Establishment of tumor patient derived cell line

CPP44-T and CPP45-M cell lines were derived respectively from primary and metastatic tumor from patients 44 and 45, as CTC44 and CTC45 lines comes from blood of patients 44 and 45. The two others cells lines CPP24 and CPP19 comes from primary and metastatic tumor respectively. For CTC44 and CTC45, respectively 4 and 5 sub-cell lines were obtained simultaneously and independently from separated wells. CTC 41 cell line was previously described as CTC-MCC-41 line (1).

Patient-derived cell culture of colon cancer cells (CPP19-24-25 and 30-44-45) were obtained from CRC biopsies provided by CHU-Carémeau (Nîmes, France) within an approved protocol (NCT01577511). Signed informed consents were obtained from patients prior to sample acquisition in accordance with all ethical and legal aspects (ethical agreement n° 2011-A01141-40).

Tumors were incubated 5 min in HBSS , bleached (0.4%) and washed 5 times in HBSS, minced into fragments (<2mm3) and digested with liberase H (0.26U/mL, Roche) diluted in Accumax (Sigma-Aldrich). After 2 hours at 37°C, cell suspensions were filtered through a 40 µm mesh to obtain a single cell suspension and plated in DMEM medium, supplemented with FBS, glutamine, antibiotics (100 µg/ml gentamicin and 40 µg/ml ciprofloxacin). For this study, patient derived cell lines have been grown in suspension and in M12 medium described above.

Subcutaneous CTC injection

200 000 cells from CTC lines were injected subcutaneously into the right flanks on female CD1 nude mice aged 6 weeks (Charles River, France) in a 1:1 mixture of Matrigel and DMEM in a final volume of 50 µL.

Additionally, 2000 cells were injected for CTC41 as a proof of concept to assess their tumor initiation ability at low concentration, which is one of the cancer stem cell features. Body weight and clinical signs of the mice were determined every second day, tumor volume [(length x width x thickness)/2] was measured using a caliper. Mice were sacrificed when tumors reached 500mm3. Procedures were carried out in accordance with the French guidelines for experimental animal studies (CEEA-LR-1026).

Intra splenic injection of CTC lines

This protocol used Female CB17 SCID (severe combined immunodeficiency) mice aged 6 weeks (Charles River, France). Procedures were carried out in accordance with the French guidelines for experimental animal studies (agreement CEEA-LR-12145).

An anesthetic solution of ketamine/medetomidine (18mg/ml; 0.1mg/ml) was injected in the peritoneum. A skin and peritoneal wall incision was performed under the rib cage on the left side of the mouse abdomen. The spleen was gently exteriorized from the abdominal cavity. Two million cells in a volume of 50 µl were injected with an insulin needle (30G) into the spleen extremity. After 2 minutes, a total splenectomy was performed by ligating the vessels and the peritoneum and the skin were closed by running sutures and wound clips. Animals were monitored daily if they started to show weight decrease or clinical signs.

Slide preparation for immunohistochemistry

Xenograft tumors

After sacrifice, tumors were harvested and fixed overnight with PFA 4%. After washing with PBS, tumors are dehydrated with serial ethanol baths followed by xylene baths before being embedded in paraffin.

Spheres

Spheres were centrifuged at 500g and washed in PBS before being fixed for 10 min in 4% PFA. Spheres were washed with PBS and are resuspended in pre-warmed histogel (Thermoscientific) and immediately placed on ice. Once solid, gel containing spheres is embedded as described for xenograft tumors above. 7µm slides were then cut using a microtome (Microm, HM335E).

Immunostaining

Embedded tumor slides were first dewaxed by heating at 56°C and immersing them in serial xylene and graded ethanol baths.

For immunofluorescence: Antigen retrieval was performed in boiling citrate buffer for 20 min and nonspecific binding sites were blocked for 1h in Phosphate Buffered Saline (PBS) containing 5% milk and 0,5% triton X-100. Slides were incubated overnight at 4°C with primary antibodies diluted in blocking buffer. Anti-Villin (1/250) was from Millipore (MAB 1671), anti-ChromograninA (1/250) was from Santa Cruz (SC-1488), anti-Mucin2 (1/100) was from Santa Cruz (SC-7314), anti-ECadherin (1/500) was from BD Biosciences (610181) and from Santa Cruz (SC-7870), anti-Cytokeratin20 (1/100) was from Novus bio (NB110-59940After washing, slides were incubated with fluorescent secondary antibodies (FluoroProbe, Interchim: 488 donkey IGG anti-rabbit, 647 goat IGG anti-rabbit, 568 donkey IGG anti-goat, 647 goat IGG anti-mouse, 594goat IGG anti-mouse, 488 goat IGG anti-mouse)), nuclei are stained with DAPI and finally samples were mounted in fluoromount G (SouthernBiotech). Slides were observed using an epifluorescent microscope (Zeiss AxioImager Z1 apotome).

For immunohistochemistry,nonspecific binding sites were blocked for 1h in Phosphate Buffered Saline (PBS) containing 5% milk and 0,5% triton X-100. Slides were incubated overnight at 4°C with primary antibodies diluted in blocking buffer anti-CD44V6 (1/50) was from R&D (2F10) and anti ALDH (1/50) was from BD Biosciences (611194) . after washing with PBS, secondary antibodies were developed using 3 3′ diaminobenzidine tetrahydrochloride (DAB). For this technique, slides were counterstained with hematoxylin and coverslipped with Permount. Slides were scanned using Hamamatsu nanozoomer and observed usind NDPI software.

Extreme Limiting Dilution Analysis (ELDA)

The frequency of cancer cells with in vitro tumorigenic potential was determined using the Extreme Limiting Dilution Analysis after cell sorting 100/50/10 of alive cells per well in M12 medium (n=20 wells/condition) in 96-well low attachment plate (Corning). Number of wells containing at least one sphere (>50µM) were counted after 7 days and analysis was performed using http://bioinf.wehi.edu.au/software/elda/.

Aldefluor assay

The aldefluor assay was performed according to the manufacturer's instructions (Stem Cell Technologies), using 2.5µl of reagent for 250,000 cells with or without 22.5µl of the ALDH inhibitor diethylaminobenzaldehyde (DEAB) for an incubation time of 30 minutes at 37°C. Sytox blue (Invitrogen) was added for dead cell exclusion. ALDH bright cells (ALDHbr) were identified in cell lines by comparing the same sample with and without DEAB.

Staining and flow cytometry analysis

CD44-APC (BD Pharmingen 559942), CD44v6-APC (R&D FAB36604), CD26-APC (BD Pharmingen 563670) or CD133/1-APC (MACS 130-098-829) antibodies (1/100) were incubated with 100 000 cells in PBS containing 5% FBS for 30 minutes at room temperature. Stainings are analyzed using the Cyan cytometer (Beckman Coulter) or MACSQUANT (Miltenyi) analyzer.

Pyrosequencing

Biotinylated PCR products were performed with an initial denaturation for 15 min at 95°C, followed by 40 cycles of denaturation for 30 s at 94°C, primer annealing for 30 s at 54°C, and extension for 30s at 72°C, followed by a final extension for 10 min at 72°C. All amplification reactions were performed in a DNA Thermal Cycler (Veriti, Thermofisher scientific, France) with Pyromark mastermix containg the Taq polymerase (Qiagen, France).Table 1 shows the set of primers used to amplify the sequence of interest.

Mutation	**PCR primer sequence**	**Sense**	**Sequencing primer**
V600	GAC AAC TGT TCA AAC TGA TG	forward	AGTAAAAATAGGTGATT

	Biotine- AGTAAAAATAGGTGATT	reverse	

Primer sequences for the amplicons used for PCR amplification and sequencing analysis

DNA product consisted of amplified tumoral DNA from control cells. Templates for the pyrosequencing analysis were carried out as recommended by the manufacturer: real-time pyrosequencing was performed at 28°C in an automated, PyromarkQ24. Prior to analysis, the enzymes and each of the four dNTPs (PyroMark Q24 Gold Reagents, Qiagen, France) were loaded into a special cartridge that was mounted in the PyromarkQ24 instrument. The sequencing primers are displayed in table 1.

Principle

The nucleotides are added one after the other according to a defined order. If the nucleotide is complementary to the strand being synthesized, it is incorporated by the polymerase and releases pyrophosphate. A ATPsulfurylase transforms it into ATP which is used by a luciferase to produce oxyluciferin and a light signal.
The pyrosequencer captures this light signal and represents the form of a peak on the pyrogram. The height of this peak is a function of the intensity of the light signal itself proportional to the number of incorporated nucleotides. One can therefore deduce the sequence from the size of the peaks obtained. Moreover, in case of point mutation, the size of the peaks provides a quantification of the proportion of carriers strands of any nucleotide.

RNA extraction and real-time PCR

Total RNA was extracted using the RNeasy mini kit (Qiagen) and treated with DNAse-1. The first strand cDNA was synthetized using MMLV (Invitrogen) and random hexamers primers, and gene expression

was measured by real-time PCR. PCR program was as follow: 95°C 6min, 45 cycles [94°C 15s; 63°C 15s; 72°C 25s]. Primer sequences are provided in Supplementary Table 6.

NGS

For CTC41, Pre-capture whole genome sequencing libraries were prepared for each of the 10 cell lines from 200ng of DNA using the NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs, CatNo.: E7370G) and corresponding index primers according to manufacturer's instructions. The resulting uniquely barcoded sequencing libraries were pooled and subjected to targeted capture using the IDT xGen Pan-Cancer Panel (Integrated DNA Technologies, CatNo.: 1055590) according to manufacturer's instructions. The final capture library was sequenced on a Illumina MiSeq Sequencer using 2x150 Paired-end V2 chemistry. Sequencing data analysis and variant calling were performed using the Targeted Enrichment Analysis workflow in Illumina MiSeq Reporter 2.4.60 analysis software. Variant annotation and interpretation was aided by Illumina Variant Studio 2.1.46 software. Only pass filter variants as determined by MiSeq Reporter default settings were considered. Candidate variants indicating heterogeneity between the ten cell lines were selected based on the following criteria: To be considered as candidate, a variant had to be categorised as high confidence (Read Depth > 50, Variant frequency > 30%, PASS in MiSeq Reporter variant call filter) in at least one of the ten cell lines to exclude artefacts. All 10 cell lines were then screened for the presence of the candidate variant, regardless of the high confidence criteria. Any candidate variant that was present in all 10 cell lines was discarded. The remaining eight candidate variants were inspected manually in all ten data-sets using IGV (2) (3), resulting in a further seven candidate variants being classified as homogenous. Variant data analysis and filtering was performed using custom scripts in R (http://www.R-project.org/).

For CTC44 and CTC 45, samples were assessed using a clinically accredited next generation amplicon sequencing approach at the Centre for Translational Pathology, University of Melbourne. Targeted regions of interest from exon 15 of the BRAF gene, exons 9 and 20 of the PIK3CA gene, exons 2, 3, and 4 of both KRAS and NRAS genes, and exons 18, 19, 20 and 21 of the EGFR gene were amplified using multiplex PCR. Sequencing of the final library was performed on an Illumina MiSeq Next

Generation Sequencer. Mutations were detected using MiSeq Reporter Software. This assay has an accredited limit of detection of approximately 3% mutation allele fraction.

RNA-sequencing

Paired-end reads were aligned to the hg19 reference genome using the TopHat2 (4) spliced aligner with the no coverage search and Bowtie2 sensitive options (5). Fragments were summarised by gene against a RefSeq annotation using the featureCounts function (6) of the SubRead package, ignoring those where both reads of the pair had mapping quality scores less than 10. Normalisation and differential expression testing was conducted using four two packages: edgeR, DESeq, DESeq2 and limma with the voom transformation, available via Bioconductor for the R programming environment. For this comparison the primary tumour samples were taken to be the reference and the CTC samples the condition of interest. The HTSFilter R package (7)was used to remove genes with consistent low expression at the stage suggested by the HTSFilter vignette for edgerR, DESeq and DESeq2. For DESeq2 this was done following normalization and testing replacing the default filtering as suggested by the HTSFilter documentation. A specific HTSFilter function for limma does not exist so filtering was performed on the raw count data prior to normalisation. The results of the four two methods were summarised and those genes found to be differentially expressed with an adjusted p-value of less than 0.05 by at least twoboth methods were taken as differentially expressed. The estimated log fold-changes from the four two methods for all genes were averaged, ignoring infinite values produced by DESeq. Genes filtered for all methods were assigned fold-changes of zero. These average fold-changes were used as input to GAGE to test the regulation of KEGG pathways in the MsigDB v5.0 database after converting to approve symbols using the HGNC symbol-checker. As GAGE expects fold-changes for all genes those that were filtered were assigned values of zero.

To compare the nature and breadth of variants detected in CPP44 vs CTC44 cells, we took the BAM files generated in the differential expression analysis and used the bam-readcount utility tool (https://github.com/genome/bam-readcount) to profile allele frequencies for all autosomal exonic bases (defined in Illumina Nextera Rapid Capture Exome v1.2 target regions). Our filtering criteria

required that mapping quality ≥20 and base call quality ≥Q20 for a base to contribute to depth of coverage. In addition, only exonic bases with ≥20x coverage in all 6 samples (3 CPP44 and 3 CTC44 replicates) were considered.

In vitro chemosensitivity assays

10 000 cell/well were plated in M12 in 96-well low adherent plates. After 24 hours cells were exposed to various concentrations of 5-FU and SN38 (FIRI) or vehicle (3 well/condition), vemurafenib or regorafenib. After 72 hours of treatment, cell viability was assessed by Cell Titer Glow assay (Promega) following manufacturer instructions.

Supplementary references

1. Cayrefourcq L, Mazard T, Joosse S, Solassol J, Ramos J, Assenat E, et al. Establishment and characterization of a cell line from human circulating colon cancer cells. Cancer Res. 2015 Mar 1;75(5):892-901.
2. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013 Mar;14(2):178-92.
3. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011 Jan;29(1):24-6.
4. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14(4):R36.
5. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Apr;9(4):357-9.
6. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014 Apr 1;30(7):923-30.

7. Rau A, Gallopin M, Celeux G, Jaffrezic F. Data-based filtering for replicated high-throughput transcriptome sequencing experiments. Bioinformatics. 2013 Sep 1;29(17):2146-52.