


Neutrophils escort circulating tumour cells to enable cell cycle progression

Journal Article

Author(s):

Szczerba, Barbara M.; Castro-Giner, Francesc; Vetter, Marcus; Krol, Ilona; Gkoutela, Sofia; Landin, Julia; Scheidmann, Manuel C.; Donato, Cinzia; Scherrer, Ramona; Singer, Jochen; Beisel, Christian; Kurzeder, Christian; Heinzelmann-Schwarz, Viola; Rochlitz, Christoph; Weber, Walter P.; [Beerenwinkel, Niko](#) ; Aceto, Nicola

Publication date:

2019-02-28

Permanent link:

<https://doi.org/10.3929/ethz-b-000330210>

Rights / license:

[In Copyright - Non-Commercial Use Permitted](#)

Originally published in:

Nature 566(7745), <https://doi.org/10.1038/s41586-019-0915-y>

Neutrophils Escort Circulating Tumor Cells to Enable Cell Cycle Progression

Barbara Maria Szczerba¹, Francesc Castro-Giner^{1,2}, Marcus Vetter^{3,4}, Ilona Krol¹, Sofia Gkountela¹, Julia Landin⁴, Manuel C. Scheidmann¹, Cinzia Donato¹, Ramona Scherrer¹, Jochen Singer^{2,5}, Christian Beisel⁵, Christian Kurzeder^{3,6}, Viola Heinzelmann-Schwarz³, Christoph Rochlitz⁴, Walter Paul Weber⁶, Niko Beerenwinkel^{2,5} & Nicola Aceto^{1,*}

¹ Department of Biomedicine, Cancer Metastasis Lab, University of Basel and University Hospital Basel, 4058 Basel, Switzerland

² SIB Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

³ Gynecologic Cancer Center, University Hospital Basel, 4056 Basel, Switzerland

⁴ Department of Medical Oncology, University Hospital Basel, 4056 Basel, Switzerland

⁵ Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland

⁶ Breast Center, University of Basel and University Hospital Basel, 4056 Basel, Switzerland

* To whom correspondence should be addressed:

Prof. Nicola Aceto
University of Basel and University Hospital Basel
Phone: +41-61-207-0773
Nicola.Aceto@unibas.ch

Summary

A better understanding of the features that define the interplay between cancer cells and immune cells is key to identify new cancer therapies¹. Yet, focus is often given to those interactions that occur within the primary tumor and its microenvironment, while the role of immune cells during cancer dissemination in patients remains largely uncharacterized^{2,3}. Circulating tumor cells (CTCs) are precursors of metastasis in several cancer types⁴⁻⁶, and are occasionally found within the bloodstream in association with non-malignant cells such as white blood cells (WBCs)^{7,8}. The identity and function of these CTC-associated WBCs, as well as the molecular features that define the interaction between WBCs and CTCs are unknown. Here, we achieve the isolation and interrogation of individual CTC-associated WBCs, alongside with corresponding cancer cells within each CTC-WBC cluster, from multiple breast cancer patients and mouse models. Single-cell RNA sequencing reveals a specific pattern of WBCs attached to CTCs, with neutrophils representing the majority of the cases. When comparing the transcriptome profiles of CTCs that were associated to neutrophils with that of CTCs alone, we detect a number of differentially expressed genes that outline cell cycle progression, leading to a higher ability to efficiently seed metastasis. Additionally, we identify cell-cell junction and cytokine-receptor pairs that define CTC-neutrophil clusters, representing key vulnerabilities of the metastatic process. Thus, the association between neutrophils and CTCs fuels cell cycle progression within the bloodstream and expands the metastatic potential of CTCs, providing a rationale for targeting this interaction in breast cancer.

Main Text

Circulating tumor cells (CTCs) are precursors of metastasis in various solid cancers including breast cancer⁶, and are occasionally found in association to white blood cells (WBCs)⁷. The role of CTC-WBC clusters in metastasis development as well as the principles that govern the interplay between CTCs and WBCs during blood-borne metastasis are largely uncharacterized.

We first sought to determine the number and composition of CTC-WBC clusters in breast cancer patients and mouse models. We obtained blood samples from 70 patients with invasive breast cancer that discontinued their treatment due to progressive disease, as well as from five different breast cancer mouse models, and we enriched for CTCs using the Parsortix microfluidic device⁹ (**Extended Data Fig. 1a-e**). Live CTCs were stained for cancer-associated cell surface markers EpCAM, HER2, and EGFR or imaged directly for the expression of GFP, as well as labeled for CD45 to identify WBCs (**Fig. 1a** and **Extended Data Fig. 1f**). Among 70 patients, 34 (48.6%) had detectable CTCs, with a mean number of 22 CTCs per 7.5ml of blood (**Supplementary Tables 1** and **2**). While the majority of CTCs were single (88.0%), we also detected CTC clusters (8.6%) and CTC-WBC clusters (3.4%) (**Fig. 1b** and **Extended Data Fig. 1g,h**). Similarly, we observed that CTC-WBC clusters were present in all tested mouse models, comprising those with immunodeficient or immunocompetent background, ranging from 0.05% to 61% of the total CTC population (**Fig. 1b** and **Extended Data Fig. 1i,j**). Importantly, CTC abundance and ratios dramatically changed when drawing blood upstream of capillary beds as opposed to more downstream locations, indicating that clustered CTCs are shed early, yet may be trapped in capillaries before reaching the periphery (**Extended Data Fig. 1k-n**). Thus, CTC-WBC clusters are rare in the peripheral circulation, yet consistently found across breast cancer patients and mouse models.

We then asked what type of WBCs is found in CTC-WBC clusters. We made use of a robotic micromanipulator to dissociate CTC-WBC clusters (**Supplementary Video 1**), enabling single-cell RNA sequencing analysis of cluster-associated WBCs and their comparison to reference WBCs from matched donors (**Fig. 1c**) using reference component analysis (RCA)¹⁰. In patients, we found that 75% of CTC-associated WBCs relate to the myeloid lineage, while the remaining ones (25%) are similar to T cells (**Fig. 1d** and **Extended Data Fig. 2a,b**). Similarly, we found that 93% of CTC-associated WBCs from mouse models are also characterized by a myeloid cell-like expression profile (**Extended Data Fig. 2c-e**). To dissect the exact proportion of CTC-associated WBCs that are neutrophils, monocytes or macrophages, we labeled CTC-WBC clusters for Ly-6G, CD11b, F4/80, as well as with Wright-Giemsa staining to define nuclear morphology (**Extended data Fig. 2f,g**). We found that the vast majority (85.5-91.7%) of CTC-associated WBCs are positive for Ly-6G

and display a nuclear morphology typical of neutrophils, while a minority (8.3-14.5%) are monocytes (CD11b⁺/F4/80⁻/Ly-6G⁻) and no F4/80⁺ macrophages are found (**Fig. 1e,f** and **Extended Data Fig. 2h-j**). Further, RNA sequencing analysis also revealed *ARG1*, *CXCL1*, *CXCL2*, *CXCL10*, *CCL2*, *CXCR2* and *VEGFA* expression in most CTC-associated neutrophils from both patients and mouse models (**Extended Data Fig. 2k**), indicating that CTC-associated neutrophils share gene expression features of pro-tumor N2-like cells¹¹.

We next asked whether the presence of CTC-neutrophil clusters in breast cancer patients could predict disease outcome. Strikingly, patients in whom at least one CTC-neutrophil cluster is detected in 7.5ml of peripheral blood are characterized by a significantly worse progression-free survival compared to patients with ≥ 5 CTCs per 7.5ml of peripheral blood (previously defined as a threshold for adverse outcome¹²) (**Fig. 1g**), as well as when compared to all patients with no CTC-neutrophil clusters, patients with at least one CTC per 7.5ml of blood, or patients in whom either single CTCs or CTC clusters are found (**Extended Data Fig. 3a-c**). Additionally, we individually micromanipulated equal numbers of CTCs from CTC-neutrophil clusters, CTC clusters and single CTCs, spontaneously generated from tumor-bearing mice, and intravenously injected 100 CTCs per mouse in tumor-free recipient mice from each of these categories. We found that mice injected with CTCs from CTC-neutrophil clusters develop overt metastasis much faster than those injected with CTCs alone, and accordingly, survive less (**Fig. 1h** and **Extended Data Fig. 3d-h**). Thus, CTC-neutrophil clusters represent the most efficient metastasis-forming cell subpopulation among breast CTCs, and their presence in the bloodstream of patients is associated with a poor prognosis.

We next sought to determine the molecular consequences of the interaction between CTCs and neutrophils by dissociating CTC-neutrophil clusters and comparing the expression profile of CTCs from CTC-neutrophil clusters to that of CTCs alone (**Fig. 2a**). We first determined differential gene expression in the Balb/c-4T1-*GFP* model, where we could retrieve the highest number of CTCs from CTC-neutrophil clusters ($n=25$). Compared to CTCs that were not associated to neutrophils, we found that CTCs from CTC-neutrophil clusters are characterized by differential expression of a set of 51 genes, of which 41 are upregulated and 10 are downregulated (**Fig. 2b** and **Supplementary Tables 3,4**). Pathway analysis with the upregulated genes revealed that CTCs from CTC-neutrophil clusters display a remarkable enrichment in positive regulators of cell cycle and DNA replication programs compared to CTCs alone (**Fig. 2c**), independently of the number of detected features or reads in each individual sample (**Extended Data Fig. 4a**). The same results were also obtained from CTCs of patients (**Fig. 2c** and **Extended Data Fig. 4b**). Immunofluorescence staining of CTCs confirmed higher levels of Ki67 expression in CTCs from CTC-neutrophil clusters (**Fig. 2d,e**), in line with the RNA

sequencing results. In contrast, no significant differences were observed for genes involved in epithelial-to-mesenchymal transition, cancer stem cell markers, or platelet-related genes¹³ (**Extended Data Fig. 4c-h**). Further, we asked whether neutrophil proximity would confer a proliferative advantage to cancer cells at the level of the primary tumor, disseminated tumor cells (DTCs) and overt metastasis. Surprisingly, we found that Ki67 expression does not increase in cancer cells that surround tumor-infiltrated neutrophils in the primary tumor or overt metastasis (**Extended Data Fig. 5a-d**). Yet, a higher Ki67 expression is retained in DTCs from CTC-neutrophil clusters (**Extended Data Fig. 5e,f**), i.e. when they are deprived of other stromal-derived signals that are typical of overt disease.

We then asked which cytokines are expressed by CTC-associated neutrophils and paralleled by simultaneous expression of matching cytokine receptor(s) in the corresponding cancer cells. We found that four cytokines (*TNF α* , *OSM*, *IL1 β* , and *IL6*) are most frequently expressed by CTC-associated neutrophils of patients or patient-derived mouse models, and matched by the expression of their receptors by the corresponding cancer cells (**Extended Data Fig. 6a**). With a reverse approach, we also found that CTCs from CTC-neutrophil clusters most frequently express cytokines such as *CSF1*, *CSF3* (a.k.a. granulocyte colony-stimulating factor, G-CSF), *TGF β 3*, and *IL15*, possibly involved in neutrophil stimulation¹⁴⁻¹⁷, while corresponding neutrophils express their receptors (**Extended Data Fig. 6b**). Consistently, we observed that a 24h *in vitro* treatment (coherent with neutrophil lifespan¹⁸) with IL6, IL1 β or both was sufficient to confer proliferative advantage to 4T1 cells upon dissemination, leading to faster metastasis development and shorter overall survival of mice (**Fig. 2f,g** and **Extended Data Fig. 6c-e**). Further, CRISPR-mediated knockout of IL6 or IL1 β receptors in cancer cells, namely *IL6ST* and *IL1R1*, did not alter the frequency of spontaneously-generated CTC-neutrophil clusters but it suppressed their proliferative advantage (**Extended Data Fig. 6f-h**).

Given recent findings highlighting that the presence of myeloid cells in the primary tumor site leads to accumulation of mutational events¹⁹, we asked whether the mutational load of CTCs obtained from CTC-neutrophil clusters is different from that of CTCs alone in patients (**Fig. 3a**). Interestingly, we found that the mutational burden is similar both between CTCs isolated from CTC-neutrophil clusters and CTC alone, as well as when comparing donors with or without CTC-neutrophil clusters (**Extended Data Fig. 7a,b**). Yet, we observed that the overall frequency of C>T mutations was increased in CTCs from CTC-neutrophil clusters compared to CTCs alone, as well as in donors with CTC-neutrophil clusters, independently of the nucleotide context (**Fig. 3b** and **Extended Data Fig. 7c,d**). While a general, age-related accumulation of C>T mutations has been reported²⁰, we did not observe any age difference between the two groups (**Extended Data Fig. 7e**). Next, considering only high-impact mutations, we asked whether specific genes are exclusively and

recurrently mutated in donors with CTC-neutrophil clusters. This scenario would be consistent with a model whereby certain genetic alterations would influence the recruitment of immune cells to the primary tumor²¹, and increase the likelihood of generating CTC-neutrophil clusters. We found that a number of genes are indeed carrying high-impact mutations exclusively in donors with CTC-neutrophil clusters (**Extended Data Fig. 7f** and **Supplementary Table 5**). We then engineered 4T1-*GFP* cells to individually express all the mutations found in two of the most frequently mutated genes, i.e. *MERTK* and *TLE1*, and injected them in the mammary gland of NSG mice. We observed that the introduction of *TLE1* mutations 1787G>A or 1509G>C leads to a higher infiltration of neutrophils in the primary tumor and a higher proportion of spontaneously-generated CTC-neutrophil clusters (33-41-fold increase), without affecting primary tumor size (**Fig. 3c** and **Extended Data Fig. 7g-k**). These results are in line with recent observations involving *TLE1* function in regulating myeloid cells infiltration into normal and neoplastic tissues²². Conversely, co-culture of cancer cells with neutrophils did not result in the acquisition of mutations within the same hotspots (**Extended Data Fig. 8a,b**). Together, our data reveals that while the overall mutational load remains unchanged, donors with CTC-neutrophil clusters feature a higher proportion of C>T substitutions and the presence of high-impact recurrent mutations in genes that promote neutrophil recruitment, such as *TLE1*.

We next tested whether the depletion or augmentation of the total neutrophil population would affect the ratio of spontaneously-generated CTC-neutrophil clusters. We depleted neutrophils by *in vivo* treatment with neutralizing antibodies against Ly-6G (α Ly-6G) or conversely, we stably overexpressed G-CSF to stimulate the production and recruitment of neutrophils to the tumor site (**Extended data Fig. 9a**). As expected, while treatment with α Ly-6G reduced neutrophil infiltration to the primary tumor site, G-CSF augmented it without altering primary tumor size (**Extended Data Fig. 9b,c**). Yet, α Ly-6G-treated mice completely lack CTC-neutrophil clusters from the circulation and display a delayed CTC shedding rate compared to control mice, while overexpression of G-CSF leads to an earlier CTC release and substantially increases the proportion of CTC-neutrophil clusters (>88-fold) (**Fig. 4a** and **Extended Data Fig. 9d-f**). Consequently, neutrophil depletion or augmentation results in a delayed or accelerated metastasis development, respectively, mirrored by a shorter or longer overall survival of treated mice (**Extended Data Fig. 9g,h**). In contrast, neutrophil depletion with α Ly-6G is not effective when cancer cells are administered directly through intravenous injection of pretreated mice (**Extended Data Fig. 9i-l**). Of note, in our cohort, G-CSF treatment of breast cancer patients occurred more often in those patients that were positive for CTCs, including CTC-WBC clusters (**Extended Data Fig. 9m**). Thus, overall neutrophil abundance impacts the likelihood that a tumor has to spontaneously shed CTC-neutrophil clusters.

We next sought to identify actionable vulnerabilities of CTC-neutrophil clusters without targeting the entire neutrophil population. To this end, we investigated cell-cell junction pairs expressed by CTC-neutrophil clusters and possibly mediating their heterotypic cell binding (**Extended Data Fig. 10a,b**). We engineered a CRISPR/Cas9-based loss-of-function minipool screen *in vivo*, whereby a pool of cells carrying individual knockouts for *F11R*, *ICAM1*, *ITGB2* and *VCAM1* are injected in the mammary gland of recipient mice, followed by CTC targeted barcode sequencing to reveal selective sgRNA dropouts, highlighting genes whose knockout does not allow CTC-neutrophil clusters formation (**Fig. 4b**). Importantly, we observed no differences in primary tumor growth and no selective sgRNA dropouts in primary tumor cells (**Extended Data Fig. 10c,d**), suggesting that knockout of *F11R*, *ICAM1*, *ITGB2* or *VCAM1* does not affect proliferation in the primary tumor. Yet, we found that 4/4 *VCAM1* sgRNAs selectively dropped out in CTCs from CTC-neutrophil clusters, while they were still present in CTCs alone (**Fig. 4c**), highlighting a possible *VCAM1* requirement for CTC-neutrophil clusters formation. We further validated this finding using individual sgRNAs (**Fig. 4d** and **Extended Data Fig. 10e**). Thus, *VCAM1* functionally mediates the interaction between CTCs and neutrophils, and its inhibition does not allow the formation of CTC-neutrophil clusters.

Altogether, our data provide new insights into the processes that define the interaction between cancer cells and immune cells during blood-borne dissemination. We propose a model whereby neutrophils directly interact with CTCs to support cell cycle progression in circulation and to accelerate metastasis seeding. This mechanism of metastatic spread and the possibility that CTC-neutrophil clusters may be targeted therapeutically provides an opportunity to reduce the spread of breast cancer.

Methods

Patients. After obtaining written informed consent, breast cancer patients donated 7.5–15ml of blood in EDTA vacutainers at least once. All blood specimens were obtained at the University Hospital Basel under the study protocols EKNZ BASEC 2016-00067 and EK 321/10, approved by the Swiss authorities (EKNZ, Ethics Committee northwest/central Switzerland) and in compliance with the Declaration of Helsinki. Involved patients were characterized by invasive breast cancer (all subtypes), high tumor load and progressive disease at the time of blood sampling.

Cell culture. MDA-MB-231 LM2 human breast cancer cells (obtained from Dr. Joan Massagué, MSKCC, NY, USA) and 4T1 murine breast cancer cells (ATCC) were grown in DMEM medium (#11330-057, Gibco) supplemented with 10% FBS (#10500064, Gibco) and antibiotic/antimycotic (#15240062, Gibco) in a humidified incubator at 37 °C with 20% O₂ and 5% CO₂. Py2T cells were a gift from Dr. Gerhard Christofori (University of Basel, Switzerland). Human CTC-derived BR16 cells were generated as previously described²³ from a breast cancer patient at the University Hospital Basel, and propagated as suspension cultures in a humidified incubator at 37 °C with 5% O₂ and 5% CO₂. Human CTC-derived Brx-50 cells were obtained from Drs. Daniel Haber and Shyamala Maheswaran (Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA). Cell lines were not authenticated, did not belong to the list of commonly misidentified cell lines (International Cell Line Authentication Committee) and resulted negative for Mycoplasma contamination. All cell lines were transduced with lentiviruses carrying GFP-Luciferase (*GFP*) at a multiplicity of infection (MOI) < 5.

Mouse experiments. All mouse experiments were carried out according to institutional and cantonal guidelines (approved mouse protocol #2781, cantonal veterinary office of Basel-City). Maximal approved tumor volumes of 2800 mm³ were never exceeded. Nod Scid Gamma (NSG) and Balb/c female mice were purchased from The Jackson Laboratory (Bar Harbor, Maine, USA) and kept in pathogen-free conditions, accordingly to institutional guidelines. Transgenic MMTV-PyMT female mice were obtained from Dr. Gerhard Christofori (University of Basel). Orthotopic breast cancer lesions were generated in 8-10 weeks old female NSG females upon the injection with either 1x10⁶ LM2-*GFP*, 0.5x10⁶ 4T1-*GFP* or 1x10⁶ BR16-*GFP* cells into the mammary fat pad. Similarly, female Balb/c mice received a syngeneic graft of 0.5x10⁶ 4T1-*GFP* cells. In all cases, breast cancer cells were inoculated in 100µl of 50% Cultrex PathClear Reduced Growth Factor Basement Membrane Extract (#3533-010-02, R&D Biosystems) in PBS. Blood draw for CTC analysis and organ dissection were performed after 3 weeks for NSG-4T1-*GFP*, 4-5 weeks for Balb/c-4T1-*GFP* and NSG-LM2-*GFP*, 5 months for NSG-BR16-*GFP* and at 13 weeks of age for MMTV-PyMT mice. Generally, immunocompetent models (Balb/c-4T1-*GFP* and MMTV-PyMT) developed a primary tumor that reached the

maximum allowed size before developing overt metastatic disease. For this reason, they were rather used throughout the manuscript as models to assess direct metastatic potential of cancer cells injected directly in the venous circulation (i.e. tail vein). In contrast, immunocompromised models (NSG-4T1-*GFP*, NSG-LM2-*GFP* and NSG-BR16-*GFP* mice) were used as the preferred system to assess spontaneous CTC and metastasis formation from the primary tumor. All experiments whereby both immunocompetent and immunocompromised mice were used side by side have led to the same conclusions. All mice were randomized before mouse experiments, blindly selected prior to injection, and sample size was chosen based on power analysis.

CTC capture. Human CTCs were captured from unprocessed peripheral blood samples with the Parsortix microfluidic device using Cell Separation Cassettes (GEN3D6.5, ANGLE), within 1 hour from blood draw. Next, in-cassette staining was performed with an antibody cocktail containing antibodies against EpCAM-AF488 (#CST5198, Cell Signaling Technology), HER2-AF488 (#324410, BioLegend), EGFR-FITC (#GTX11400, GeneTex) and CD45-BV605 (#304042, BioLegend). For mouse experiments, 250–1000 μ l of blood was collected through cardiac puncture and processed immediately on the Parsortix microfluidic device. For tumor-draining vessel experiments, the tumor was first exposed by opening the mouse flank. The largest tumor-associated vessel was then localized and approximately 2 μ l of blood was collected upon a small incision. CTCs from the MMTV-PyMT mouse model were stained with antibodies against mouse EpCAM-AF488 (#118210, BioLegend) and CD45-BV605 (#103140, BioLegend). For all other models (xenografts and syngeneic), carrying cancer cells stably expressing a GFP-Luciferase reporter, only anti-CD45 staining was performed, while CTCs were identified based on GFP expression. The number of captured CTCs, including single CTCs, CTC clusters and CTC-WBC clusters, was determined while cells were still in the cassette. CTCs were then released from the cassette in DPBS (#14190169, Gibco) onto ultra-low attachment plates (#3471-COR, Corning). Representative pictures were taken at 40x magnification with Leica DMI4000 fluorescent microscope using Leica LAS and analyzed with ImageJ.

Assessment of the direct metastatic potential of CTCs. 8-10 weeks old NSG females were injected with 0.5×10^6 4T1-*GFP* or 1×10^6 BR16-*GFP* cells. 8-10 weeks old Balb/c females were injected with 0.5×10^6 4T1-*GFP* cells. Upon tumor development, blood was collected via heart puncture and run through the Parsortix device. Single CTCs, CTC clusters and CTC-neutrophil clusters were individually micromanipulated and 100 cells per mouse (for NSG-4T1-*GFP* model) or 500 cells per mouse (for NSG-BR16-*GFP* and Balb/c-4T1-*GFP* models) from each category were injected into the tail vein of recipient mice (NSG or Balb/c, respectively). Metastasis onset and growth rate in lungs was noninvasively monitored on a weekly schedule with the IVIS system, or through lung immunohistochemical staining of pan-cytokeratin (#GTX27753, Genetex) at the time of experiment termination. The experiment was terminated at the time of overt metastatic disease for 4T1-*GFP* CTCs

in NSG mice, 7 weeks after injection of BR16-*GFP* CTCs in NSG mice, and 7 weeks after injection of 4T1-*GFP* CTCs in Balb/c mice.

White blood cell sorting. Reference WBCs were obtained from the peripheral blood of breast cancer patients ($n=5$) and healthy individuals ($n=3$) after signing informed consent, naïve NSG and Balb/c mice (females at 8-12 weeks), Balb/c-4T1-*GFP* and NSG-CDX-BR16-*GFP* mouse models at the time of experiment termination. In brief, red blood cells, granulocytes and mononuclear cells were separated by gradient centrifugation with Lymphoprep (#1114545, STEMCELL Technologies). Desired fractions were manually isolated and washed with 1% BSA/PBS buffer. Additionally, the granulocyte fraction was purified from contaminating red blood cells by 10 minutes incubation in 0.16M ammonium chloride. Unspecific antibody binding was prevented by blocking the Fc receptor for 15 minutes (human: #422301, BioLegend; mouse: #101320, BioLegend). Cells were stained with white blood cell markers: human – anti-CD14-APC (#301808, BioLegend), anti-CD66b-FITC (#305104, BioLegend), anti-CD3-BV421 (#317344, BioLegend), anti-CD19-FITC (#302206, BioLegend), anti-CD335-PE (#331908, BioLegend) anti-CD41-PE/Cy5 (#303708, BioLegend); mouse – anti-Gr-1-APC/Cy7 (#108423, BioLegend), anti-CD11b-APC (#101211, BioLegend), anti-CD3-BV421 (#100227, BioLegend), anti-CD19-FITC (#115505, BioLegend) or anti-CD19-BV605 (#115539, BioLegend; for mouse models with GFP reporter), anti-CD49b-PE (#108907, BioLegend), anti-CD41-PE/Cy5 (#133921, BioLegend). Cell populations were determined by the expression of characteristic markers: for human granulocytes (CD66b⁺CD41⁻), monocytes (CD14⁺CD3⁻CD19⁻CD335⁻CD41⁻), T cells (CD14⁻CD3⁺CD19⁻CD335⁻CD41⁻), B cells (CD14⁻CD3⁻CD19⁺CD335⁻CD41⁻), NK cells (CD14⁻CD3⁻CD19⁻CD335⁺CD41⁻); for mouse granulocytes (Gr-1⁺CD41⁻), monocytes (CD11b⁺CD3⁻CD19⁻CD49b⁻CD41⁻), T cells (CD11b⁻CD3⁺CD19⁻CD49b⁻CD41⁻), B cells (CD11b⁻CD3⁻CD19⁺CD49b⁻CD41⁻), NK cells (CD11b⁻CD3⁻CD19⁻CD49b⁺CD41⁻). One hundred cells from each population were sorted (FACS Aria III, BD Biosciences) directly into microcentrifuge tubes containing 2.5µl RLT Plus lysis buffer (#1053393, Qiagen).

Neutrophil co-culture with tumor cells. Human neutrophils were purified from healthy donor blood upon gradient centrifugation with LymphoprepTM (Stemcell Technologies). 8'000 neutrophils were added to 100'000 LM2, BR16 or Brx50 cells and co-cultured for 72 hours. Then, gDNA was isolated from tumor cells (or untreated control cells) and processed for whole exome sequencing.

Exome and Transcriptome Sequencing. Individual cells from CTCs alone or CTC-WBC clusters were mechanically separated with gentle micromanipulation (CellCelector, ALS). AF488/FITC-positive (or GFP-positive) and BV605-negative CTCs or AF488/FITC-negative and BV605-positive WBCs were immediately transferred into individual tubes (#321-032-501, Axygen) containing 2.5µl RLT Plus lysis buffer (#1053393, Qiagen) and 1U/µl SUPERase In RNase Inhibitor (#AM2694, Invitrogen). Samples were immediately frozen on dry ice and kept at -80°C until

further processing. Following previously published protocol for parallel DNA and RNA sequencing from individual cells²⁴, genomes and transcriptomes of lysed cells were separated and amplified (#25-6601-97, GE Healthcare for genome and Smart-seq2 from for transcriptome). Reference white blood cells were prepared solely with Smart-seq2 protocol. Libraries were prepared with Nextera XT (Illumina), exomes were enriched using SureSelect XT Human All Exon v6 + Cosmic kit (Agilent technologies) and sequenced on HiSeq 2500 (Illumina) in 100bp paired-end mode for DNA sequencing and on NextSeq 500 (Illumina) 75bp single read mode for RNA sequencing.

Differential white blood cell staining on CTC-WBC clusters. Live CTCs captured within the Parsortix microfluidic cassette were stained with anti-Biotin-CD45 (#103104, BioLegend) and detected with Streptavidin-BV421 (#405226, BioLegend), anti-mouse Ly-6G-AF594 (#127636, BioLegend) and anti-CD11b-AF647 (clone M1/70, kind gift from Dr. Roxane Tussiwand, University of Basel) or with anti-F4/80-AF594 (#123140, BioLegend) and CD11b-AF647. Additionally, MMTV-PyMT-derived CTCs were marked with EpCAM-AF488 (#118210, BioLegend). Next, cells were gently released from the microfluidic system into ultra-low attachment plate and immediately imaged (Leica DMI400). The number of CTC-WBC clusters with neutrophils (Ly-6G⁺CD11b^{med}), monocytes (Ly-6G⁺CD11b^{med/high}) and macrophages (F4/80⁺CD11b⁺) was assessed. Immediately after imaging, cells were centrifuged (500rpm, 3 minutes) on a glass slide and fixed in methanol for 1 minute. After brief air-drying, slides were stained using Wright-Giemsa stain kit (#9990710, ThermoFisher) to visualize nuclear morphology of captured cells, following the manufacturer's instructions.

Immunofluorescence staining. Formalin-fixed, paraffin-embedded (FFPE) sections were obtained from primary tumors and metastatic sites of patients with ER/PR-positive breast cancer (Department of Pathology, University Hospital Basel) who had detectable CTC-WBC clusters. Similarly, mouse-derived primary tumors and metastases were fixed in 4% paraformaldehyde and prepared according to a standard paraffin embedding protocol. Human and mouse sections were handled according to a standard immunofluorescent paraffin-embedded tissue staining protocol. Briefly, after deparaffinization in xylene and re-hydration, antigen retrieval was carried out in 10mM sodium citrate (pH 6.0) at 95°C for 25 minutes. For CTC and DTC staining, cell suspension was centrifuged (3min, 500 rpm) on a coated glass slide (#5991056, ThermoFisher) and air-dried. Cells were fixed in 4% paraformaldehyde for 12 min and stored in PBS until needed. For both FFPE sections and cells, after 1 hour of blocking with 10% horse serum, specimens were co-stained for pan-cytokeratin (#GTX27753, Genetex) detected with anti-mouse IgG-AF488 (#A-21202, ThermoFisher), myeloperoxidase (#AF3667-SP, R&D) detected with anti-goat IgG-AF568 (A-11057, ThermoFisher), Ki67 (#ab15580, Abcam) detected with anti-rabbit IgG-AF647 (A-31573, ThermoFisher) and DAPI (#D1306, ThermoFisher).

***In vitro* cytokine treatment.** 100'000 4T1-*GFP* cells per well were seeded in a 6-well plate and cultured in growth medium overnight. Next morning, cells were washed 3 times with PBS and given starvation medium (0.1% FBS). After 48h, the medium was supplemented with 25ng/ml recombinant mouse IL6 (#575702, BioLegend), IL1 β (#575102, BioLegend), TNF α (#575202, BioLegend) and OSM (#762802, BioLegend), either individually or in combination. Cells were stimulated for 24h and then harvested upon trypsinization, enumerated using automatic cell counter and 300'000 cells were injected intravenously into 8-10 weeks old female mice.

***Mertk* and *Tle1* mutagenesis.** Lentiviral vectors with human *MERTK* (CCSB-Broad LentiORF, CloneId: ccsbBroad304_11503, Dharmacon) and human *TLE1* (Precision LentiORF, CloneId: PLOHS_100005903, Dharmacon) served as base for introduction of specific mutations using QuikChange II XL site-directed mutagenesis kit (#200522, Agilent Technologies). Lentiviral particles were then prepared with Dharmacon Transduction Starter Kit and upon transduction, 4T1-*GFP* cells were selected with 9 μ g/ml Blasticidin S for 6 days.

Myeloid cells depletion. For neutrophil depletion studies in primary tumor models, mice were injected intraperitoneally with 400 μ g of anti-Ly-6G IgG2a (#127650, BioLegend) or control IgG2a (#400566, BioLegend) when tumors were palpable (day two after injection of 4T1-*GFP* cells, day six after injection of LM2-*GFP* cells and day 30 after injection of BR16-*GFP* cells). Efficiency of immune cell depletion was monitored after 48 hours with Advia120 Hematology Analyzer (Siemens) using Multispecies version 5.9.0-MS software (Bayer). Additionally, NSG-4T1-*GFP* mice received a second dosage of anti-Ly-6G or control IgG2a antibodies (100 μ g) on day 19, NSG-LM2-*GFP* mice on day 25 and NSG-BR16-*GFP* mice on day 45. Tumor size was determined with caliper measurements every seven days and tumor volume was calculated using modified ellipsoid formula: $V=1/2(\text{Length} \times \text{Width}^2)$. At termination, lung metastases were measured with IVIS Lumina II (Perkin Elmer) and metastatic index was determined by normalizing the photon/second count of the metastasis with that of the primary tumor. For neutrophil pre-depletion experiments, a single dose of 400 μ g of anti-Ly-6G IgG2a was injected intraperitoneally 24h before the intravenous inoculation of cancer cells. Mice were sacrificed in accordance to our approved protocol and the survival data was inferred accordingly.

G-CSF overexpression. Human *G-CSF* was transduced into 4T1-*GFP*, LM2-*GFP* and BR16-*GFP* cells using the Precision LentiORF (GE Healthcare) system. Construct-positive cells were selected with 9 μ g/ml Blasticidin S for 4 days (4T1) or 7 days (LM2, BR16). Overexpression of G-CSF was confirmed by qPCR using human-specific primers for LM2 and BR16 cells (Forward: 5'GAGTTGGGTCCCACCTTG3', Reverse: 5'TGGAAAGCAGAGGCGAAG3') or primers recognizing both mouse and human transcripts for 4T1 cells (Forward:

5'TGTGCCACCTACAAGCTGTG3', Reverse:
5'CCATCTGCTGCCAGATGGTGGT3').

Generation of 4T1-Cas9-GFP cells. 4T1 cells were transduced with the lentiviruses carrying pLentiCas9-EGFP (#63592, Addgene) at a MOI of 1. GFP-positive cells were sorted as single cells into 96-well plates and cultured as clonal cell lines. Lines with 100% GFP-positivity were kept and Cas9 expression was confirmed by western blotting (#844301, Biolegend).

sgRNA minipool design, transduction and *in vivo* transplantation. All sgRNAs were designed using the GPP Web Portal (<https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design>) and sgRNA oligos were synthesized by Microsynth. Each sgRNA was individually cloned into the pLentiGuide-Puro vector (#52963, Addgene). 4T1-Cas9-GFP cells were then transduced separately with each individual sgRNA vector at MOI=0.4. Upon seven days of puromycin selection, 4T1-Cas9-GFP cells carrying individual sgRNAs were mixed in equal cell numbers, taken for genomic DNA extraction and, at the same time, subcutaneously injected (≥ 1000 cells per sgRNA; 500'000 total cells) into the mammary fat pad of NSG mice.

sgRNA sequencing. gDNA was extracted from cells at different stages (prior to injection, upon primary tumor growth and from spontaneously formed CTCs) using salt precipitation. The library preparation was carried out using two-step PCR, where the first PCR amplifies a broad region including the sgRNA sequence cassette and the second PCR adds Illumina sequencing adapters to the products from the first PCR, as described previously²⁵. Samples were then sequenced on NextSeq 500 SR75 sequencers. After quality control using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc>), reads were trimmed using cutadapt (v1.9.1) and aligned to the sgRNA sequences using bowtie2 (v2.2.9), allowing for one mismatch. The normalized counts of each sgRNA were computed by dividing the number of reads for each sgRNA by the library size.

Single-cell RNA-seq data processing. After sequencing, initial quality assessment for RNA-seq data was performed using FastQC, FastQ Screen (https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen), and visualized with MultiQC (v0.8). Adaptor sequences, first 9 base pairs and low-quality ends were removed with Trim Galore (v0.4.2, http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/; parameters : --phred33 --length 36 --clip_R1 9). Trimmed reads were aligned to a combined human (GRCh38) and mouse (GRCm38) genome reference using STAR (v 2.5.2a; parameters: --runMode alignReads --genomeLoad LoadAndExit). Quality control of resulting BAM files was performed with RSeQC (v2.6.4). The gene-level expression counts were computed with featureCounts (v1.5.1) using the gene annotations obtained from RefSeq (release 70). Samples with less than 800 features detected

(threshold ≥ 1 mapped read) or showing more than 5% of contamination from the other species were removed from further analysis. To normalize gene counts for cell-specific biases, we used size factors computed utilizing the deconvolution implemented in the scran package (v1.6.5) available on R/Bioconductor. After normalization, the effect of technical factors (library size and number of detected features) on variance was evaluated using t-distributed stochastic neighbor embedding (t-SNE) adjusted by patient or mouse model. CTCs showing a substantial contribution of stromal genes and the absence of cancer-associated genes, and CTC-associated WBCs showing no expression of CD45 were removed from the analysis. scRNA-seq data processing, quality control, and visualization was performed with the help of the R/Bioconductor package scater (v 1.6.0).

Reference Component Analysis (RCA). RCA was utilized to identify single cell types using reference transcriptomes. For human samples, the reference bulk transcriptomes were obtained from the Human U133A/GNF1H Gene Atlas and the Primary Cell Atlas (<http://biogps.org/>), averaging expression levels when multiple replicates were present. Mouse transcriptomes were obtained from the Mouse GNF1M and MOE430 Gene Atlas (<http://biogps.org/>). The initial gene selection for the reference transcriptome panel was performed as previously described¹⁰. An additional filtering of genes was achieved by removing genes specific to CTCs from the human panel and by selecting highly variable genes (HVGs) from the mouse panel. A gene was defined as CTC-specific if its normalized expression (log-counts) relative to the median across the reference WBCs set exceeded 5 in at least 5% of CTC samples. In mouse, only genes that showed high variability in their expression across reference WBCs were included. In order to select HVGs in mouse, gene-specific variance of expression across reference WBCs was estimated using trendVar and decomposed into biological and technical components using decomposeVar from scran package. Highly variable genes were selected on the basis of their biological component (biological variance ≥ 5) and adjusted *P*-value (threshold ≤ 0.05). A total of 5,279 genes were selected for the human reference panel and 655 for the mouse panel. Projection of each sample onto the reference transcriptome was performed as previously described¹⁰, calculating the Pearson correlation between the \log_{10} (FPKM) values of the scRNA-seq samples and the \log_{10} expression values of the global panel using the functions provided by the RCA R package (v1.0; <https://github.com/GIS-SP-Group/RCA>). For visualization, reference cell types with a low correlation with query samples and non-immune related features were removed. Hierarchical clustering was performed to cluster samples based on their projection values.

Differential expression and gene set enrichment analyses. We determined differentially expressed genes by the edgeR likelihood ratio test method (v3.20.1) using the normalized counts with the deconvolution approach and the robust dispersion of estimates options. Gene set over-representation analysis of KEGG pathways in the differentially expressed genes (adjusted p-value threshold ≤ 0.25) was performed with the kegg method implemented in the edgeR R/Bioconductor package

(v3.20.1). Enrichment of the KEGG pathways ‘Cell cycle’ (hsa04110) and ‘DNA replication’ (hsa03030) in patient samples was tested with the self-contained rotation gene set test (ROAST) from the limma R/Bioconductor package (3.34.2) using the msq option as a gene set summary statistic and 5’000 rotations to compute p-values.

Cytokine and cytokine ligand analysis. A comprehensive collection of cytokines and their receptors was obtained from KEGG pathway ‘Cytokine-cytokine receptor interaction’ (accession codes hsa04060 and mmu04060 for human and mouse, respectively). Next, human one-to-one orthologous genes for the mouse gene set were obtained from Ensembl (v91) using the biomaRt (v2.34) R/Bioconductor package in order to combine human and mouse datasets. A cytokine-receptor pair was considered to be expressed in a CTC-neutrophil cluster if the cytokine gene in the neutrophil sample and its corresponding receptor in the CTC were expressed at log₂ normalized counts per million mapped reads (CPM) ≥ 5 . For CTC-neutrophil clusters containing more than one detached CTC, all possible CTC-neutrophil pairs were considered.

Single-cell DNA-seq data processing. Paired-end reads were aligned to the GRCh38 human or GRCm38 mouse reference genomes using BWA-mem algorithm (v0.7.13; parameters : -M) (<https://arxiv.org/abs/1303.3997>) and sorted using SAMtools (v1.3.1). Reads were then deduplicated using Picard MarkDuplicates (v2.9.0; <http://picard.sourceforge.net/>) on a per-sample basis and local realignment was performed using the Genome Analysis Toolkit (GATK) IndelRealigner (v3.7.0) at the sample and donor level to improve alignment accuracy around indels. Quality control and coverage and exome enrichment statistics were generated using FastQC, CollectHsMetrics from Picard suite, and QualiMap (v 2.2.1) and visualized using MultiQC (v0.8).

Somatic mutation calling and mutation spectrum. Mpileup files were generated with SAMtools (v1.3.1; parameters : -B -q 40) and variants were called using Monovar (v2016-05-14) on all samples from the same donor simultaneously. Resulting variants were annotated using SnpEff on ENSEMBL v86 (www.ensembl.org), dbSNP (build 150), 1000 genomes project (phase 1), and coding mutations from cosmic (v81) using SnpSift (v4.3p). Somatic mutation rates were calculated as the ratio of the number of somatic variants and the number of nucleotides covered in the exome at $\geq 2x$. Putative damaging somatic mutations were identified exclusively in donors with matched WBC sequenced using an empirical filtering strategy removing (1) variants present in public databases (dbSNP, 1000 genomes project) at a frequency $\geq 1\%$ or found in 2 or more founders, (2) variants present in at least one reference WBC sample from the same donor, (3) variant loci not covered in reference WBC samples (threshold ≥ 3 reads), and (4) likely damaging events (truncating, frameshift or splice site variant). VCF processing, downstream filtering, and analysis was performed using the VariantAnnotation and vcfR R/Bioconductor packages. Trinucleotide context of the somatic mutation spectrum was generated and visualized with the SomaticSignatures package (v2.14.0).

Survival analyses. Survival analyses were performed using the survival R package (v 2.41-3). Kaplan-Meier curves were generated and Log-Rank test was used to estimate the significance of the difference in survival between groups. For patients, progression-free survival was defined as the period of time between primary tumor diagnosis and first relapse. For NSG-4T1-*GFP* mouse model analysis, death was selected as the endpoint for the analysis and defined as the moment a given animal had to be euthanized according to our mouse protocol guidelines.

Data availability. Data analysis, statistical testing and visualization were conducted in R (version 3.4.0; R Foundation for Statistical Computing, Vienna, Austria). RNA and exome sequencing data have been deposited to Gene Expression Omnibus (GEO, NCBI; accession number GSE109761) and to the European Nucleotide Archive (ENA, EMBL-EBI; accession number PRJEB24623), respectively. Original R scripts to reproduce data analysis have been deposited to GitHub (CA, USA; accession URL <https://github.com/CMETlab/CTC-WBC>). Source data for all mouse experiments are provided. All data are available from the corresponding author upon reasonable request.

Acknowledgements

We thank all patients that donated blood for our study, as well as all involved clinicians and study nurses. We thank Dr. Joan Massagué (Memorial Sloan Kettering Cancer Center, New York, USA), Dr. Daniel Haber and Dr. Shyamala Maheswaran (Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA) for donating cell lines. We thank Dr. Gerhard Christofori for MMTV-PyMT mice and comments on the manuscript, and all members of the Aceto lab for feedback and discussions. We thank Katja Eschbach and Elodie Burcklen from the Genomics Facility Basel (D-BSSE of the ETH Zürich) for generating sequencing libraries and performing next-generation sequencing. We thank Stefan Arnold (D-BSSE of the ETH Zürich) and Simone Müntz Soysal (University Hospital Basel) for support with sample acquisition and processing. We thank Till Ryser (Aceto lab, University of Basel) for help with CRISPR/Cas9-related experiments. Calculations were performed at sciCORE (<http://scicore.unibas.ch/>) scientific computing center of the University of Basel. Research in the Aceto lab is supported by the European Research Council, the Swiss National Science Foundation, the Swiss Cancer League, the Basel Cancer League, the two Cantons of Basel through the ETH Zürich, and the University of Basel.

Author information

Affiliations

Department of Biomedicine, Cancer Metastasis Lab, University of Basel and University Hospital Basel, 4058 Basel, Switzerland

Barbara Szczerba, Francesc Castro-Giner, Sofia Gkoutela, Ilona Krol, Cinzia Donato, Ramona Scherrer, Manuel C. Scheidmann & Nicola Aceto

SIB Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

Francesc Castro-Giner, Jochen Singer & Niko Beerenwinkel

Gynecologic Cancer Center, University Hospital Basel, 4056 Basel, Switzerland

Marcus Vetter, Christian Kurzeder & Viola Heinzelmann-Schwarz

Department of Medical Oncology, University Hospital Basel, 4056 Basel, Switzerland

Marcus Vetter, Julia Landin & Christoph Rochlitz

Breast Center, University of Basel and University Hospital Basel, 4056 Basel, Switzerland

Christian Kurzeder, & Walter Paul Weber

Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland

Jochen Singer, Christian Beisel & Niko Beerenwinkel

Contributions

B.S. and N.A. designed the study, performed the experiments and wrote the manuscript. F.C.-G. performed the computational analysis. S.G., I.K., C.D. and R.S. processed blood samples, mouse tissues and performed immunofluorescence staining. M.C.S. performed CRISPR/Cas9-related experiments. M.V., J.L., C.K., V.H.-S., C.R., and W.P.W. provided patient samples and clinical input throughout the project. C.B. generated sequencing data. J.S. and N.B. provided input during computational methods development and data analysis. All authors have read, commented and approved the manuscript in its final form.

Competing financial interests

N.A. and B.M.S. declare competing financial interests, as they are listed as inventors in patent applications that are related to CTC clusters and CTC-neutrophil clusters.

Figure Legends

Main Figures

Figure 1: CTC-neutrophil clusters are highly-efficient metastatic precursors. a, Representative images of a CTC-WBC cluster, a CTC cluster and a single CTC from NSG-CDX-BR16-*GFP* mice. CTCs are expressing GFP (*green*), while CTC-associated WBCs are labeled with anti-CD45 antibodies (*red*). $n=8$. **b,** Pie charts displaying the mean percentage of single CTCs (*grey*), CTC clusters (*green*) and CTC-WBC clusters (*red*) in breast cancer patients and mouse models. The number of independent biological replicates (n) is shown for each model. **c,** Schematic of the experimental design. CTC-WBC clusters are dissociated into individual cells and processed for RNA-sequencing (RNA-seq). The transcriptome of CTC-associated WBCs is then compared to reference (REF) WBCs. **d,** Reference component analysis clustering of CTC-associated WBCs and reference WBCs from breast cancer patients, displaying projection scores of cells (*columns*; $n=50$) on the immune reference panel (*rows*). **e,** Representative images of CTC-WBC clusters stained for Ly-6G (neutrophils; *gold*) together with GFP (cancer cells; *green*) (*top*) or processed with the Wright-Giemsa (WG) assay to define nuclear morphology (*bottom*). $n=8$. **f,** The pie charts show the mean percentage of CTC-neutrophil clusters and CTC-monocyte clusters in individual models. The number of independent biological replicates (n) is shown for each model. **g,** Kaplan-Meier plot showing progression-free survival of breast cancer patients. $n=9$ for patients with ≥ 1 CTC-neutrophil clusters and $n=10$ for patients with ≥ 5 CTCs; P value by two-sided Log-Rank test is shown. **h,** Schematic of the experimental design. 100 CTCs from CTC-neutrophil clusters or CTCs alone are injected in the tail vein of recipient mice to measure their metastatic potential (*left*). Kaplan-Meier plot showing overall survival of mice (*right*). $n=5$ for CTC-neutrophil clusters and $n=10$ for CTCs alone; P value by two-sided Log-Rank test is shown.

Figure 2: CTCs from CTC-neutrophil clusters are highly proliferative. a, Schematic of the experimental design. CTC-neutrophil clusters are dissociated and then processed for RNA-sequencing (RNA-seq). The transcriptome of CTCs from CTC-neutrophil clusters is compared to that of CTCs alone. **b,** Heatmap of genes differentially expressed between CTCs from CTC-neutrophil clusters ($n=25$) and CTCs alone ($n=4$), isolated from Balb/c-4T1-*GFP* mice. The heatmap displays gene-scaled (Z -score) \log_2 counts per million mapped reads (CPM) values after normalization, with columns representing samples ($n=29$) and rows representing genes. $q < 0.05$ by edgeR likelihood ratio test. **c,** KEGG pathways over-represented ($P < 0.05$ by one-sided hypergeometric test) among upregulated genes in CTCs of

CTC-neutrophil clusters from Balb/c-4T1-*GFP* mice (*left*) and test of selected pathways in patients ($P < 0.05$ by ROAST test; *right*). **d**, Representative pictures of CTC-neutrophil clusters and CTCs alone from NSG-CDX-BR16-*GFP* mice, stained for pan-cytokeratin (pCK; *green*), myeloperoxidase (MPO; *gold*), Ki67 (*purple*) and DAPI (nuclei; *blue*). $n=3$. **e**, The plots show the mean percent of Ki67-positive CTCs. $n=3$ for all; error bars represent S.E.M.; P values by two-sided Student's t test are shown. **f**, Schematic of the experimental design. 4T1-*GFP* cells are stimulated for 24h with IL6, IL1 β or both (pool), then injected in recipient mice to assess their metastatic potential. **g**, Kaplan-Meier survival analysis of NSG (*left*) or Balb/c (*right*) mice injected with cytokine-treated 4T1-*GFP* cells. $n=4$ for NSG-4T1-*GFP* and $n=3$ for Balb/c-4T1-*GFP*; P values by two-sided Log-Rank test are shown.

Figure 3: Whole-exome sequencing highlights recurrent mutational events in CTCs from CTC-neutrophil clusters. **a**, Schematic of the experimental design.

CTC-neutrophil clusters are dissociated and then processed for whole-exome sequencing (WES). CTCs that were associated to neutrophils are compared to CTCs alone. **b**, Mutation distribution in CTCs from CTC-neutrophil clusters ($n=14$) compared to CTCs alone ($n=56$). Lines within the violin plots show the 25th, 50th and 75th percentile, respectively, while dots represent individual CTCs. P value by two-sided Wilcoxon sign-ranked test is shown. **c**, Pie charts displaying the mean percentage of single CTCs (*grey*), CTC clusters (*green*) and CTC-neutrophil clusters (*gold*) in mice injected with 4T1-pLOC, 4T1-*TLE1* WT, 4T1-*TLE1* G1787A or 4T1-*TLE1* G1509C cells (*top*). The number of independent biological replicates (n) is shown for each condition. The plots show the mean fold change of CTC ratios (*bottom*). $n=3$; error bars represent S.E.M.; P values by two-sided Student's t test are shown.

Figure 4: Identification of vulnerabilities of CTC-neutrophil clusters. **a**, Pie charts displaying the mean percentage of single CTCs (*grey*), CTC clusters (*green*) and CTC-neutrophil clusters (*gold*) in NSG-4T1-*GFP* mice treated with α Ly-6G antibodies or G-CSF overexpression (*top*). W= weeks upon tumor development. The number of independent biological replicates (n) is shown for each condition. The plots show the mean fold change of CTC ratios from NSG-4T1-*GFP* mice treated with α Ly-6G antibodies or G-CSF overexpression (*bottom*). Error bars represent S.E.M.; P values by two-sided Student's t test are shown. **b**, Schematic of the experimental design. 4T1-Cas9-*GFP* cells are transduced with a pool of sgRNAs targeting cell-cell junctions and injected in NSG mice. Upon tumor development, spontaneously-generated CTC-neutrophil clusters, CTC clusters and single CTCs are sequenced to identify sgRNAs dropouts. **c**, Bar plot showing the log₂ fold change (FC) of individual sgRNAs (numbered 1 to 4) found in CTCs from CTC-neutrophil clusters versus CTCs alone. sgRNAs whose representation was reduced in CTCs from CTC-neutrophil clusters are shown in *dark red*. $n=3$. **d**, Pie charts displaying the mean

percentage of single CTCs (*grey*), CTC clusters (*green*) and CTC-neutrophil clusters (*gold*) in NSG mice carrying 4T1-Cas9-*GFP* tumors expressing a control sgRNA (Ctrl sgRNA) or individual sgRNAs targeting *Vcam1* (*Vcam1* KO) (*top*). The number of independent biological replicates (*n*) is shown for each condition. The plot shows the mean fold change of CTC ratios from mice carrying 4T1-Cas9-*GFP* tumors expressing a control sgRNA (Ctrl sgRNA) or individual sgRNAs targeting *Vcam1* (*Vcam1* KO) (*bottom*). Error bars represent S.E.M.; *P* value by two-sided Student's *t* test is shown.

Extended Data Figures

Extended Data Fig. 1. CTC capture in breast cancer patients and mouse models.

a, Schematic of the CTC capture strategy with the Parsortix device. **b**, Schematic of the experimental design. 50 single CTCs and 50 CTC clusters are spiked into blood to assess capture rate. **c**, Representative pictures of CTCs and WBCs captured on the Parsortix device and stained for EpCAM, HER2, EGFR (*green*) and CD45 (*red*) (*left*); *n*=3. The plot shows the mean CTC capture efficiency (*right*); *n*=3; error bars represent S.E.M. **d**, Schematic of the experimental design. 50 single CTCs are spiked into blood to evaluate artificial CTC aggregation rate during processing. **e**, Representative picture of captured CTCs (*left*); *n*=3. The plot shows the mean percent of captured single CTCs, CTC clusters and CTC-WBC clusters (*right*); *n*=3; error bars represent S.E.M. **f**, Representative pictures of CTCs from patients and mouse models. *n*=34 for patients; *n*=8 for NSG-LM2-*GFP* and NSG-4T1-*GFP*; *n*=7 for Balb/c-4T1-*GFP*; *n*=5 for MMTV-PyMT. **g**, The plot shows mean CTC counts in patients. *n*=34; error bars represent S.E.M. **h**, The plot shows the number of CTCs in each patient-derived CTC-WBC cluster. The *red* line represents the mean. **i**, The plots show mean CTC counts in mouse models. *n*=8 for NSG-CDX-BR16-*GFP*, NSG-LM2-*GFP* and NSG-4T1-*GFP*; *n*=7 for Balb/c-4T1-*GFP*; *n*=5 for MMTV-PyMT. Error bars represent S.E.M. **j**, The plot shows the number of CTCs in each mouse model-derived CTC-WBC cluster. The *red* line represents the mean. **k**, Pie charts displaying the mean percentage of single CTCs (*grey*), CTC clusters (*green*) and CTC-WBC clusters (*red*) in mice upon blood draw *via* heart puncture (HP) or tumor-draining vessel (TDV) (*left*). The plots show the mean number of CTCs from the same experiment (*right*). Error bars represent S.E.M.; *n*=5 for NSG-CDX-BR16-*GFP*, *n*=3 for NSG-LM2-*GFP*; *P* values by two-sided Student's *t* test are shown. **l**, The plots show fold change of CTC counts, comparing HP *versus* TDV blood draw. Error bars represent S.E.M.; *n*=5 for NSG-CDX-BR16-*GFP*, *n*=3 for NSG-LM2-*GFP*; *P* values by two-sided Student's *t* test are shown. **m**, The plots show the number of CTCs in each mouse model-derived CTC-WBC cluster, isolated *via* HP or TDV. The *red* lines represent the mean. *P* values by two-sided Student's *t* test are shown. **n**, The plot shows the mean number of CTCs at day 10 after tumor inoculation, collected

from HP, TDV or peripheral circulation (i.e. tail vein; PC). Error bars represent S.E.M.; $n=3$; P values by two-sided Student's t test are shown.

Extended Data Fig. 2. Characterization of CTC-associated WBCs. **a**, Bar plot showing the expression levels of white blood cell (WBC) marker CD45 in patient samples, including CTC-associated WBCs (*red*), free-floating peripheral WBCs (*blue*) and CTCs alone (*green*). **b**, Principal component analysis (PCA) of CTC-associated WBCs of patients and five reference WBC populations. $n=50$. **c**, Bar plot showing the expression levels of CD45 in mouse samples, including CTC-associated WBCs (*red*) and CTCs alone (*green*). **d**, PCA of CTC-associated WBCs from all mouse models and five reference WBC populations. $n=47$. **e**, Reference component analysis clustering of CTC-associated WBCs (*red*) and reference WBCs from mouse models, displaying projection scores of cells (*columns*; $n=47$) on the immune reference panel (*rows*). **f**, Representative immunofluorescence images of mouse immune cells stained for CD45, Ly-6G and CD11b. Mouse lymphocytes ($CD45^+/Ly-6G^-/CD11b^-$), mouse monocytes ($CD45^+/Ly-6G^-/CD11b^+$), and mouse granulocytes ($CD45^+/Ly-6G^+/CD11b^{low}$) are shown (*top*). Representative immunofluorescence images of mouse lymphocytes and macrophages (peritoneum-derived) stained for F4/80 are shown (*bottom*). Mouse macrophages display a $F4/80^+$ phenotype, while lymphocytes display a $F4/80^-$ phenotype. $n=3$ for all. **g**, Representative images of human and mouse T cells, B cells, NK cells, monocytes and granulocytes stained with the Wright-Giemsa protocol to highlight nuclear morphology (*left*). Wright-Giemsa staining of the human CTC-derived cell line BR16 is also shown (*right*). $n=3$ for all. **h**, Representative immunofluorescence images of CTC-neutrophil clusters and CTC-monocyte clusters isolated from mouse models and stained for Ly-6G (*gold*) and CD11b (*blue*). CTCs stably express *GFP* (*green*). $n=8$ for NSG-CDX-BR16-*GFP*, NSG-LM2-*GFP* and NSG-4T1-*GFP*; $n=7$ for Balb/c-4T1-*GFP*. **i**, Representative images of CTC-neutrophil clusters stained with the Wright-Giemsa protocol to highlight nuclear morphology. $n=8$ for NSG-CDX-BR16-*GFP*, NSG-LM2-*GFP* and NSG-4T1-*GFP*. **j**, Bar graph showing the mean number of CTC-neutrophil clusters and CTC-monocyte clusters in mouse models. Error bars represent S.E.M.; $n=8$ for NSG-CDX-BR16-*GFP*, NSG-LM2-*GFP*, NSG-4T1-*GFP* and $n=7$ for Balb/c-4T1-*GFP*. **k**, Heatmaps showing the projection scores of mouse-derived (*left*) and patient-derived (*right*) CTC-associated neutrophils (*columns*) on pro-tumoral (N2) neutrophil markers (*rows*).

Extended Data Fig. 3. Progression-free survival analysis in breast cancer patients and mouse models. **a**, Kaplan-Meier progression-free survival (PFS) analysis comparing patients with ≥ 1 CTC-neutrophil cluster per 7.5ml of peripheral blood ($n=9$) versus all patients with no CTC-neutrophil clusters ($n=48$); P value by two-sided Log-Rank test is shown. **b**, Kaplan-Meier PFS analysis comparing patients with ≥ 1 CTC-neutrophil cluster per 7.5ml of peripheral blood ($n=9$) versus patients

with ≥ 1 CTC per 7.5ml of peripheral blood but no CTC-neutrophil clusters ($n=21$); P value by two-sided Log-Rank test is shown. **c**, Kaplan-Meier PFS analysis comparing patients with ≥ 1 CTC-neutrophil cluster per 7.5ml of peripheral blood ($n=9$), patients with ≥ 1 single CTC per 7.5ml of peripheral blood but without CTC-neutrophil clusters ($n=14$), and patients with ≥ 1 CTC cluster per 7.5ml of peripheral blood but without CTC-neutrophil clusters ($n=7$).; P value by two-sided Log-Rank test is shown. Of note, these results are consistent with our previous observations whereby PFS differences in patients with single CTCs *versus* CTC-clusters were visible only when CTC clusters were present for multiple time points along disease progression. **d**, Schematic of the experimental design. 100 CTCs from CTC-neutrophil clusters, CTC clusters or single CTCs are injected in the tail vein of tumor-free recipient mice to measure their metastatic potential. **e**, The plot shows normalized bioluminescence signal from the lungs of injected mice. $n=5$ for all; Error bars represent S.E.M.; $*P < 0.05$ by two-sided Student's t test. **f**, Kaplan-Meier plot showing overall survival of injected mice. $n=5$ for all; P values by two-sided Log-Rank test are shown. **g**, Representative picture of a metastatic lesion in NSG mice injected intravenously with either with CTC-neutrophil clusters, CTC clusters or single CTCs from NSG-BR16-*GFP* mice. Metastases are stained for pan-cytokeratin (pCK; *green*) and DAPI (nuclei; *blue*) (*left*); $n=3$. The plot shows the mean number of metastatic foci per field of view (*right*). $n=3$; Error bars represent S.E.M; P values by two-sided Student's t test are shown. **h**, Representative picture of a metastatic lesion in the lungs of Balb/c mice injected intravenously with either with CTC-neutrophil clusters, CTC clusters or single CTCs from Balb/c-4T1-*GFP* mice. Metastases are stained for pan-cytokeratin (pCK; *green*) and DAPI (nuclei; *blue*) (*left*); $n=3$. The plot shows the mean number of metastatic foci per field of view (*right*). $n=3$; Error bars represent S.E.M; P values two-sided Student's t test are shown.

Extended Data Fig. 4. Gene expression analysis of single-cell RNA sequencing data. **a**, t-distributed stochastic neighbor embedding (t-SNE) analysis of CTCs from CTC-neutrophil clusters and CTCs alone using the 500 most variable genes. t-SNE plots for Balb/c-4T1-*GFP* samples are colored by number of detected genes (*left*) and number of reads per sample (*right*). $n=29$. **b**, t-SNE plots for patient samples colored by number of detected genes (*left*) and number of reads per sample (*right*). $n=68$. **c**, Heatmap showing the projection scores of mouse model-derived CTCs from CTC-neutrophil clusters and CTCs alone in relation to epithelial and mesenchymal genes. $n=59$. **d**, Heatmap showing the projection scores of patient-derived CTCs from CTC-neutrophil clusters and CTCs alone in relation to epithelial and mesenchymal genes. $n=68$. **e**, Heatmap showing the projection scores of mouse model-derived CTCs from CTC-neutrophil clusters and CTCs alone in relation to cancer stem cell genes. $n=59$. **f**, Heatmap showing the projection scores of patient-derived CTCs from CTC-neutrophil clusters and CTCs alone in relation to cancer stem cell genes. $n=68$. **g**, Heatmap showing the projection scores of mouse model-derived CTCs from CTC-neutrophil clusters and CTCs alone in relation to platelet genes. $n=59$. **h**, Heatmap

showing the projection scores of patient-derived CTCs from CTC-neutrophil clusters and CTCs alone in relation to platelet genes. $n=68$.

Extended Data Fig. 5. Proliferation of tumor cells adjacent to neutrophils in primary and metastatic tissues. **a**, Representative immunofluorescence images of NSG-LM2-*GFP* primary tumor and matched lung metastasis, stained for pan cytokeratin (pCK; *green*), MPO (*gold*), Ki67 (*purple*) and DAPI (nuclei; *blue*). $n=3$. **b**, The plots show the mean percent of Ki67-positive cancer cells in the primary tumor and metastatic sites (lung or brain) of mouse models, both overall and when considering only those cells that are adjacent to neutrophils. $n=3$ for all; error bars represent S.E.M; ns= not significant by two-sided Student's *t* test. **c**, Representative immunofluorescence images of BR57 primary tumor and matched liver metastasis, stained for pan cytokeratin (pCK; *green*), MPO (*gold*), Ki67 (*purple*) and DAPI (nuclei; *blue*). $n=3$. **d**, The plots show the mean percent of Ki67-positive cancer cells, both overall and when considering only those cells that are adjacent to neutrophils, in matched primary and metastatic sites of nine breast cancer patients. Error bars represent S.E.M.; ns= not significant by two-sided Student's *t* test. **e**, Schematic of the experimental design. 100 CTCs from CTC-neutrophil clusters or CTC alone are injected in the tail vein of recipient mice to measure disseminated tumor cells (DTC) proliferation. **f**, Representative pictures of DTCs stained for pan-cytokeratin (pCK; *green*), Ki67 (*purple*) and DAPI (nuclei; *blue*) (*left*). The plot shows the mean percent of Ki67-positive DTCs (*right*). $n=3$. Error bars represent S.E.M; $*P=0.001$ by two-sided Student's *t* test.

Extended Data Fig. 6. Characterization of cytokine-mediated crosstalk within CTC-neutrophil clusters. **a**, Schematic of the experimental design (*top*). The heatmap shows the transcriptional landscape of cytokines and corresponding receptors expressed in at least 20% of CTC-neutrophil clusters (*bottom*). The cytokine-receptor pairs that are most frequently expressed in human cells are shown in *red*. **b**, Schematic of the experimental design (*top*). The heatmap shows the transcriptional landscape of cytokine receptors and corresponding cytokines expressed in at least 20% of CTC-neutrophil clusters (*bottom*). The cytokine-receptor pairs that are expressed in at least 40% of CTC-neutrophil clusters are shown in *red*. **c**, The plot shows the mean 4T1-*GFP* cell number upon starvation and stimulation with IL6, IL1 β , TNF α , OSM or all four cytokines together (cytokine pool). $n=3$; error bars represent S.E.M.; $*P < 0.05$ by two-sided Student's *t* test. **d**, Plots showing the mean percentage of Ki67-positive disseminated tumor cells (DTCs) in the bone marrow of injected mice. $n=3$ for all; error bars represent S.E.M.; $*P < 0.05$ by two-sided Student's *t* test. **e**, The plots show normalized bioluminescence signal from the lungs of injected mice. $n=4$ for all; error bars represent S.E.M.; $*P < 0.05$ by two-sided Student's *t* test. **f**, Tumor growth curves of NSG mice injected with 4T1-Cas9-*GFP* cells expressing a control vector (Ctrl sgRNA) or sgRNAs targeting *Il1r1* or *Il6st*.

$n=3$; Error bars represent S.E.M.; ns= not significant by two-sided Student's t test. **g**, Pie charts displaying the mean percentage of single CTCs (*grey*), CTC clusters (*green*) and CTC-neutrophil clusters (*gold*) of injected mice (*left*); $n=3$. The plots show the mean fold change of CTC ratios from injected mice (*right*); $n=3$; error bars represent S.E.M.; ns= not significant by two-sided Student's t test. **h**, The plot shows the mean percent of Ki67-positive CTCs from injected mice. $n=3$; error bars represent S.E.M.; * $P=0.001$ by two-sided Student's t test.

Extended Data Fig. 7. Mutation analysis of single-cell whole exome sequencing data. **a**, Somatic mutation rate (mutations/Mb) of CTCs from CTC-neutrophil clusters ($n=14$) versus CTCs alone ($n=56$), normalized by donor. Lines within the violin plots show the 25th, 50th and 75th percentile, respectively, while dots represent individual CTCs. P value by two-sided Wilcoxon sign-ranked test is shown. **b**, Somatic mutation rate (mutations/Mb) in all CTCs isolated from donors with CTC-neutrophil clusters (Donors (+); $n=6$) and donors without CTC-neutrophil clusters (Donors (-); $n=5$). Lines within the violin plots show the 25th, 50th and 75th percentile, respectively, while dots represent individual CTCs. ns= not significant by two-sided Wilcoxon sign-ranked test. **c**, Nucleotide substitution pattern among putative somatic mutations in CTCs isolated from donors with CTC-neutrophil clusters (Donors (+)) versus donors without CTC-neutrophil clusters (Donors (-)). $n=6$ for Donors (+) and $n=5$ for Donors (-). Lines within the violin plots show the 25th, 50th and 75th percentile, respectively, while dots represent individual CTCs. P value by two-sided Wilcoxon sign-ranked test is shown. **d**, The bar plots show the nucleotide context of given mutations in CTCs alone versus CTCs from CTC-neutrophil clusters. **e**, Plot showing the age distribution of Donors (+) ($n=10$) and Donors (-) ($n=24$). The *red* lines represent the mean. ns= not significant by two-sided Student's t test. **f**, The tile plot represent genes (*columns*) containing predicted high-impact mutations in at least two Donors (+) and in none of the Donors (-). **g**, The plots show the mean fold change for *MERTK* and *TLE1* (WT or mutated) transcripts compared to control (Ctrl) cells. $n=3$; error bars represent S.E.M.; * $P<0.004$ by two-sided Student's t test. **h**, Tumor growth curves representing mean tumor volume measurements of NSG mice injected with 4T1 cells carrying an empty vector (pLOC), WT or mutated *MERTK* (*left*) or *TLE1* (*right*). $n=3$; error bars represent S.E.M. ns= not significant by two-sided Student's t test. **i**, Representative images of the primary tumor of injected mice, stained for pan cytokeratin (pCK, *green*), myeloperoxidase (MPO, *gold*) and DAPI (nuclei, *blue*) (*top*); $n=3$. The plot shows the mean number of infiltrated neutrophils per field of view within the primary tumor (*bottom*). Error bars represent S.E.M. $n=3$; ns= not significant by two-sided Student's t test. **j**, Representative images of the primary tumor of injected mice, stained for pan cytokeratin (pCK, *green*), myeloperoxidase (MPO, *gold*) and DAPI (nuclei, *blue*) (*top*); $n=3$. The plot shows the mean number of infiltrated neutrophils per field of view within the primary tumor (*bottom*). Error bars represent S.E.M. $n=3$; * $P=0.002$ ** $P=0.0007$ by two-sided Student's t test. **k**, Pie charts displaying the mean percentage of single CTCs (*grey*), CTC clusters (*green*)

and CTC-neutrophil clusters (*gold*) in injected mice. The number of independent biological replicates (*n*) is shown for each condition.

Extended Data Fig. 8. Co-culture of cancer cells and neutrophils does not lead to the accumulation of key mutational events. **a**, Schematic of the experimental design. Neutrophils were purified from healthy donor blood and cultured with either CTC-derived cell lines (BR16, Brx50) or LM2 cells for 72h. Tumor cells were harvested, and isolated gDNA was processed for whole exome sequencing (WES). **b**, Tile plot showing the mutation status of all key loci found mutated in patients with CTC-neutrophil clusters. None of the CTC-neutrophil clusters-associated mutations were detected upon co-culture of cancer cells with neutrophils.

Extended Data Fig. 9. Effects of neutrophil depletion or augmentation in mice. **a**, The plots show the mean number of neutrophils in the circulation of mice treated with Ly-6G neutralizing antibodies (α Ly-6G) (*left*), or carrying G-CSF overexpressing tumors (*right*). Error bars represent S.E.M.; the number of independent biological replicates (*n*) is provided for simplicity directly within the source data. NA=not available; * $P < 0.03$, ** $P < 0.0001$ by two-sided Student's *t* test. **b**, Representative images of the primary tumor of NSG-LM2-*GFP* mice stained for pan cytokeratin (pCK, *green*), myeloperoxidase (MPO, *gold*) and DAPI (nuclei, *blue*) (*left*); *n*=3. The plots show the mean number of infiltrated neutrophils per field of view within the tumor (*right*). *W*= weeks upon tumor development. Error bars represent S.E.M. *n*=3, * $P < 0.03$, ** $P < 0.0001$ by two-sided Student's *t* test. **c**, Tumor growth curves representing mean tumor volume measurements in the presence or absence of α Ly-6G antibodies or G-CSF overexpression. Error bars represent S.E.M. The number of independent biological replicates (*n*) is provided for simplicity directly within the source data. ns= not significant by two-sided Student's *t* test. **d**, The plots show the mean counts of single CTCs, CTC clusters and CTC-neutrophil clusters in mice. Error bars represent S.E.M. The number of independent biological replicates (*n*) is provided for simplicity directly within the source data. ns= not significant; ND=not detected; * $P < 0.05$ by two-sided Student's *t* test. **e**, Pie charts displaying the mean percentage of single CTCs (*grey*), CTC clusters (*green*) and CTC-neutrophil clusters (*gold*) in NSG-LM2-*GFP* and NSG-CDX-BR16-*GFP* mice treated with α Ly-6G antibodies or G-CSF overexpression. *W*= weeks upon tumor development; the number of independent biological replicates (*n*) is shown for each condition. **f**, The plots show the mean fold change of CTC ratios from NSG-LM2-*GFP* and NSG-CDX-BR16-*GFP* mice treated with α Ly-6G antibodies or G-CSF overexpression. Error bars represent S.E.M. The number of independent biological replicates (*n*) is provided for simplicity directly within the source data. * $P = 0.045$, ** $P = 0.01$, *** $P = 0.004$ by two-sided Student's *t* test. **g**, Representative bioluminescence images of lungs from mice treated with α Ly-6G antibodies or G-CSF overexpression (*left*); the number of independent biological replicates (*n*) is provided for simplicity directly within the source data. *W* = weeks

upon tumor development. The plots show the mean metastatic index of mice treated with α Ly-6G antibodies or G-CSF overexpression (*right*). The number of independent biological replicates (*n*) is provided for simplicity directly within the source data; error bars represent S.E.M.; * $P < 0.03$ ** $P < 0.01$ by two-sided Student's *t* test. **h**, Kaplan-Meier survival plots showing overall survival rates of mice. The number of independent biological replicates (*n*) is provided for simplicity directly within the source data; * $P < 0.02$ by two-sided Log-rank test. **i**, Schematic of the experiment. NSG, FVB and Balb/c mice were pre-treated with α Ly-6G antibodies or control IgG. 4T1-*GFP* cells or Py2T-*GFP* cells were then injected into the tail vein to assess metastasis development. **j**, The plots show mean normalized bioluminescence signal in the lungs of injected mice. *n*=3; error bars represent S.E.M.; ns= not significant by two-sided Student's *t* test. **k**, Kaplan-Meier survival plot of injected mice. *n*=3; ns= not significant by two-sided Student's *t* test. **l**, The plots show the mean percentage of Ki67-positive disseminated tumor cells (DTCs) collected from the bone marrow of injected mice. *n*=3; error bars represent S.E.M.; ns= not significant by two-sided Student's *t* test. **m**, The bar graph shows the proportion of breast cancer patients that were treated with G-CSF, related to their CTC status. *n*=42 for no CTCs, *n*=23 for CTCs, *n*=9 for CTC-neutrophil clusters. *P* value by two-sided Fisher's exact test is shown.

Extended Data Fig. 10. Expression of cell-adhesion molecules (CAMs)-receptor pairs on CTC-neutrophil clusters. **a**, Schematic of the experimental design (*top*). The heatmap shows the expression landscape of cell-adhesion molecules (CAMs) and corresponding receptors that are expressed in at least 20% of CTC-neutrophil clusters (*bottom*). The CAM-receptor pairs that are expressed in at least 50% of CTC-neutrophil clusters are shown in *red*. **b**, Schematic of the experiment (*top*). The heatmap shows the expression landscape of CAM receptors and corresponding CAMs that are expressed in at least 20% of CTC-neutrophil clusters (*bottom*). The CAM-receptor pairs that are expressed in at least 50% of CTC-neutrophil clusters are shown in *red*. **c**, Tumor growth curves representing mean tumor volume measurements of mice injected with 4T1-Cas9-*GFP* cells expressing a control vector (CTRL sgRNA) or sgRNA pools targeting *F11r*, *Icam1*, *Itgb2* and *Vcam1* (CRISPR pool). *n*=3; error bars represent S.E.M.; ns= not significant by two-sided Student's *t* test. **d**, The plot shows the proportion of reads derived from sgRNAs targeting *F11r*, *Icam1*, *Itgb2* and *Vcam1* (4 sgRNAs each) in the 4T1-Cas9-*GFP* cell line upon library transduction as well as in three primary tumors from NSG-4T1-Cas9-*GFP* mice. All sgRNAs were represented in the tumor until the end of the experiment. **e**, Tumor growth curves representing mean tumor volume measurements of mice injected with 4T1-Cas9-*GFP* cells expressing a control vector (CTRL sgRNA) or individual sgRNAs targeting *Vcam1*. *n*=3; error bars represent S.E.M.; ns= not significant by two-sided Student's *t* test.

References

- 1 Khalil, D. N., Smith, E. L., Brentjens, R. J. & Wolchok, J. D. The future of cancer treatment: immunomodulation, CARs and combination immunotherapy. *Nat Rev Clin Oncol* **13**, 273-290, doi:10.1038/nrclinonc.2016.25 (2016).
- 2 Mohme, M., Riethdorf, S. & Pantel, K. Circulating and disseminated tumour cells - mechanisms of immune surveillance and escape. *Nat Rev Clin Oncol* **14**, 155-167, doi:10.1038/nrclinonc.2016.144 (2017).
- 3 Lambert, A. W., Pattabiraman, D. R. & Weinberg, R. A. Emerging Biological Principles of Metastasis. *Cell* **168**, 670-691, doi:10.1016/j.cell.2016.11.037 (2017).
- 4 Aceto, N. *et al.* Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* **158**, 1110-1122, doi:10.1016/j.cell.2014.07.013 (2014).
- 5 Cheung, K. J. *et al.* Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E854-863, doi:10.1073/pnas.1508541113 (2016).
- 6 Pantel, K. & Speicher, M. R. The biology of circulating tumor cells. *Oncogene* **35**, 1216-1224, doi:10.1038/onc.2015.192 (2016).
- 7 Aceto, N., Toner, M., Maheswaran, S. & Haber, D. A. En Route to Metastasis: Circulating Tumor Cell Clusters and Epithelial-to-Mesenchymal Transition. *Trends Cancer* **1**, 44-52, doi:10.1016/j.trecan.2015.07.006 (2015).
- 8 Stott, S. L. *et al.* Isolation of circulating tumor cells using a microvortex-generating herringbone-chip. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 18392-18397, doi:10.1073/pnas.1012539107 (2010).
- 9 Xu, L. *et al.* Optimization and Evaluation of a Novel Size Based Circulating Tumor Cell Isolation System. *PLoS One* **10**, e0138032, doi:10.1371/journal.pone.0138032 (2015).
- 10 Li, H. *et al.* Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat Genet* **49**, 708-718, doi:10.1038/ng.3818 (2017).
- 11 Fridlender, Z. G. *et al.* Polarization of tumor-associated neutrophil phenotype by TGF-beta: "N1" versus "N2" TAN. *Cancer cell* **16**, 183-194, doi:10.1016/j.ccr.2009.06.017 (2009).
- 12 Cristofanilli, M. *et al.* Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N Engl J Med* **351**, 781-791, doi:10.1056/NEJMoa040766 (2004).
- 13 Labelle, M., Begum, S. & Hynes, R. O. Direct signaling between platelets and cancer cells induces an epithelial-mesenchymal-like transition and promotes metastasis. *Cancer cell* **20**, 576-590, doi:10.1016/j.ccr.2011.09.009 (2011).
- 14 Rothstein, G. *et al.* Stimulation of neutrophil production in CSF-1-responsive clones. *Blood* **72**, 898-902 (1988).

- 15 He, J. Q. *et al.* Association of genetic variations in the CSF2 and CSF3 genes with lung function in smoking-induced COPD. *Eur Respir J* **32**, 25-34, doi:10.1183/09031936.00040307 (2008).
- 16 Verri, W. A., Jr. *et al.* IL-15 mediates antigen-induced neutrophil migration by triggering IL-18 production. *Eur J Immunol* **37**, 3373-3380, doi:10.1002/eji.200737488 (2007).
- 17 Kacinski, B. M. *et al.* FMS (CSF-1 receptor) and CSF-1 transcripts and protein are expressed by human breast carcinomas in vivo and in vitro. *Oncogene* **6**, 941-952 (1991).
- 18 McCracken, J. M. & Allen, L. A. Regulation of human neutrophil apoptosis and lifespan in health and disease. *J Cell Death* **7**, 15-23, doi:10.4137/JCD.S11038 (2014).
- 19 Canli, O. *et al.* Myeloid Cell-Derived Reactive Oxygen Species Induce Epithelial Mutagenesis. *Cancer cell* **32**, 869-883 e865, doi:10.1016/j.ccell.2017.11.004 (2017).
- 20 Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-421, doi:10.1038/nature12477 (2013).
- 21 Wellenstein, M. D. & de Visser, K. E. Cancer-Cell-Intrinsic Mechanisms Shaping the Tumor Immune Landscape. *Immunity* **48**, 399-416, doi:10.1016/j.immuni.2018.03.004 (2018).
- 22 Ramasamy, S. *et al.* Tle1 tumor suppressor negatively regulates inflammation in vivo and modulates NF-kappaB inflammatory pathway. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 1871-1876, doi:10.1073/pnas.1511380113 (2016).
- 23 Yu, M. *et al.* Cancer therapy. Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility. *Science* **345**, 216-220, doi:10.1126/science.1253533 (2014).
- 24 Macaulay, I. C. *et al.* Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. *Nat Protoc* **11**, 2081-2103, doi:10.1038/nprot.2016.138 (2016).
- 25 Chen, S. *et al.* Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**, 1246-1260, doi:10.1016/j.cell.2015.02.038 (2015).

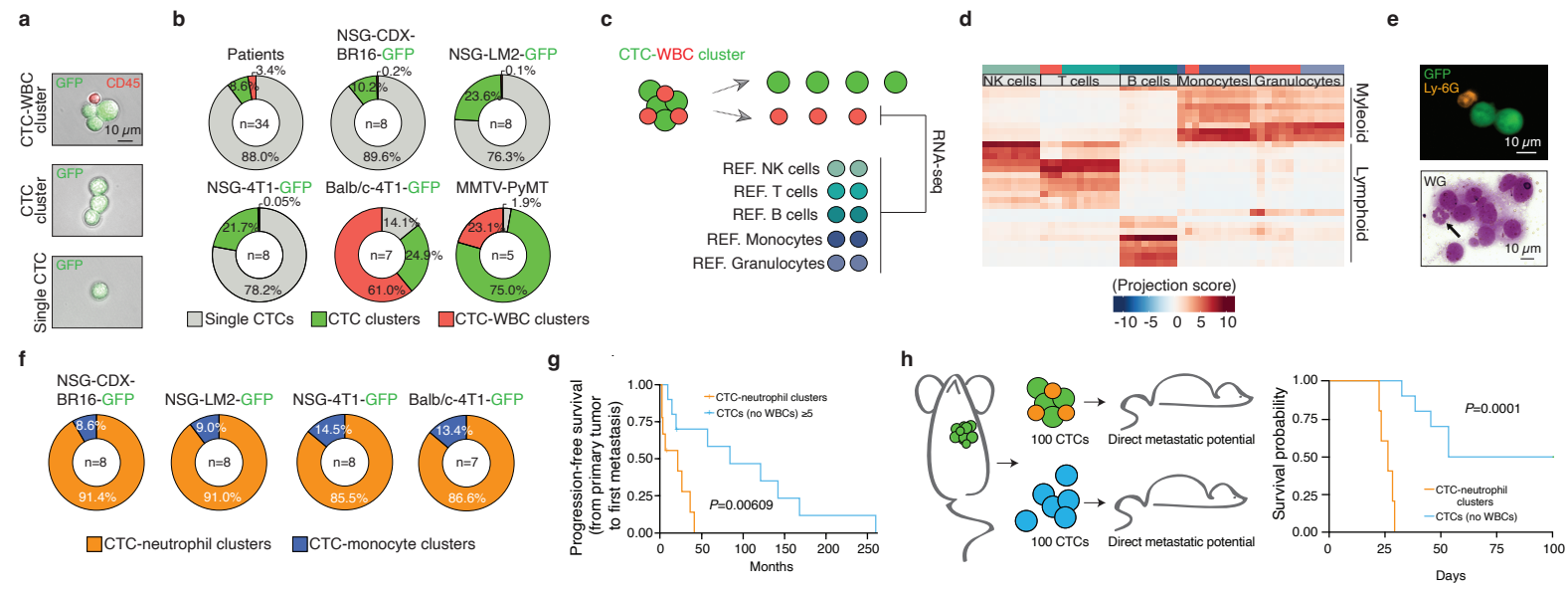


Figure 1

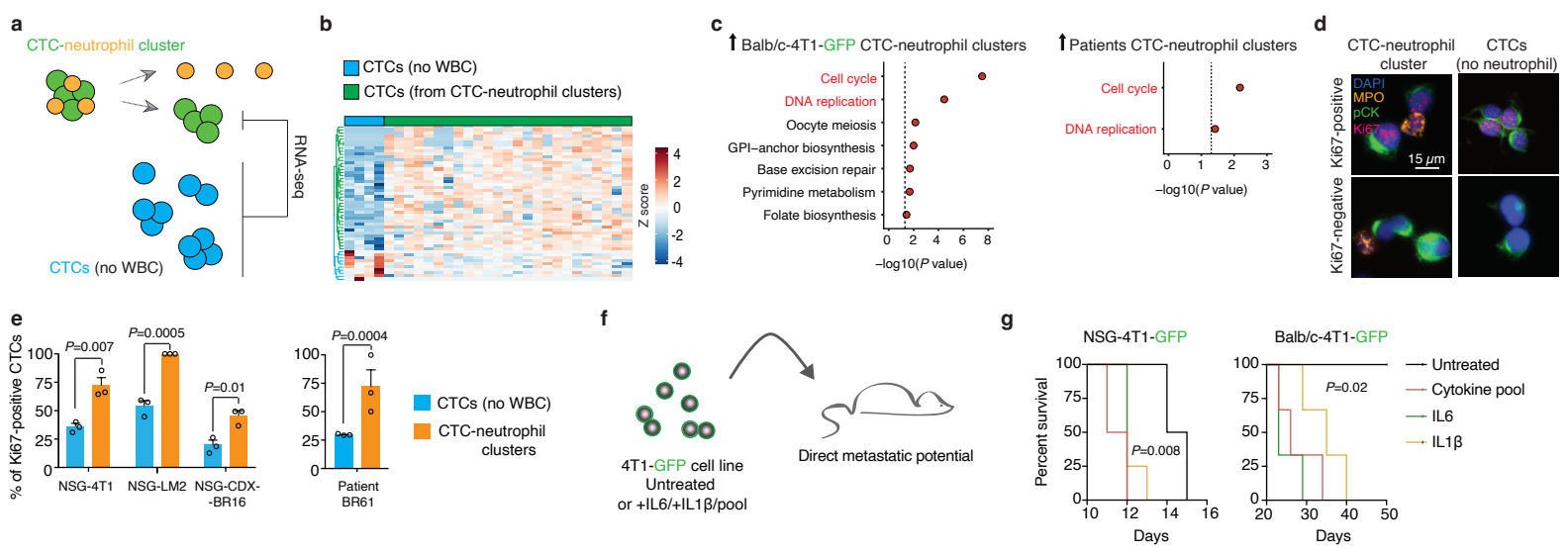


Figure 2

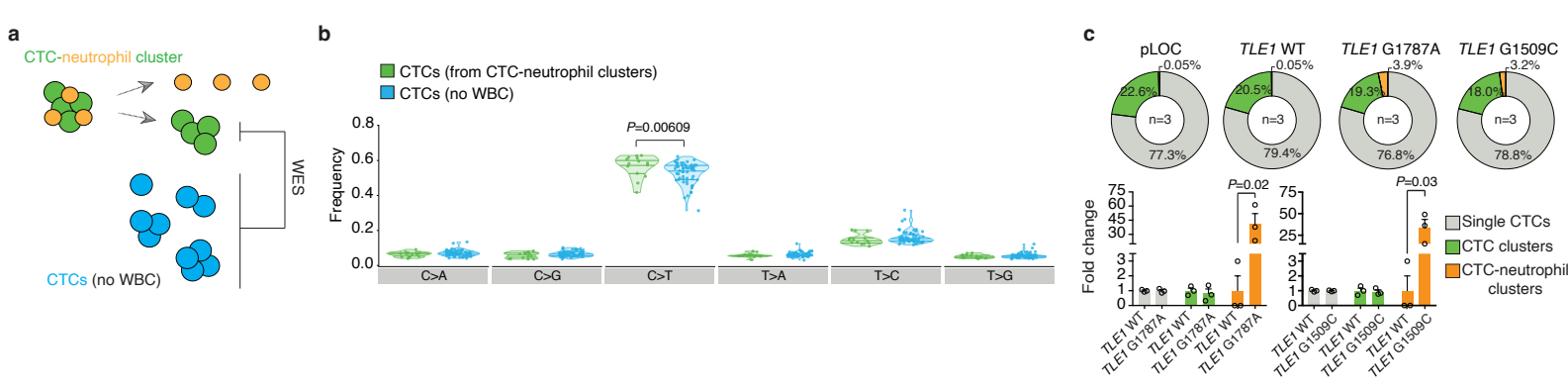


Figure 3

