Lung endothelium instructs dormancy of susceptible metastatic tumour cells

Moritz Jakab*[,#,1,2,3], Ki Hong Lee*[1,2,3], Alexey Uvarovskii[4,6,§], Svetlana Ovchinnikova[4,5], Shubhada R Kulkarni[1,2], Till Rostalski[1,2], Simon Anders[4,5], and Hellmut G Augustin[#,1,2]

[1]European Center for Angioscience (ECAS), Medical Faculty Mannheim, Heidelberg University; [2]Division of Vascular Oncology and Metastasis, German Cancer Research Center Heidelberg (DKFZ-ZMBH Alliance); [3]Faculty of Biosciences, Heidelberg University; [4]Center for Molecular Biology (ZMBH), Heidelberg University; [5]Bioquant Center, Heidelberg University

[§]present address: Evotec SE, Göttingen

Equally contributing first authors; [#]Corresponding authors

Correspondence should be addressed to:

Moritz Jakab or Hellmut G. Augustin
European Center for Angioscience, Medical Faculty Mannheim, Heidelberg University, and German Cancer Research Center Heidelberg
Im Neuenheimer Feld 280
D-69120 Heidelberg, Germany
Phone: +49-6221-421500
Email: m.jakab@dkfz.de or augustin@angioscience.de

Jakab, Lee, *et al*.

Abstract

During metastasis, cancer cells hijack blood vessels and travel via the circulation to colonize distant sites[1,2]. Due to the rarity of these events, the immediate cell fate decisions of arrested circulating tumour cells (aCTC) are poorly understood and the role of the endothelium, as the interface of dissemination, remains elusive[3,4]. Here, we developed a novel strategy to specifically enrich for aCTC subpopulations capturing all cell states of the extravasation process and, in combination with single cell RNA-sequencing, provide a first blueprint of the transcriptional basis of early aCTC decisions. Upon their arrest at the metastatic site, tumour cells either started proliferating intravascularly or extravasated and preferably reached a state of quiescence. Endothelial-derived angiocrine Wnt factors were found to drive this bifurcation by inducing a mesenchymal-like phenotype in aCTCs instructing them to follow the extravasation-dormancy branch. Surprisingly, homogenous tumour cell pools showed an unexpected baseline heterogeneity in Wnt signalling activity and epithelial-to-mesenchyme-transition (EMT) states. This heterogeneity was established at the epigenetic level and served as the driving force of aCTC behaviour. Hypomethylation enabled high baseline Wnt and EMT activity in tumour cells leading them to preferably follow the extravasation-dormancy route, whereas methylated tumour cells had low activity and proliferated intravascularly. The data identify the pre-determined methylation status of disseminated tumour cells as a key regulator of aCTC behaviour in the metastatic niche. While metastatic niche-derived factors per default instruct the acquisition of quiescence, aCTCs unwind a default proliferation program and only deviate from it if hypomethylation in key gene families renders them responsive towards the microenvironment.

Main:

Tumour cell (TC) dormancy poses a major hurdle for the treatment of cancer[1-3]. During metastasis, dormant tumour cells (DTC) reside in close proximity to blood vessels and acquire a stem-like phenotype[4]. Yet, metastasizing TCs show a remarkable heterogeneity in their genetic and molecular makeup, which can be attributed to why certain cancer cells reach a state of tumour dormancy, whereas others outgrow immediately to form macro-metastases[5-8]. This differential behaviour is not solely driven by TC intrinsic properties, but also influenced by the niche, as certain niches in principle favour TC proliferation, whereas others are primarily tumour-suppressive[9-11]. This argues for a scenario in which the induction of dormancy depends on cell-intrinsic properties and matching microenvironmental factors. This is also corroborated by the finding that DTCs, once committed to their fate, require dramatic events to be awakened[12,13]. We therefore hypothesized that TC behaviour is primed during the initial arrival of CTCs at the metastatic niche and that the vascular endothelium, as the interface of dissemination, is a crucial fate instructor.

Wnt and EMT pathways are drivers of tumour extravasation and dormancy

To test our hypothesis, we developed a novel experimental model to temporally assess TC and endothelial cell (EC) interactions *in vivo* at single cell resolution for the first time. Wildtype Balb/c mice were intravenously injected with GFP-expressing 4T1 breast cancer cells (4T1-GFP). Lung-seeded TCs and corresponding total lung ECs were isolated at day 0 (baseline), day 1.5 (peak phase of extravasation) and day 3.5 (induction of proliferation) and purified by

43   fluorescence-activated cell sorting (FACS) (**Fig. 1a**). Plate-based single cell RNA-sequencing
44   (scRNAseq) was used to analyse the transcriptional signatures of TC-EC interactions during
45   metastatic colonization. To discriminate extravascular from intravascular TCs, mice were
46   intravenously injected at day 1.5 with fluorescently labelled anti-H2-Kd antibody labelling all
47   body cells, including syngeneic 4T1-GFP cells that were exposed to the circulation. We further
48   discriminated proliferative from quiescent TCs at day 3.5 based on the dilution of CellTracer
49   dye and considered TCs exhibiting dye retention as dormant (**Fig. 1a**). For each TC
50   subpopulation and matched ECs, equal numbers were sorted from at least three biological
51   replicates. The cells in the dataset showed homogenous distribution for counts and detected
52   genes and the replicates interlaced well in a Uniform Manifold Approximation and Projection
53   (UMAP), demonstrating the robustness of the experimental approach. Moreover, substructures
54   in the UMAP were found to be specifically enriched for cells from the respective FACS gates,
55   suggesting that the gating strategy was suitable to enrich rare TC subpopulations, even though
56   it did not yield high purity. Correspondingly, *H2-K1* expression did not differ between the
57   intravascular and extravascular fraction, highlighting that differences in staining intensity
58   reflected exposure to circulation and not gene regulation.

59   Clustering of the combined TC dataset identified a total of five clusters that reflected the
60   respective FACS gates (**Fig. 1b**). Trajectory analysis to reconstruct the colonization process
61   with TCs of the intravascular cluster set as starting point identified three main trajectories
62   transitioning 1) from intravascular to proliferative cells, 2) to a subset of extravascular cells,
63   and 3) through a subset of extravascular cells to dormant cells (**Fig. 1c**). These findings
64   indicated that TC extravasation was a pre-requisite for tumour dormancy, but dispensable for
65   TC proliferation, which molecularly defines earlier microscopy-based concepts[14,15]. Based on
66   this finding, we next compared proliferative vs. dormant cells and scored each TC from the day
67   3.5 dataset for the expression of G2M-phase and S-phase genes. Only TCs that showed dye
68   retention and were not in cycle were considered as bona fide dormant to exclude cells that had
69   proliferated but dropped out of cycle (**Fig. 1d**). After regression of cell cycle-associated genes,
70   differential gene expression analysis (DGEA) was performed with subsequent gene set
71   enrichment analysis (GSEA) (**Fig. 1e**). In line with previous reports[16,17], transforming growth
72   factor beta (TGFβ) signalling and epithelial-to-mesenchyme transition (EMT) gene sets were
73   enriched in DTCs. Surprisingly, beta-Catenin-mediated canonical Wnt signalling was identified
74   as one of the most significantly enriched pathways (**Fig. 1e**). Wnt ligands have extensively been
75   characterized as protumorigenic growth factors[18], promoting proliferation in primary tumours
76   and metastases, as well as CTC survival[19-22]. Unexpectedly, expression of Wnt pathway and
77   EMT-associated genes was enriched alongside the extravascular-dormancy trajectory (**Fig. 1f,
78   g**), supporting the hypothesis that niche-derived Wnt ligands may drive tumour dormancy. In
79   conclusion, these data provide a first transcriptional blueprint of aCTC fate decisions in the
80   metastatic lung.

81   **The lung endothelium displays a bimodal response towards arriving CTCs**

82   It was previously established that the endothelium serves as a systemic amplifier of primary
83   tumour-derived signals[23,24]. Here, clustering of lung ECs revealed the emergence of cycling
84   ECs as well as a shift of general capillary ECs (gCap)[25,26] towards an activated phenotype as a
85   response towards arriving TCs. To determine the basis of gCap activation, DGEA on filtered

86   capillary EC pseudo-bulks was performed. For this, large vessel EC were manually annotated
87   and removed from the dataset based on congruent marker gene expression. Capillary ECs
88   displayed an immediate response pattern with genes being mostly regulated at day 1.5. These
89   immediate-response genes involved immune modulatory and cell cycle genes and were
90   regulated in a bimodal-manner with systemic upregulation of secreted EC factors (angiokines)
91   and focal enrichment of biosynthesis genes in activated gCap. This led us to conclude that the
92   lung endothelium, while exerting important immune regulatory functions at the systems level,
93   also served as a local producer of biomass thereby generating a conducive micro-niche.

Niche-derived angiocrine Wnt ligands are instructors of tumour dormancy

95   Next, we analysed the consequences of the enriched Wnt signalling in DTCs. For this purpose,
96   4T1-GFP breast cancer cells were treated *in vitro* for 2 weeks with canonical Wnt pathway
97   agonists prior to injection in a gain-of-function (G-O-F) approach. Conversely, mice were
98   treated with a Porcupine inhibitor (LGK974) to create a Wnt-deficient environment (**Fig. 2a**).
99   As expected, Wnt G-O-F programmed TCs to follow the extravasation-dormancy route, which
100   resulted in enhanced extravasation and higher incidence of dormancy with overall reduced
101   short-term metastatic burden (**Fig 2b, c**). In contrast, Wnt depletion enhanced metastatic
102   outgrowth, but did not affect extravasation (**Fig 2b, c**). As modulating Wnt signalling was
103   sufficient to alter TC behaviour *in vivo*, we probed for sources of Wnt ligands in the metastatic
104   niche. The endothelium was found to robustly express Wnt ligands across the experimental
105   timeline. Yet, their expression was not changed (**Fig. 2d**). Nevertheless, depleting Wnt ligands
106   specifically from the vascular niche by EC-specific knockout of the Wnt cargo receptor *Wntless*
107   (*Wls*) led to a significantly increased short-term metastatic burden, thereby phenocopying the
108   systemic pharmacological inhibition. This was observed for E0771-GFP breast cancer cells, but
109   also for B16F10 melanoma cells (**Fig. 2e**), highlighting the endothelium as a major source of
110   dormancy-inducing Wnt ligands. The Wnt dependency of metastatic dormancy acquisition
111   could also be observed in a clinically relevant spontaneous metastasis model involving surgical
112   removal of the primary tumour (**Fig. 2f**). Specifically, systemic treatment with LGK974 did not
113   affect primary tumour growth or size, mouse body weight or the tumour vasculature (**Fig. 2g**),
114   but resulted in a significantly increased metastatic burden (**Fig. 2h**). Collectively, these data
115   establish that angiocrine Wnt ligands are crucial instructors of extravasated TC dormancy.

Metastatic tumour cell behaviour is not regulated at the receptor-ligand level or by niche occupancy

118   As the expression of angiocrine Wnt ligands was not changed, we reasoned that Wnt signalling
119   activity differences may result from distinct TC receptor repertoires. Surprisingly, prediction
120   of EC-TC interactions based on DGEA of TC pseudo-bulks (intravascular vs extravascular and
121   proliferative vs. dormant) using CellPhoneDB[26] did not reveal Wnt pathway components.
122   Repeating the analysis specifically with TC-expressed Wnt receptors led to the identification
123   of five (co)-receptors that were differentially expressed for either the intravascular vs
124   extravascular or the dormant vs proliferative comparison, but not for both. Moreover, receptor
125   expression was not enriched in the intravascular-proliferative or the extravasation-dormancy
126   branch of the trajectory-analysis, suggesting that the observed differences in Wnt signalling
127   activity were not established at the receptor-ligand level.

128 As the lung endothelium harbours two distinct vascular beds that are defined by less penetrable
129 gCaps and more permissive aerocytes (aCap)[23, 24] (**Fig. 3a**), we tested whether distinct vascular
130 niche occupancy could drive the observed differential TC behaviour. Employing an *in vivo*
131 niche-labelling system[28], we specifically enriched for ECs interacting with mostly proliferative
132 4T1-GFP cells or quiescent D2.0R-GFP breast cancer cells that served as a proxy for dormant
133 4T1 cells. Labelled niche ECs, as well as matched unlabelled total ECs were FACS-purified
134 and subjected to bulk RNAseq (**Fig. 3b, c**). Aerocyte and gCap-specific genes that showed
135 robust and stable expression were used to deconvolute the cellular composition of the bulk
136 samples (**Fig. 3d**). While a general bias towards the aCap signature could be observed for all
137 tumour-bearing samples, no differences were detected between the dormant and proliferative
138 niche or their unlabelled counterparts (**Fig. 3e**), indicating that DTCs and proliferative TCs
139 occupy the same vascular niches.

140 Interestingly, proliferative TCs, in contrast to DTCs, induced the production of extracellular
141 matrix (ECM), immune response genes and proliferation in ECs. While matrix-remodelling
142 processes were specific to the niche, proliferative and pro-inflammatory programs were part of
143 a systemic response. We sought to deduce a marker gene set from the bulk comparisons that
144 was specific for ECs extracted from the proliferative niche and tested all conditions against
145 each other. The resulting gene panel was used to predict tumour-interacting ECs in the
146 scRNAseq data. Predicted tumour-interacting ECs co-localized with the previously identified
147 biosynthetic ECs in the UMAP, confirming the hypothesis that the endothelium elicited a
148 bimodal response. To test the generality of the gene panel, a publicly available scRNAseq
149 dataset[29] was utilized and a similar enrichment was found specifically for primary lung tumour
150 ECs compared to non-tumorous matched samples.

151 **Heterogenous methylation states pre-determine tumour cell behaviour**

152 As DTC fates were established independently of differential exogenous factors, we next probed
153 for TC-intrinsic properties. Surprisingly, cultured TCs exhibited a heterogeneous but
154 correlating baseline expression of EMT- and Wnt pathway-associated genes (**Fig. 4a, b**). Such
155 state differences were also identified in freshly isolated CTC of breast cancer patients[30,31],
156 indicating baseline tumour cell-intrinsic differences. Interestingly, overnight pulse treatment
157 with Wnt agonists failed to programme the extravasation-dormancy shift observed for long-
158 term treatments (**Fig. 4c**) and the expression profile of key EMT-associated transcription
159 factors was not markedly changed between pulse-treated and reprogrammed cells. This led us
160 to hypothesize that TC were restricted in their responsiveness towards niche-derived factors by
161 an epigenetic barrier. In agreement, pulse-treatment of 4T1-GFP with de-methylating agent
162 (decitabine) enabled TCs to preferably follow the extravasation-dormancy route (**Fig. 4d**). This
163 was not driven by cellular fitness or changes in homing capacity. Similar to the Wnt G-O-F,
164 hypomethylation did not induce EMT *in vitro* and dormancy-induction was still dependent on
165 Wnt (**Fig. 4e**). Notably, pulse-treating hypomethylated cells with Wnt agonist prior to injection
166 did not alter the *in vivo* phenotype, indicating that niche-derived signalling was saturated and
167 sufficient. Moreover, none of the *in vitro* treatments affected the proliferation rate of
168 proliferation-committed TCs *in vivo*, clearly showing that differences in short-term metastatic
169 burden were a consequence of dormancy-induction.

Jakab, Lee, *et al*.

170   We then assessed the methylation state of dormant and proliferative TCs by whole genome
171   bisulfite-sequencing. Global methylation levels were not changed in DTCs, but promoter
172   sequences and gene bodies displayed considerable hypomethylation, while enhancer sequences
173   were unaffected (**Fig. 4f**). Remarkably, hypomethylation mainly occurred in genes and
174   promoters that were epigenetically sealed in proliferative TCs (>70% methylation). We
175   computed the overlap of genes with >10% hypomethylation in DTC and gene ontology terms
176   from the mouse signature database (MSigDB)[32] and found transcription factor binding and cell
177   fate processes as top hits. This confirmed the hypothesis that hypomethylation underlay cellular
178   plasticity and was the driving force of TC responsiveness towards niche-derived dormancy-
179   inducing factors.

Discussion:

181   Tumour cell pre-determination is an emerging concept[33-35]. Here, we identified the epigenetic
182   pre-coding of disseminated TC behaviour in the metastatic niche. DTC were characterized by
183   hypomethylation in promoter sequences and gene bodies, whereas proliferative TCs were
184   epigenetically sealed. We envision that the plastic-dormant and the sealed-proliferative state
185   form a dynamic equilibrium. Long-term treatment with a Wnt agonist would direct the
186   equilibrium towards the plastic-dormant state, without affecting the cell-state itself. This is
187   supported by reports of similar phenotype transitions caused by long-term *in vitro* treatments
188   or targeted genetic manipulation of signalling pathways[36,37]. However, such state-transitions
189   were also reported to occur spontaneously and were found to be a pre-requisite for metastatic
190   outgrowth[38]. TC hypomethylation was reflected at the transcriptomic level by an elevated
191   baseline expression of EMT and Wnt pathway-associated genes. A similar heterogeneous
192   expression was found in freshly isolated CTCs from breast cancer patients and could be linked
193   directly to their metastatic potential[30,39]. In this context, the primary tumour could be viewed as
194   a heterogenous amplifier in which high selective pressure forces the acquisition of distinct TC
195   states. Recent lineage tracing experiments highlighted this phenomenon and revealed hybrid
196   EMT TC states as the underlying principle of metastatic dissemination[5-8,35]. While EMT was
197   needed for migration and intravasation, too much of it limited metastatic outgrowth.
198   Interestingly, hybrid EMT states were not discrete but formed a continuum that correlated with
199   the metastatic outcome[6-8]. Similar gradual states could be observed in cultured TCs and could
200   be a direct consequence of epigenetic plasticity. Plastic cells would show high EMT and follow
201   the extravasation-dormancy route, whereas sealed TCs would form macro-metastases. Probing
202   the epigenetic and transcriptomic state of CTCs could therefore serve as predictive tool to assess
203   the likelihood of metastatic relapse in patients.

204   Besides cell-intrinsic properties, disseminated TC phenotypes are established as a consequence
205   of instructive niche-derived factors. Here, we identified endothelium-derived angiocrine Wnt
206   signalling as a prototypic example of such dependency. However, other factors and other
207   cellular sources have been identified previously and are most likely to act synergistically[9, 16, 17,
208   40-44] . Most surprisingly, homeostatic angiocrine Wnt signalling was found to be sufficient to
209   drive dormancy-induction, suggesting a default tumour-suppressive lung niche. Similar default
210   programs could occur in other organs in which ECs comprise a major Wnt source and were
211   reported previously in different contexts[9,11]. Moreover, primary tumour-instructed remodelling

212   of the niche could change the default state[45-49]. Additionally, the data strongly suggest that
213   metastatic TCs actively inflicted a niche-EC gene program that resembled primary tumour EC
214   signatures[29,50] and that could fuel TC proliferation by altering the biophysical properties of the
215   micro-niche.

216   Collectively, these data provide a first insight into the establishment of arrested CTC fates. We
217   show that susceptible epigenetic states render TCs responsive towards niche-derived default
218   factors, thus, opening the opportunity to probe for TC-niche interdependencies at the systems
219   level.

220

References:

1   Risson, E., Nobre, A. R., Maguer-Satta, V. & Aguirre-Ghiso, J. A. The current paradigm and challenges ahead for the dormancy of disseminated tumor cells. *Nat. Cancer* 1, 672-680, doi:10.1038/s43018-020-0088-5 (2020).

2   Polzer, B. & Klein, C. A. Metastasis awakening: the challenges of targeting minimal residual cancer. *Nat. Med.* 19, 274-275, doi:10.1038/nm.3121 (2013).

3   Labelle, M. & Hynes, R. O. The initial hours of metastasis: the importance of cooperative host–tumor cell interactions during hematogenous dissemination. *Cancer Discovery* 2, 1091-1099, doi:10.1158/2159-8290.Cd-12-0329 (2012).

4   Ghajar, C. M. Metastasis prevention by targeting the dormant niche. *Nat. Rev. Cancer* 15, 238-247, doi:10.1038/nrc3910 (2015).

5   Fluegen, G. et al. Phenotypic heterogeneity of disseminated tumour cells is preset by primary tumour hypoxic microenvironments. *Nat. Cell Biol.* 19, 120-132, doi:10.1038/ncb3465 (2017).

6   Simeonov, K. P. et al. Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. *Cancer Cell* 39, 1150-1162.e1159, doi:10.1016/j.ccell.2021.05.005 (2021).

7   Lüönd, F. et al. Distinct contributions of partial and full EMT to breast cancer malignancy. *Dev. Cell* 56, 3203-3221.e3211, doi:10.1016/j.devcel.2021.11.006 (2021).

8   Yang, D. et al. Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. *Cell*, doi:10.1016/j.cell.2022.04.015 (2022).

9   Ghajar, C. M. et al. The perivascular niche regulates breast tumour dormancy. *Nat. Cell Biol.* 15, 807-817, doi:10.1038/ncb2767 (2013).

10 Baumann, Z., Auf der Maur, P. & Bentires-Alj, M. Feed-forward loops between metastatic cancer cells and their microenvironment-the stage of escalation. *EMBO Mol. Med.*, e14283, doi:10.15252/emmm.202114283 (2022).

11 Crist, S. B. et al. Unchecked oxidative stress in skeletal muscle prevents outgrowth of disseminated tumour cells. *Nat. Cell Biol.*. 24, 538-553, doi:10.1038/s41556-022-00881-4 (2022).

12 Albrengues, J. et al. Neutrophil extracellular traps produced during inflammation awaken dormant cancer cells in mice. *Science* (New York, N.Y.) 361, doi:10.1126/science.aao4227 (2018).

13 Dai, J. et al. Astrocytic laminin-211 drives disseminated breast tumor cell dormancy in brain. *Nat. Cancer* 3, 25-42, doi:10.1038/s43018-021-00297-3 (2022).

14 Al-Mehdi, A. B. et al. Intravascular origin of metastasis from the proliferation of endothelium-attached tumor cells: A new model for metastasis. *Nat. Med.* 6, 100-102, doi:10.1038/71429 (2000).

15 Wong, C. W. et al. Intravascular location of breast cancer cells after spontaneous metastasis to the lung. *Am. J. Pathol.* 161, 749-753, doi:10.1016/s0002-9440(10)64233-2 (2002).

16 Bragado, P. et al. TGF-β2 dictates disseminated tumour cell fate in target organs through TGF-β-RIII and p38α/β signalling. *Nat. Cell Biol.* 15, 1351-1361, doi:10.1038/ncb2861 (2013).

17 Nobre, A. R. et al. Bone marrow NG2(+)/Nestin(+) mesenchymal stem cells drive DTC dormancy via TGFβ2. *Nat. Cancer* 2, 327-339, doi:10.1038/s43018-021-00179-8 (2021).

265   18   Niehrs, C. The complex world of WNT receptor signalling. *Nat. Rev. Mol. Cell Biol*. 13,
266   767-779, doi:10.1038/nrm3470 (2012).
267   19   Yu, M. et al. RNA sequencing of pancreatic circulating tumour cells implicates WNT
268   signalling in metastasis. *Nature* 487, 510-513, doi:10.1038/nature11217 (2012).
269   20   Miyamoto, D. T. et al. RNA-Seq of single prostate CTCs implicates noncanonical Wnt
270   signaling in antiandrogen resistance. *Science* (New York, N.Y.) 349, 1351-1356,
271   doi:10.1126/science.aab0917 (2015).
272   21   Malladi, S. et al. Metastatic latency and immune evasion through autocrine inhibition of
273   WNT. *Cell* 165, 45-60, doi:10.1016/j.cell.2016.02.025 (2016).
274   22   Tammela, T. et al. A Wnt-producing niche drives proliferative potential and progression in
275   lung adenocarcinoma. *Nature* 545, 355-359, doi:10.1038/nature22334 (2017).
276   23   Singhal, M. et al. Temporal multi-omics identifies LRG1 as a vascular niche instructor of
277   metastasis. *Sci. Trans. Med.* 13, eabe6805, doi:10.1126/scitranslmed.abe6805 (2021).
278   24   Hongu, T. et al. Perivascular tenascin C triggers sequential activation of macrophages and
279   endothelial cells to generate a pro-metastatic vascular niche in the lungs. *Nat. Cancer* 3,
280   486-504, doi:10.1038/s43018-022-00353-6 (2022).
281   25   Vila Ellis, L. et al. Epithelial Vegfa specifies a distinct endothelial population in the mouse
282   lung. *Dev. Cell* 52, 617-630.e616, doi:10.1016/j.devcel.2020.01.009 (2020).
283   26   Gillich, A. et al. Capillary cell-type specialization in the alveolus. *Nature* 586, 785-789,
284   doi:10.1038/s41586-020-2822-7 (2020).
285   27   Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB:
286   inferring cell–cell communication from combined expression of multi-subunit ligand–
287   receptor complexes. *Nat. Protoc*. 15, 1484-1506, doi:10.1038/s41596-020-0292-x (2020).
288   28   Ombrato, L. et al. Metastatic-niche labelling reveals parenchymal cells with stem features.
289   *Nature* 572, 603-608, doi:10.1038/s41586-019-1487-6 (2019).
290   29   Goveia, J. et al. An integrated gene expression landscape profiling approach to identify
291   lung tumor endothelial cell heterogeneity and angiogenic candidates. *Cancer Cell* 37, 21-
292   36.e13, doi:10.1016/j.ccell.2019.12.001 (2020).
293   30   Cheng, Y. H. et al. Hydro-Seq enables contamination-free high-throughput single-cell
294   RNA-sequencing for circulating tumor cells. *Nat. Commun*. 10, 2163, doi:10.1038/s41467-
295   019-10122-2 (2019).
296   31   Szczerba, B. M. et al. Neutrophils escort circulating tumour cells to enable cell cycle
297   progression. *Nature* 566, 553-557, doi:10.1038/s41586-019-0915-y (2019).
298   32   Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739-
299   1740, doi:10.1093/bioinformatics/btr260 (2011).
300   33   Shaffer, S. M. et al. Rare cell variability and drug-induced reprogramming as a mode of
301   cancer drug resistance. *Nature* 546, 431-435, doi:10.1038/nature22794 (2017).
302   34   Emert, B. L. et al. Variability within rare cell states enables multiple paths toward drug
303   resistance. *Nat. Biotechnol*. 39, 865-876, doi:10.1038/s41587-021-00837-3 (2021).
304   35   Quinn, J. J. et al. Single-cell lineages reveal the rates, routes, and drivers of metastasis in
305   cancer xenografts. *Science* (New York, N.Y.) 371, doi:10.1126/science.abc1944 (2021).
306   36   Pattabiraman, D. R. et al. Activation of PKA leads to mesenchymal-to-epithelial transition
307   and loss of tumor-initiating ability. *Science* (New York, N.Y.) 351, aad3680, doi:10.1126/
308   science.aad3680 (2016).

9

309 37 Zhang, Y. et al. Genome-wide CRISPR screen identifies PRC2 and KMT2D-COMPASS
310   as regulators of distinct EMT trajectories that contribute differentially to metastasis. *Nat.*
311   *Cell Biol.* 24, 554-564, doi:10.1038/s41556-022-00877-0 (2022).
312 38 Fumagalli, A. et al. Plasticity of Lgr5-negative cancer cells drives metastasis in colorectal
313   cancer. *Cell Stem Cell* 26, 569-578.e567, doi:10.1016/j.stem.2020.02.008 (2020).
314 39 Gkountela, S. et al. Circulating tumor cell clustering shapes DNA methylation to enable
315   metastasis seeding. *Cell* 176, 98-112.e114, doi:10.1016/j.cell.2018.11.046 (2019).
316 40 Werner-Klein, M. et al. Interleukin-6 trans-signaling is a candidate mechanism to drive
317   progression of human DCCs during clinical latency. *Nat. Commun*. 11, 4977,
318   doi:10.1038/s41467-020-18701-4 (2020).
319 41 Ren, D. et al. Wnt5a induces and maintains prostate cancer cells dormancy in bone. *J. Exp.*
320   *Med.* 216, 428-449, doi:10.1084/jem.20180661 (2019).
321 42 Montagner, M. et al. Crosstalk with lung epithelial cells regulates Sfrp2-mediated latency
322   in breast cancer dissemination. *Nat. Cell Biol*. 22, 289-296, doi:10.1038/s41556-020-0474-
323   3 (2020).
324 43 Pein, M. et al. Metastasis-initiating cells induce and exploit a fibroblast niche to fuel
325   malignant colonization of the lungs. *Nat. Commun*. 11, 1494, doi:10.1038/s41467-020-
326   15188-x (2020).
327 44 Correia, A. L. et al. Hepatic stellate cells suppress NK cell-sustained breast cancer
328   dormancy. *Nature* 594, 566-571, doi:10.1038/s41586-021-03614-z (2021).
329 45 Peinado, H. et al. Pre-metastatic niches: organ-specific homes for metastases. *Nat. Rev.*
330   *Cancer* 17, 302-317, doi:10.1038/nrc.2017.6 (2017).
331 46 Borriello, L. et al. Primary tumor associated macrophages activate programs of invasion
332   and dormancy in disseminating tumor cells. *Nat. Commun*. 13, 626, doi:10.1038/s41467-
333   022-28076-3 (2022).
334 47 Costa-Silva, B. et al. Pancreatic cancer exosomes initiate pre-metastatic niche formation in
335   the liver. *Nat. Cell Biol*. 17, 816-826, doi:10.1038/ncb3169 (2015).
336 48 Di Martino, J. S. et al. A tumor-derived type III collagen-rich ECM niche regulates tumor
337   cell dormancy. *Nat. Cancer* 3, 90-107, doi:10.1038/s43018-021-00291-9 (2022).
338 49 Liu, Y. & Cao, X. Characteristics and significance of the pre-metastatic niche. *Cancer Cell*
339   30, 668-681, doi:10.1016/j.ccell.2016.09.011 (2016).
340 50 Lambrechts, D. et al. Phenotype molding of stromal cells in the lung tumor micro-
341   environment. *Nat. Med.* 24, 1277-1289, doi:10.1038/s41591-018-0096-5 (2018).

342

Acknowledgements:

## Data availability:

All raw sequencing data, annotated and filtered count matrices generated in this study, will be made available to the reviewers upon request and publicly available upon final acceptance of the manuscript.

Code availability:

All code generated in this study will be made available to the reviewers upon request and publicly available upon final acceptance of the manuscript.

364 **Figure 1**
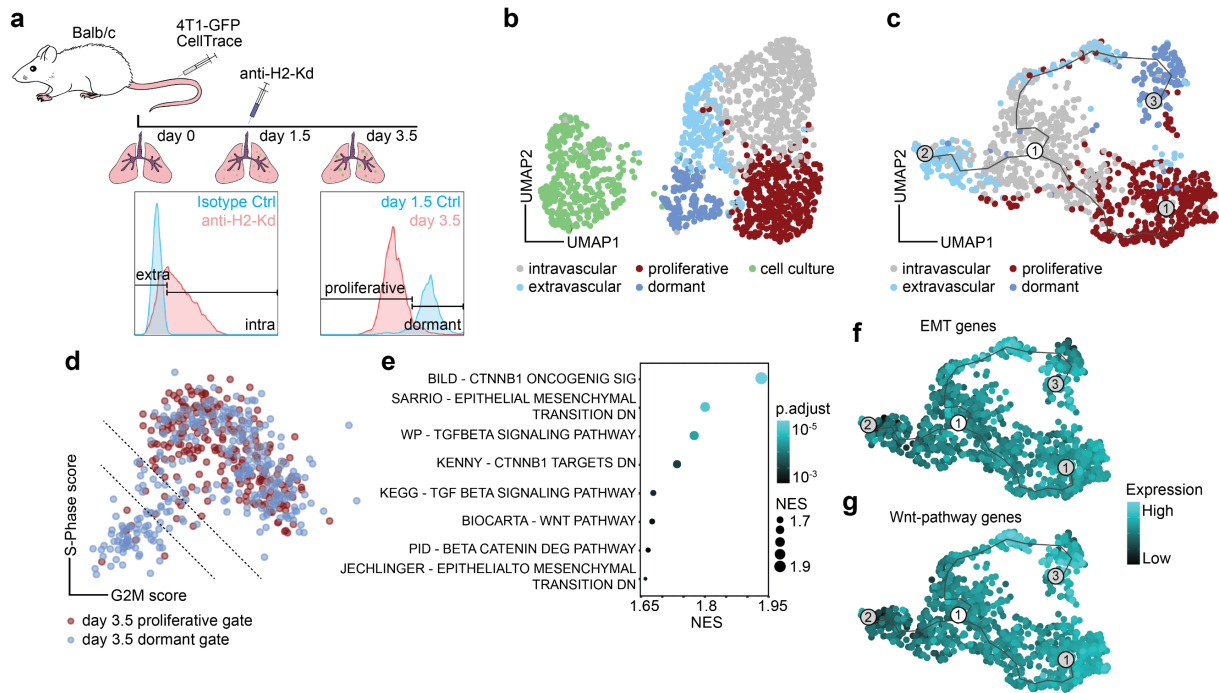
365

366 **Fig. 1: Extravasating and dormant tumour cells are defined by a Wnt and EMT signature. a**, Schematic
367 of the experimental design. Recipient Balb/c mice received two injections of 1×10^6 4T1-GFP cells stained
368 with CellTracer dye into the tail vein. At day 1.5 post-injection, mice were injected intravenously with 5μg
369 of fluorescently labelled anti-H2-Kd antibody. Tumour cells (TC) and endothelial cells (EC) were sampled
370 at day 0 (uninjected baseline), day 1.5 and day 3.5. TCs were discriminated based on extravasation status on
371 day 1.5 and based on proliferation status on day 3.5. **b**, Uniform manifold approximation and projection
372 (UMAP) of total TC dataset. Shared nearest neighbour (SNN-) based clustering resolves transcriptomes of
373 TCs into 5 clusters. **c**, Trajectory analysis of extracted TCs reveals three transition branches. **d**, Scatter plot
374 of S-Phase and G2M gene expression scores for individual cells extracted on day 3.5 and coloured by
375 respective FACS gates. Dotted line indicates thresholds of cells with score sums < -1 (lower line) and score
376 sums < 0 (upper line). **e**, Gene set enrichment analysis (GSEA) of genes upregulated in bona fide dormant
377 tumour cells ranked by fold change. NES, normalised enrichment score. **f**, Summed expression of 384 genes
378 upregulated during epithelial-to-mesenchyme transition (EMT) and **g**, 146 genes associated with Wnt
379 pathway plotted on the trajectory graph from **c**.
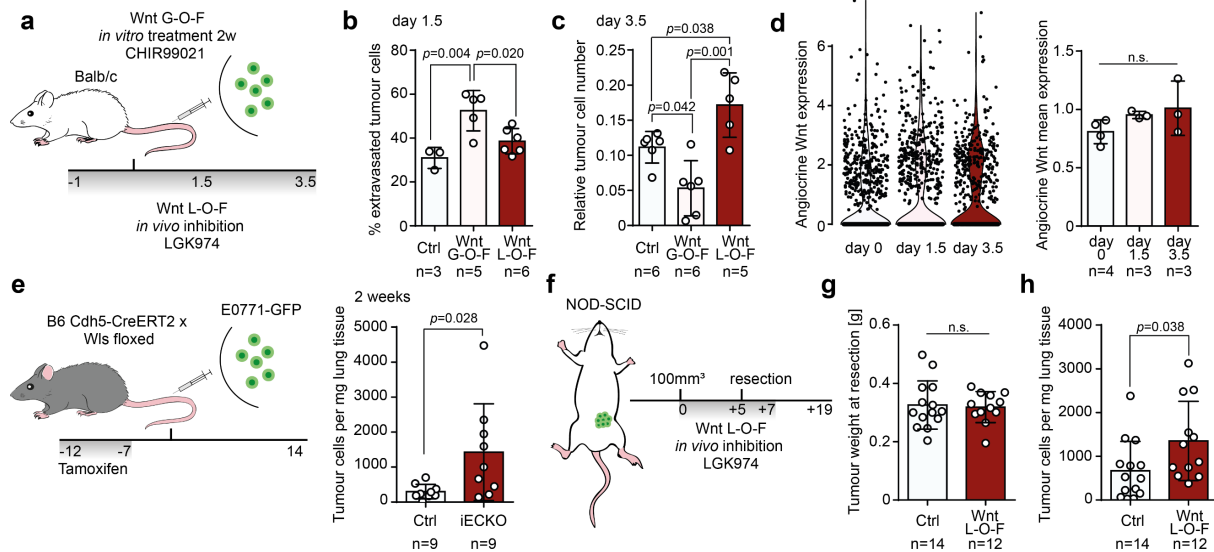
Figure 2

Fig. 2: Lung endothelial cells are a major source of dormancy-inducing Wnt ligands. **a**, Schematic of the experimental gain-of-function (G-O-F) and loss-of-function (L-O-F) strategy. Grey bar indicates timespan (in days) of daily treatment with LGK974. **b**, Percentage of extravasated TCs for control treatment, G-O-F or L-O-F approach 1.5 days post-injection. Error bars s.d., p values by one-way ANOVA with Tukey post-test are shown. **c**, Quantification of relative TC number normalised to EC abundance in lungs 3.5 days post-injection for control treatment, G-O-F or L-O-F approach. Error bars s.d., p values by one-way ANOVA with Tukey post-test are shown. **d**, Summed expression of Wnt ligands per timepoint in individual ECs (left panel) or as mean expression for pseudo-bulks of biological replicates. Error bars s.d., p values were calculated by one-way ANOVA with Tukey post-test. n.s., not significant. **e**, Schematic of experiment. Gene recombination was induced by tamoxifen administration. 2×10^5 E0771-GFP cells were injected into the tail vein of EC-specific knockout (iECKO) and control animals. Grey bar indicates timespan (in days) of daily tamoxifen treatment (left panel). Total number of TCs per milligram lung tissue of control and iECKO mice two weeks post-injection of E0771-GFP. Error bars s.d., p value by two-tailed t-test is shown. **f**, Schematic of the experiment. 1×10^6 4T1-GFP cells were implanted into the mammary fat pad of NOD-SCID mice. Once tumours reached a size of 100mm³, mice were treated with LGK974 for five days until tumour-resection. After resection, mice were treated for an additional two days and left to develop metastases. Grey bar indicates timespan (in days) of daily treatment with LGK974. **g**, Weights of resected primary tumours. Error bars s.d., p value was calculated by two-tailed t-test. n.s., not significant. **h**, Total number of TCs per milligram lung tissue of control and LGK974 treated mice two weeks post-resection. Error bars s.d., p value by two-tailed t-test is shown.

13

402 **Figure 3**

403
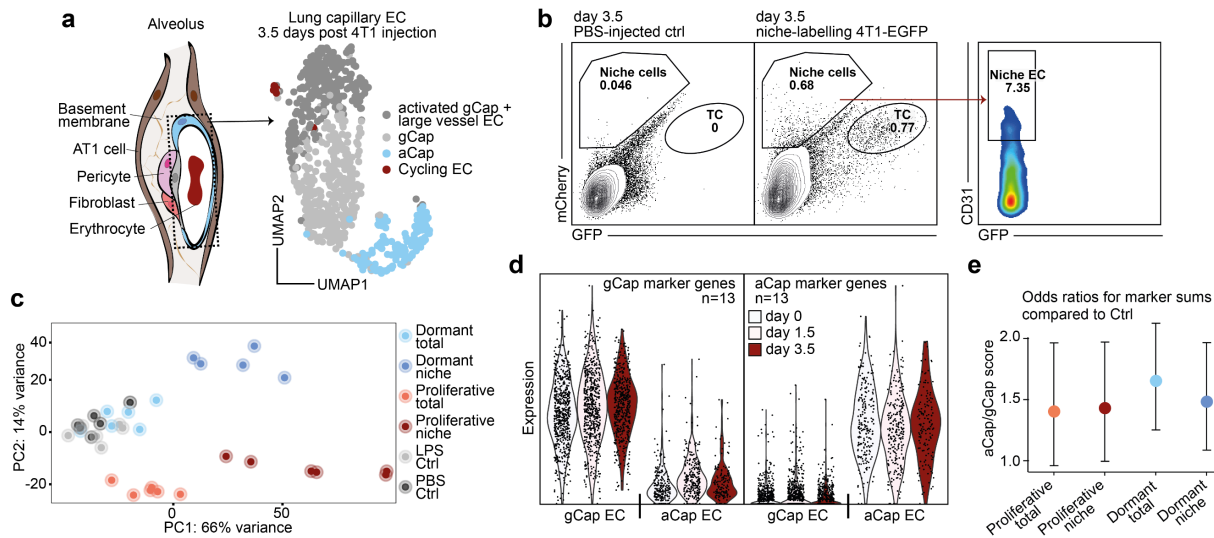
Fig. 3. Dormant tumour cells do not occupy distinct vascular niches in the lung. a, Schematic of lung alveolus (left panel). Dotted box highlights ECs. UMAP of EC transcriptomes reflecting the composition of bulk EC samples 3.5 days post-injection of 4T1-GFP cells. **b**, Representative FACS gates for purifying labelled niche ECs 3.5 days post-injection of niche-labelling 4T1-GFP. **c**, Principal component analysis of samples included in the experiment. Total samples refer to unlabelled CD31+ ECs, niche samples refer to labelled CD31+ ECs, dormant samples refer to injections of niche-labelling D2.0R-GFP, proliferative samples refer to injections of niche-labelling 4T1-GFP, LPS controls were injected intraperitoneally with LPS 24 hours prior to euthanasia, PBS controls were injected intravenously with PBS 3.5 days prior to euthanasia. **d**, Summed expression of gCap and aCap marker genes in individual ECs split by timepoints and by EC identity. **e**, Log2 fold changes of aCap marker genes to gCap marker genes odds ratios normalised to PBS-injected control samples. Error bars indicate 95% confidence interval.

14

Figure 4
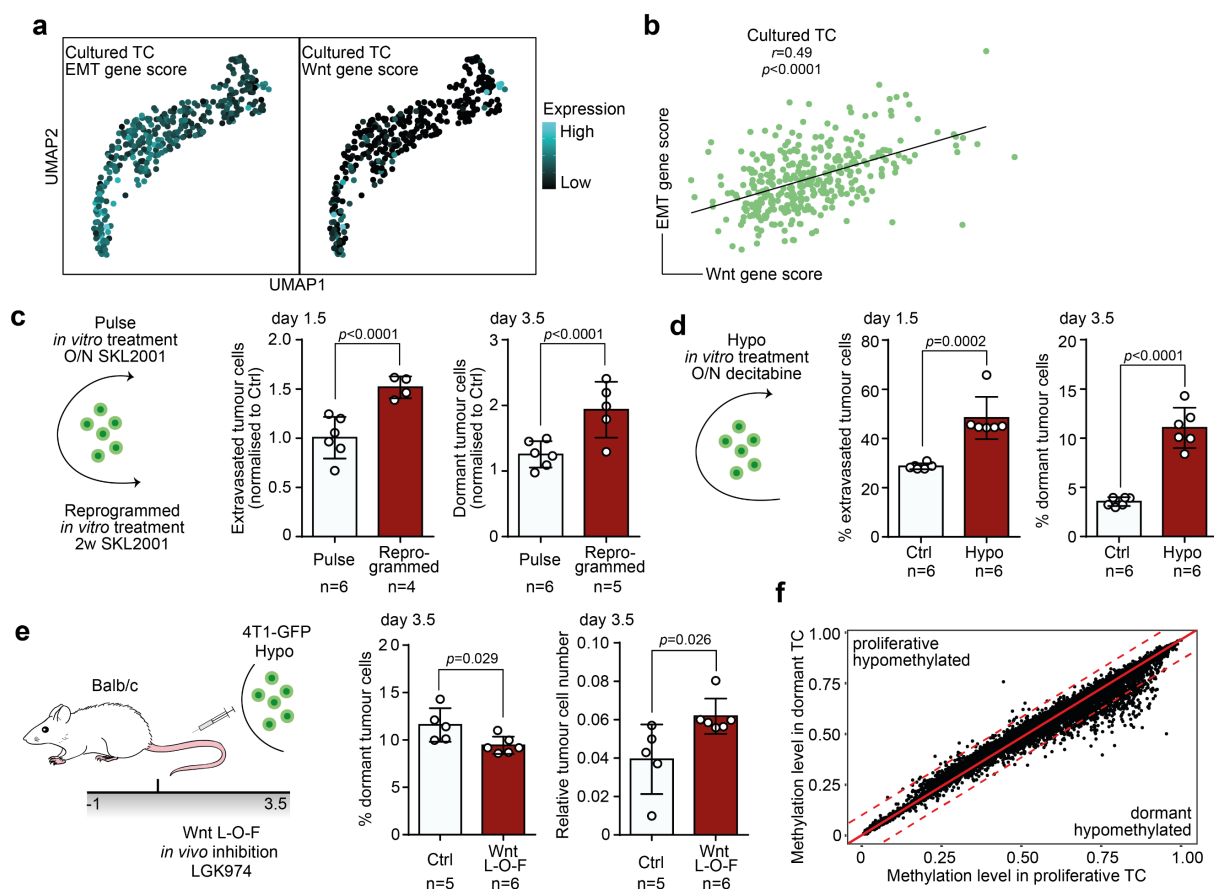
Fig. 4. Tumour cell behaviour in the metastatic niche is pre-determined by methylation state. a, Gene scores of EMT (left panel) and Wnt pathway-associated genes (right panel) in cultured TCs visualised in a UMAP and **b**, correlation of the gene scores. *P* value and *r* value by Pearson correlation are shown. **c**, Schematic of experiment (left panel). 4T1-GFP cells were either treated overnight with Wnt agonist (pulse) or for two weeks (reprogrammed) prior to injection. Relative fraction of extravasated TCs normalised to the respective control 1.5 days post-injection (mid panel) and relative fraction of dormant TCs normalised to the respective control 3.5 days post-injection (right panel). Error bars, s.d., *p* values by two-way ANOVA with Sidak post-test are shown. **d**, Schematic of experiment (left panel). 4T1-GFP cells were treated overnight with de-methylating agent decitabine (hypo). Percentage of extravasated TCs for control and hypomethylation treatment 1.5 days post-injection (mid panel) and percentage of dormant TCs 3.5 days post-injection (right panel). Error bars, s.d., *p* values by two-tailed t-test are shown. **e**, Schematic of experimental L-O-F approach (left panel). 4T1-GFP cells were treated overnight with decitabine. Grey bar indicates timespan (in days) of daily treatment with LGK974. Percentage of dormant TCs 3.5 days post-injection for control and LGK974 treated animals (mid panel) and relative tumour cell number normalised to EC abundance in lungs (right panel). Error bars, s.d., *p* values by two-tailed t-test are shown. **f**, Scatter plot of methylation level (fraction of methylated CpG islands) for gene bodies in dormant TCs and proliferative TCs. Red line indicates no differences in methylation, dotted red lines indicate thresholds for >10% differences in methylation.

15

Materials & Methods:

Animal studies

Female NOD-SCID and BALB/c mice were acquired from Janvier Labs. B6 *Cdh5*-CreERT2 x *Wls* floxed mice were published previously[51] and bred in barrier animal facilities of the German Cancer Research Centre. All animal work was performed in accordance with German national guidelines on animal welfare and the regulations of the regional council Karlsruhe under permit numbers G-164/16, G-107/18, DKFZ305 and DKFZ370. Mice were housed in sterile cages, maintained in a temperature-controlled room and fed autoclaved water and food *ad libitum*. All animals were monitored daily for signs of disease and ear punches were used for genotyping the mice. Imported mice were allowed to acclimatize for a minimum of seven days before each experiment. For all experiments, 8-12 weeks old mice were used and euthanized via rapid cervical dislocation of the spinal cord at the experimental endpoint.

For experimental metastasis, tumour cells were resuspended in 200 µl PBS and injected into the tail vein of mice. For transcriptomic and epigenomic screening experiments (**Fig. 1a, Fig. 3a**), female Balb/c mice were injected twice with $1x10^6$ tumour cells (4T1-GFP, niche-labelling 4T1-GFP, niche-labelling D2.0R-GFP) with a 30 minute break between injections. For pharmacological treatment studies, female Balb/c mice were injected once with $1x10^6$ 4T1-GFP cells. For genetic knockout experiments, male and female B6 *Cdh5*-CreERT2 x *Wls* floxed mice were injected with $2x10^5$ E0771-GFP or B16F10, respectively. To stain intravascular cells, 5 µg of fluorescently labelled anti-H2-Kd antibody in 50 µl PBS was injected retro-orbitally into mice 2 min prior to euthanasia.

Pharmacological systemic depletion of WNT was achieved by daily oral gavage of 10 mg/kg body weight LGK974 resuspended in 0.5% methylcellulose (Sigma-Aldrich), 0.5% TWEEN 80 (Sigma-Aldrich) in PBS as described previously[52].

EC-specific depletion of Wnt ligands was achieved using B6 *Cdh5*-CreERT2 x *Wls* floxed mice. Genetic recombination was initiated by intraperitoneal delivery of 2 mg tamoxifen (Sigma Aldrich) dissolved in 50 µl corn oil with 5% ethanol. Both Cre+ and Cre- littermates received 5 consecutive daily injections and were subjected to a one-week washout period before the start of the experiment.

To model spontaneous dissemination, $1x10^6$ 4T1-GFP cells in 100 µl PBS were injected into the inguinal mammary fat pad of female NOD-SCID mice. Tumour volumes were assessed by calliper measurements (tumour volume = ½ x length x width x width). Upon reaching tumour sizes of 100 mm³, mice were treated daily with LGK974 for 5 days with subsequent resection of the primary tumour. LGK974 treatment continued for 2 days post-resection and mice were left to develop metastases for 2 weeks.

To account for general inflammatory signatures in lung EC in the niche-labelling experiment (**Fig. 3b**, **c**), mice were injected with 1 mg/kg LPS (Sigma Aldrich) in 0.9% NaCl (Braun) intraperitonially 24 h prior to euthanasia.

After euthanizing the mice, lungs were collected in PBS, metastatic foci were counted (B16F10 experiments), lungs were imaged using a stereomicroscope (Leica) (E0771 experiments,

475 primary tumour experiments) and processed for flow cytometry. Resected primary tumours
476 were rinsed in PBS and fixed in formalin free Zn-buffer.

477 **Cell Culture**

478 4T1-GFP, D2.0R and E0771-GFP cells were a gift from the laboratories of Dr. Robert
479 Weinberg (Whitehead Institute, Cambridge, MA), Dr. Jonathan Sleeman (Heidelberg
480 University, Mannheim, Germany) and Dr. Kairbaan Hodivala-Dilke (Barts Cancer Institute,
481 London, England), respectively. B16F10 cells were purchased from ATCC. All cells were
482 maintained at 37°C and 5% CO_2 in high humidity and cultured in high glucose DMEM (Gibco)
483 supplemented with 10% (vol/vol) FCS and 100 U/mL penicillin/streptomycin (Sigma Aldrich).
484 D2.0R-GFP and 4T1-mCherry cells were generated by lentiviral transduction with TurboGFP
485 and mCherry reporter, respectively. Niche-labelling cells were generated by lentiviral
486 transduction of 4T1-GFP and D2.0R-GFP cells. Niche-labelling lentivirus was provided by the
487 laboratory of Dr. Ilaria Malanchi (Francis Crick Institute, London, England). Cells were
488 checked regularly for mycoplasma contamination by PCR and cell identity was confirmed by
489 cell morphology. Cells were subcultured upon reaching 80-90% confluency by trypsin-EDTA
490 (Sigma Aldrich) treatment.

491 For *in vitro* treatments, cell media were supplemented with 20 mM SKL2001 in DMSO
492 (Selleckchem), 3 mM in DMSO CHIR99021 (Selleckchem) or 1 µM in PBS decitabine (Sigma
493 Aldrich). Cells were treated overnight (~17 h) for pulse treatments (SKL2001, CHIR99021,
494 decitabine) or for 2 weeks with daily media changes for reprogramming (SKL2001,
495 CHIR99021) with drugs or vehicle (solvent only).

496 **Isolation of lung cells**

497 Isolated lungs were minced on ice using curved serrated scissors. The minced tissue was
498 resuspended in DMEM supplemented with Liberase Thermolysin Medium enzyme mix (0.2
499 mg/ml, Roche) and DNAse I (0.2mg/ml, Sigma Aldrich) and incubated at 37°C first for 15 min
500 and then again for 12 minutes. After each incubation, minced tissues were passed through 18G
501 cannula syringes 30 times. After the second incubation, digested tissues were passed through
502 100 µm cell strainer to remove tissue debris and cell clumps. The following steps were
503 performed on ice. The digestion reaction was quenched by adding FCS and samples were
504 centrifuged at 4°C and 400 *g* for 4 min. Erythrocytes were lysed by resuspending the cell pellet
505 in pre-chilled 1x ammonium chloride potassium (ACK) buffer. The reaction was quenched by
506 adding ice-cold PBS, followed by centrifugation.

507 **Flow cytometry analysis and FACS sorting**

508 Whole lung single cell suspensions were passed through a 40 µm cell strainer and preincubated
509 with anti-mouse CD16/CD32 Fc block (1:100, Thermo Fisher Scientific) for 15 min in flow
510 buffer (PBS supplemented with 5% (vol/vol) FCS) and, subsequently, with the appropriate
511 antibody-mix for 20 min on ice.

512 For cell sorting and flow cytometry analysis, dead cells were excluded by staining with
513 FxCycle[TM] Violet Stain (1:1000, Thermo Fisher Scientific) or Fixable Viability Dye
514 eFluor[TM] 780 (1:1000, Thermo Fisher Scientific) according to the manufacturer's instructions.
515 All samples were gated on viable cells followed by exclusion of cell doublets and CD45+,
516 LYVE1+, PDPN+ and TER119+ cells using the BD FACS Diva Software (BD Biosciences).

517   For flow cytometry, samples were recorded on the BD LSR Fortessa or BD FACSCanto II cell
518   analyser (both BD Biosciences) and flow data was analysed with FlowJo software (BD
519   Biosciences, v10). Tumour cells frequencies were calculated either as percentage of sample-
520   matched lung endothelial cells (relative tumour cell number) or as total tumour cell counts per
521   mg lung tissue using CountBrightTM Absolute Counting Beads according to the
522   manufacturer's protocol. Cells were sorted using a BD bioscience Aria cell sorting platform
523   (BD Biosciences) with 100 µm nozzle.

Single cell RNA-sequencing

525   ScRNAseq on tumour cell subpopulations and endothelial cells was performed using a modified
526   SMART-Seq2 protocol[53]. In brief, single cells were sorted directly into 96 well-plates
527   containing 1 µl of lysis buffer per well, centrifuged and snap frozen in liquid nitrogen. For
528   CD45- PDPN- LYVE- TER119- CD31+ EC, four plates (384 cells) were sorted for 3 biological
529   replicates from day 1.5 and day 3.5 (total of 1152 cells per timepoint). Day 0 control samples
530   were split and three plates (288 cells) were sorted for 2 biological replicates on day 1.5, as well
531   as on day 3.5 (total of 1152 cells), respectively, to account for technical batch-effects. For TC,
532   1 plate of matched intravascular and extravascular fractions (each 96 cells) was sorted from 4
533   biological replicates (total of 384 cells per fraction) on day 1.5. Similarly, 1 plate of matched
534   dormant and proliferative fractions was sorted from 4 biological replicates on day 3.5. Frozen
535   plates were thawed on ice and oligo-dT-primer were annealed at 70°C for 3 min. 1.3 µl of
536   reverse transcription mix with template-switching oligo was added to each well and isolated
537   mRNA was transcribed to full-length cDNA. Full-length cDNA was then amplified by adding
538   2.4 µl of PCR mastermix to each well. Due to their low RNA-content, EC cDNA was amplified
539   using 22 cycles, while TC cDNA was amplified with 18 cycles. Amplified cDNA was purified
540   using AMPure XP beads (Beckman Coulter) and random wells were selected for quality control
541   using 2100 Bioanalyzer (Agilent) and Qubit fluorometer (Thermo Fisher Scientific). DNA
542   concentration for each well was measured using Quant-iT^{TM} high sensitivity kit (Thermo Fisher
543   Scientific) and concentrations were manually adjusted to 0.1 - 0.3 ng/µl. Tagmentation was
544   performed using the Nextera XT DNA library preparation kit (Illumina) and a mosquito liquid
545   handler (SPT Labtech). For this purpose, 1.2 µl of Nextera XT – TD buffer mix was added to
546   each well of a 384 well plate with 0.4 µl of cDNA. For EC, 96 well plates from the individual
547   biological replicates were pooled into one 384 well plate, whereas TC replicates were pooled
548   according to sort gates. After tagmentation, customized i5 and i7 index primers were added,
549   resulting in unique labelling of each well in the 384 well plate and tagmented cDNA was
550   amplified using 14 PCR cycles. All uniquely labelled wells from each plate were pooled and
551   multiplexed libraries were purified and quality controlled using TapeStation (Agilent) and
552   Qubit fluorometer (Thermo Fisher Scientific), resulting in one multiplex per biological replicate
553   for EC and one multiplex per FACS-sorted TC fraction. Multiplexes were then sequenced on
554   individual lanes on a HiSeq2000 (Illumina) using V4 50 cycle single read kit generating
555   approximately 500.000 reads per cell.

Bulk RNA-sequencing of labelled niche EC

557   For the niche-labelling experiment 50.000 unlabelled lung EC and sample-matched total
558   labelled EC were directly sorted as described above into RNase-free 1.5 ml microcentrifuge
559   tubes containing 100 µl lysis buffer and immediately snap frozen on dry ice. For each condition

18

560 6 biological replicates were included. For control samples (LPS control and PBS control)
561 50.000 lung EC were sorted. Snap frozen RNA was extracted using Arcturus PicoPure RNA
562 Isolation Kit (Thermo Fisher Scientific) according to manufacturer instructions. RNA was
563 quality controlled by Qubit fluorometer (Thermo Fisher Scientific) and 2100 Bioanalyzer
564 (Agilent). Samples with RNA integrity number (RIN)-values below 8 were discarded. RNA
565 was transcribed to full-length cDNA using the SMART-Seq2[53] protocol and RNA-sequencing
566 libraries were generated using the NEBNext® Ultra™ II FS DNA library preparation kit (New
567 England Biolabs) according the manufacturer's protocol with DNA input below 100 ng.
568 Libraries were indexed using unique i5 and i7 combinations and equimolarly pooled into one
569 multiplex. The multiplex was sequenced over two lanes on a NovaSeq 6000 using the S1 100
570 cycle paired-end kit generating approximately 35×10^6 reads per sample.

Whole genome bisulfite sequencing

572 For whole genome bisulfite sequencing (WGBS) analysis of dormant and proliferative TC,
573 lungs of 6 mice were pooled into one sample and 200.000 proliferative TC as well as total
574 dormant TC were sorted from 4 pools. Sorted samples were centrifuged and cell pellets were
575 snap frozen on dry ice. Genomic DNA was extracted using the NucleoSpin tissue mini kit for
576 DNA from cells and tissue (Macherey-Nagel). DNA integrity was assessed by TapeStation
577 (Agilent) and samples with DNA integrity number (DIN)-values below 7 were discarded.
578 WGBS libraries were prepared using the xGen™ Methyl-Seq DNA Library Prep Kit (IDT)
579 with partially modified steps in bead clean-up/size selection. Briefly, 200ng genomic DNA was
580 fragmented to 700 – 1000 bp using Covaris ultrasonicator (Covaris , Inc.) and quality checked
581 using TapeStation (Agilent Technologies). Fragmented DNA samples were treated with
582 bisulfite using the EpiTect Bisulfite Kit (Qiagen) following the instructions in the Illumina
583 WGBS for Methylation Analysis Guide (Part # 15021861 Rev. B). After bisulfite conversion
584 adapters were attached to 3`ends of single-stranded DNA fragments which are then extended
585 to generate complementary uracil-free molecules as described in manufactures protocol. The
586 double-stranded DNA fragments were subsequently cleaned up using 1.6x AMPure XP beads
587 (Beckman Coulter) and size selected with a bead ratio of 0.6x and 0.2x, followed by ligation of
588 truncated adapter 2 to uracil-free strand. The adapter-ligated libraries were enriched and
589 indexed using 6 cycles of PCR and purified using magnetic beads according to the protocol.
590 Amplified libraries were quality checked using Qubit fluorometer (Thermo Fisher Scientific)
591 and TapeStation (Agilent). Libraries were pooled equimolarly into one multiplex and
592 sequenced over two lanes on a NovaSeq 6000 using the S1 150 cycle paired-end kit enabling
593 an average genomic coverage of >15.

Niche-labelling RNA-Sequencing analysis

595 Raw sequencing data were demultiplexed and FASTQ files were generated using bcl2fastq
596 software (Illumina, v2.20.0.422). FASTQ files were mapped to the GRCm38 mouse reference
597 genome using salmon (v0.7.2)[54] and count matrices were constructed with the R package
598 tximport (v1.18.0)[55]. Differential gene expression analysis was performed using DESeq2
599 (v.1.30.1)[56]. Each condition was tested against each condition and differentially expressed
600 genes were used for gene set enrichment analysis (GSEA). GSEA[57,58] was performed using the
601 R package clusterProfiler (v3.18.1)[59] or the GSEA java desktop application and the Molecular
602 Signatures Database (MSigDB, v7.4)[32] provided by the Broad Institute.

603 For the proliferative niche EC gene panel, DEG were filtered for genes that were specifically
604 upregulated in proliferative niche EC in at least three comparisons ($p < 0.01$ & log2 fold change
605 > 0.5) and were not regulated in non-proliferative niche EC comparisons. For deconvolution of
606 bulk samples aCap and gCap marker genes were defined using the scRNAseq dataset.
607 Expression coefficients (aCap/gCap) of summed marker genes were calculated for each bulk
608 sample using quasibinomal fitting and normalised to PBS injected control samples. Resulting
609 ratios were exponentiated for plotting.

610 **Single-cell RNA-seq analysis**

611 *Pre-processing and normalisation*

612 Raw sequencing data were processed as described above. Gene expression was normalised to
613 the mean expression of a housekeeping gene panel (*Actb*, *Gapdh*, *Tubb5*, *Ppia*, *Ywhaz*, *B2m*,
614 *Pgk1*, *Tbo*, *Arbp*, *Gusb*, *Hprt1*) for each cell, scaled with factor 10.000 and log10 normalised.
615 Normalised count matrices were further analysed using the R package Seurat (v4.0.1)[60,61]. Gene
616 counts per cell, read counts per cell and percentage of mitochondrial transcripts were computed
617 using the respective functions of the Seurat package. For the EC dataset, cells with a percentage
618 of mitochondrial transcripts greater than 5%, along with those with fewer than 1000 genes were
619 excluded. For the TC dataset, only cells with a mitochondrial transcript percentage less than
620 5% and more than 2500 genes were kept for further analysis.

621 *Dimension reduction and clustering*

622 Shared nearest neighbour (SNN)-based clustering and UMAP visualization were performed
623 using the FindClusters and RunUMAP functions within the Seurat package. Each of these were
624 performed on the basis of a principal component analysis, which was performed using the
625 RunPCA function of the Seurat package. For the EC dataset dimensional reduction was
626 performed on 10 principal components (PC) with the resolution parameter set to 0.2 for
627 clustering, whereas for the TC dataset 15 PCs and 0.5 resolution were used. Clusters were
628 annotated using congruent marker expression or according to enrichment for cells derived from
629 a specific FACS gate. Contaminating cells were removed from the dataset based on expression
630 of immune marker genes (*Ptprc*, *Itgam*, *Itgax*, *Adgre1*, *Cd3e*, *Cd19*, *Cd56*) or stromal cell and
631 vessel mural cell marker genes (*Pdgfrb*, *Des*, *Myh11*, *Col1a2*, *Pdgfra*, *Cspg4*, *Pdpn*, *Acta2*)

632 *Cell cycle scoring and differential gene expression*

633 The cell-cycle state was assessed using the gene set and scoring system described previously[62].
634 Briefly, the S and G_2M scores were calculated based on a list of 43 S phase-specific and 54
635 G_2 or M phase-specific genes. Cells that originated from the dormant FACS gate and had
636 summed scores of less than -1 were tested against cells from the proliferative gate that had score
637 sums greater than 0. Differentially expressed genes were calculated by Wilcoxon rank sum test
638 using FindMarkers-function in Seurat and used for GSEA as described above.

639 EMT and Wnt gene sets were compiled from MSigDB and gene scores for each cell were
640 calculated using the AddModuleScore-function in Seurat.

641 *Trajectory analysis*

642 Trajectory analysis of lung resident TC was performed using the R package Monocle (v.3
643 alpha)[63,64]. The filtered and normalised TC count matrix was subset from the TC Seurat object.
644 Clustering and dimension reduction was performed using default parameters in Monocle3. The

645 trajectory graph was built by setting cells from the intravascular sorting-gate as starting point.
646 Cells were coloured according to cluster identities as identified in Seurat. Gene expression of
647 EMT and Wnt gene sets was visualised using the plot_cells function.

Analysis of TC-EC interactions

649 Leveraging the biological replicates, TC pseudo-bulks were formed, for which all counts from
650 cells of a specific FACS gate and biological replicate were summed and differentially expressed
651 genes (DEG) were computed for the experimental timepoints using DESeq2[56]. DEG were
652 filtered against CellPhoneDB[27] database to retrieve putative ligands and receptors.
653 CellPhoneDB ligands and receptors were further filtered for expression of interaction partners
654 in the day 1.5 EC dataset (for intravascular versus extravascular comparison) or for expression
655 in the day 3.5 EC dataset (for dormant versus proliferative comparison), respectively. Log2 fold
656 changes of TC-expressed ligands or receptors were plotted against each. Receptors or ligands
657 with upregulation in extravasated and dormant TC were considered trajectory defining, as well
658 as receptors or ligands upregulated in intravascular and proliferative TC.

Analysis of publicly available human CTC datasets
660 Normalised and filtered count matrices were downloaded from provided source data[30] or the
661 gene expression omnibus (GEO) under the accession code GSE109761[31]. Dimension reduction
662 and visualization were performed in Seurat as described above using default parameters. Gene
663 scores of human orthologues of the EMT and Wnt gene lists were computed as described above.

WGBS data analysis

665 Raw sequencing data were demultiplexed and FASTQ files were generated using bcl2fastq
666 software (Illumina, v2.20.0.422). FASTQ files were trimmed using Trimgalore (v0.6.6)[65] and
667 mapped to the GRCm39 mouse reference genome using Bismark (v0.22.3)[66]. Forward and
668 reverse strands were collapsed and methylation sites were called in Bismark. Differentially
669 methylated regions were determined with the R package bsseq using default parameters
670 (v1.26.0)[67]. For this, biological replicates were summed and methylation fractions for annotated
671 genomic regions between the two conditions were compared. Regulatory elements, promoters
672 and gene bodies were annotated with annotation sheets downloaded from Ensembl database
673 (release 105)[68].

Real time quantitative PCR

675 Total RNA of cell cultured tumour cells was isolated using the GenElute Mammalian Total
676 RNA Purification Kit (Merck) according to the manufacturer's instructions. 1000 ng of RNA
677 was reverse transcribed using the QuantiTect Reverse Transcription Kit (Qiagen) according to
678 the manufacturer's protocol. Gene expression analysis was performed by quantitative PCR
679 using TaqMan reactions (Thermo Fisher Scientific) and Lightcycler 480 (Roche). Gene
680 expression levels were assessed using the C_t-method and normalised to the expression of *Actb*,
681 resulting in ΔC_t-values. Relative gene expression was assessed by normalizing ΔC_t-values of
682 individual samples to the average control ΔC_t-value, resulting in $\Delta \Delta C_t$-values. Relative fold
683 changes to control were then calculated as $2^{-\Delta \Delta Ct}$.

21

Histology

Zinc-fixed primary tumours were paraffin embedded and cut into 7 µm sections. Sections were deparaffinized and rehydrated and antigen retrieval was performed by incubation with Proteinase K (20 µg/ml, Gerbu Biotechnik) for 5 min at 37°C. Tissues were blocked in 10% ready-to-use goat serum (Zymed) for 1 hour at room temperature, followed by overnight incubation with rat anti-CD31 (1:100, BD Bioscience) and rabbit anti-Desmin (1:100, Abcam) diluted in blocking buffer at 4 °C. After three washes in TBS-T, slides were stained with anti-rat Alexa647 and anti-rabbit Alexa546 antibody at room temperature for 1 hr. Cell nuclei were counterstained with 1:2000 Hoechst 33342 (Sigma-Aldrich) and sections were mounted with DAKO mounting medium (Agilent). Images were acquired as whole-area tile scans using an Axio Scan.Z1 slide scanner (Zeiss). Image analysis was performed using Fiji software (ImageJ, 1.53q). After region-of-interest (ROI) selection, CD31, Desmin and DAPI channels were binarized using thresholding. For vessel area, percentage of CD31+ area within the ROI was calculated. For vessel coverage, CD31 channel was masked and Desmin overlap with CD31 was calculated. CD31+/Desmin+ double-positive vessel were considered covered and coverage was calculated as the ratio of covered/uncovered vessels.

Statistical Analysis

Statistical analysis was performed using GraphPad Prism (v6) and R (v4.0.5).

Method specific references:

51 Korn, C. et al. Endothelial cell-derived non-canonical Wnt ligands control vascular pruning in angiogenesis. *Development* (Cambridge, England) 141, 1757-1766, doi:10.1242/dev.104422 (2014).

52 Liu, J. et al. Targeting Wnt-driven cancer through the inhibition of Porcupine by LGK974. *Proc. Natl. Acad. Sci. USA* 110, 20224-20229, doi:10.1073/pnas.1314239110 (2013).

53 Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc*. 9, 171-181, doi:10.1038/nprot.2014.006 (2014).

54 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417-419, doi:10.1038/nmeth.4197 (2017).

55 Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000research* 4, 1521, doi:10.12688/f1000research. 7563.2 (2015).

56 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15, 550, doi:10.1186/s13059-014-0550-8 (2014).

57 Mootha, V. K. et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet*. 34, 267-273, doi:10.1038/ng1180 (2003).

58 Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545-15550, doi:doi:10.1073/ pnas.0506580102 (2005).

59 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284-287, doi:10.1089/omi.2011.0118 (2012).

60 Stuart, T. et al. Comprehensive integration of single-cell Data. *Cell* 177, 1888-1902.e1821, doi:https://doi.org/10.1016/j.cell.2019.05.031 (2019).

728   61   Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* 184, 3573-3587.e3529,
729   doi:10.1016/j.cell.2021.04.048 (2021)

730   62   Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-
731   seq. *Science* (New York, N.Y.) 352, 189-196, doi:doi:10.1126/science.aad0501 (2016).

732   63   Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudo-
733   temporal ordering of single cells. *Nat. Biotech.* 32, 381-386, doi:10.1038/nbt.2859 (2014).

734   64   Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods*
735   14, 979-982, doi:10.1038/nmeth.4402 (2017).

736   65   Krueger F, Trimgalore (2021), *GitHub repository*, https://github.com/FelixKrueger/TrimGalore,
737   doi: 10.5281/zenodo.5127899

738   66   Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq
739   applications. *Bioinformatics* 27, 1571-1572, doi:10.1093/bioinformatics/btr167 (2011).

740   67   Hansen, K. D., Langmead, B. & Irizarry, R. A. BSmooth: from whole genome bisulfite sequencing
741   reads to differentially methylated regions. *Genome Biol.* 13, R83, doi:10.1186/gb-2012-13-10-r83
742   (2012).

743   68   Cunningham, F. et al. Ensembl 2022. *Nucleic Acids Res*. 50, D988-D995, doi:10.1093/nar/
744   gkab1049 (2021).