tumor exomes were 80X, 135X and 135X, respectively. Whole-exome sequencing data have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-10766. Somatic variant discovery and variant annotation workflow are described in the supplemental material & methods.

Pathway analysis and comparison of genomic mutational patterns between Biopsy and paired CD56+CTC samples

Initially, to unravel the common mutational patterns between CTCs and matched biopsies, a simple Venn diagram/overlap analysis was performed for each patient separately at the variant level (based on SNVs & INDELs) and at the gene level. The final variants that were common in the remaining variant lists of CTC and paired biopsy samples for each patient, were imported to the BioInfoMiner platform (https://bioinfominer.com) for the functional interrogation of the perturbed pathways in the CTC and biopsy samples utilizing different hierarchical and ontological vocabularies (Gene Ontology & Reactome Pathways) (16). Additionally, in order to investigate the mutational relatedness of CTC and biopsy samples, based on their mutational profiles in the six base substitution catalogues, the R package MutationalPatterns (v 1.10.0) was used (17). Finally, semantic similarity (GO functional similarity) based on the aforementioned pathway enrichment analysis was performed between CTC and biopsy samples for each patient, to identify any common molecular mechanisms.

Derivation of an informative gene signature of CD56+CTC samples

For the derivation of a minimal gene set, expanding the molecular profile of the circulating tumor cell samples in the SCLC patients, we applied a data mining approach based on the aforementioned results of the WES computational analysis. Briefly, the first step included the integration of two distinct gene lists. The one, included a subset of recurrent somatic alterations in the CTC samples, which were present in at least in three out of the four patients (at the gene level, union of SNVs & InDels), whereas the other included the union of the copy number alterations (gain or loss), as described in Somatic variants discovery section (see supplementals). The final merged gene list, was then further filtered using an independent microarray dataset of human SCLC cell lines (18). The raw microarray data were analyzed through R/Bioconductor software (R version 3.6.1) (19,20). Normalization and pre-processing were performed using the R package *affycoretools* (v 1.56.0) (21). Additionally, to remove unexpressed genes, we applied a non-specific intensity filtering based on the density of the probeset expression values, to remove genes that were unexpressed in less than 20% of the total samples. It resulted in 2343 unique genes, representing the total genomic alterations of the CTC samples. Next, an additional dimensionality reduction of the merged feature set was conducted by selecting the "altered genes", which were also identified as "expressed" based on various SCLC cell lines gene expression microarray dataset (9731 genes). The resulting gene list, was used as an input to the BioInfoMiner web platform, resulted in a final set of 75 hub genes. A subsequent query of this signature to various cancer genomics databases and data portals-oriented to SCLC highlighted specific gene components, with pivotal role to lung cancer progression and manifestation. In parallel, we compared the final signature with the cBioPortal for Cancer Genomics open-source resource, utilizing 3 available SCLC studies (Cancer Gene Census, Intogen driver genes and MycancerGenome) with 229 patients in total.

In vivo xenograft experiments

NSG (*NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ)* mice were purchased from Charles River Laboratories (Wilmington, MA). Mice were subcutaneously injected with a cell suspension (obtained after circulating tumor cell isolation) on the left flank with Matrigel (v/v) (Corning, Corning NY), as previously described (9). Tumor size was monitored using calliper twice per week. All animal procedures met the European Community Directive guidelines (Agreement B35-238-40 Biosit Rennes, France) and were approved by the local ethics committee (Agreement APAFIS # 7163; regional ethics committee of Brittany; France), ensuring the breeding and the daily monitoring of the animals in the best conditions of well-being according to the law and the rule of 3R (Reduce-Refine-Replace). After tumor growth mice were euthanized when tumors reached 1000 mm3 and tumors were harvested and cut in two pieces (one used for cell culture and the other for formalin-fixation). Refer to supplemental material & methods for RNAseq analysis and identification of SCLC patient subtype.

Cell culture

Tumors were dissociated using the human tumor dissociation kit (Miltenyi) associated to the gentle MACS dissociator (Miltenyi Biotec, Paris) according to manufacturer's recommendations and cells were directly cultured in DMEM medium (Gibco, EnglandThermoFisher Scientific, Waltham, MA) supplemented with 5% fetal bovine serum (SVF), 0.5% Penicillin-Streptomycin, 8,4 ng/mL cholera toxin, 5ug/mL insulin, 0,4 ug/mL hydrocortisone, 24 ug/mL adenine and 10 ng/mL EGF.

Immunohistochemistry

Five-micron sections of formalin-fixed paraffin-embedded tissues were transferred to glass slides and incubated in TBST in the presence of 5% BSA. Immunostaining was performed using an automated slide staining system (Discovery XT -Ventana Medical Systems). Whole

slide image acquisition was performed using a Nanozoomer 2.0-HT software (Hamamatsu Photonics K.K., Hamamatsu, Japan). Antibodies against human CD56 (123C3.D5), Synaptophysine and Chromogramine A Ab-3 (clone LK2H10 PHE5) were obtained from ThermoFischer Scientific.

Statistical analysis

Demographic and descriptive data are given as the median with the range. Categorical variables were compared with the Fisher exact test or the Pearson X^2 test, and quantitative variables were compared with the Mann-Whitney U test when appropriate. Two-tailed P-values were reported, with $P < 0.05$ considered as statistically significant. The Kaplan-Meier method with log-rank test was used to perform survival analysis. Finally, binomial logistic regression model was built with the $CD56^+$CTC count (□7 CD56+CTC or ≤7 CD56+CTC/mL of blood) as the dependent variable. The pre-specified cut-off of 7 CD56+CTC/mL was chosen, based on study investigating prognosis of CTC in SCLC which shows that more than 50 CTC/7.5mL (equivalent to 6.6 CTC/mL) correlate with survival (6). GraphPad Prism for Windows software (version 5.03; Graph Pad Software Inc, San Diego, CA) was used for the statistical and survival analyses.

Results

CD56+CTC isolated from whole blood samples are tumorigenic

To isolate CD56+CTC from SCLC patients, we have set up a two steps protocol. First, hematopoietic cells and platelets were depleted and then CD56+CTC were isolated by FACS sorting, using CD56 as a positive selection marker and CD45 as a negative selection marker to avoid NK cells contamination (Fig 1B). To validate and assess the efficiency of our EpCAM independent CTC isolation method, we performed a preliminary experiment using 20mL of whole blood samples obtained from healthy donor. Blood samples were contaminated with 750, 1500 or 3000 NCI-H69 cells and CD56+/CD45- cells were then isolated. Average recovery efficiency of CD56+/CD45- cells from whole blood sample was 57.1% (from 52.26% to 62.2%) and was stable over a wide range of NCI-H69 cells added to the samples (Supplementary Fig. S2A). We observed that a delay to process the blood samples led to decrease in the ability to capture CD56+/CD45- NCI-H69 cells, especially after 48 hours (Supplementary Fig. S2B). Based on these results, we processed all the samples for this study immediately after collection. A representative example of flow cytometry gating strategy used for cell sorting is shown in Figure 1B (right panel). In order to establish the tumorigenic proprieties of CD56+CTC, we performed CDX generation through implantation of isolated CD56+CTC on the flank of nude mice, according to previous published methodology (9). Out of five attempts, CD56+CTC from two patients generated tumors in mice (Figure 1C). Subsequent immunohistochemistry analysis on these CDX tumors show positivity for commonly used markers for SCLC diagnosis: CD56, synaptophysin and chromogranin A, demonstrating the stability of the patient's neuroendocrine tumor lineage (Fig. 1D). Moreover, *in vitro* CTC-derived cell lines were obtained from two patient's samples and amplified in culture (Fig. 1D). Consistently with the recent classification of SCLC, CDX model derived from patient 15 was successfully classified

with an ASCL1 transcriptomic profile (Fig. 1E). Altogether, these results confirmed the tumorigenic properties of CD56+CTC isolated from whole blood samples of patients with a treatment-naïve SCLC.

CD56+CTC share common genomic alteration with SCLC tissue biopsies but display a distinct mutational signature

To further characterize CD56+CTC at the genomic level, we submitted the CTC samples, matched PBMC and tumor biopsies to whole-exome-sequencing for four patients in our cohort. The average sequencing coverage (up to 135X and with a minimum depth of 20X) was 0.70, 0.87 and 0.88 for CTC, biopsy and PBMC respectively (Table 1). Many genes commonly found to be altered according to large scale genomic studies of small cell lung cancer (22), were also detected in our genomic analysis (Fig. 2A). Alterations in *TP53* or *RB1*, the two most consistently altered genes in SCLC, were found in $CD56^{+}$CTC of 3 patients (75%). CD56+CTC from one patient didn't show any *TP53* or *RB1* mutation but CNV analysis revealed a mono-allelic deletion of both genes (Fig. 2B). Moreover, our analysis showed that the altered pathway commonly found in SCLC, such as NOTCH signaling, were also observed in our CD56+CTC samples (Supplemental Fig. 3B). These results confirmed that the genetic profiles of CD56+CTC are similar to the known genomic landscape of SCLC.

Interestingly, the mutational load in CD56+CTC was much higher than in matched tumor biopsies for all patients (Table 1). The mean mutational load in CD56+CTC and tumor biopsies was 188.02 and 6.44 mutations/Mb respectively. The purity of CTC samples was higher than the corresponding biopsies. The majority of mutations found in the tissues were confirmed in liquid biopsy (except for patient 06 were only 39/103 mutations where shared between samples) (Fig. 2C and supplemental M&M). Moreover, CD56+CTC and tissue showed a different mutational signature profile (Fig. 2D). Indeed C>A transversion,

classically associated with tobacco exposure, was the dominant mutation type in the tissue biopsies. Conversely, C>T transversion (outside CpG sites) was dominant in blood samples, suggesting that drivers of mutagenesis might be distinct in CD56+CTC. In line with this hypothesis, hierarchical clustering of samples based on their mutational patterns using COSMIC mutational signatures (v2-March 2015) tended to discriminate between blood and biopsy samples (Supplemental Fig.3A). Even if CTC and biopsy samples showed high similarity with signature 3 (defect of homologous recombination), signature 25 (transcriptional strand bias for T>A mutations) and signature 5 (transcriptional strand bias for T>C substitutions at ApTpN), CD56+CTC were significantly associated with signature 15 (defective DNA mismatch repair). On the other hand, biopsies showed a higher correlation with signature 4, related to tobacco mutagens, a mutational process strongly related to SCLC. Overall, the highest difference was observed in signatures 19 & 30 (etiology unknown; mostly abundant in CTC) and signatures 24 & 29 (both associated with transcriptional strand bias for C>A mutations; major enrichment in biopsies), supporting the idea of a distinct "circulating" drivers of mutagenesis in CD56+CTC clones.

To explore altered signaling pathways in CD56+CTC, we exploited BioInfoMiner web platform using our sequencing analysis as an input. The BioInfoMiner (BIM) web platform implements robust statistical and network analysis for functional enrichment investigation in order to highlight the most important biological mechanisms related to the phenotype under study. Thus, we identified a genomic signature comprised of 75 hub genes for the CD56+CTC at the time of diagnosis (Fig.2E). In order to capture holistically the perturbed biological components that are related to SCLC pathophysiology and progression, we utilized one of the most comprehensive biomedical vocabularies REACTOME pathway analysis. Subsequent functional enrichment analysis of the compact gene set revealed altered biological mechanisms related to neural cell adhesion, extracellular matrix organization, hypoxia,

signaling kinase cascaded and immune response (Fig.2F). Notably, the majority of these processes are biologically either directly or indirectly associated with lung cancer progression or invasive potential. This analysis also stressed the importance of understudied processes so far in SCLC: sumoylation and transcriptional regulation by AP-2 and RUNX2. Subsequently, three clinical databases of SCLC (Cancer Gene Census, Intogen driver genes and MycancerGenome) were analyzed with BIM leading to the identification of 22, 17 and 9 hub genes respectively (Supplemental Table 1). Interestingly, 23 of the 75 hub genes from CD56+CTC, such as *ERBB2*, *TP53*, *CREBBP* and *NF1*, were found enriched in at least 1 of the 3 queried databases, illustrating the high affinity of the liquid biopsies to recapitulate the molecular genomic landscape of small cell lung cancer. The REACTOME pathway analysis of the 23 hub genes further confirmed that the altered biological mechanisms found in CD56+CTCs correspond to those found in SCLC, with the addition of the p53 signaling pathways (Supplemental Fig.3D).

Altogether, these observations support the fact that CD56+CTC can reliably recapitulate the mutational status of treatment-naive SCLC tumors. Moreover, as it can detect more genomic alteration than tumor tissue biopsy, which is inherently limited to a fraction of the clonal population, analysis of CD56+CTC may represent a more comprehensive way of capturing the whole mutational picture of SCLC. Moreover, our functional enrichment analysis helps to clarify the biological specificities of CD56+CTC.

Cohort demographics

From April 2016 to April 2021, 46 patients were eligible for inclusion in the study. Seven patients met the exclusion criteria and 39 patients had CD56+CTC isolation from blood samples. For five patients, isolated $CD56^{+}$CTC were used for *in vivo*/*ex vivo* culture and one patient withdrew consent. Finally, the number of CD56+CTC at diagnosis was available for 33 patients who were therefore included in the statistical analysis (Supplementary Fig.4). The

demographic and descriptive statistics of the cohort are presented in Table 2. As expected, patients were predominantly male (69.7%), active smoker (69.7%) and had mainly an extensive-stage disease at diagnosis (66.7%). Of note, a large proportion of the patients progressed within 3 months after the end of the first-line of chemotherapy and was considered chemorefractory (45.5%). Based on the pre-specified cut-off of 7 CD56+CTC/mL, two groups of patients were compared. The two groups were well balanced with respect with age, sex, smoking exposure, chemosensitivity or median follow-up (Table 2). However, the group with more than 7 CD56+CTC/mL at diagnosis was enriched in patients with ES-SCLC (85.7% vs 52.7%), without reaching statistical significance (Fisher's exact test; p= 0.067). Remarkably, the five patients without any metastatic localization at diagnosis show a low CD56+CTC count (mean: 3.23±3.30 cells/mL). Two patients show no detectable CD56+CTC in blood samples, whereas their IHC tumor biopsy was positive for CD56. Among these patients, one had LS-SCLC and the other one had a low thoracic tumor burden associated with multiple cerebral and medullary metastases.

Clinical and prognosis significance of CD56+CTC

In order to correlate tumor burden and propensity to detect CTC in SCLC patients, we compared the number of CD56+CTC according to the initial stage of the disease. Patients with ES-SCLC show significantly higher number of CD56+CTC compared with patient diagnosed with LS-SCLC (respective median of 7.95 vs 2.00; p=0.014) (Fig. 3A). To gain further insight into the specific clinical parameters associated with higher numbers of CTCs detected, we looked at each descriptor of the TNM classification (Fig 3B). Even if we observed a trend toward a higher count of CD56+CTC in patient with more advanced T, N or M status, statistical significance was not reached. As of December 2021, a relapse or death event occurred in all patients in our cohort after first line-treatment. The median PFS was 5.2 months in the overall cohort with 5.2 months and 5.3 months in the low CD56+CTC group

and high CD56+CTC group, respectively (HR=0.80 CI 95% [0.39-1.56]; p=0.5) (figure 3C, left panel). Median overall survival was 8.1 months in the overall cohort. No statistically significant difference was observed between the two groups: 10.16 months in the low CD56+CTC group versus 8.7 months in the high $CD56^{+}$CTC group (HR=0.92 CI 95% [0.43-1.93]; p=0.8) (figure 3C, right panel). Importantly, no difference in survival was observed between ES-SCLC and LS-SCLC in our cohort (Supplemental Fig.5).

These results demonstrate that >7 CD56+CTC/mL at diagnosis is associated with extensive stage disease and may reflect global tumor burden. However, as it does not translate into a poorer prognosis, the initial count of CD56+CTC might not be predictive of response under first-line systemic treatment.

Somatic evolution of CD56+CTC at progression after first-line treatment

In order to decipher the tumor biology of SCLC at relapse, we conducted CD56+CTC isolation for patient 04 at clinical progression (costal metastasis after 3 cycles of carboplatin-etoposide). CD56+CTC counts were higher at relapse compared to diagnosis: 42 cells/mL versus 8 cells/mL respectively. However, WES showed that the mutation load decreased almost two fold in CD56+CTC at relapse (Fig. 4A). Interestingly, most of the mutations (1667/2497=66.7%) observed in relapsed CD56+CTC were also detected in CD56+CTC isolated at diagnosis (Fig. 4B). Conducting functional analysis enrichment, we were able to identify altered pathways shared between diagnosis and relapse of circulating CD56+ tumor clones. Interestingly, oncogenic pathways known to drive tumorigenesis in SCLC were significantly enriched in common genomic alteration (Fig. 4C). Notably, signaling pathways currently investigated in targeted therapy trials (DLL3-targeting molecules for the NOTCH pathway, or PI3KCA/AKT inhibitors for the MAPK pathway) were implicated in both types of CD56+CTC (23,24). These results suggest that overcoming relapse after first-line systemic

treatment may require targeting these signaling pathways, as we speculate this gives a selective advantage to its persistence in resurgent SCLC clones.

Discussion

This prospective study demonstrates the feasibility of an EpCAM-independent and immunofluorescence-based CTC isolation in SCLC and its application in clinical practice. We confirmed that CD56+CTC display a neuroendocrine tumor lineage and conserve tumorigenic properties *in vivo* and *in vitro*. Coherently, whole-exome sequencing validated that CD56+CTC show typical genomic alterations and classical altered pathways commonly found in SCLC (e.g. TP53 mutation, RB1 loss) but they are also characterized by distinct mutational patterns. Altogether, these results support the specificity of our immunofluorescence-based isolation method. Although CD56+CTC numeration at diagnosis did not significantly impact prognosis in our cohort, we prospectively observed that ES-SCLC display a higher CD56+CTC count at diagnosis compared to LS-SCLC. Finally, we show for one patient of our cohort, that detection of CD56+CTC at progression after first-line treatment uncover shared oncogenic pathways between diagnosis and relapse circulating tumor cells.

As already reported in non-small cell lung cancer, EpCAM-independent detection of CTC show a high sensitivity (25). Indeed, median number of CTCs at baseline using CellSearch® approach in a large SCLC cohort was 14 CTC/7.5ml of blood (14), compared to 28 CTC/7.5ml of blood in our cohort. Interestingly, in a recent cohort restricted to ES-SCLC, the median number of CTC at diagnosis using CellSearch® was 30 CTC/7.5 ml of blood (26), whereas it reaches 225 CTC/7.5 ml of blood in the ES-SCLC patients in our cohort. As EMT is expected to be a more potent phenomenon when metastatic spread occurs, we speculate that our EpCAM-independent method allow us to capture CTCs from SCLC that remains undetected using CellSearch® approach, especially in late-stage SCLC. Supporting this hypothesis, similar observations were made in a large cohort of 108 SCLC patients where CellSearch® failed to detect CTCs for 22 patients whereas immunofluorescence-based approach succeeded (14). Moreover, intra- and inter-patient heterogeneity was observed for

EMT markers in both CTCs and circulating tumor microemboli in lung cancer (27) and could partially explain these discrepancies.

To our knowledge, this study is the first to report genomic features of EpCAM-independent CTC in SCLC. Beyond the confirmation of typical genomic alteration of SCLC in CD56+CTC, we demonstrate that the mutational load was high compared to paired tumor biopsies or published data on CTC of SCLC. This suggests that our EpCAM-independent method allow us to capture more exhaustively CTC heterogeneity. To support this hypothesis we report a set of 75 hub genes altered in CD56+CTC, comprising 23 core genes shared with public SCLC genomic databases. Coherently, unique CTC genomic signatures have been described in others solid malignancies (28,29), highlighting cell populations with different functional or metastatic potential. In the context of SCLC, larger genomic studies considering single-cell sequencing of EpCAM-negative CTC are needed to confirm the new pathways we identified as affected in our dataset. Of note, a higher number of patients from our cohort will be further sequenced and analyzed with the same workflow, in order to confirm our results. Moreover, further longitudinal data from sequencing of CTC at relapse are needed to establish firmly the relevant common mechanisms that might drive resistance to therapy.

Prognostic value of CTC from SCLC has been extensively reported before. Indeed, Hou et al. reported that more than 50 CTCs/7.5 ml of blood at diagnosis has a detrimental effect on PFS and OS (6). Similarly, Su et al. and Carter et al. reported that copy-number-alteration (CNA) analysis from single cell CTC could predict with chemosensitity status and PFS (7,10). In our study, no difference of CD56+CTC numeration was observed based on the chemosensitivity status of tumors and no association with prognosis was found based on the pre-specified cut-off value of 7 CD56+CTC/mL. In line with this result, a recent work from Masseratakis et using an immunofluorescence CTC-isolation method did not show correlation between CD56+CTC and clinical outcomes (14). Altogether, we hypothesized that CD56+CTC might

not correlate with the invasive potential or the aggressiveness of the primitive tumor and/or might not dictate tumor sensitivity to systemic therapy. One caveat to this hypothesis is the absence of survival impact of the stage of the disease in our cohort. Some confounding factors and the relative small size of our cohort might have limited the power of the survival analysis. A better understanding of the biology of EpCAM-negative CTC and of the processes of EMT in SCLC is needed to definitely address this question.

Our study has several limitations. The cohort is relatively small and patients were recruited in a single institution. A direct comparison of EpCAM-independent isolation method and CellSearch® could not be performed as this technology is not available in our institution. However, Masseratakis et al, confirmed the favorable comparison of an immunofluorescence-based CTC isolation from SCLC compared to the CellSearch® technology (14). Moreover, the tumor CD56 status was unknown for 33.3% (11/33) of patients in our cohort. Only one of all tumors (n=22) tested for CD56 using immunohistochemistry was negative in our cohort, but we were able to detect CD56+CTC for this patient. Of note, as combination of anti PD-L1 with chemotherapy was not at a standard of care during the larger inclusion period of the study, we couldn't assess the relevance of CD56+CTC upon immunotherapy. However, as EMT has been associated with inflammatory tumor microenvironment and elevation of multiple targetable immune checkpoint molecules in lung cancer (30), exploring the relation between CD56+CTC and EMT in the context of the new immunotherapeutic strategies in SCLC remains an interesting avenue to explore.

Conclusion

Isolation of CD56+CTC using an immunofluorescence-based EpCAM-independent method at diagnosis is feasible in SCLC and might capture more efficiently tumor genomic heterogeneity. CD56+CTC are characterized by a high mutation load and a distinct mutational signature from paired tumor biopsies. Above classical pathways altered in SCLC, we identify new biological processes specifically affected in CD56+CTC. High numeration of CD56+CTC is associated with ES-SCLC. Finally, detection of CD56+CTC at progression after first-line treatment might help to understand somatic evolution of SCLC and elaborate strategies overcoming therapeutic resistance.

Bibliography

1. Torre LA, Siegel RL, Jemal A. Lung Cancer Statistics. Adv Exp Med Biol. 2016;893:1–19.

2. van Meerbeeck JP, Fennell DA, De Ruysscher DKM. Small-cell lung cancer. Lancet Lond Engl. 2011;378:1741–55.

3. Horn L, Mansfield AS, Szczęsna A, Havel L, Krzakowski M, Hochmair MJ, et al. First-Line Atezolizumab plus Chemotherapy in Extensive-Stage Small-Cell Lung Cancer. N Engl J Med. 2018;379:2220–9.

4. Paz-Ares L, Dvorkin M, Chen Y, Reinmuth N, Hotta K, Trukhin D, et al. Durvalumab plus platinum-etoposide versus platinum-etoposide in first-line treatment of extensive-stage small-cell lung cancer (CASPIAN): a randomised, controlled, open-label, phase 3 trial. Lancet Lond Engl. 2019;394:1929–39.

5. Faivre-Finn C, Snee M, Ashcroft L, Appel W, Barlesi F, Bhatnagar A, et al. Concurrent once-daily versus twice-daily chemoradiotherapy in patients with limited-stage small-cell lung cancer (CONVERT): an open-label, phase 3, randomised, superiority trial. Lancet Oncol. 2017;18:1116–25.

6. Hou J-M, Krebs MG, Lancashire L, Sloane R, Backen A, Swain RK, et al. Clinical significance and molecular characteristics of circulating tumor cells and circulating tumor microemboli in patients with small-cell lung cancer. J Clin Oncol Off J Am Soc Clin Oncol. 2012;30:525–32.

7. Su Z, Wang Z, Ni X, Duan J, Gao Y, Zhuo M, et al. Inferring the Evolution and Progression of Small-Cell Lung Cancer by Single-Cell Sequencing of Circulating Tumor Cells. Clin Cancer Res Off J Am Assoc Cancer Res. 2019;25:5049–60.

8. Tay RY, Fernández-Gutiérrez F, Foy V, Burns K, Pierce J, Morris K, et al. Prognostic value of circulating tumour cells in limited-stage small-cell lung cancer: analysis of the concurrent once-daily versus twice-daily radiotherapy (CONVERT) randomised controlled trial. Ann Oncol Off J Eur Soc Med Oncol. 2019;30:1114–20.

9. Hodgkinson CL, Morrow CJ, Li Y, Metcalf RL, Rothwell DG, Trapani F, et al. Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. Nat Med. 2014;20:897–903.

10. Carter L, Rothwell DG, Mesquita B, Smowton C, Leong HS, Fernandez-Gutierrez F, et al. Molecular analysis of circulating tumor cells identifies distinct copy-number profiles in patients with chemosensitive and chemorefractory small-cell lung cancer. Nat Med. 2017;23:114–9.

11. Polioudaki H, Agelaki S, Chiotaki R, Politaki E, Mavroudis D, Matikas A, et al. Variable expression levels of keratin and vimentin reveal differential EMT status of circulating tumor cells and correlation with clinical characteristics and outcome of patients with metastatic breast cancer. BMC Cancer. 2015;15:399.

12. Krawczyk N, Meier-Stiegen F, Banys M, Neubauer H, Ruckhaeberle E, Fehm T. Expression of stem cell and epithelial-mesenchymal transition markers in circulating tumor cells of breast cancer patients. BioMed Res Int. 2014;2014:415721.

13. Hanssen A, Loges S, Pantel K, Wikman H. Detection of Circulating Tumor Cells in Non-Small Cell Lung Cancer. Front Oncol. 2015;5:207.

14. Messaritakis I, Stoltidis D, Kotsakis A, Dermitzaki E-K, Koinis F, Lagoudaki E, et al. TTF-1- and/or CD56-positive Circulating Tumor Cells in patients with small cell lung cancer (SCLC). Sci Rep. 2017;7:45351.

15. Nicholson AG, Chansky K, Crowley J, Beyruti R, Kubota K, Turrisi A, et al. The International Association for the Study of Lung Cancer Lung Cancer Staging Project: Proposals for the Revision of the Clinical and Pathologic Staging of Small Cell Lung Cancer in the Forthcoming Eighth Edition of the TNM Classification for Lung Cancer. J Thorac Oncol. 2016;11:300–11.

16. Dual IRE1 RNase functions dictate glioblastoma development. EMBO Mol Med. John Wiley & Sons, Ltd; 2018;10:e7929.

17. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. 2018;10:33.

18. Polley E, Kunkel M, Evans D, Silvers T, Delosh R, Laudeman J, et al. Small Cell Lung Cancer Screen of Oncology Drugs, Investigational Agents, and Gene and microRNA Expression. JNCI J Natl Cancer Inst. 2016;108:djw122.

19. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5:R80.

20. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods. Nature Publishing Group; 2015;12:115–21.

21. MacDonald JW. affycoretools: Functions useful for those doing repetitive analyses with Affymetrix GeneChips [Internet]. Bioconductor version: Release (3.14); 2022 [cited 2022 Apr 27]. Available from: https://bioconductor.org/packages/affycoretools/

22. George J, Lim JS, Jang SJ, Cun Y, Ozretić L, Kong G, et al. Comprehensive genomic profiles of small cell lung cancer. Nature. 2015;524:47–53.

23. Jin Y, Chen Y, Tang H, Hu X, Hubert SM, Li Q, et al. Activation of PI3K/AKT Pathway Is a Potential Mechanism of Treatment Resistance in Small Cell Lung Cancer. Clin Cancer Res. 2022;28:526–39.

24. Giffin MJ, Cooke K, Lobenhofer EK, Estrada J, Zhan J, Deegen P, et al. AMG 757, a Half-Life Extended, DLL3-Targeted Bispecific T-Cell Engager, Shows High Potency and Sensitivity in Preclinical Models of Small-Cell Lung Cancer. Clin Cancer Res Off J Am Assoc Cancer Res. 2021;27:1526–37.

25. Hofman V, Ilie MI, Long E, Selva E, Bonnetaud C, Molina T, et al. Detection of circulating tumor cells as a prognostic factor in patients undergoing radical surgery for non-small-cell lung carcinoma: comparison of the efficacy of the CellSearch Assay™ and the isolation by size of epithelial tumor cell method. Int J Cancer J Int Cancer. 2011;129:1651–60.

26. Cheng Y, Liu X-Q, Fan Y, Liu Y-P, Liu Y, Liu Y, et al. Circulating tumor cell counts/change for outcome prediction in patients with extensive-stage small-cell lung cancer. Future Oncol Lond Engl. 2016;12:789–99.

27. Hou J-M, Krebs M, Ward T, Sloane R, Priest L, Hughes A, et al. Circulating tumor cells as a window on metastasis biology in lung cancer. Am J Pathol. 2011;178:989–96.

28. Kanwar N, Hu P, Bedard P, Clemons M, McCready D, Done SJ. Identification of genomic signatures in circulating tumor cells from breast cancer. Int J Cancer. 2015;137:332–44.

29. Heitzer E, Auer M, Gasch C, Pichler M, Ulz P, Hoffmann EM, et al. Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing. Cancer Res. 2013;73:2965–75.

30. Lou Y, Diao L, Cuentas ERP, Denning WL, Chen L, Fan YH, et al. Epithelial-Mesenchymal Transition Is Associated with a Distinct Tumor Microenvironment Including Elevation of Inflammatory Signals and Multiple Immune Checkpoints in Lung Adenocarcinoma. Clin Cancer Res Off J Am Assoc Cancer Res. 2016;22:3630–42.

Table 1

		# of isolated cells/mL	# of mutations detected	Mutational load (/Mb)	WES coverage (%)	Purity (%)
Patient04	CD56+ CTC	8,0	13507	201	0.76	0.79
	Biopsy	-	633	9	0.89	0.76
Patient 14	CD56+ CTC	7,9	19523	290	0.71	0.92
	Biopsy	-	310	4.6	0.79	0.86
Patient 19	CD56+ CTC	83,3	10769	160	0.87	0.7
	Biopsy	-	388	5.8	0.88	0.29
Patient 06	CD56+ CTC	5,8	6870	102	0.46	0.95
	Biopsy	-	432	6.5	0.92	0.86

Table 2

		Total	CTC-CD56+ ≤7/mL	CTC-CD56+ >7/mL
N		**33**	**19**	**14**
Age	Median (Range)	63 (49-82)	60 (49-82)	67.5 (49-81)
Gender	Male (n,%)	23 (69,7%)	14 (73.7%)	9 (64.3%)
	Female (n,%)	10 (30,3%)	5 (26.3%)	5 (35.7%)
Smoking status	Active (n,%)	23 (69,7%)	13 (68.5%)	10 (71.4%)
	Former (n,%)	9 (27.3%)	6 (31.5%)	3 (21.4%)
	None (n,%)	1 (3%)	0 (0%)	1 (7.2%)
Stage	LS-SCLC (n,%)	11 (33.3%)	9 (47.3%)	2 (14.3%)
	ES-SCLC (n,%)	22 (66.7%)	10 (52.7%)	12 (85.7%)
Metastatic sites	Cerebral (n,%)	7 (21.2%)	6 (31.5%)	1 (7.15%)
	Exclusively non-cerebral (n,%)	21 (63.6%)	9 (47.4%)	12 (85.7%)
	None (n,%)	5 (15.2%)	4 (21.1%)	1 (7.15%)
CD56+CTC count (#/mL)	Median (Range)	3.76 (0-571.4)	1.76 (0-6.76)	30.02 (7.85-571.4)
Largest tumor diameter (mm)	Mean (±SD)	62.8 (±25.3)	66.6 (±26.7)	57.3 (±21.9)
Chemosensitivity	Chemosensitive (n,%)	18 (54.5%)	11 (57.9%)	7 (50%)
	Chemorefractory (n,%)	15 (45.5%)	8 (42.1%)	7 (50%)
Folow-up (days)	Median (Range)	244 (16-1742)	244 (20-1742)	250.5 (16-1731)

Figure and tables legends:

Figure 1. Isolation of CD56+CTCs and generation of pre-clinical models. **A**. Schematic workflow of the CTC-CPC study. **B**. Isolation procedure of CD56+CTC from whole blood sample (left panel); Representative result of flow cytometry cell sorting (right panel). **C**. Circulating tumor cells-derived xenograft (CDX) obtained from isolated CD56+CTC. **D**. IHC phenotype of *in vivo* models obtained from isolated CD56+CTC (upper panel); *Ex vivo* cell lines obtained from isolated CD56+CTC (middle panel); IF phenotype of *in vivo* models obtained from isolated CD56+CTC after flow cytometry cell sorting (lower panel). **E**. Heatmap and hierarchical clustering based on the expression data of the four (established) key transcription regulators that are (commonly) used to identify the SCLC molecular subtype. Samples are derived from 50 SCLC cell lines, from CCLE with known molecular subtypes and from one patient of unknown SCLC subtype. For every gene, the color scale indicates its relative expression, from blue (low) to red (high).

Figure 2. Genomic characterization of CD56+CTC. A. Heatmap of genomic alterations of SCLC candidate genes (CTC samples of 4 patients). Mutation rates are displayed in the top panel. **B.** Somatic copy number alterations in CD56+CTC (amplifications in red and deletions in blue); lower panel shows copy number alteration in published SCLC genomic studies. **C.** Venn diagrams represent non synonymous mutations (SNPs & InDels) in biopsies and CD56+CTCs. **D.** Mutation spectrum plot illustrates the relative contribution of the 6 base substitution types to the point mutation spectrum for each patient sample. Bars show the mean relative contribution of each mutation type over all patient samples per sample type (biopsy and CD56+CTC) and error bars indicate the standard deviation. **E**. Derived genomic signature comprised of 75 genes for the CD56+CTC at the time of diagnosis. **F**. Analysis of the Reactome pathway of the 75 genes CD56+CTC genomic signature (ranked according p-value).

Figure 3. Clinical significance of CD56+CTC at diagnosis. A. Diagnosis numeration of CD56+CTC according stage to SCLC staging (*: $p<0.05$; # : two values outside the limits of the graph). **B.** Numeration of CD56+CTC according to the 8th TNM IASCL classification (ns : non-significant) **C.** Kaplan-Meier curve of progression-free-survival (left panel) or overall survival (right panel) according to the numeration of CD56+CTC with a pre-specified cut-off value of 7 CTC/mL.

Figure 4. CD56+CTCat diagnosis and relapse share common genomic alterations. A. Mutational load of all samples from patient 04. **B.** Venn diagrams representing non synonymous mutations (SNPs & InDels) in CTCs at diagnosis and after relapse (right) for

patient 04. **C.** Analysis of the Gene Ontology of the common non synonymous genes (1667) at diagnosis and after relapse.

Table 1. Genomic analysis of tumor biopsies and CD56+CTC.

Table 2. Cohort demographics.

Supplemental Figure and Table legends:

Supplemental Figure 1. Schematic illustration of the Variant Annotation workflow employed in this study. The multi-step analysis procedure includes elimination of variants with low quality read, variant annotation, filtering of the genetic variants based on the effect prediction, as well as functional and frequency filtering of the final lists of variants. The final lists of genes from each sample occurred from the indels and rare SNPs heating exonic and splice site regions. The workflow also provides information concerning the somatic status in lung cancer datasets that are recorded in COSMIC database and the occurrence, which indicates the relative number of the recorded COSMIC samples for the resulted variants.

Supplemental Figure 2. Efficiency of the EpCAM-independent immunofluorescence-based CD56+CTC isolation method. A. Representative images of flow cytometry plot (upper panel) with indicated number of NCI-H69 cells contaminating 20mL of whole blood sample. Results are represented in a column graph with standard deviation (lower panel; n=3). **B.** Representative images of flow cytometry plot (upper panel) after indicated lapse time of contamination of 20mL of whole blood sample with NCI-H69 cells. Results are represented in a column graph with standard deviation (lower panel; n=3).

Supplemental Figure 3. Pathways altered in CD56+CTC and mutational signature. A. Heatmap of pairwise cosine similarity between the mutational profile of each sample and available COSMIC signatures. The signatures have been ordered according to hierarchical clustering (average linkage) using the cosine similarity between signatures, such that similar signatures are displayed close together. **B.** Signaling pathways recurrently affected and found significantly enriched after pathway analysis with cBioPortal in CD56+CTC. Red and blue boxes denote genes with activating and inactivating alterations, respectively. **C**. Derived genomic signature comprised of 75 genes for the CD56+CTC at the time of diagnosis (left panel). The respective color denotes the relative overlap of the 75 genes with 3 clinical databases (Cancer Gene Census, Intogen driver genes and MycancerGenome; red:≥1/3 incidence, blue:3/3 incidence) and are summarized in the derived genomic signature comprised of 23 hub genes for the CD56+CTC at the time of diagnosis (left panel)
C. Analysis of the Reactome pathway of the 23 genes hub CTCs genomic signature.

Supplemental Figure 4. Flow chart of CTC-CPC study (data cut-off : 31/12/21)

Supplemental Figure 5. Kaplan-Meier curve of overall survival according to the stage of the SCLC (ES: extensive-stage; LS : limited-stage).

Supplemental Table 1. CD56+CTC gene signature compared to SCLC tissue database. The respective color denotes the relative overlap of the 75 genes with 3 clinical databases (Cancer Gene Census, Intogen driver genes and MycancerGenome; red:≥1/3 incidence, blue:3/3 incidence)

Acknowledgement:

We'd like to thank, la Region Bretagne, La Ligue Contre le Cancer (Grand Ouest), l'Association pour la Recherche sur le Cancer (ARC) and Roche® company for their financial supports to the CTC-CPC study. This work was also supported by ANR through the Labcom Oncotrial projet 2014–2019 (Université de Rennes 1 – UAR Biosit – BIOTRIAL, Rennes, France), notably through grants to UJ, RP & TG, and by the "Région Bretagne" as part of a collaborative project with Biotrial supported by "Biotech Santé Bretagne". We thank the cell sorting facility (Cytometry, UAR Biosit, Rennes, France), the histopathology facility (H2P2, UAR Biosit, Rennes, France) for their technical expertise and the animal facility the animal facility (Arche, UAR Biosit, Rennes, France) for animal husbandry and care. We thank the DRCI department of the CHU Rennes for administrative support.

Data and code availability

Whole-exome sequencing data have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-10766. The computational integrative workflow and code used is available upon request.

Contributions:

CR conceived and designed the CTC-CPC study. CR, LC, TD, FJ and UJ performed the experiments. CR, RC, YLG and HL included patients in the study and collected clinical data. MA helped for the management of the genomic data collection. EV, ML, GL, AJ, MA, GK, CM, RP and AC were responsible for bioinformatics analysis. ML, CM and GK contributed also to the variant annotation workflow creation and copy number analysis performed. CR, EV, TG, HL, AC and RP analyzed the data. All authors approved the final manuscript.

Conflict of interest:

CTC-CPC study was initially funded by the JOTGO 2015 research grant with CR as a recipient. This grant was sponsored by Roche® company.

Figure 1

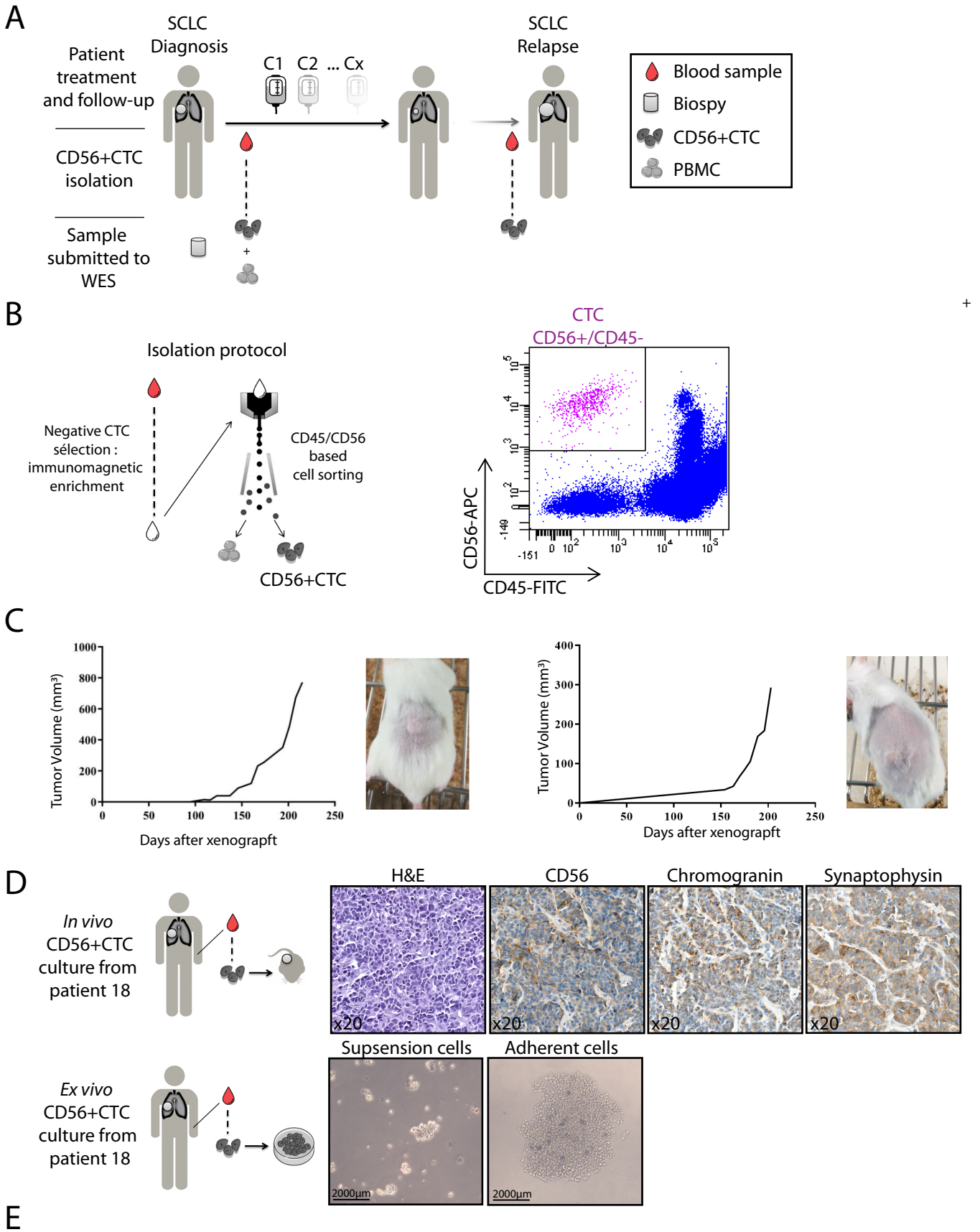
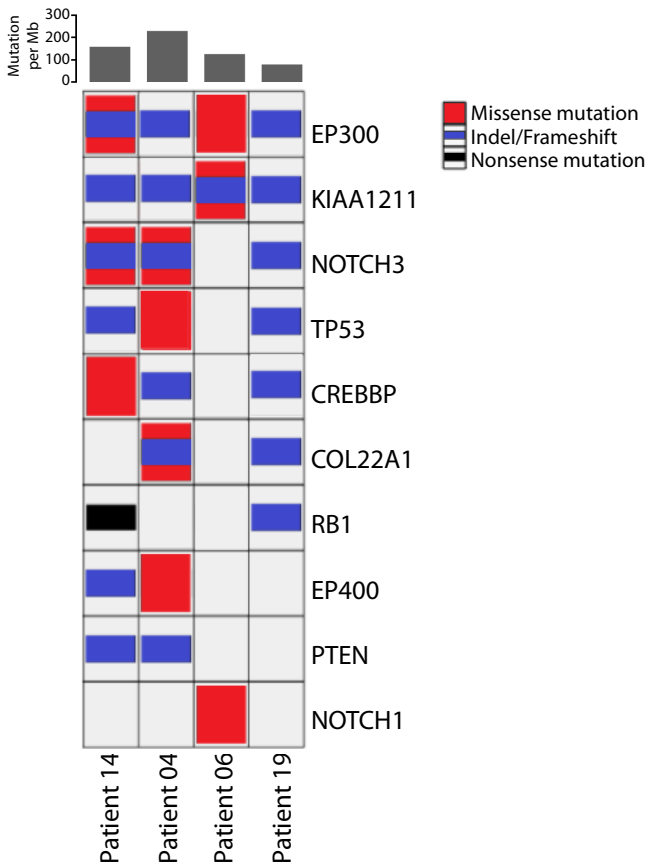
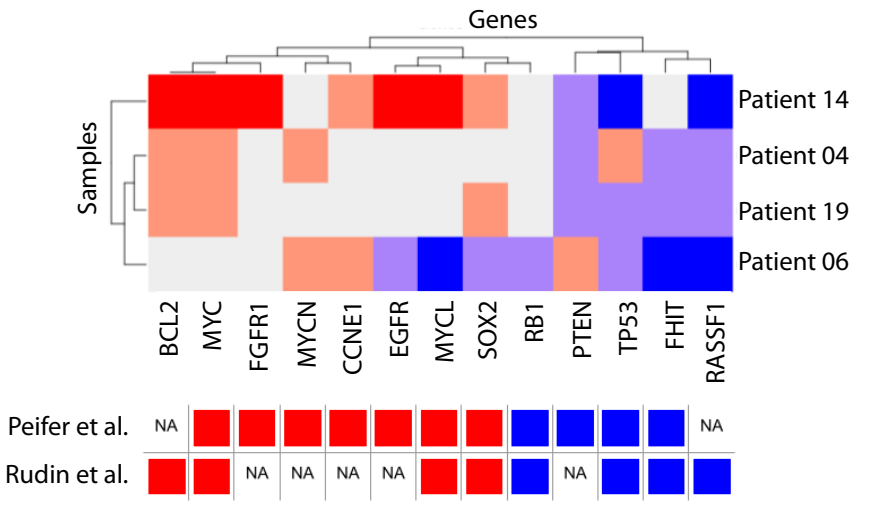

A
Patient treatment and follow-up
CD56+CTC isolation
Sample submitted to WES
SCLC Diagnosis
C1 C2 ... Cx
SCLC Relapse
Blood sample
Biospy
CD56+CTC
PBMC
B
Isolation protocol
Negative CTC sélection : immunomagnetic enrichment
CD45/CD56 based cell sorting
CD56+CTC
CTC
CD56+/CD45-
CD56-APC
CD45-FITC
C
Tumor Volume (mm³)
Days after xenograpft
Tumor Volume (mm³)
Days after xenograpft
D
In vivo CD56+CTC culture from patient 18
H&E
CD56
Chromogranin
Synaptophysin
x20
x20
x20
x20
Ex vivo CD56+CTC culture from patient 18
Supsension cells
Adherent cells
2000µm
2000µm
E
POU2F3
YAP1
NEUROD1
ASCL1
SCLC subtype
SCLC-A
SCLC-N
SCLC-P
SCLC-Y
Unknown
Gene expression
+2
+1
0
-1
-2
NCIH1048 NCIH526 CORL311 NCIH211 SW1271 NCIH2286 NCIH841 NCIH339 NCIH196 SBC5 NCIH1341 DMS114 DMS53 NCIH1876 DMS454 NCIH1184 NCIH889 NCIH1618 CORL88 NCIH1836 NCIH2029 CORL95 NCIH1105 CEM.18-03 NCIH510 CORL47 NCIH2196 SHP77 NCIH146 NCIH1930 NCIH1436 NCIH1963 NCIH2081 DMS153 DMS79 COLO688 NCIH209 NCIH1092 NCIH69 NCIH2066 CORL279 NCIH2171 SCLC21H NCIH82 NCIH1694 NCIH524 DMS273 CORL24 NCIH446 NCIH2227 HCC33

Figure 2

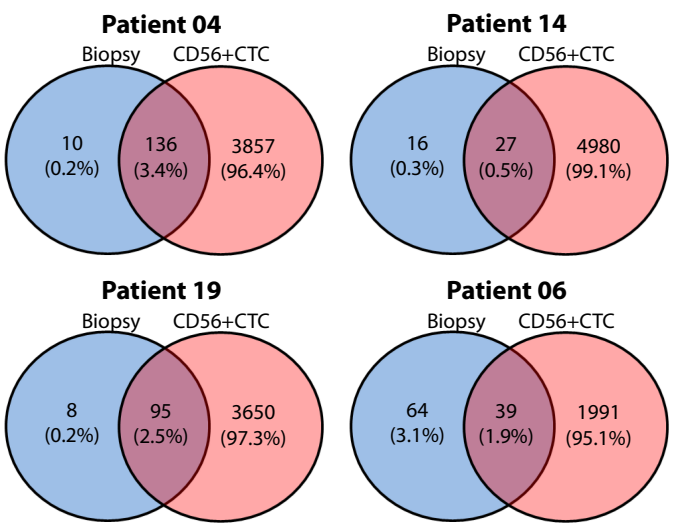
A



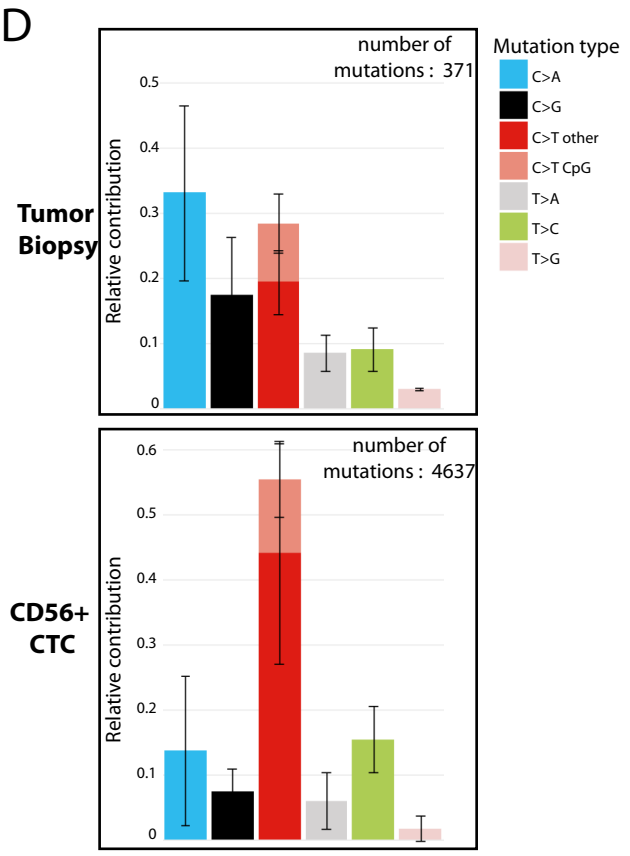
B



C



D



E

CD56+CTC 75 genes signature

AR	COL5A1	HCK	LRP6	PROX1
ARHGAP35	COL9A1	HDAC6	MAP3K12	PTK7
ARNT	COL9A3	HIF1A	MDM2	PTPRS
AXL	CREBBP	HIF3A	MECP2	ROCK2
BAX	CSF1	HIPK2	MED1	RUNX2
BMI1	DNM2	INSR	MEIS1	SIX4
BRAF	DSCAM	INSRR	MTOR	SMAD4
BTRC	EGLN1	ITGAV	NF1	STAT3
CACNA1A	EGR1	ITGB1	NR4A3	STAT5B
CACNA1C	EIF2AK3	JARID2	NRG1	TP53
CHD3	EP300	KALRN	PARP1	TRIM28
CITED2	EPHB1	KDM5B	PIP5K1C	TSC1
COL27A1	ERBB2	KMT2A	PLCB1	WNT1
COL4A2	ESR1	KMT2E	PRKCG	YAP1
COL4A3	FZD5	LAMC3	PRKDC	ZBTB7A

F

Rank	Term	Definition	Enrichment	Pvalue
1	1234158	Regulation of gene expression by Hypoxia-inducible Factor	6/11	0.0026
2	9006934	Signaling by Receptor Tyrosine Kinases	23/455	0.0067
3	8864260	Transcriptional regulation by the AP-2 (TFAP2) family of transcription factors	6/38	0.0094
4	3000178	ECM proteoglycans	8/76	0.0095
5	186797	Signaling by PDGF	7/58	0.0151
6	419037	NCAM1 interactions	6/42	0.0173
7	8948216	Collagen chain trimerization	6/44	0.0194
8	3108232	SUMO E3 ligases SUMOylate target proteins	11/181	0.0244
9	3000157	Laminin interactions	5/30	0.0244
10	2990846	SUMOylation	11/187	0.0281
11	8941326	RUNX2 regulates bone development	5/31	0.033
12	1234174	Cellular response to hypoxia	7/75	0.0383
13	3000171	Non-integrin membrane-ECM interactions	6/59	0.0387

Figure 3

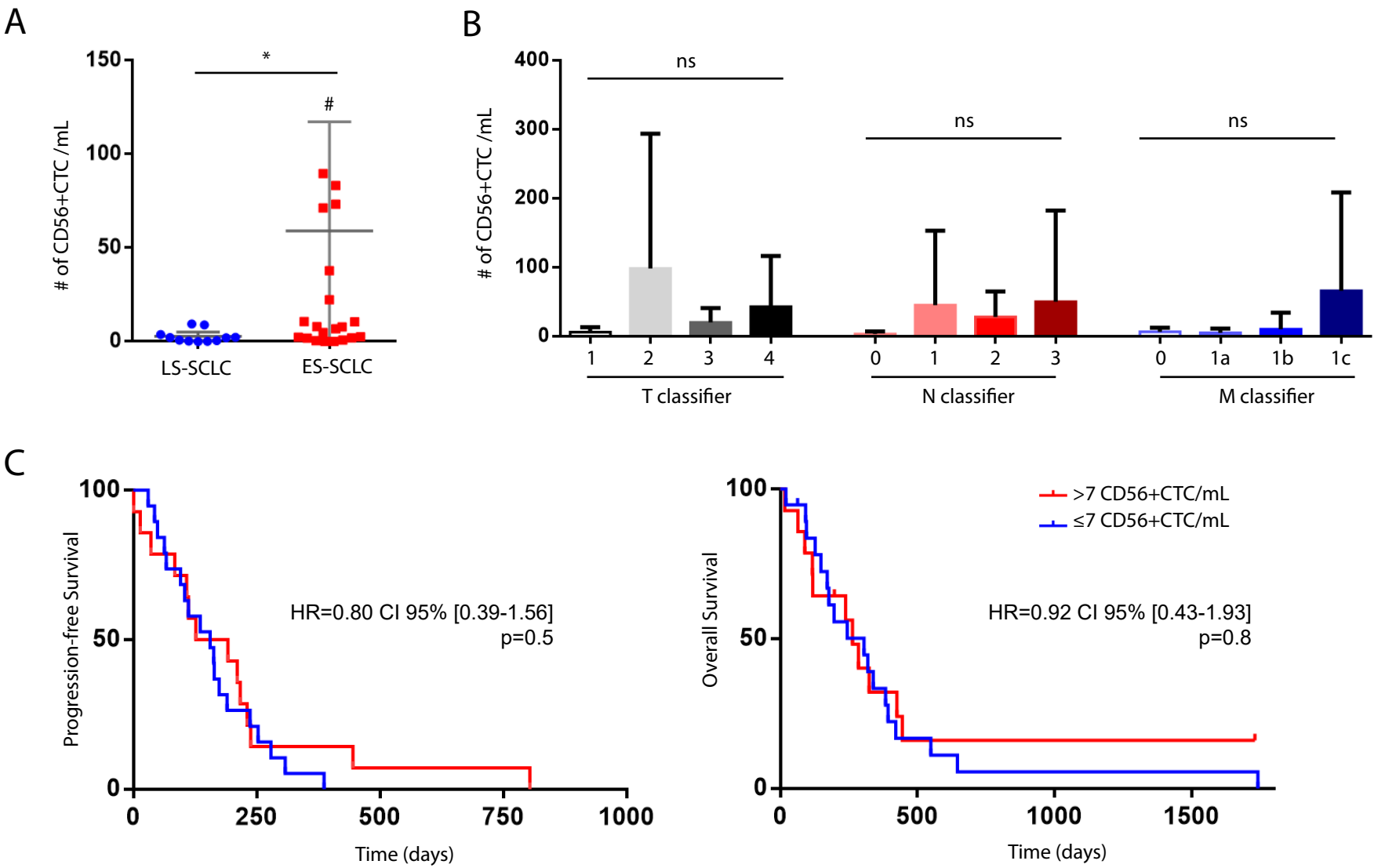
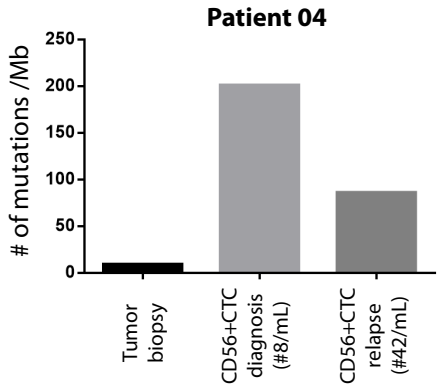
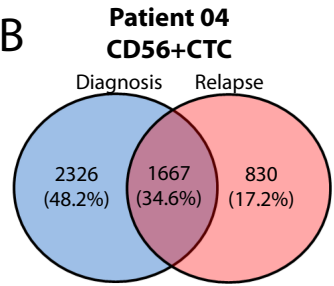

A
150
100
50
0
# of CD56+CTC /mL
*
#
LS-SCLC
ES-SCLC
B
400
300
200
100
0
# of CD56+CTC /mL
ns
ns
ns
1
2
3
4
0
1
2
3
0
1a
1b
1c
T classifier
N classifier
M classifier
C
100
50
0
Progression-free Survival
HR=0.80 CI 95% [0.39-1.56]
p=0.5
0
250
500
750
1000
Time (days)
100
50
0
Overall Survival
>7 CD56+CTC/mL
≤7 CD56+CTC/mL
HR=0.92 CI 95% [0.43-1.93]
p=0.8
0
500
1000
1500
Time (days)

Figure 4

A



B



C

Rank	Term	Definition	Enrichment	Pvalue
1	452723	Transcriptional regulation of pluripotent stem cells	6/31	0.0017
2	212676	Dopamine Neurotransmitter Release Cycle	5/23	0.0039
3	6794362	Protein-protein interactions at synapses	10/88	0.0055
4	210747	Regulation of gene expression in early pancreatic precursor cells	3/8	0.0082
5	8874211	CREB3 factors activate genes	3/9	0.0113
6	8856828	Clathrin-mediated endocytosis	13/145	0.0114
7	879518	Transport of organic anions	3/12	0.0146
8	199418	Negative regulation of the PI3K/AKT network	10/110	0.0161
9	9012852	Signaling by NOTCH3	6/49	0.0161
10	1257604	PIP3 activates AKT signaling	18/264	0.0194
11	9013507	NOTCH3 Activation and Transmission of Signal to the Nucleus	4/25	0.0211
12	9006925	Intracellular signaling by second messengers	20/303	0.0229
13	5675221	Negative regulation of MAPK pathway	5/40	0.0279
14	6811558	PI5P, PP2A and IER3 Regulate PI3K/AKT Signaling	9/103	0.03