ARTICLE

doi:10.1038/nature20785

Early dissemination seeds metastasis in breast cancer

Hedayatollah Hosseini[1], Milan M. S. Obradović[1]†, Martin Hoffmann[2], Kathryn L. Harper[3], Maria Soledad Sosa[3]†, Melanie Werner-Klein[4], Lahiri Kanth Nanduri[1]†, Christian Werno[2], Carolin Ehrl[1], Matthias Maneck[1], Nina Patwary[1], Gundula Haunschild[1], Miodrag Gužvić[1], Christian Reimelt[1], Michael Grauvogl[5], Norbert Eichner[6], Florian Weber[7], Andreas D. Hartkopf[8], Florin-Andrei Taran[8], Sara Y. Brucker[8], Tanja Fehm[9], Brigitte Rack[10], Stefan Buchholz[11], Rainer Spang[5], Gunter Meister[6], Julio A. Aguirre-Ghiso[3] & Christoph A. Klein[1,2]

Accumulating data suggest that metastatic dissemination often occurs early during tumour formation, but the mechanisms of early metastatic spread have not yet been addressed. Here, by studying metastasis in a HER2-driven mouse breast cancer model, we show that progesterone-induced signalling triggers migration of cancer cells from early lesions shortly after HER2 activation, but promotes proliferation in advanced primary tumour cells. The switch from migration to proliferation was regulated by increased HER2 expression and tumour-cell density involving microRNA-mediated progesterone receptor downregulation, and was reversible. Cells from early, low-density lesions displayed more stemness features, migrated more and founded more metastases than cells from dense, advanced tumours. Notably, we found that at least 80% of metastases were derived from early disseminated cancer cells. Karyotypic and phenotypic analysis of human disseminated cancer cells and primary tumours corroborated the relevance of these findings for human metastatic dissemination.

Systemic cancer (the dissemination and subsequent distant outgrowth of cells from a solid tumour) occurs in two phases: a clinically latent stage of hidden cancer spread, and then manifest metastasis. Manifest metastasis remains mostly incurable. The period of clinically undetectable minimal residual disease, defined by disseminated cancer cells (DCCs) that are left behind after primary tumour surgery, offers a time window to prevent metastasis[1,2]. However, only circumstantial knowledge is available about minimal residual disease, and consequently, systemic (adjuvant) therapies improve outcome in only about 20% of patients[3,4]. This situation indicates that our current understanding of early systemic cancer is insufficient to prevent metastasis.

The first direct evidence for a characteristic biology of early disseminated cancer and minimal residual disease came from analyses of DCCs isolated from bone marrow of patients with breast cancer before (M0 stage, according to Union for International Cancer Control guidelines) and after (M1 stage) manifestation of metastasis[5,6], indicating that M0-DCCs might have disseminated early and evolved in parallel with the primary tumour[7]. Studies in transgenic mouse models[8–10] and in patients with pre-malignant lesions or *in situ* carcinomas[8,11,12] corroborated this concept but the relevance of DCCs remains contested[13].

We therefore addressed the issue of breast-cancer-cell dissemination soon after cancer initiation and investigated whether mechanisms exist that reduce metastatic seeding from advanced cancers. Finally, we investigated whether early DCCs (eDCCs) are able to form metastases. We report on a mechanism involving cell density, HER2 and PGR signalling that reconciles early and late dissemination models.

PGR and HER2 regulate gene expression in early lesions

In BALB-NeuT mice, dissemination starts shortly after expression of the *Her2* transgene (also known as *Erbb2*) at puberty (around 4 weeks of age), when the first hyperplastic lesions become apparent[8]. From 4–9 weeks of age, we observed micro-invasion[8], and a sharp decline in the ratio of DCCs to total tumour area (a measure of cell numbers at risk of dissemination) during primary tumour growth (Extended Data Fig. 1a). The genetic program governing dissemination from early lesions in microdissected tissue samples (Extended Data Fig. 1b and Supplementary Table 1) showed a signature gene expression profile compared to healthy mammary glands, primary tumours and lung metastases (Fig. 1a). We defined 1,278 gene transcripts unique to early lesions of which 300 were highly conserved between mouse and human (Supplementary Data 1).

We confirmed differential expression of selected transcripts by quantitative PCR (qPCR) (Extended Data Fig. 1c) and analysed transcript expression of steroid hormone receptors (strong candidate regulators; Supplementary Tables 2, 3), all of which, except *Esr1* (also known as *ERalpha*), showed the highest expression in early lesions (Extended Data Fig. 1d). When we assessed the expression of *Ahnak*, *Baz2a*, *Nfatc3*, *Nr3c1* and *Nr3c2* genes, which were used as surrogate markers of the early lesion signature (Fig. 1b, Supplementary Tables 2, 3 and Supplementary Data 1), progesterone was the only steroid hormone that activated a similar expression profile (Extended Data Fig. 1e), but only in early lesion cells of 9-week-old BALB-NeuT mice, and not in wild-type mammary or primary tumour cells (Fig. 1c). Moreover, the expression of progesterone receptor B (PGR-B), which is the main isoform expressed for mammary gland development[14–16], correlated

[1]Experimental Medicine and Therapy Research, University of Regensburg, 93053 Regensburg, Germany. [2]Project group 'Personalized Tumour Therapy', Fraunhofer Institute for Toxicology und Experimental Medicine, 93053 Regensburg, Germany. [3]Division of Hematology and Oncology, Department of Medicine, Department of Otolaryngology, Department of Oncological Sciences, Tisch Cancer Institute, Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York 10029, USA. [4]Institute of Immunology, University of Regensburg, 93053 Regensburg, Germany. [5]Department of Statistical Bioinformatics, Institute of Functional Genomics, University of Regensburg, 93053 Regensburg, Germany. [6]Biochemistry Center Regensburg (BZR), Laboratory for RNA Biology, University of Regensburg, 93053 Regensburg, Germany. [7]Institute of Pathology, University of Regensburg, 93053 Regensburg, Germany. [8]Department of Gynecology and Obstetrics, University of Tübingen, 72076 Tübingen, Germany. [9]Department of Gynecology and Obstetrics, University of Düsseldorf, 40225 Düsseldorf, Germany. [10]Department of Gynecology and Obstetrics, University Munich, 80337 Munich, Germany. [11]Department of Gynecology and Obstetrics, University Medical Center Regensburg, 93053 Regensburg, Germany.
†Present addresses: Tumor Heterogeneity, Metastasis and Resistance, Department of Biomedicine, University of Basel, University Hospital Basel, CH-4031 Basel, Switzerland (M.M.S.O.); Department of Pharmacological Sciences, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York 10029, USA (M.S.S.); Department of Gastrointestinal, Thoracic and Vascular Surgery, Medizinische Fakultät Carl Gustav Carus, Technische Universität Dresden, 01307 Dresden, Germany (L.K.N.).

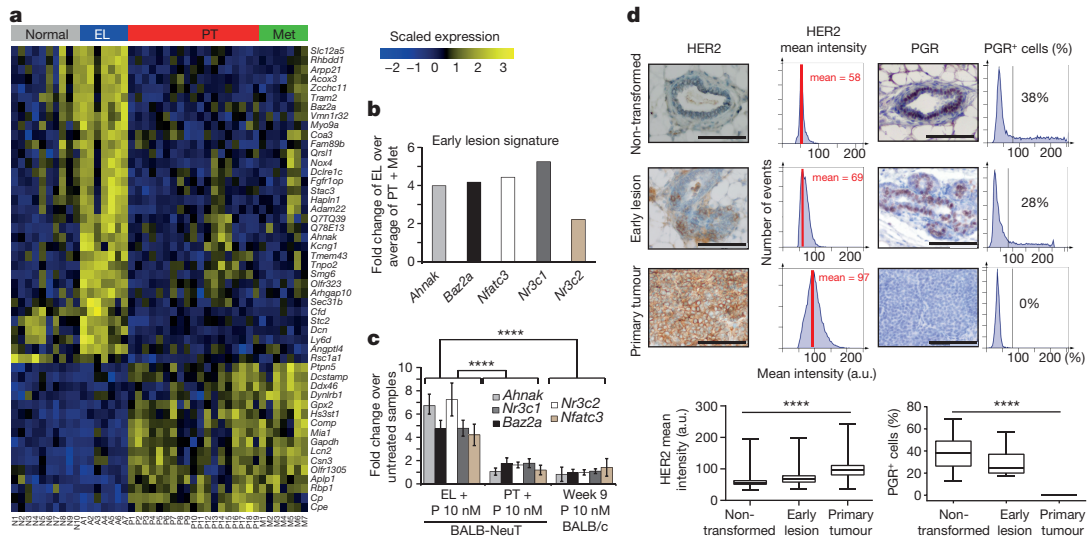


Figure 1 | Identification of a gene expression signature linked to early dissemination. a, Heatmap of genes that were differentially expressed in different sample types: normal mammary glands from BALB/c (N1–N10), early lesions (EL, A1–A7), primary tumours (PT, P1–P19) and lung metastases (Met, M1–M7) from BALB-NeuT mice. Yellow, upregulation; blue, downregulation. **b**, Five-gene surrogate signature (qPCR) for the early lesion profile. **c**, Progesterone (P) activates the early lesion signature *in vitro*. Data are mean ± s.d. **d**, TissueFAX cytometric quantification of HER2 and PGR protein expression. Images, representative staining of HER2 (left) and PGR (right). Scale bars, 100 μm. Mean HER2 staining intensity (red line, left histograms) in arbitrary units (a.u.) and percentage of PGR^+ cells (right histogram) and box plots. Boxes show lower quartile, median and upper quartile and whiskers indicate minimum and maximum; $^{****}P \leq 0.0001$ (*t*-test and Stouffer's combined probability test (**c**) and one-way ANOVA(**d**)).

with low to moderate HER2 expression in early lesions and wild-type mammary tissue (Extended Data Fig. 1f). We therefore quantified HER2 and PGR protein expression in mammary cells displaying normal, hyperplastic or advanced tumour morphology. The number and staining intensity of $HER2^+$ cells steadily increased with advancing tumour morphology, in which early lesions had intermediate cell numbers and HER2 expression levels. PGR staining intensity was constant, but the number of PGR^+ cells declined from approximately 40% in normal glands to 0% in primary tumours (Fig. 1d).

The early lesion signature could be activated in 4T1, an MMTV–HER2-negative mouse tumorigenic cell line (Extended Data Fig. 1g) that displays weak endogenous HER2 expression (Extended Data Fig. 1h). Furthermore, HER2-negative mouse tumour cells (67NR and MM3MG cells) did not express the early lesion signature (Extended Data Fig. 1g), but progesterone treatment or transduction of the B isoform of PGR (*Pgr-B*) induced upregulation of HER2 in 4T1 or MM3MG cells, respectively (Extended Data Fig. 1h-i). Collectively, these results suggested that the genetic program of early lesions depends on the combined activation of progesterone and HER2 pathways.

Progesterone induces migration of early lesion cells

Because progesterone mediates branching[14] in mammary gland development, we investigated the role of the progesterone-induced early lesion signature in cancer cell migration. We found that the mRNA levels of the progesterone-induced paracrine signals (PIPS) *Rankl* (also known as *Tnfsf11*) and *Wnt4* were upregulated in early lesion samples (Extended Data Fig. 2a). Treatment of early-lesion-derived cells with PIPS mimicked the effect of progesterone (Extended Data Fig. 2b), suggesting that early lesions exploit the mechanisms of mammary branching for metastasis. Consistent with this, PGR^+ cells were enriched in distal ducts of normal mammary glands (advancing the branching tree away from the nipple during developmental fat pad invasion) compared to proximal ducts closer to the origin (more differentiated ducts; Extended Data Fig. 2c, d).

Furthermore, progesterone and PIPS induced migration of mammary cells from early-lesion-derived samples (freshly prepared or early-lesion-derived mammospheres) and suppressed migration in cells from primary tumours (Fig. 2a and Extended Data Fig. 2e, f).

PIPS also activate mammary stem cells during mammary gland development[17], prompting us to analyse mammosphere formation[18]. Consistent with previous reports on HER2-stimulated stemness[19,20] BALB-NeuT-derived samples generated significantly higher numbers of spheres than controls (Fig. 2b). Notably, normal mammary cells derived from young (4–9-week-old) mice generated more spheres than cells derived from older mice (Fig. 2b). Furthermore, early-lesion-derived cells generated higher sphere numbers in response to progesterone and PIPS than primary tumour samples ($P \leq 0.01$; Fig. 2c). Neutralizing antibodies against RANKL and a WNT inhibitor abrogated the effect of progesterone on early lesion cell migration and sphere formation (Fig. 2d).

Migrating early lesion cells (that is, those arriving on the other side of the transwell membrane) formed increased sphere numbers in response to progesterone (Fig. 2e and Extended Data Fig. 2g). Oestrogen also induced migration and sphere formation, however, the progesterone inhibitor RU486 inhibited this pro-migratory and pro-sphere-forming effect (Extended Data Fig. 2h), possibly because oestrogen acts via transcriptional induction of *Pgr*[21]. Together, these results suggested that moderate HER2 expression with progesterone or PIPS availability promote sphere-formation and migratory responses in mammary epithelial cells.

HER2 expression levels determine cellular responses

We next performed a series of mechanistic *in vitro* experiments using mouse mammary epithelial MM3MG cells (oestrogen receptor (ERα)-negative but expressing low levels of PGR and HER2). MM3MG cells transduced with *Pgr-B* or *Her2* were subjected to sphere-formation and migration assays (Extended Data Fig. 3). PIPS-responsive cells (in both migration and sphere-formation assays) were $HER2^{low}/PGR^-$, whereas PGR^+ cells themselves were not migrating and $HER2^{high}$ cells were similar to non-migrating primary tumour cells and showed enhanced proliferation.

Cell density regulates HER2 and PGR expression

Notably, individual primary-tumour-derived cells re-expressed *Pgr* mRNA and protein when cultured at low cellular densities (Extended Data Fig. 4a, b). We therefore plated the BALB-NeuT primary-tumour-derived TUBO cell line at different cell densities and found PGR

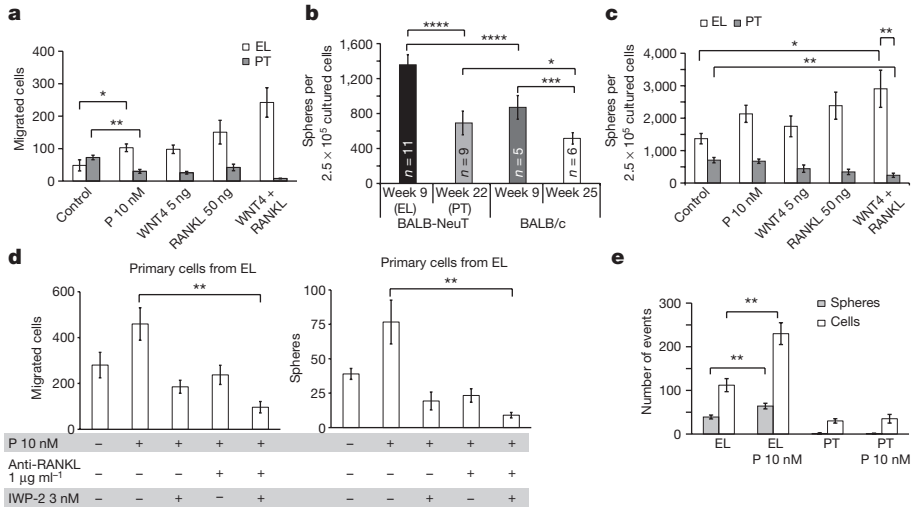


Figure 2 | Progesterone induces migration and sphere formation of early lesion cells. a, Early lesion and primary tumour cells respond to progesterone or PIPS (WNT4, RANKL) with increased (early lesions) or decreased (primary tumours) migration. **b**, Mammosphere formation depends on age and HER2 expression. **c**, Cells from early lesions and primary tumours respond to progesterone or PIPS with increased (early lesions) or decreased (primary tumours) sphere formation. **d**, Depletion of PIPS by IWP-2 (WNT inhibitor) or anti-RANKL (neutralizing antibody) reduces migration (left) and sphere formation (right) of early lesion cells. **e**, Migrating cells activated by PIPS form spheres (see also Extended Data Fig. 2g). Data are mean ± s.d.; $*P \leq 0.05$; $**P \leq 0.01$; $***P \leq 0.001$; $****P \leq 0.0001$ (Student's *t*-test).

expression in 10 ± 5% (mean ± s.d.) of cells grown at low density, but undetectable PGR expression in cells grown at high density (Fig. 3a). Several experiments suggested that there was a soluble factor with PGR-suppressing activity in the vesicular fraction of cell culture supernatants, and that this activity was conserved between mouse and human (Extended Data Fig. 4b–e). To identify this activity, we analysed the microRNA (miRNA) profiles of supernatants with and without exosomes from TUBO cells and the HER2-overexpressing cell line MM3MG–Her2. miRNA sequencing (Supplementary Data 2) and bioinformatic prediction of *Pgr* regulators (Extended Data Fig. 4f)

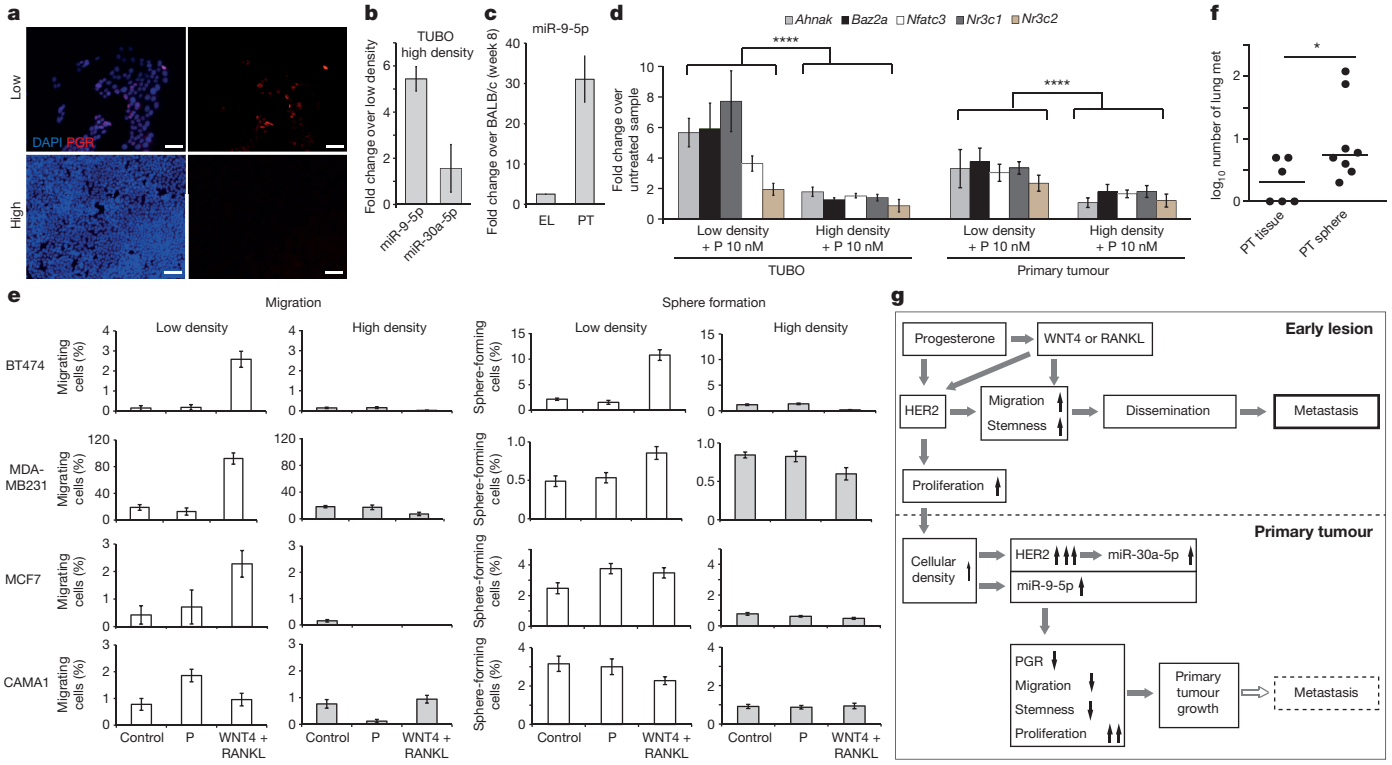


Figure 3 | Cell density regulates PGR expression and early lesion phenotype. a, TUBO cells re-express PGR at low cell density. Scale bars, 50 μm. **b**, TUBO cells grown at high density upregulate miR-9-5p. **c**, Expression of miR-9-5p in early lesion and primary tumour samples. **d**, Primary tumour and TUBO cells generate the early lesion signature only when grown at low density. **e**, Migration and sphere-formation of four human cell lines grown at low and high densities and treated with PIPS (WNT4 and RANKL) or progesterone (P) (see also Extended Data Fig. 5). **f**, Number of lung macro-metastases (17 weeks after tumour resection) after tumour formation from transplanted tumour pieces (1 mm^3; high density) or 50 spheres in 40 μl Matrigel (low density) and primary tumour surgery (shown are median and individual values). **g**, Mechanisms of local tumour and distant metastasis formation as derived from *in vitro* and *in vivo* (see Fig. 4) data. $*P \leq 0.05$; $****P \leq 0.0001$ (*t*-test and Stouffer's combined probability test (**d**); mean ± s.d. (**b**–**e**); Mann–Whitney *U*-test (**f**)).

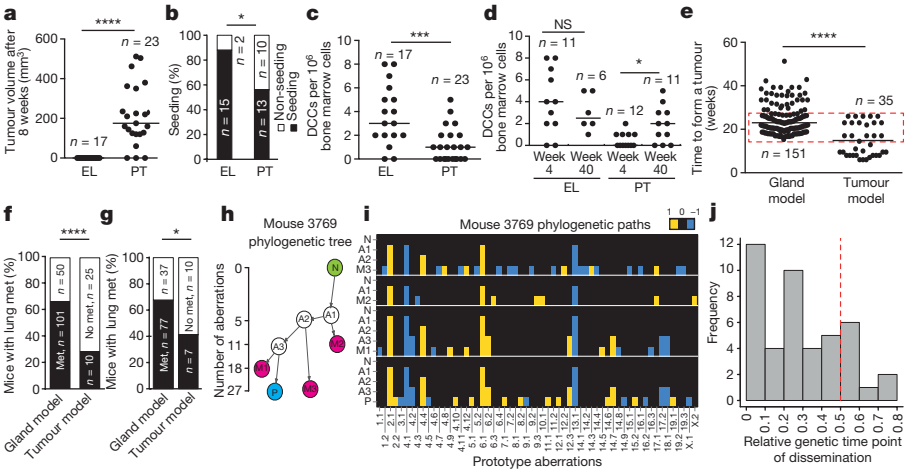


Figure 4 | Progesterone signalling regulates tumour formation and dissemination *in vivo*. **a**, Tumour formation 8 weeks after transplantation of spheres derived from primary tumours or early lesions into mammary fat pads of wild-type BALB/c siblings. **b**, Percentage of mice with DCCs (detected by cytokeratin staining) in bone marrow 8 weeks after transplantation. **c**, Number of DCCs in bone marrow of mice 8 weeks after transplantation. **d**, DCC counts in bone marrow in recipients transplanted at different ages. **e**, Time to tumour formation after transplantation of mammary glands (from 4-week-old mice; gland model) or tumour pieces (from 20–22-week-old mice; tumour model). The red box highlights mice from both models with similar tumour growth kinetics, which are analysed separately in **g** and Extended Data Fig. 6o–q. **f**, Percentage of mice with lung macro-metastasis from gland or primary tumour models. **g,** Macro-metastasis formation in recipient mice with similar tumour growth kinetics (mice from the red box in **e**; see Extended Data Fig. 6o). **h**, Example of a phylogenetic tree (mouse 3769). A1–A3, inferred common ancestors; M1–M3, metastases 1–3; N, normal cell; P, primary tumour. The ordinate indicates the number of aberrations per profile on a square root scale. **i**, Aberration profiles along tree paths from N via A1–A3 to P or M1–M3 in terms of aberration prototypes (see Extended Data Figs 7–9). **j**, Distribution of relative time points of dissemination on a genetic scale for all 44 primary tumour–metastases pairs. The red line indicates dissemination after which 50% of primary tumour changes were acquired as an arbitrary threshold for early versus late dissemination; see Extended Data Fig. 7d. $^{*}P \leq 0.05$; $^{***}P \leq 0.001$; $^{****}P \leq 0.0001$; NS, not significant (Student's *t*-test (**a**, **c**, **d**); Fisher's exact test (**b**); χ^2 test (**f**, **g**); Mann–Whitney *U*-test (**e**)). Lines in **a**, **c**, **d** and **e** denote the median.

identified miR-30a-5p and miR-9-5p as abundantly expressed in TUBO cells and able to downregulate *Pgr* in T47D cells (Extended Data Fig. 4g, h). miR-30a-5p expression was upregulated by *Her2* (Extended Data Fig. 4f, i) and miR-9-5p expression was sensitive to cell density (Fig. 3b, c).

Low cell density induced early-lesion-like features (such as induction of the early lesion signature, and reduced HER2 expression, migration and sphere formation) in TUBO or primary-tumour-derived cells from the BALB-NeuT model (Fig. 3d and Extended Data Fig. 4j–m). Notably, several human breast cancer cell lines shared density regulation of HER2- and PGR-regulating miRNAs (Extended Data Fig. 5a–d). Moreover, we observed density regulation of migration and sphere formation in 10 out of 10 human breast cancer cell lines tested (7 out of 10 by PIPS and 3 out of 10 by progesterone induction; Fig. 3e and Extended Data Fig. 5e).

To assess the contribution of cell density *in vivo*, we compared metastasis formation of transplanted primary tumour pieces (very high cell density) with primary tumourspheres (50 spheres injected in 40 µl Matrigel, that is, very low density). After primary tumour formation, we performed curative surgery and evaluated metastasis formation. No difference was found in the percentage of mice with metastases (8 out of 19 for tumour pieces compared with 6 out of 16 for spheres), however, animals transplanted with primary tumourspheres had a higher number of metastases ($P \leq 0.05$; Fig. 3f). In summary, we obtained support for a model of metastatic dissemination regulated by cell density, HER2 expression and progesterone signalling (Fig. 3g).

PGR signalling regulates dissemination *in vivo*

To validate these findings, we analysed physiological conditions of reduced (higher age; Extended Data Figs 1f, 2a, 6a, b) and increased (pregnancy) PGR signalling. We transplanted spheres generated from early lesions and primary tumour samples into young (4 weeks of age; PGR-rich) and old (40 weeks of age; PGR-reduced) wild-type recipients. All mice were killed when the first mice developed tumours of 5–10 mm diameter (8 weeks later). Most mice transplanted with primary tumours had palpable tumours, but none of the mice transplanted with early lesions (Fig. 4a), although these nevertheless harboured microscopic early lesions and/or ductal carcinoma *in situ* (DCIS) (Extended Data Fig. 6c, d). Notably, transplantation of spheres from early lesions increased the number of animals containing DCCs, and resulted in higher numbers of DCCs in bone marrow compared to transplantation of primary tumourspheres (Fig. 4b, c).

There was suppressed dissemination and stimulated tumour formation from primary tumourspheres in the PGR-rich microenvironment of young mammary glands (Fig. 4d and Extended Data Fig. 6e) as expected. Dissemination from early lesions was not reduced in old recipients (Fig. 4d). However, the transplanted early lesions generated a PGR-rich microenvironment in old recipients (Extended Data Fig. 6d), consistent with the observation that cells from early lesions, but not from primary tumours, could generate PGR-positive cells in 3D culture (Extended Data Fig. 6f, g). Co-transplantation of BALB-NeuT primary tumourspheres with MM3MG–Pgr-B into 40-week-old mice resulted in reduction of dissemination and stimulation of tumour formation (Extended Data Fig. 6h, i).

Because progesterone levels are physiologically increased 10- to 20-fold during pregnancy, we mated female transgenic mice at the age when early lesions (week 7) and early tumour formation (week 15) occurred. Mice were killed at term, and those mated at the early lesion age displayed higher numbers of DCCs (Extended Data Fig. 6j); whereas those mated at week 15 formed large tumours within 3 weeks, faster than unmated controls (Extended Data Fig. 6k).

To assess metastasis from early lesions and primary tumour lesions, we transplanted pieces of mammary glands (gland model) of 4–5-week-old transgenic mice or from primary tumours (primary tumour model) into the cleared mammary fat pad of 4-week-old wild-type recipients (Extended Data Fig. 6l, m). Tumour growth to 5–10 mm was fast in animals transplanted with primary tumour pieces (indicating their high viability) and took longer in animals in the gland model (Fig. 4e). After surgical removal of primary tumours (Extended Data Fig. 6l, m), more mice displayed metastasis in the gland model (Fig. 4f; $P < 0.0001$),

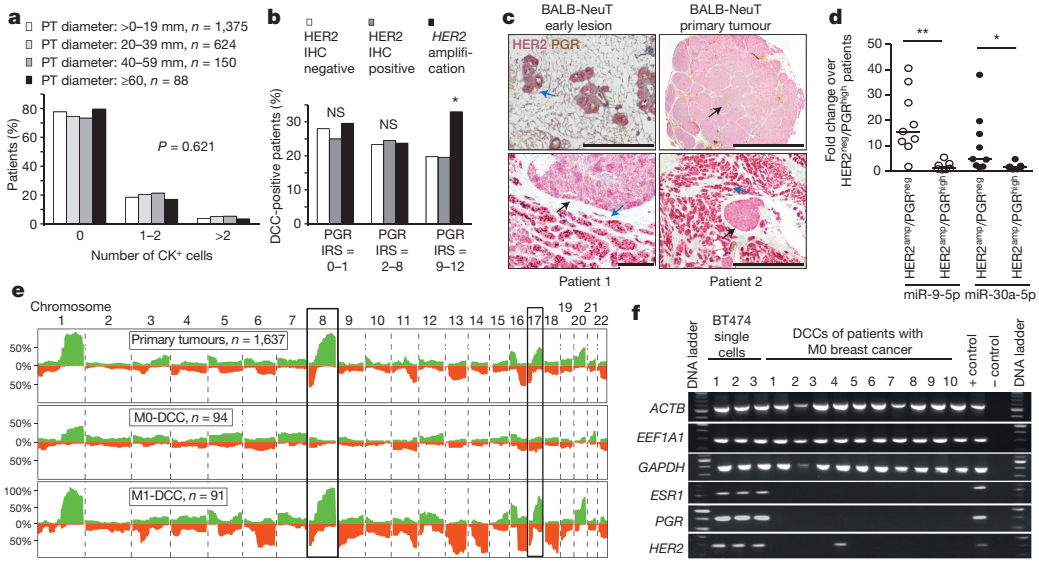


Figure 5 | PGR and HER2 signalling, and dissemination in patients with breast cancer. a, An increase in tumour diameter is not accompanied by an increase in DCCs in bone marrow. **b**, PGR and HER2 expression identifies a HER2^{high}/PGR^{high} subgroup of patients with the highest seeding rates. **c**, Comparison of HER2^{high}/PGR^{high} human breast cancers and primary BALB-NeuT mouse tumours for HER2/PGR staining. Top, representative images of an early lesion (left) and primary tumour (right) from the BALB-NeuT model. Bottom, representative images of two patients. High-density regions (strong HER2 and low PGR expression, black arrows) and regions of invasive cells (strong PGR and HER2 expression, blue arrows) are shown. Scale bars, 1 mm (top, bottom right) and 200 μm (bottom left). **d**, *PGR*-downregulating miRNAs are repressed in HER2^{high}/PGR^{high} human mammary carcinomas. **e**, Copy-number alterations in human primary breast cancers (from Progenetix database) and DCCs isolated from bone marrow of patients with breast cancer with and without metastasis (M0, $n = 94$; M1, $n = 91$). The *y* axis depicts the percentage of samples with aberrations (green, gain; red, loss) for each chromosomal region. **f**, Oestrogen receptor (*ESR1*) and progesterone receptor (*PGR*) transcript expression in human breast cancer DCCs (10 out of 26 DCCs from 19 M0 patients are shown; see Supplementary Table 8). *ACTB*, *EEF1A1* and *GAPDH* denote controls for sample quality. BT474 single cells are a positive control. $^*P \leq 0.05$; $^{**}P \leq 0.01$; NS, not significant (χ^2 test (**a** and **b**); Mann–Whitney *U*-test (**d**)).

although numbers of metastatic foci were similar in both cases (Extended Data Fig. 6n). Since the slower growth of early lesions until surgery might have extended the available time for metastasis formation, assuming dissemination occurred early, we restricted our analysis to mice with similar kinetics of primary tumour growth (red box in Fig. 4e and Extended Data Fig. 6o). Again, more mice developed lung metastases in the gland model (Fig. 4g), although in the analysis of this subgroup, follow-up time after tumour surgery was significantly longer for mice from the primary tumour model (Extended Data Fig. 6p). The number of metastases was similar (Extended Data Fig. 6q). Together, the *in vivo* results are in line with *in vitro* findings that PIPS induce migration and stemness of early lesion cells and proliferation of advanced primary tumour cells.

We used the gland model to determine which metastases were derived from early rather than late DCCs and performed phylogenetic analyses of 28 primary tumours with 1 or more lung metastases in the same mouse. Phylogenetic trees were generated from copy-number alterations, since BALB-NeuT tumours rarely display point mutations, similar to human breast cancer[22]. Total numbers of copy-number alterations were indistinguishable between primary tumours and metastases (Extended Data Fig. 7a) and no individual change was significantly associated with primary tumour or metastasis origin (Extended Data Fig. 7b). This suggested that aberrations shared between primary tumours and metastases were acquired earlier, indicating the time of genetic divergence (Fig. 4h, i). In all cases, we observed branching evolution with none or one (single metastasis) or several (multiple metastases; Fig. 4i and Extended Data Figs 7–9) ancestors. To assign derivation of primary tumours and metastases as from early or late DCCs, we assessed the proportion of primary tumour alterations that were already present in the last common ancestor. In linear progression, this value would be 1 (the primary tumour had acquired all mutations before the metastatic precursor left the site). Here, we used the very conservative threshold of ≥0.5 for late dissemination. Notably, we found that for 35 out of 44 individual primary tumour–metastasis pairs (79.5%), lung metastases were derived from eDCCs that disseminated before the primary tumour had acquired 50% of its alterations (Fig. 4j).

HER2 and PGR cooperation in human metastasis

The mechanisms of breast cancer dissemination described above cannot be studied directly in patients, as the event occurs before diagnosis. We therefore checked whether human breast cancers could also seed relatively fewer bone-marrow DCCs (detected by cytokeratin staining[23]) with growing primary tumour size, as seen in the BALB-NeuT model (Extended Data Fig. 1a). Indeed, the percentage of DCC-positive patients or DCC numbers in bone marrow of 2,239 patients with breast cancer did not increase with tumour diameter (Fig. 5a). We then investigated whether DCC numbers were associated with HER2 and PGR expression and categorized primary tumours according to their expression levels of PGR and HER2 (Supplementary Table 4). Notably, in breast cancers with a high PGR score, genetic activation of *HER2* increased the dissemination rate ($P \leq 0.05$; Fig. 5b), akin to early lesions in the BALB-NeuT model. This subgroup of patients (HER2^{amp}/PGR^{high}) comprised 3.7% of all patients (85 out of 2,239) or 24.6% of patients with *HER2*-amplified tumours (85 out of 345). HER2^{amp}/PGR^{high} (Supplementary Table 5) primary tumours contained HER2 and PGR single- and double-positive and double-negative cells (Extended Data Fig. 10a). Areas of high cell density lacked PGR expression and invasive regions of lower density contained strongly double-positive cells (Fig. 5c), suggesting density-mediated *PGR* regulation within the same samples. We therefore analysed the expression of *PGR*-regulatory miRNAs identified in the BALB-NeuT mouse model (Supplementary Table 6). HER2^{amp}/PGR^{high} tumours displayed lower levels of *PGR*-downregulating miRNAs compared to HER2^{amp}/PGR^{neg} tumours (Fig. 5d), similar to the human HER2^{high}/PGR^{high} cell lines BT474 and T47D (Extended Data Fig. 5a, b). PGR-negative, high-density regions from HER2^{amp}/PGR^{high} samples (Fig. 5c) displayed strong *PGR*-regulating miRNA overexpression (Extended Data Fig. 10b).

DCCs from patients with breast cancer without metastasis (stage M0; $n = 94$ cells; Supplementary Table 7) and with metastasis (stage M1; $n = 91$ cells) were compared to bulk primary breast cancers ($n = 1{,}637$) from a large database (Progenetix database; http://www.progenetix.net). Comparative genome hybridization profiles of primary tumours were very similar to those of single DCCs isolated from M1-stage patients (Fig. 5e). By contrast, single DCCs from M0-stage patients, although these displayed clearly aberrant profiles, lacked the chromosomal gains and losses characteristic of primary tumours, such as 8p loss or 8q gain (Fig. 5e). Thus we can conclude that: (i) DCCs often disseminate before the acquisition of typical breast cancer copy-number alterations; (ii) tumour cells displaying the typical karyotype of established tumours are rarely found at M0-stage of disease in the bone marrow; (iii) DCCs displaying M1-like genomes must replace DCCs with M0-like genomes to generate metastatic disease, even when the primary tumour was surgically removed.

Finally, we tested whether human DCCs ($n = 26$) isolated from the bone marrow of 18 patients with luminal and 1 patient with triple-negative breast cancer (Supplementary Table 8) lacked PGR expression as predicted by the BALB-NeuT model (Extended Data Fig. 3o). We identified and profiled DCCs by combined genome and transcriptome analyses of single cells[24]. None of the 26 DCCs displaying genomic aberrations expressed *PGR* (Fig. 5f). Notably, the only DCC expressing *HER2* transcripts originated from one of two patients diagnosed with hormone receptor positive DCIS only.

Discussion

This work provides evidence that mouse and human mammary cancer cells migrate and disseminate from morphologically very early lesions. The mechanism, which is shut down as primary tumours grow and overt lesions develop, consists of three major components: cell density, PGR signalling and HER2 signalling. While the specific molecular details are more likely to be tissue dependent than universal, our proposed mechanism may provide a general framework for the understanding of metastasis with cancer cells undergoing a switch from a dissemination to proliferation mode (Fig. 3g).

Our findings challenge the concept that late-disseminating (that is, shortly before surgery), fully mature cancer cells necessarily have a higher ability to form metastases[13]. Indeed, our genetic analyses showed that 80% of metastases in BALB-NeuT mice were derived from eDCCs. The genomic profiles of human DCCs isolated intraoperatively months to years before metastasis represented early cancer cells, not the predominant primary tumour clones, indicating that eDCCs have yet to acquire critical alterations such as gains on chromosome 8q to form metastases. The time to acquire such changes may largely account for the long latency periods and late relapses, which are becoming clinically more and more relevant[25]. Therefore, parallel progression[7] seems to be typical rather than exceptional. This is supported by a recent sequencing study, in which not a single case of linear progression from primary tumour to metastasis was found[26].

Our data indicate that breast cancer hijacks a developmental program, in which progesterone and its paracrine signals regulate mammary epithelial branching, fat pad invasion[14,27,28] and mammary stem-cell expansion[14,17,18,29–31] during development and pregnancy, for dissemination, independently of breast cancer subtype. Relevance to human disease is highlighted by a careful analysis of mortality from DCIS, hitherto defined as pre-invasive lesion[32]. Of the 3% of DCIS fatalities, more than 50% die of metastasis without local relapse, indicating lethal dissemination before surgery of the pre-invasive lesion[32]. Evidently, dissemination will also occur early if the tumour is not diagnosed as DCIS but later as invasive cancer. Moreover, death from DCIS increased significantly to 8% for young women[32], possibly because in young women the epithelial compartment and the microenvironment have high PGR expression.

The gradual generation of a $HER2^{high}/PGR^{-}$ phenotype may explain why early lesions of the BALB-NeuT model, but not advanced tumours, were found to represent human luminal tumours[33]. The BALB-NeuT model apparently represents a mixed luminal and HER2 phenotype and models several breast cancer subtypes. Cell density, progesterone, PIPS and HER2 signalling regulated dissemination and sphere formation in breast cancer cell lines of all subtypes similar to the BALB-NeuT model. Notably, HER2-positive circulating tumour cells were detected in DCIS and lobular carcinoma *in situ* or patients with M0-stage breast cancer irrespective of the HER2 status of the primary tumour[34].

High cell density and HER2 expression were responsible for the proliferative switch of mammary epithelial cells. Oncogenic mutations characterize benign tumours that do not metastasize[35–37]; indeed, strong activation of oncogenic pathways represses metastasis while increasing proliferation[38,39]. However, our experiments do not exclude the possibility that metastases form from advanced tumours because the early lesion dissemination program may become re-activated in areas of low cell density.

Our findings have implications for the understanding of metastasis and development of therapies: systemically spread cancer cells probably comprise cells derived from different stages of primary tumour evolution, including the earliest. Since DCCs from early and later stages have metastatic potential, therapies targeting the seed of metastasis need to address this heterogeneity.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 October 2015; accepted 11 November 2016.
Published online 14 December 2016.

1. Aguirre-Ghiso, J. A., Bragado, P. & Sosa, M. S. Metastasis awakening: targeting dormant cancer. *Nat. Med.* **19,** 276–277 (2013).
2. Polzer, B. & Klein, C. A. Metastasis awakening: the challenges of targeting minimal residual cancer. *Nat. Med.* **19,** 274–275 (2013).
3. Cole, B. F., Gelber, R. D., Gelber, S., Coates, A. S. & Goldhirsch, A. Polychemotherapy for early breast cancer: an overview of the randomised clinical trials with quality-adjusted survival analysis. *Lancet* **358,** 277–286 (2001).
4. Gianni, L. *et al.* Treatment with trastuzumab for 1 year after adjuvant chemotherapy in patients with HER2-positive early breast cancer: a 4-year follow-up of a randomised controlled trial. *Lancet Oncol.* **12,** 236–244 (2011).
5. Klein, C. A. *et al.* Genetic heterogeneity of single disseminated tumour cells in minimal residual cancer. *Lancet* **360,** 683–689 (2002).
6. Schmidt-Kittler, O. *et al.* From latent disseminated cells to overt metastasis: genetic analysis of systemic breast cancer progression. *Proc. Natl Acad. Sci. USA* **100,** 7737–7742 (2003).
7. Klein, C. A. Parallel progression of primary tumours and metastases. *Nat. Rev. Cancer* **9,** 302–312 (2009).
8. Hüsemann, Y. *et al.* Systemic spread is an early step in breast cancer. *Cancer Cell* **13,** 58–68 (2008).
9. Rhim, A. D. *et al.* EMT and dissemination precede pancreatic tumor formation. *Cell* **148,** 349–361 (2012).
10. Eyles, J. *et al.* Tumor cells disseminate early, but immunosurveillance limits metastatic outgrowth, in a mouse model of melanoma. *J. Clin. Invest.* **120,** 2030–2039 (2010).
11. Banys, M. *et al.* Hematogenous and lymphatic tumor cell dissemination may be detected in patients diagnosed with ductal carcinoma *in situ* of the breast. *Breast Cancer Res. Treat.* **131,** 801–808 (2012).
12. Sänger, N. *et al.* Disseminated tumor cells in the bone marrow of patients with ductal carcinoma *in situ*. *Int. J. Cancer* **129,** 2522–2526 (2011).
13. Valastyan, S. & Weinberg, R. A. Tumor metastasis: molecular insights and evolving paradigms. *Cell* **147,** 275–292 (2011).
14. Fernandez-Valdivia, R. *et al.* Revealing progesterone's role in uterine and mammary gland biology: insights from the mouse. *Semin. Reprod. Med.* **23,** 22–37 (2005).
15. Mulac-Jericevic, B., Lydon, J. P., DeMayo, F. J. & Conneely, O. M. Defective mammary gland morphogenesis in mice lacking the progesterone receptor B isoform. *Proc. Natl Acad. Sci. USA* **100,** 9744–9749 (2003).
16. Aupperlee, M. D., Smith, K. T., Kariagina, A. & Haslam, S. Z. Progesterone receptor isoforms A and B: temporal and spatial differences in expression during murine mammary gland development. *Endocrinology* **146,** 3577–3588 (2005).
17. Brisken, C. Progesterone signalling in breast cancer: a neglected hormone coming into the limelight. *Nat. Rev. Cancer* **13,** 385–396 (2013).
18. Liao, M. J. *et al.* Enrichment of a population of mammary gland cells that form mammospheres and have *in vivo* repopulating activity. *Cancer Res.* **67,** 8131–8138 (2007).

19. Korkaya, H., Paulson, A., Iovino, F. & Wicha, M. S. HER2 regulates the mammary stem/progenitor cell population driving tumorigenesis and invasion. *Oncogene* **27,** 6120–6130 (2008).
20. Ginestier, C. *et al.* ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* **1,** 555–567 (2007).
21. Lee, Y. J. & Gorski, J. Estrogen-induced transcription of the progesterone receptor gene does not parallel estrogen receptor occupancy. *Proc. Natl Acad. Sci. USA* **93,** 15180–15184 (1996).
22. Stephens, P. J. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486,** 400–404 (2012).
23. Fehm, T. *et al.* A concept for the standardized detection of disseminated tumor cells in bone marrow from patients with primary breast cancer and its clinical implementation. *Cancer* **107,** 885–892 (2006).
24. Gužvić, M. *et al.* Combined genome and transcriptome analysis of single disseminated cancer cells from bone marrow of prostate cancer patients reveals unexpected transcriptomes. *Cancer Res.* **74,** 7383–7394 (2014).
25. Klein, C. A. Framework models of tumor dormancy from patient-derived observations. *Curr. Opin. Genet. Dev.* **21,** 42–49 (2011).
26. Brastianos, P. K. *et al.* Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov.* **5,** 1164–1177 (2015).
27. Brisken, C. *et al.* Essential function of Wnt-4 in mammary gland development downstream of progesterone signaling. *Genes Dev.* **14,** 650–654 (2000).
28. Fernandez-Valdivia, R. *et al.* The RANKL signaling axis is sufficient to elicit ductal side-branching and alveologenesis in the mammary gland of the virgin mouse. *Dev. Biol.* **328,** 127–139 (2009).
29. Asselin-Labat, M. L. *et al.* Steroid hormone receptor status of mouse mammary stem cells. *J. Natl. Cancer Inst.* **98,** 1011–1014 (2006).
30. Joshi, P. A. *et al.* Progesterone induces adult mammary stem cell expansion. *Nature* **465,** 803–807 (2010).
31. Mukherjee, A. *et al.* Targeting RANKL to a specific subset of murine mammary epithelial cells induces ordered branching morphogenesis and alveologenesis in the absence of progesterone receptor expression. *FASEB J.* **24,** 4408–4419 (2010).
32. Narod, S. A., Iqbal, J., Giannakeas, V., Sopik, V. & Sun, P. Breast cancer mortality after a diagnosis of ductal carcinoma *in situ*. *JAMA Oncol.* **1,** 888–896 (2015).
33. Astolfi, A. *et al.* Gene expression analysis of immune-mediated arrest of tumorigenesis in a transgenic mouse model of HER-2/neu-positive basal-like mammary arcinoma. *Am. J. Pathol.* **166,** 1205–1216 (2005).
34. Ignatiadis, M. *et al.* HER2-positive circulating tumor cells in breast cancer. *PLoS One* **6,** e15624 (2011).
35. Hafner, C. *et al.* Oncogenic PIK3CA mutations occur in epidermal nevi and seborrheic keratoses with a characteristic mutation pattern. *Proc. Natl Acad. Sci. USA* **104,** 13450–13454 (2007).
36. Hafner, C. *et al.* Multiple oncogenic mutations and clonal relationship in spatially distinct benign human epidermal tumors. *Proc. Natl Acad. Sci. USA* **107,** 20780–20785 (2010).
37. Pollock, P. M. *et al.* High frequency of BRAF mutations in nevi. *Nat. Genet.* **33,** 19–20 (2003).
38. Hutchinson, J. N., Jin, J., Cardiff, R. D., Woodgett, J. R. & Muller, W. J. Activation of Akt-1 (PKB-α) can accelerate ErbB-2-mediated mammary tumorigenesis but suppresses tumor invasion. *Cancer Res.* **64,** 3171–3178 (2004).
39. Liu, H. *et al.* MYC suppresses cancer metastasis by direct transcriptional silencing of α_V and β_3 integrin subunits. *Nat. Cell Biol.* **14,** 567–574 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. Perry for his critical reading of the manuscript. This work was supported by grants from the DFG (KI 1233/2-1, KL 1233/3-1, KL 1233/10-1 (C.A.K.); Me2064/4-1, (G.M.); SP 938/2-1 (R.S.); INST 89/341-1 FUGG (TissueFAX)); the Dr Josef Steiner Foundation and ERC (322602) to C.A.K.; the SWCRF, CA109182, CA196521, CA163131, BC132674 (J.A.A.-G.) F31 CA183185 (K.L.H.) BC112380 (M.S.S.), NIH 1S10RR024745 Microscopy CoRE at ISMMS and by a donation from A. Jungmayer.

Author Contributions C.A.K. and H.H. designed and evaluated core experiments. H.H. performed core experiments. M.M.S.O., C.W., L.K.N., C.E., C.R. and M.Gu. helped with *in vivo*, *in vitro* and primary culture experiments. M.H., M.M., M.Gr., R.S. and H.H. performed bioinformatic and statistical analyses. M.W.-K., K.L.H., M.S.S. and F.W. performed staining and pathological analysis. N.E. and G.M. performed miRNA sequencing and analysis. G.H., N.P., A.K.H., F.-A.T., S.Y.B., B.R., S.B. and T.F. performed DCC analysis or collected patient data. C.A.K. and H.H. wrote the manuscript with input from J.A.A.-G.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.A.K. (christoph.klein@ukr.de).

METHODS

Mice. BALB-NeuT transgenic mice were obtained through collaboration with G. Forni and maintained in our facilities according to the European Union guidelines. All animal experiments were performed according to the EU and national institutional regulations. Mice were screened at 3–4 weeks of age for hemizygosity ($neuT^+/neuT^-$), and negative littermates served as wild-type BALB/c mice controls. Mammary glands of BALB-NeuT female mice were inspected twice a week and arising tumours were measured in two perpendicular diameters. Data acquisition for bone-marrow DCCs was performed in a blinded manner, whereas enumeration of lung metastasis was performed unblinded by two observers. All experimental animal procedures were approved and conducted according to German federal and state regulations (Government of Upper Palatinate, 55.2-2532.1-27/14).

Surgery and transplantation experiments. Mice were anaesthetized with midazolam 5 mg kg^{-1}, fentanyl 0.05 mg kg^{-1}, medetomidin 0.5 mg kg^{-1} by intraperitoneal injection. The thorax and abdomen were shaved; the skin was incised caudal to cranial along the midline. Fifty spheres were mixed with Matrigel (BD Biosciences: 356231, final concentration, 40%) and injected into the fourth right mammary gland of BALB/c mice (4 and 40-week-old mice). For tissue transplantation, a piece (approximately 1 mm^3) of donor mammary tissue from 4-week-old BALB-NeuT mice (gland model) or primary tumours (primary tumour model) was implanted in the cleared mammary fat pad of recipient mice (4-week-old BALB/c mice). The skin was closed by a suture using polygelatin string (Ethicon) and anaesthesia was antagonized with flumazenil 0.5 mg kg^{-1}, atipamezol 2.5 mg kg^{-1}, naloxon 1.2 mg kg^{-1} by subcutaneous injection. Postoperative analgesia was achieved by buprenorphin (0.1 mg kg^{-1}) by subcutaneous injection. Curative surgery or dissection was done when the diameter of tumours was between 5–10 mm. After surgery, mice were kept until we observed the first general signs of reduced health. After dissection, lungs were macroscopically inspected and individual metastases counted.

Extent of dissemination relative to tumour area. Data were taken from ref. 8. Briefly, gland or tumour areas were calculated from 270 mammary glands or tumours of 27 mice assuming the shape of an ellipse or circle for each tumour. The tumour area of mammary glands without palpable (that is, not measurable by a caliper) tumours was set to 0.1 mm^2 (that is, assuming a diameter of 350 μm of a total, circular hyperplastic lesion within a mammary gland) for lesions from 4–9-week-old mice and 0.4 mm^2 for 11-week-old mice. The adjustment for 11-week-old mice was based on a microscopic evaluation showing an about 4-fold increase in hyperplastic lesions. Dissemination to the bone marrow was determined by the number of cytokeratin-positive cells per 10^6 bone-marrow cells (Extended Data Fig. 1a).

Mouse bone marrow preparation and staining for DCCs. Bone marrow was collected from femurs and tibiae. The bone marrow was rinsed with a 26-G needle with 1 ml of PBS. After density gradient centrifugation, 5×10^5 interphase cells were put on adhesion slides (Menzel). At least 10^6 cells per mouse were stained to detect positive cells. Blocking solution (5% rabbit serum in 1× TBS (50 mM Tris-base, 150 mM Nacl, pH 7.4)) was added to the slides to rehydrate the cells and to block unspecific binding of antibodies to the cells. After 20 min the blocking solution was discarded and primary antibody against cytokeratins 8 and 18 (CK8/18; all antibodies and working concentrations are in Supplementary Table 9) or guinea pig serum (the CK antibody originated from guinea pig) as control, was added and slides were incubated for 60 min. The primary antibody was discarded and the slides were washed 3× for 3 min in 1× TBS. The slides were incubated with the secondary antibody for 25 min, and then washed 3× for 3 min in 1× TBS followed by incubation with the ABC complex (Vector Laboratory) for 25 min. Finally, the development system of the BCIP/NBT (AP Conjugate Substrate Kit, Bio-Rad Laboratories GmbH, 1706432; Levamisol hydrochloride, Sigma-Aldrich GmbH, L-9756) for alkaline phosphatase enzymatic substrate was added for 10 min. The slides were washed 3× for 3 min and screened for CK8/18-positive cells. The positive cells were typically violet-to-black in colour. TUBO, a tumour cell line derived from a mouse primary mammary tumour of BALB-NeuT and known to express CK8/18, was used as a positive control.

Laser microdissection and microarray analysis. Laser microdissection (PALM MicroBeam from Carl Zeiss MicroImaging GmbH) was performed to dissect metastatic lesions from lungs, primary tumours and epithelial layers of mammary glands of BALB-NeuT mice at the time point of early lesions (7–9-week-old mice; examples are shown in Extended Data Fig. 1b), and BALB/c mice at different ages (description of samples is given in Supplementary Table 1). Small pieces adding up to 100,000 μm^2 for each sample were catapulted into a cap with 10 μl paramagnetic, biotinylated, oligo-dT-peptide, nucleic-acid, bead suspension and lysis buffer (Active Motif, 29011). Extraction of mRNA and microarray experiments were performed as described previously[40]. Heatmaps in Fig. 1a were generated using Euclidean distance and complete linkage agglomerative clustering on row (gene)-wise standardized expression data (zero mean, unit standard deviation).

Cell lines, cell culture and cell stimulation. Breast cancer cell lines (4T1, 66cl4, and 67NR) were provided by F. Miller. These cell lines were derived from a single mammary tumour that arose spontaneously in a wild-type BALB/cfC3H mouse. The MM3MG mouse mammary epithelial cell line derived from a BALB/c background was purchased from ATCC (ATCC CRL6376). The TUBO cell line is a cloned cell line established *in vitro* from a lobular carcinoma that arose spontaneously in a BALB-NeuT mouse (gift from G. Forni). All mouse and stably transduced cell lines were grown in DMEM medium (Pan-Biotech, P04-03500) supplemented with 10% (20% for TUBO cell line) FCS (Pan-Biotech: P30-3702), 2 mM L-glutamine (Pan-Biotech, P04-80100), 10 U ml^{-1} penicillin/streptomycin (Pan-Biotech, P06-07050). All human cell lines were purchased from ATCC and each cell line was maintained in medium recommended by ATCC. The origin of the cell lines was confirmed by short tandem repeat (STR) analysis (Cell-ID, Promega). All cells were incubated at 37 °C with 5% CO_2. Steroid hormones (progesterone, aldosterone, β-oestradiol, testosterone and hydrocortisone; all from Sigma-Aldrich) and RU486 (Sigma-Aldrich) were dissolved in ethanol. RANKL (mouse Rankl, Abcam, ab151200; human RANKL, Abcam, ab9958); WNT4 (mouse WNT4, R&D systems, 475-WN; human WNT4, Abnova, H00054361-P01); lapatinib (Santa Cruz Biotechnology, SC202205); IWP-2 (Sigma-Aldrich, I0536); RANKL-neutralizing antibody (Lifespan Biotech, LS-C150261) were dissolved according to the manufacturer's instructions. All cell lines were routinely tested for mycoplasma and were found to be negative.

Primary cultures and sphere formation assay. Fresh mammary glands or primary tumours were digested with 200 units per ml collagenase I (Worthington Biotech, LS004196) and 1 μg ml^{-1} hyaluronidase (Sigma-Aldrich, 4272) in basal medium for 2 h at 37 °C. The basal medium consisted of DMEM/F12 (PAN biotech, P04-41450) supplemented with 10 mM HEPES buffer (Sigma-Aldrich, H0887), penicillin/streptomycin (Pan Biotech, P1-010) and 10 μg ml^{-1} insulin (Sigma-Aldrich, I9278). Digested tissue cells were centrifuged and re-suspended in basal medium. The cells were subsequently cultured at a density of 5×10^4 cells per ml in ultra-low adherent plates coated with 1.2% poly-HEMA (Sigma-Aldrich, P3932) or at a density of 2.6×10^4 cells per cm^2 for adherent culture in DMEM medium (Pan-Biotech, P04-03500) supplemented with 10% FCS (Pan-Biotech, P30-3702), 2 mM L-glutamine (Pan-Biotech, P04-80100), 10 U ml^{-1} penicillin/streptomycin (Pan-Biotech, P06-07050). Sphere culture medium was basal medium supplemented with 2% B27 (Gibco, 17504044), 10 μg ml^{-1} EGF (Sigma-Aldrich, E9644), 10 ng ml^{-1} bFGF (Sigma-Aldrich, F0291), 20 ng ml^{-1} hIL6 (gift from S. Rose-John), 4 ng ml^{-1} heparin (Sigma-Aldrich, H3149), 5 ng ml^{-1} GRO-α (R&D systems, 275-GR). Concentrations of activators, inhibitors and other molecules are given in the main text, figures or legends. Sphere cultures were incubated at 37 °C with 5% CO_2 and 7% O_2 and cultures were screened for spheres after 10 days. Only spheres with a diameter over 50 μm were counted. The size of mammospheres was inspected under a light microscope and measured using Zeiss Axiovision software (Carl Zeiss) after 10 days.

Cell density experiments. TUBO cells were cultured at 3×10^4 cells per cm^2 for high density and 5.2×10^3 cells per cm^2 for low-density experiments. Primary cells derived from primary tumours were cultured in 10.6×10^4 cells per cm^2 for high-density and 2.2×10^3 cells per cm^2 for low-density experiments. Density criteria for human cell lines were 100% confluency for high density and 20–30% confluency for low-density. For hormone treatment and comparisons between low and high-density experiments, cells were incubated for 76 h with fresh hormone treatment and washes (2× with PBS) at 24-h intervals. We avoided changing medium and washing during incubation of cells for miRNA analyses. In migration experiments we seeded 10^4 cells per well (24-well migration chambers) for low-density experiments for all cell lines and 5×10^4 cells per well for high-density experiments with TUBO cells and 4×10^4 for the other cell lines.

Transwell assay. Transwell inserts (Corning, 3419) with a microporous membrane of 0.4 μm were used to separate the upper and lower compartments. The microporous membrane allows only soluble factors to pass between the compartments. Early lesion cells were cultured in the lower chamber and primary tumour cells were cultured in the upper chamber. Both were cultured at a density of 10^6 cells per well of 6-well plates (DMEM with 10% FCS).

Migration/sphere-formation assays. Transwell inserts (Corning: 3422) with 8-μm pores were coated with 30% matrigel. 4×10^4 from cell lines and 10^5 cells isolated from tissue or dissociated spheres were resuspended in FCS-free medium (DMEM) before seeding. Cells were then seeded in 200 μl of FCS-free medium on top of the Matrigel layer and FCS medium (DMEM containing FCS) was added to the lower chamber. For additional treatments medium in both upper and lower compartments was supplemented with the reagents at concentrations specified in the text and figures. After incubating cells isolated from mammospheres or from freshly digested tissues for 72 h, inserts were removed and cells were fixed with methanol (−20 °C for 10 min) and stained with trypan blue. Cells were counted from 3 fields (4× magnification) when visualized under the microscope.

For the combined migration/sphere formation assay, cells were placed on a layer of 30% Matrigel in the upper chamber and the lower chamber was coated with poly-HEMA. The mammosphere medium used is described above. After 72 h, inserts were removed, fixed and stained with trypan blue for single cell migration analysis. 600 µl fresh sphere medium was added to the lower chamber and cells were incubated for 11 days when spheres were counted (Fig. 2e and Extended Data Fig. 2g).

Proliferation assay. Single-cell suspensions were cultured in 96-well plates (Corning Inc) and proliferation was evaluated by a XTT-colorimetric assay kit (Roche, 11465015001) based on the manufacturer's instructions. Seeding concentration of cells was 3,000 cells per well. The experiment was performed with 6 technical replicates. The medium was supplemented with the tested factors or hormones and vehicle (see corresponding experiments) and was changed every second day.

Immunohistochemistry. For PGR and HER2 immunohistochemistry of tissue sections, we used 5-µm sections of paraffin blocks placed onto poly-L-lysine-coated slides. Samples were dewaxed by two 5-min washes in xylene and rehydrated with graded alcohol by 5-min washes and a final wash in water. A standard Tris–EDTA buffer and pressure-cooking was used for antigen retrieval and then sections were blocked in 0.3% H_2O_2 in TBS and 10% normal goat serum. Sections were incubated for 1 h with primary antibodies and, after washing, secondary antibodies (Vector laboratory, PK4001 or PK5000) were added based on the manufacturer's recommended dilution (see Supplementary Table 9). After washing with PBS, sections were stained using the ABC detection system (Vector Laboratory) according to the manufacturer's instruction. Visualization was performed with chromogen reagent (Dako, 10046560) according to manufacturer's instructions.

Immunofluorescence staining. For staining of cells from monolayer cell cultures, cells were seeded onto 24-well culture plates at an appropriate density. After 72 h of incubation, cells were washed with PBS and fixed with 4% PFA for 10 min. Then, cells were permeabilized with 0.2% Triton X-100 followed by washing steps and blocking with 1% BSA in PBS at 37 °C and incubated with primary antibody (see Supplementary Table 9) for 1 h at room temperature. Cells were then washed 3× with PBS and incubated with labelled secondary antibody (Jackson ImmunoResearch Laboratory Inc) for 1 h at room temperature. For nuclear counterstaining, cells were incubated for 10 min with 0.5 µg ml^{-1} DAPI (Sigma-Aldrich). For the staining of spheres in differentiation experiments, mammospheres were picked and transferred to a 24-well cell-culture plate and incubated for 8 h in sphere medium in order to fix them to the surface. The subsequent staining protocol was as for monolayer cell culture staining. For staining of cells attached to the inserts from migration experiments, inserts were used directly after migration (see migration assay), for the blocking step and immunofluorescence staining, the monolayer cell culture staining procedure was applied. Images were captured on an AxioVert 200M microscopy (Carl Zeiss Microscopy).

Quantification of HER2 and PGR staining in tissue sections by TissueFAX cytometry. Tissue sections were stained with an automated staining machine (Ventana, BenchMark ULTRA). Tissue sections used for analysis were stained within the same run. Images of stained tissue sections were scanned with the TissueFAXSi-plus imaging system (TissueGnostics, Vienna, Austria; acquisition software: TissueFAXS v3.5.129) equipped with a digital Pixelink colour camera (PCO AG). Images for the analysis of HER2 and PGR staining were analysed with HistoQuest software v3.5.3.0185 (TissueGnostics). Using the HistoQuest software, two markers were created: haematoxylin as master marker (nucleus) and HER2 or PGR as non-master marker. To achieve optimal cell detection, the following parameters were adjusted: (i) nuclei size; (ii) discrimination by area; (iii) discrimination by grey and (iv) background threshold. For the evaluation of the HER2 staining intensity of cells or the percentage of PGR-expressing cells, histograms were created, allowing the visualization of corresponding cells in the source region of interest using the real-time back-gating feature. The cut-off discriminated between false events and specific signals according to cell size and intensity of staining. For HER2 staining, 38,675 primary tumour cells (6 regions, 1.99 mm^2), 28,850 cells from hyperplastic regions (25 regions, 1.55 mm^2) and 14,938 cells from non-transformed ducts (30 regions, 0.93 mm^2) were analysed. For PGR, 12,269 cells of early lesions (hyperplasia, 7 weeks, 11 regions, 0.5 mm^2), 12,702 cells of non-transformed (normal duct, 7 weeks, 56 regions, 0.7 mm^2) and 25,357 primary tumour cells (9 regions, 1.3 mm^2) were analysed (Fig. 1d).

Quantitative PCR. All mRNA extractions were performed using the RNeasy kit (Qiagen, 74104) according to the manufacturer's instructions. For miRNA extraction, the miScript II RT Kit (Qiagen, 217004) was used. cDNA was generated using a reverse transcriptase kit (Qiagen, 205311 for total RNA and 218161 for miRNA). Finally, 25 ng of cDNA was used for qPCR. qPCR was performed using a LightCycler instrument (Roche) and Fast Start Master SYBR Green Kits (Roche). Data analysis was done using the RelQuant software (Roche) with a reference gene and a calibrator (reference) sample in every run. Mouse reference cDNA served as a positive control. Samples with unspecific products in the melting curve analysis were discarded from further analysis. Expression levels are given relative to *Actb* (β-actin) for gene expression analyses and *Rnu6* for miRNA analyses (primer sequences are provided in Supplementary Table 10). All primers for mRNA analyses were synthesized by Eurofins MWG Operon, and by Qiagen for miRNA analyses.

For comparison of miRNA levels in high-density and low-density regions (Extended Data Fig. 10b) of formalin-fixed paraffin-embedded samples, regions were punched out using a 1.5-mm puncher (PFM medical; 48115). Samples were incubated for 10 min at 70 °C followed by xylene–ethanol de-paraffinization and overnight proteinase K (0.5 µg µl^{-1}, Roche 03115828001) digestion. Then miRNA extraction was performed using the miRNeasy kit. For comparison of miR-30a-5p and miR-9-5p between HER^{high}/PGR^{high} human mammary carcinomas and HER2^{high}/PGR^{-} carcinomas, miRNAs were extracted from freshly frozen samples using the miRNeasy kit. Expression of miR-9-5p and miR-30a-5p was normalized to HER2^{-}/PGR^{high} breast cancers (see Supplementary Table 6 for details on patients).

Lenti- and gamma retroviral transduction. *PGR* expression was carried out with a lentiviral construct encoding human *PGR-B* (GeneCopeia, Z5911). Lentiviral packaging was conducted as previously described[41]. Helper vectors were pSPAX2 and pMD2.G (Addgene). Selection was performed using 10 µg ml^{-1} of puromycin (Sigma-Aldrich, P8833). For *Her2* expression pLXSN–NNeu (rat wild-type *Neu*/*Her2*) was used (obtained from L. Petti)[42]. Retroviral delivery of transgenes was performed as described previously[43]. Helper vectors were pCMV–VSV-G and pUMVC3 obtained from Addgene. Selection was performed using 1,000 µg ml^{-1} of G418 (Sigma-Aldrich, G9516). MM3MG cells were transduced with lentiviruses and/or retroviral vectors and cell colonies were selected using antibiotics. Positively transduced clones were expanded and screened for PGR and/or HER2 levels by western blot analysis and qPCR.

Western blot analysis. Cell lysates were prepared using RIPA buffer (Sigma-Aldrich, R0278) and were analysed with the BCA protein assay kit (Thermo Scientific, 23227) to measure and their protein concentration was equalized. Quantified protein lysates were resolved on 6.5% SDS–PAGE gels, and transferred onto a polyvinylidene difluoride membrane (Millipore), and immunoblotted with the primary antibodies overnight followed by incubation with the horseradish peroxidase-conjugated secondary antibodies. The blots were visualized using a substrate kit (GE Healthcare, RPN2109) and bands were visualized by Imagequant LAS 4000 (GE Healthcare). The full blot images are shown in Supplementary Fig. 1.

Exosome isolation, miRNA analyses and sequencing. To prepare conditioned medium, TUBO cells were seeded at a density of 3×10^4 cells per cm^2. After 4 days, medium was collected, centrifuged and filtered and used as conditioned medium. For exosome isolation we used an ultracentrifugation method as previously described[44]. Exosome pellets were resuspended in fresh medium and used for T47D cell line treatment. PGR expression was checked at different time points (4, 8, 24, and 48 h). For miRNA sequencing we used 4×10^6 cells and exosomes were isolated from confluent medium from TUBO cells. The miRNA cloning and sequencing was done as described previously[45]. All pooled samples were sequenced on a MiSeq system (Illumina) in a single-end run with 80 cycles using the MiSeq reagent kit v3. Data analysis was performed using in-house written scripts. Sequences were mapped—without any mismatches allowed—against mouse miRNAs listed in the miRBase v20 (June 2013; http://www.mirbase.org). The minimum length of reads was set to 18 nucleotides. Annotated miRNA-reads were normalized as RPM values according to the total number of mapped reads in the respective library. Mimic miRNAs were ordered from Eurofins MWG Operon Company and all sequences are listed in Supplementary Table 11. For miRNA transfection we used reverse transfection protocols according to instructions for RNAiMAX (Life Technology, 13778030) and with a 50 nM concentration of miRNA.

Mouse genome comparative genome hybridization microarrays. DNA samples were extracted from freshly frozen samples using the DNeasy Blood & Tissue Kit (Qiagen, 69504). Genomic DNA labelling was done using the Agilent SureTag DNA Labelling Kit (Agilent, 5190-3400). Array CGH was performed on oligonucleotide-based SurePrint G3 Mouse CGH Microarray Kit, 4x180K (design code: 027411) according to the protocol provided by the manufacturer (Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis, v.7.2, July 2012).

Phylogeny of primary tumours and metastases. Ancestral relations among matched samples of primary tumour and metastases were inferred using array CGH profiles. The array comparative genomic hybridization dataset consisted of 28 primary tumour samples with 1–3 corresponding metastasis samples (18 primary tumours with 1 metastasis; 4 with 2 metastases and 6 with 3 metastases).

Positions of the probes on the array were mapped to the current mouse reference genome (mm10) using the liftOver tool[46]. No background correction was applied to the data[47]. The data were first normalized within arrays using Loess[48]. Then, log ratios were corrected for spatial artefacts using a median filter with an

11×11 block of probes[49] and a between array scale normalization was applied[50]. Duplicate probes (having the same genomic position) were summarized by their median log ratio. The R package limma[51] (v.3.28.1) was used for normalization within arrays and between arrays (with default parameters). In a final step, wavy patterns were removed from the data using an approach similar to ref. 52, but with modifications to account for broad copy-number alterations. For every sample and chromosome, correction was carried out as follows. Because the maximum number of broad copy-number alterations on any chromosome observed in the data was two, a piecewise constant function with two pieces was fitted to the log ratios to estimate these broad alterations. Each piece was required to be longer than 5% of the chromosome length to avoid spurious small pieces. The wavy pattern was estimated by fitting a Loess curve with a window size of 100 probes to the residuals of the piecewise constant function fit. To avoid smoothing true focal alterations, the weight of probes with an absolute log ratio deviation greater than 0.5 from the piecewise constant function were set to 0 for the Loess fit. The estimated wavy pattern was then subtracted from the log ratios resulting in corrected values.

After normalization the log ratios were segmented using Circular Binary Segmentation as implemented in the R package DNAcopy[53] (v.1.46.0). The default parameters were used except for α, which was set to 0.001. Segments with a length of 5 or less probes were merged with the closest adjacent segment. For every sample, states with means closer than 0.05 were merged iteratively beginning with the two states with closest means. When two states were merged the new mean was given by the mean of the two old state means weighted by the number of probes in every state. After state merging, the remaining segment means were adjusted to have a median of 0. These segmented copy number profiles were then deconstructed into underlying copy number events using Ziggurat Deconstruction[54]. All these steps where performed using R (v.3.3.0).

Aberration events as defined by left and right change points, and aberration type were pooled across the matched samples of a single mouse to form a mouse-specific base set of aberration prototypes. For this, amplifications and deletions that were similar by more than 80% as measured by their Jaccard-index regarding probe support were merged into single prototypes using Jaccard-distance-based complete linkage clustering and union of supports. Individual primary tumour and metastases samples were then encoded according to presence (absence) of the prototypic aberrations, whereby the present prototype was called by the minimum Jaccard-distance. The resulting feature vectors were then used for phylogenetic tree inference. Phylogenetic trees were generated by assuming ideal (that is, error-free) data and inferring plausible common ancestors (intermediates) of aberration profiles by extracting shared features of an increasing number of samples, that is, evaluating common aberration events in sample pairs, triplets, quads, and so on, and organizing these ancestors according to hierarchical levels. Subsequently, admissible edges were constructed top-down between vertices allowing for two re-losses of acquired gains and no re-gains of any losses (this condition was also ensured globally for each path). Then all simple paths from the normal cell to the samples were generated using the Igraph R-package (v.1.0.0), combined into a directed acyclic graph and filtered for the fewest genomic changes along the graph and lowest number of intermediates (maximum parsimony). This resulted in one unique phylogenetic tree for each mouse.

Detection of human DCCs. For CGH analysis of single DCCs, bone-marrow sampling of patients with M1 stage cancer was performed within the study protocol of the GEBDIS study at the Central Hospital in Augsburg after informed, written consent of patients was obtained. The ethics committees of the University of Munich (ethics vote number 007/02) and Regensburg (ethics vote number 07-079) approved bone marrow sampling (including patients with M0-stage cancer) and genomic analysis of isolated cells. For all patients informed, written consent was obtained. For bone-marrow sampling and analysis for cytokeratin-positive cells of patients from Tübingen (approval by the ethics committee of Tübingen University, reference number 560/2012R) all specimens were obtained after written, informed consent. Bone-marrow sample preparation, slide preparation, cytokeratin staining and cell isolation was performed as previously described[5].

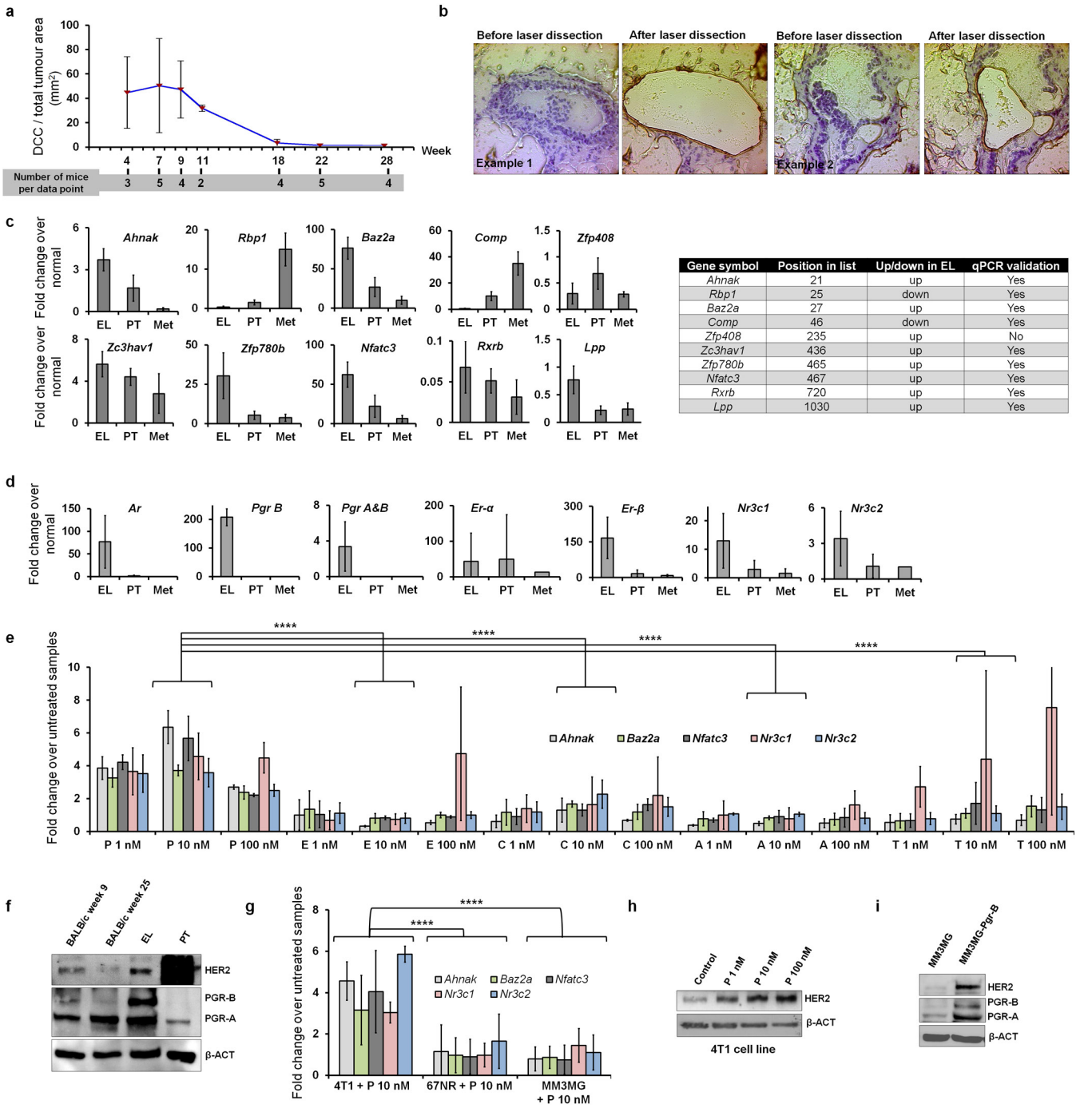
Whole genome amplification and single cell comparative genomic hybridization. Whole genome amplification was performed as previously described[5,55]. The method has become commercially available as a kit (*Ampli1*, Silicon Biosystems). A histogram of copy-number alterations (Fig. 5e) was generated for human primary breast cancers ($n = 1{,}637$) derived from the Progenetix database (http://www.progenetix.net) and DCCs isolated from bone marrow of breast cancer patients without (M0, $n = 94$; see Supplementary Table 7 for clinical details of patients) and with metastasis (M1, $n = 91$).

Analysis of patients with breast cancer for association of DCC and primary tumour receptor expression. We analysed data of 2,239 patients from the Department of Oncology and Obstetrics, University of Tübingen. DCC status was assessed according to the consensus protocol[23], using the anti-cytokeratin antibody A45B/B3 and by evaluating 2×10^6 bone-marrow cells. PGR expression of primary tumours was categorized into PGR staining scores 0–1 for absent expression; 2–8 for intermediate expression and 9–12 for high expression. HER2 status of primary tumours was categorized into the staining score 0 for absence of HER2 staining (IHC negative); score 1 and score 2 without *HER2* amplification (IHC positive) and score 2 with *HER2* amplification and 3 (which is known to be caused by *HER2* amplification; FISH positive).

Bioinformatics and statistical analysis. Statistical analyses and estimation of variation within each group of data were performed using GraphPad Prism v.6 and R v.3.3.1. For *in vivo* DCC experiments of primary tumour compared to early lesion spheres, sample size was estimated using G*Power (v3). No statistical methods were used to predetermine sample size for other experiments. For each experiment, mouse numbers are given in the figures or the text. All *in vitro* and primary culture experiments were performed at least in triplicate and Student's *t*-test was used for comparisons. For all other experiments we applied the D'Agostino-Pearson omnibus normality test. When sample size was sufficiently large ($n \geq 8$) and were not distributed normally according to the D'Agostino-Pearson test ($P \leq 0.05$) we applied the Mann–Whitney *U*-test. For gene signature evaluation in Figs 1c, 3d and Extended Data Fig. 1e, g, gene wise *t*-test *P* values were combined using Stouffer's method. A linear regression test (*F*-test for slopes) was used to compare proliferation curves and tumour growth. For comparing numbers between different groups we applied Fisher's exact test or if the sample numbers were at least 5 in each condition the χ^2 test. In Fig. 1d one-way ANOVA was used. All *P* values are two-tailed. All *P* values ($0.05 \leq P \leq 0.0001$) and statistical tests are mentioned in either figures or legends. Genomatix (v2.0) (https://www.genomatix.de) was used for signalling pathway analysis and oPOSSUM (v1) (http://opossum.cisreg.ca/oPOSSUM3/) for transcription-factor binding-site enrichment. For miRNA-binding enrichment, we used DIANALAB (http://diana.cslab.ece.ntua.gr/) and for the identification of target miRNAs for single target genes the miRANDA software (http://www.microrna.org/microrna/home.do) was used. The experiments were not randomized.

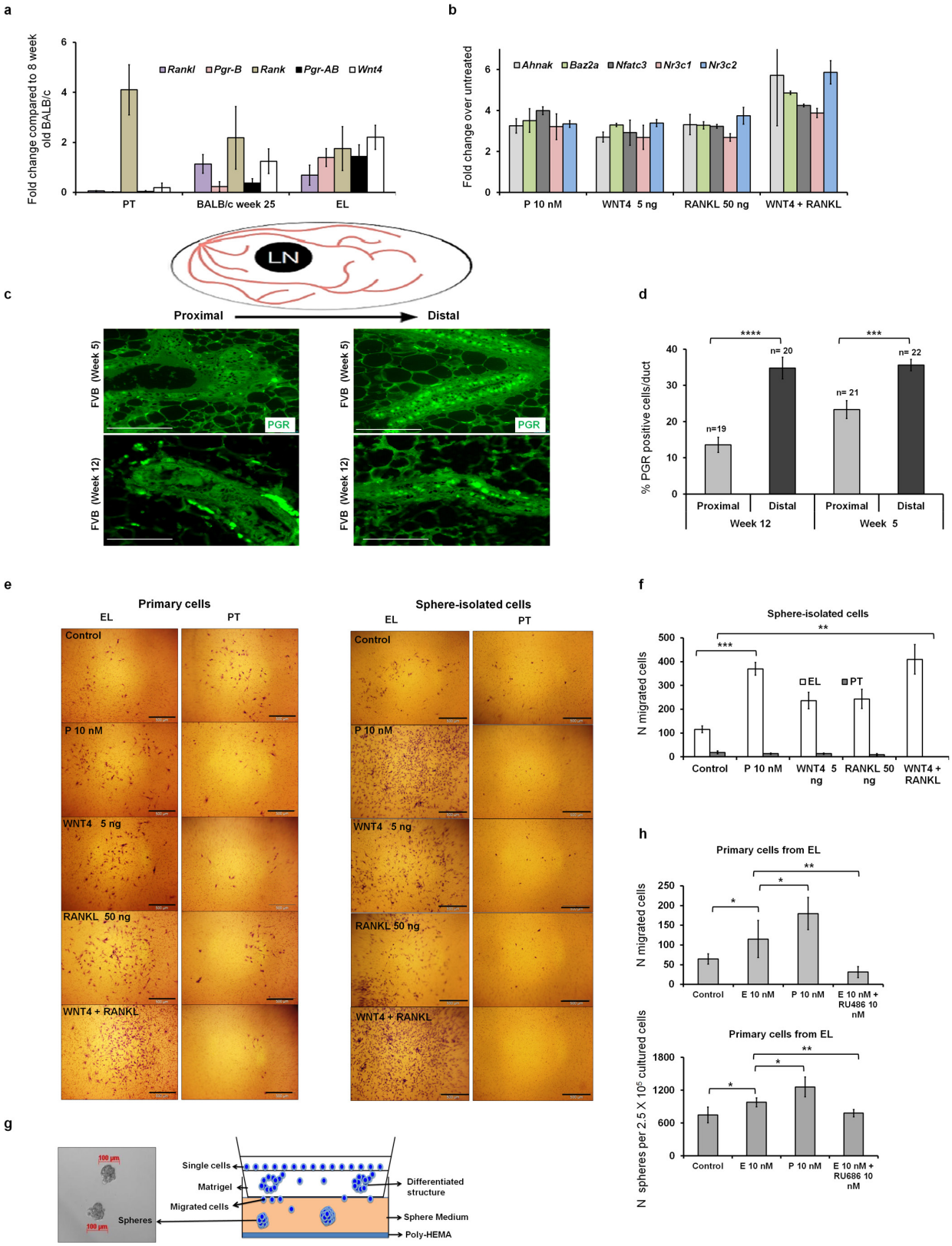
Data availability. The miRNA sequencing data and microarray data are deposited at the Gene Expression Omnibus (GEO) database under accession number GSE68683. Analysed data for microarray and miRNAs can be found in Supplementary Data 1 and 2, respectively. The mouse ancestral CGH data are deposited at the GEO database under accession number GSE87469. All raw data for presented graphs and statistics are deposited in Source Data files. Further material and data other than what is presented here can be obtained from the corresponding author (C.A.K.) upon request.

40. VerMilyea, M. D. *et al.* Transcriptome asymmetry within mouse zygotes but not between early embryonic sister blastomeres. *EMBO J.* **30,** 1841–1851 (2011).
41. Dull, T. *et al.* A third-generation lentivirus vector with a conditional packaging system. *J. Virol.* **72,** 8463–8471 (1998).
42. Petti, L. M. & Ray, F. A. Transformation of mortal human fibroblasts and activation of a growth inhibitory pathway by the bovine papillomavirus E5 oncoprotein. *Cell Growth Differ.* **11,** 395–408 (2000).
43. Stewart, S. A. *et al.* Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA* **9,** 493–501 (2003).
44. Thery, C., Amigorena, S., Raposo, G. & Clayton, A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. *Current Protoc. Cell Biology* Chapter 3, Unit 3.22 (2006).
45. Dueck, A., Eichner, A., Sixt, M. & Meister, G. A miR-155-dependent microRNA hierarchy in dendritic cell maturation and macrophage activation. *FEBS Lett.* **588,** 632–640 (2014).
46. Meyer, L. R. *et al.* The UCSC genome browser database: extensions and updates 2013. *Nucleic Acids Res.* **41,** D64–D69 (2013).
47. Yang, Y. H., Buckley, M. J. & Speed, T. P. Analysis of cDNA microarray images. *Brief. Bioinform.* **2,** 341–349 (2001).
48. Yang, Y. H. *et al.* Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30,** e15 (2002).
49. Khojasteh, M., Lam, W. L., Ward, R. K. & MacAulay, C. A stepwise framework for the normalization of array CGH data. *BMC Bioinformatics* **6,** 274 (2005).
50. Smyth, G. K. & Speed, T. Normalization of cDNA microarray data. *Methods* **31,** 265–273 (2003).
51. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43,** e47 (2015).
52. Marioni, J. C. *et al.* Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization. *Genome Biol.* **8,** R228 (2007).
53. Venkatraman, E. S. & Olshen, A. B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23,** 657–663 (2007).
54. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12,** R41 (2011).
55. Klein, C. A. *et al.* Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc. Natl Acad. Sci. USA* **96,** 4494–4499 (1999).



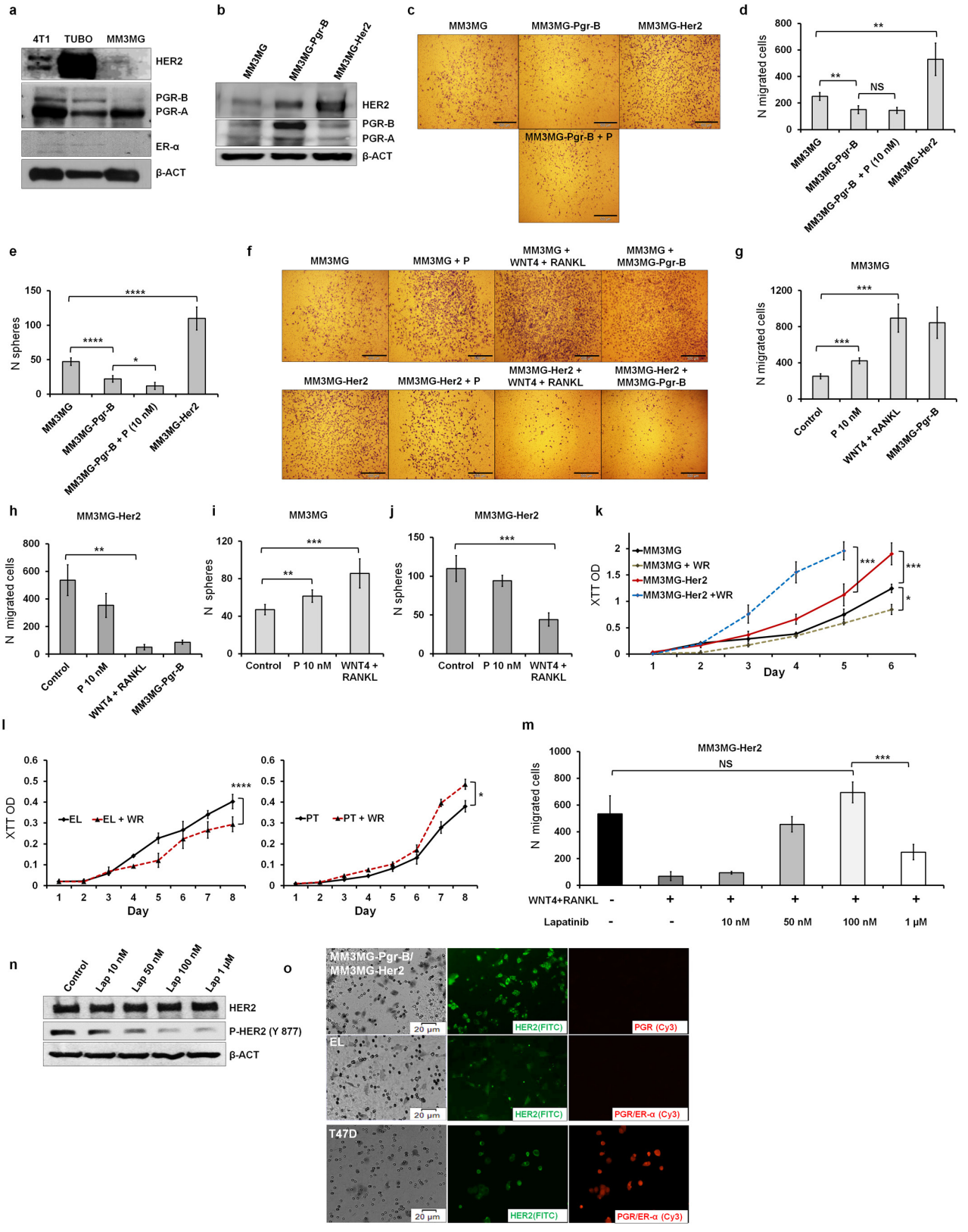
Gene symbol	Position in list	Up/down in EL	qPCR validation
Ahnak	21	up	Yes
Rbp1	25	down	Yes
Baz2a	27	up	Yes
Comp	46	down	Yes
Zfp408	235	up	No
Zc3hav1	436	up	Yes
Zfp780b	465	up	Yes
Nfatc3	467	up	Yes
Rxrb	720	up	Yes
Lpp	1030	up	Yes

Extended Data Figure 1 | Early lesion signature induction and expression of HER2 and PGR. a, The proportion of cancer cells disseminating to the bone marrow in BALB-NeuT mice decreases with increasing primary tumour area. DCCs were identified using anti-cytokeratin antibodies in bone marrow samples. The *y* axis displays the number of detected DCCs per 10^6 bone marrow cells divided by the total tumour area in mm^2. The number of mice used per data point is written below the graph. **b**, Laser microdissection of epithelial structures: two examples of 7–9-week-old BALB-NeuT mammary glands showing microdissection of regions with incipient epithelial hyperplasia. For all samples similar amounts of tissue (up to 100,000 μm^2) were isolated. **c**, qPCR validation of microarray profiles. qPCR was performed for 10 genes, upregulated or downregulated in microarray samples of early lesions and all changes, except one (*Zfp408*), were confirmed. **d**, qPCR for the mRNA level of all steroid hormone receptors (EL, early lesions; PT, advanced primary tumour; Met, lung metastasis). **e**, Primary cultures from mammary tissue of 7–9-week-old BALB-NeuT mice, were treated with different concentrations (1, 10, and 100 nM) of progesterone (P), oestrogen (E), aldosterone (A), cortisol (C), testosterone (T) or vehicle (ethanol; untreated) for 75 h. Only progesterone induces upregulation of the complete early lesion signature. **f**, Increased expression of PGR-B in young mammary glands (9-week old BALB/c and BALB-NeuT mice with early lesions compared to 25-week-old BALB/c mice), but not primary tumours correlates with increased HER2 expression. **g**, Progesterone induces the early lesion signature in 4T1 cells (highly aggressive and metastatic cell line derived from a spontaneous BALB/c mammary tumour), but not in 67NR (tumorigenic and non-metastatic cell line derived from a spontaneous BALB/c mammary tumour) and MM3MG cells (normal mammary epithelial cell line derived from a BALB/c mouse). **h**, Progesterone treatment induces upregulation of HER2 expression in 4T1 cells. **i**, Overexpression of PGR-B in MM3MG cells induces upregulation of HER2 expression. ****$P \leq 0.0001$ (Student's *t*-test and Stouffer's combined probability test); data are mean ± s.d. For gel source data, see Supplementary Fig. 1.



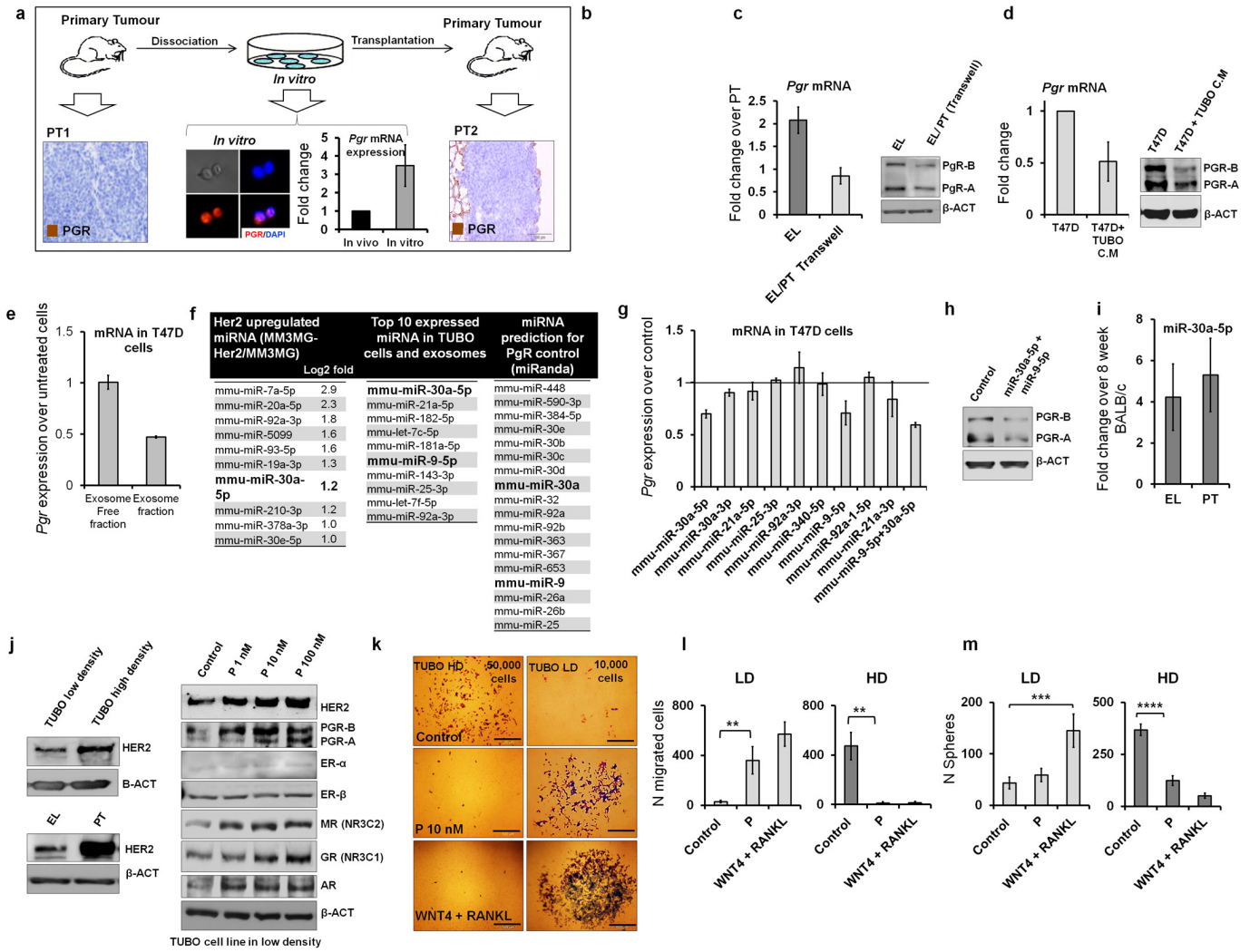
Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Progesterone regulates migration and is linked to branching morphogenesis. a, qPCR analysis of *Pgr*, *Rank* (also known as *Tnfrsf11a*), *Rankl* and *Wnt4* in normal and transgenic mouse mammary tissue or tumours. Note the increased expression of *Pgr*, *Wnt4* and *Rankl* in early lesions compared to primary tumours. Only *Rank* (the Rankl receptor) is strongly expressed in primary tumours. **b**, Primary cultures of early lesions treated with progesterone, WNT4, and RANKL. WNT4 and RANKL treatment induce the early lesion signature and act synergistically. **c**, **d**, PGR (green) staining at 5 and 12 weeks of age (FVB wild-type mice). Scale bars, 100 μm. The percentage of PGR^+ cells per duct was quantified ($n = 2$ mice per age group) in the distal and proximal portions of the gland (relative to the origin). PGR-expressing cells were more abundant in distal ducts (number of analysed ducts per group is displayed above each column). LN, lymph node.**e**, **f**, Photomicrograph of migration assay (**e**) and quantification of migrating cells (**f**) derived from fresh tissue (**e**, left) or dissociated spheres derived from primary tumours or early lesions (**e**, right; **f**, quantification). Progesterone, WNT4 and RANKL induce migration of early-lesion-derived but not primary-tumour-derived cells (see also Fig. 2a). **g**, Scheme of combined migration and sphere assay. The lower chamber is filled with serum-free sphere medium and the bottom is covered with poly-HEMA to prevent adhesion and enable sphere formation. After 72 h migration, the insert is removed and the lower chamber is analysed (after 11 days) for mammosphere formation (see Methods). **h**, Effect of oestrogen and progesterone on migration and sphere formation of mammary cells derived from early lesions. Cells were exposed to 10 nM oestrogen or progesterone or 10 nM oestrogen + 10 nM progesterone inhibitor (RU486). $*P \leq 0.05$; $**P \leq 0.01$; $***P \leq 0.001$; $****P \leq 0.0001$ (Student's *t*-test); Data are mean ± s.e.m. (**d**) or mean ± s.d. (other panels).

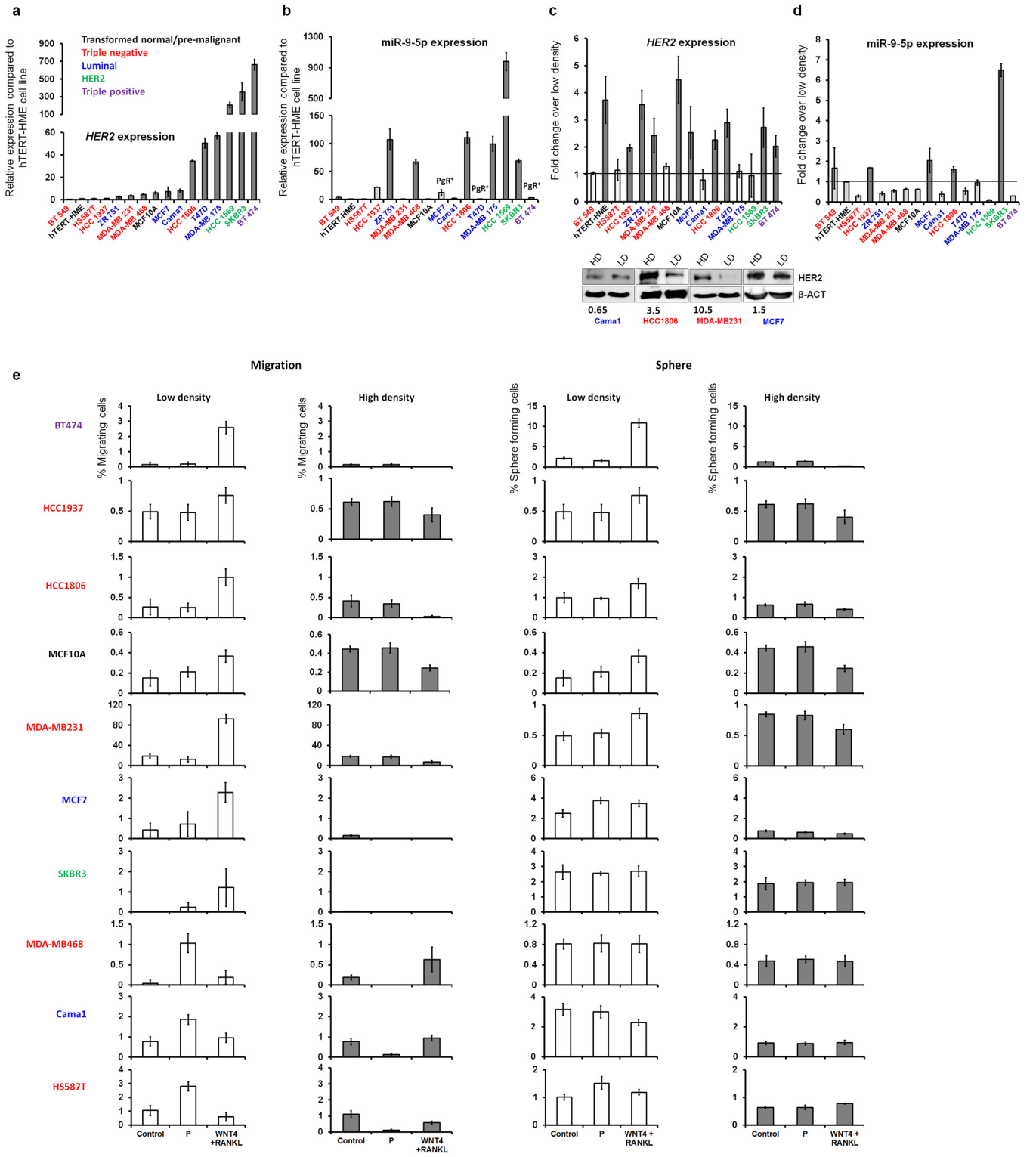


Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | HER2 expression levels regulate migration and proliferation. a, Parental MM3MG cells, a cell line derived from mammary epithelial cells of wild-type BALB/c mice, do not express ERα but express low levels of HER2 and PGR-B compared to 4T1 and TUBO cells (TUBO cells were grown at low density; see Fig. 3 and Extended Data Figs 4, 5). **b**, Immunoblot confirming successful transduction of the MM3MG cell line with *Her2* and *Pgr-B*. Note that transduction of PGR-B increases HER2 levels. **c**, **d**, Overexpression of *Pgr-B* in MM3MG cells (MM3MG-Pgr-B) reduces migration, whereas *Her2* overexpression (MM3MG–Her2) increases migration of cells. Addition of progesterone does not alter migration of *Pgr-B*-overexpressing cells (MM3MG–Pgr-B). **e**, Overexpression of *Pgr-B* in MM3MG cells reduces sphere formation, whereas *Her2* overexpression increases sphere formation. **c**–**e,** Migrating/sphere forming cells are not from the PGR^{+}, but the PGR^{-} population, which are responsive to PIPS. **f**–**j**, To investigate which PGR^{-} cells were the target population of progesterone signalling, we exposed parental MM3MG cells and *Her2*-transduced cells to progesterone, PIPS or mixed the cells with PGR^{+} cells (only for migration experiments). Progesterone, WNT4 and RANKL, and co-culture with MM3MG–Pgr-B induced sphere formation and migration of MM3MG cells, but decreased these responses in MM3MG–Her2 cells. **k**, Overexpression of *Her2* increases proliferation of MM3MG cells (MM3MG–Her2). WNT4 and RANKL (WR) further increase proliferation of MM3MG–Her2 cells, but decreases proliferation of the parental (MM3MG) cells. Therefore based on expression of HER2, cells either migrate ($HER2^{low/-}$) or proliferate ($HER2^{high}$). **l**, WNT4 and RANKL treatment induces proliferation of primary cultured cells derived from primary tumours, but reduces it in cells derived from early lesions. **m**, Reduction of HER2 signalling by lapatinib overrides the inhibitory effect of WNT4 and RANKL, and increases migration in MM3MG–Her2 cells. However, strong inhibition of HER2 signalling reduces migration. **n**, Lapatinib inhibits HER2 signalling by preventing phosphorylation. **o**, Cells that migrated through the pores of the migration chamber insert were stained for HER2 (FITC, green) and PGR or ERα (Cy3, red). In 1:1 co-culture of MM3MG–Pgr-B and MM3MG–Her2 (top) only HER2-expressing cells migrate. Migrated primary cells derived from early lesions (middle) do not express PGR and display faint HER2 staining (brightness of HER2 and PGR staining increased by 50% for better visibility). HER2 and PGR double-positive T47D cells fixed onto the filters of migration chambers served as positive control of staining. **m**–**o**, Cells with low/intermediate signalling of HER2 show the highest response in migration and sphere formation induction by PIPS. $*P \leq 0.05$; $**P \leq 0.01$; $***P \leq 0.001$; $****P \leq 0.0001$; NS, not significant; (*F*-test of the slope (**k**, **l**) or Student's *t*-test (other panels)); data are mean ± s.d. For gel source data, see Supplementary Fig. 1.

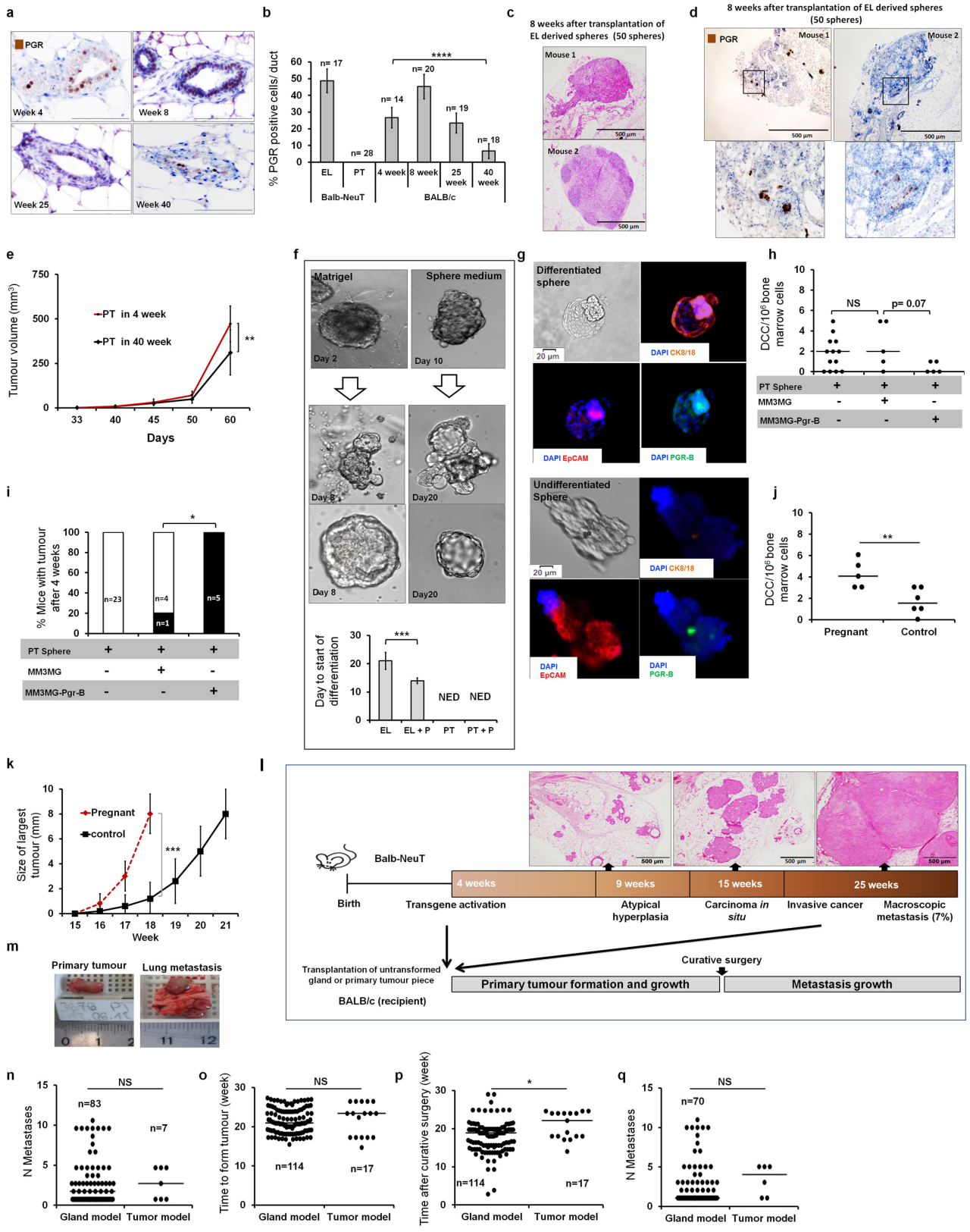


Extended Data Figure 4 | Cell density and regulation of PGR and HER2 signalling. a, PGR expression silenced in tumours can be re-activated in culture and re-silenced *in vivo*. **b**, PGR re-expression only occurs in TUBO cells grown at low density and high density after frequent medium change. **c**, Downregulation of PGR in early lesion cells cultured in a transwell assay next to primary tumour cells suggests the existence of a secreted factor passing through the membrane of the transwell insert and downregulating *Pgr* mRNA (left) and protein (right). **d**, T47D cells exposed to conditioned medium from TUBO cells display reduced *PGR* mRNA (left) and protein (right). **e**, Exosomes derived from the cell culture medium of TUBO cells grown at high density (exosome fraction) induce downregulation of PGR in T47D cells. **f**, miRNA sequencing to identify PGR regulating miRNAs. Left, top 10 upregulated miRNAs in *Her2*-overexpressing cells (MM3MG–Her2) compared to control (MM3MG). Middle, top 10 expressed miRNAs in TUBO cells and TUBO-cell-derived exosomes. Right, miRNAs predicted by the miRANDA web software to regulate *Pgr* **g**, Among all candidate miRNAs only miR-30a-5p and miR-9-5p induce downregulation of *Pgr* mRNA in T47D cells. **h**, Downregulation of PGR in T47D cells treated with miR-30a-5p and miR-9-5p. **i**, Expression of miR-30-5p in early lesions and primary tumour samples compared to mammary glands from 8-week-old BALB/c mice. **j**, Density induced upregulation of HER2 in TUBO cells grown at low or high density and early lesions compared to primary tumour samples (left) and progesterone-responsiveness of low-density TUBO cells (right). Note that levels of HER2, PGR, the glucocorticoid (GR), mineralocorticoid (MR) and androgen receptors (AR) are regulated by progesterone in a dose-dependent manner except ER-α (see related result in Extended Data Fig 1d). **k**–**m**, TUBO cells grown at low density and exposed to progesterone or PIPS migrated more (**k**, **l**) and produced more spheres (**m**), similar to cells derived from early lesions (see Fig. 2). $^{**}P \leq 0.01$; $^{***}P \leq 0.001$; $^{****}P \leq 0.0001$(Student's *t*-test); data are mean ± s.d. For gel source data, see Supplementary Fig. 1.



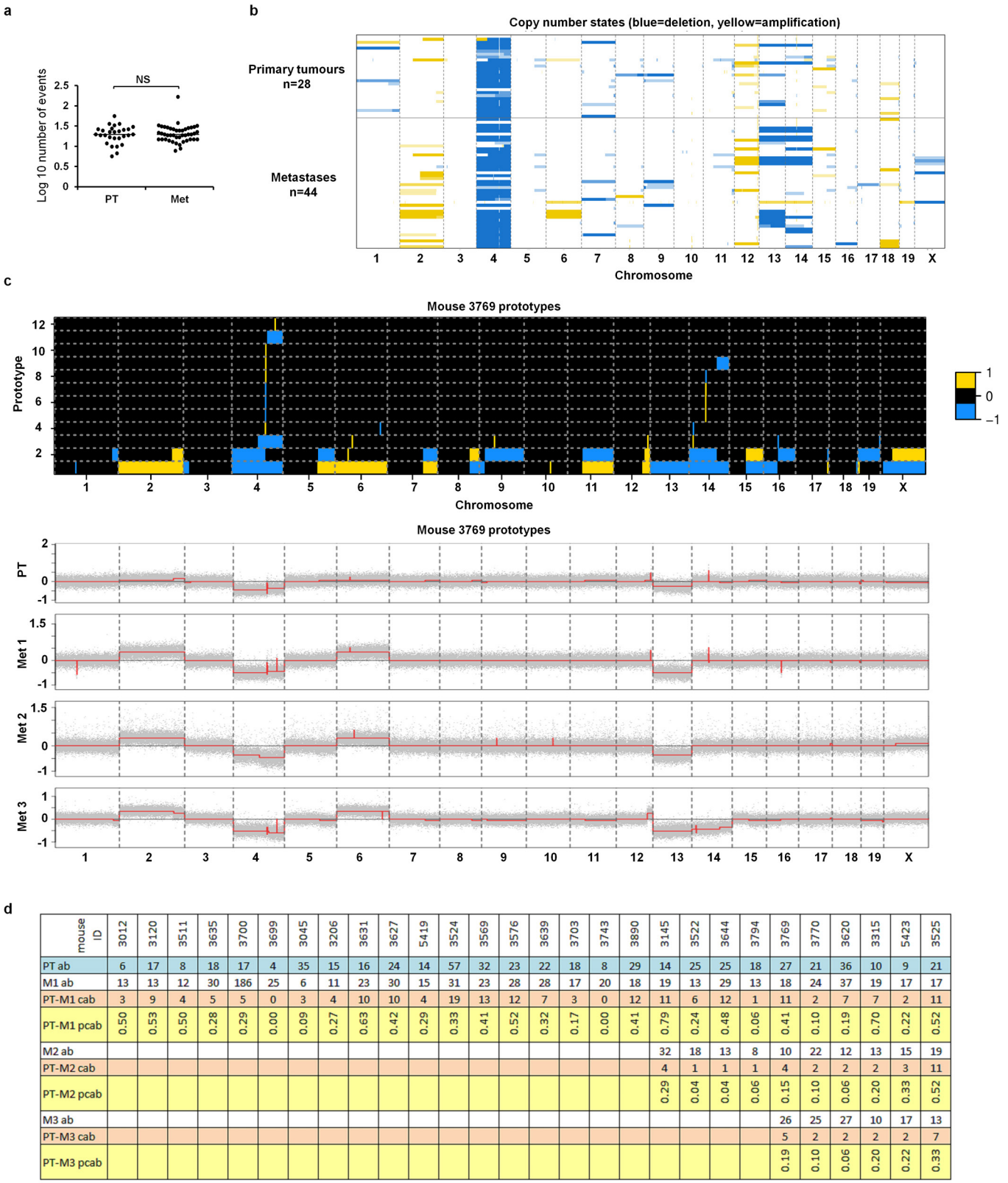
Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Cell density and regulation of PGR and HER2 signalling in human cell lines. To investigate whether human breast cancer cells display similar regulatory circuits as found in mouse cells, we selected 16 cell lines of different breast cancer subtypes. **a**, *HER2* mRNA expression levels in 15 human cell lines compared to the hTERT-HME cell line. Different colours indicate the subtype of breast cancer of each cell line. **b**, The expression of miR-9-5p in human breast cancer cell lines compared to hTERT-HME cells. Note that cell lines highly expressing *HER2* (see **a**) express more miR-9-5p similar to primary tumours of BALB-NeuT and TUBO cells (see Fig. 3b, c), whereas two $HER2^{high}$/PGR^{high} cell lines, BT474 and T47D, do not express miR-9-5p similar to human $HER2^{high}$/PGR^{high} samples (see Fig. 5d). **c**, High cell density leads to upregulation of *HER2* mRNA (top) or protein levels (bottom) in several cell lines. Only four cell lines were analysed for protein level (HER2 level was not influenced by cell density in CAMA1; HER2 level regulated by cell density in HCC1806, MDA-MB-231 and MCF7). Numbers below the blots denote fold change of HER2 in high density compared to low density normalized over β-actin. **d**, Expression of miR-9-5p is upregulated by cell density in SKBR3, HCC1937, HCC1806 and MCF7 cell lines. **e**, Migration and sphere-forming potential of 10 out of 16 cell lines grown at low and high densities, and treated with PIPS or progesterone. The first 7 cell lines regulate *HER2* transcripts by density (see **c**) and their response to PIPS is similar to TUBO cells and primary mammary cell cultures of BALB-NeuT mice (see Fig. 2 and Extended Data Fig. 4l, m). The remaining three cell lines do not regulate *HER2* transcripts by cell density, but respond to progesterone similarly to the TUBO cell line and primary mammary cell cultures of BALB-NeuT mice (see Fig. 2 and Extended Data Fig. 4l, m). We did not perform functional assays with BT549 (triple negative subtype), T47D (luminal, MCF7-like), MDA-MB-175, ZR75-1 (luminal, CAMA1-like), hTERT-HME (transformed normal, similar to MCF10A) because of breast cancer subtype redundancy or poor growth (HCC1569). *y* axes show the percentage of migrating cells (left) and observed spheres (right) relative to seeded cells. Data are mean ± s.d. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | Differentiation ability and metastasis formation. a, **b**, Representative images (**a**) and quantification (**b**) of PGR expression of mammary epithelial cells from wild-type BALB/c mice at 4, 8, 25 and 40 weeks of age. Scale bars, 100 μm. PGR expression was reduced by 75% in 40-week-old wild-type mammary gland compared to 4-week-old mice and disappeared in primary tumours (see also Extended Data Figs 1f, 2a). *n*, number of ducts or glands (in early lesions and normal tissue) or visual fields in primary tumours. **c**, **d**, Representative micrographs of lesions 8 weeks after transplantation of early lesion spheres resembling DCIS (**c**) or less-advanced early lesions (**d**) displaying PGR expression (brown nuclear staining). **e**, Tumour growth from primary tumourspheres in young and old recipients. **f**, Differentiation of cells from early lesions but not of primary tumour cells in Matrigel (left) or in sphere culture (right) into acinus-like structures. Progesterone stimulation accelerated formation of acinus-like structures by early lesion cells, under mammosphere conditions. NED, no evidence of differentiation. **g**, Staining for CK8/18, PGR and EpCAM (epithelial cell adhesion molecule) shows expression of PGR and CK8/18 only in differentiated structures (top) compared to undifferentiated spheres (bottom). **h**, **i**, Primary tumourspheres were transplanted alone ($n = 23$) or co-transplanted with MM3MG–Pgr-B ($n = 5$) or MM3MG spheres ($n = 5$). DCCs numbers in bone marrow (**h**) and the number of mice with tumours (**i**) were checked 4 weeks later. *Pgr-B*-transduced mammary epithelia suppressed metastatic dissemination and accelerated tumour formation from primary tumourspheres. **j**, Pregnancy at the early lesion stage induces dissemination. A group of young BALB-NeuT mice mated ($n = 5$) at the early lesion stage (7-week-old) and were killed at the end of pregnancy. These mice did not form palpable tumours, but had a higher number of DCCs compared to unmated control mice ($n = 6$). **k**, Pregnancy at the advanced tumour stage. A group of BALB-NeuT mice ($n = 5$) were mated at the time of *in situ* carcinoma (15-week-old) and killed at the end of pregnancy. All pregnant mice had faster growing tumours compared to unmated control mice. **l**, Schematic of transplantation protocol for mammary gland or primary tumour tissue pieces into wild-type recipients. **m**, Example of primary tumour and macro-metastasis assessment. **n**, Number of metastatic foci in transplanted mice. 18 mice from the gland model and 3 mice from the tumour model were excluded from analyses owing to the fusion of metastatic lesions, which made it difficult to count individual lesions. **o**, Similar growth kinetics of primary tumours from gland and tumour piece models for samples from the red box in Fig. 4e. **p**, **q**, Mice from **o** were compared for the duration of the follow-up period after surgery. Mice from both groups were killed at the first signs of general health deterioration, which occurred earlier in gland-model mice (**p**). Longer follow-up time after curative surgery did not result in more metastases in recipients transplanted with primary tumour pieces (**q**). $*P \leq 0.05$; $**P \leq 0.01$; $***P \leq 0.001$; $****P \leq 0.0001$; NS, not significant; (Student's *t*-test (**b**, **f**, **h**, **j**); Fisher's exact test (**i**); *F*-test for the slopes (**c**, **k**); Mann–Whitney *U*-test (**n**–**q**)). Data are shown as mean ± s.d. (**b**, **e**, **f**, **k**) or median (**h**, **j**, **n**–**q**).

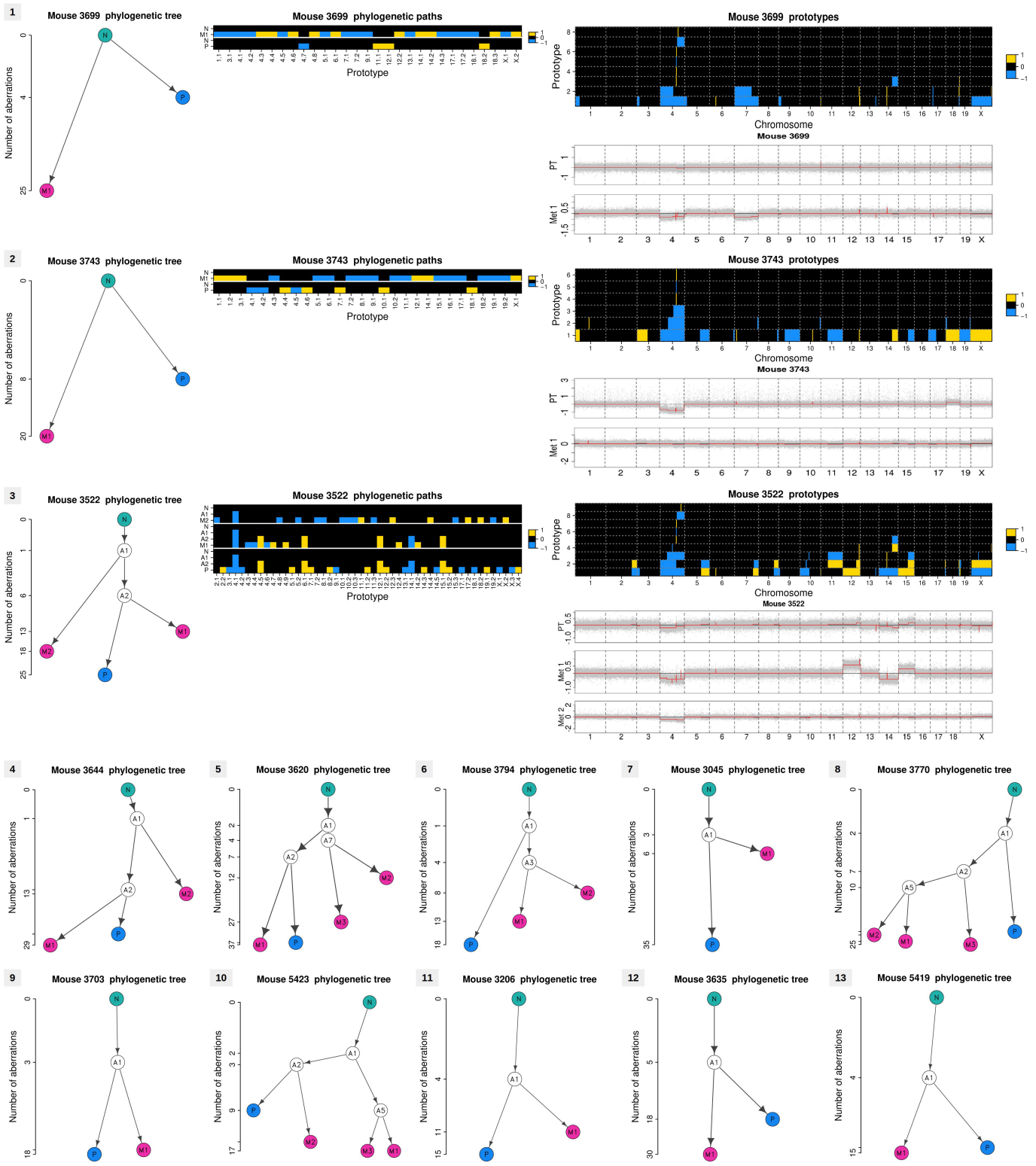


d

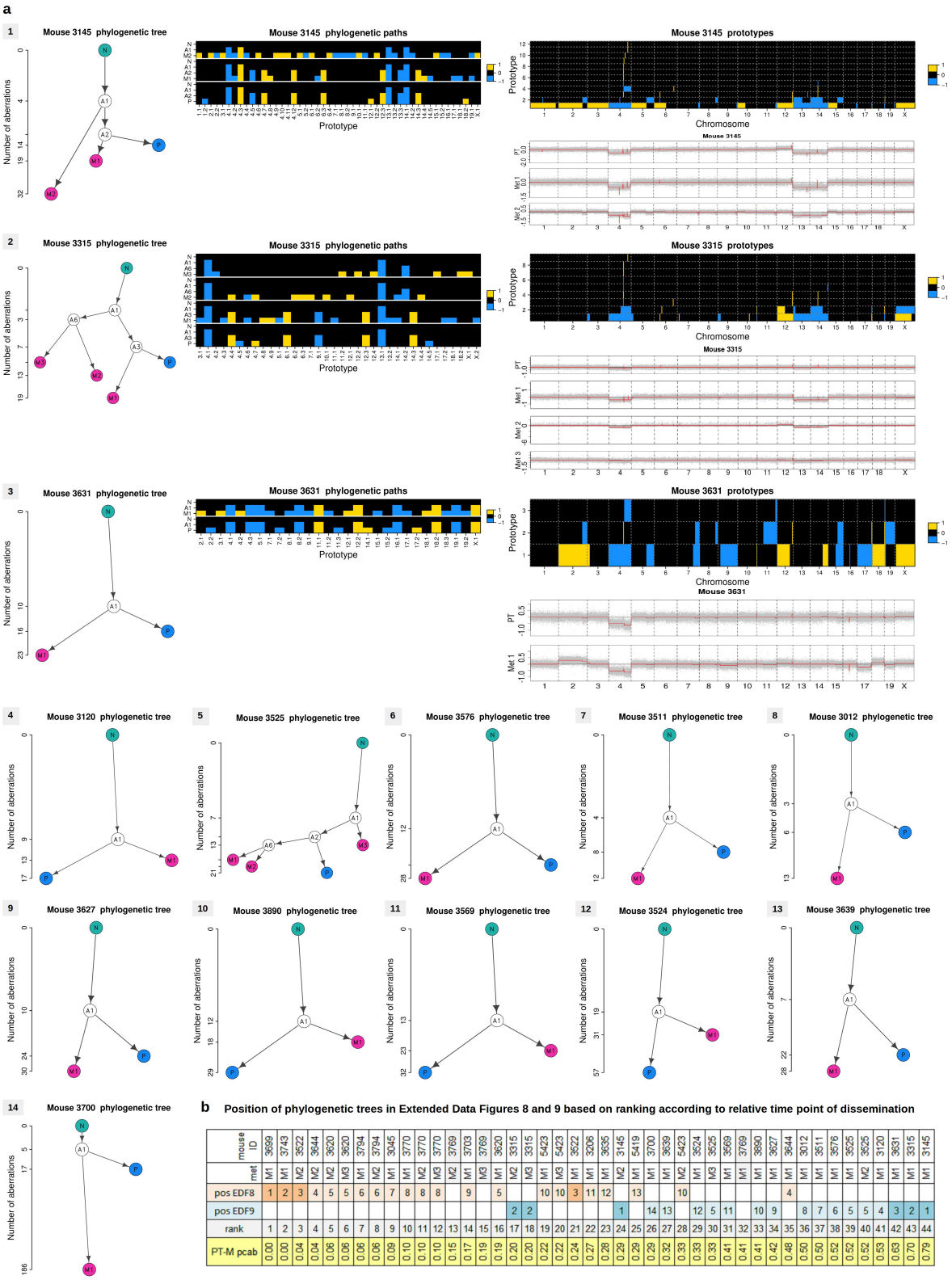
mouse ID	3012	3120	3511	3635	3700	3699	3045	3206	3631	3627	5419	3524	3569	3576	3639	3703	3743	3890	3145	3522	3644	3794	3769	3770	3620	3315	5423	3525
PT ab	6	17	8	18	17	4	35	15	16	24	14	57	32	23	22	18	8	29	14	25	25	18	27	21	36	10	9	21
M1 ab	13	13	12	30	186	25	6	11	23	30	15	31	23	28	28	17	20	18	19	13	29	13	18	24	37	19	17	17
PT-M1 cab	3	9	4	5	5	0	3	4	10	10	4	19	13	12	7	3	0	12	11	6	12	1	11	2	7	7	2	11
PT-M1 pcab	0.50	0.53	0.50	0.28	0.29	0.00	0.09	0.27	0.63	0.42	0.29	0.33	0.41	0.52	0.32	0.17	0.00	0.41	0.79	0.24	0.48	0.06	0.41	0.10	0.19	0.70	0.22	0.52
M2 ab																			32	18	13	8	10	22	12	13	15	19
PT-M2 cab																			4	1	1	1	4	2	2	2	3	11
PT-M2 pcab																			0.29	0.04	0.04	0.06	0.15	0.10	0.06	0.20	0.33	0.52
M3 ab																							26	25	27	10	17	13
PT-M3 cab																							5	2	2	2	2	7
PT-M3 pcab																							0.19	0.10	0.06	0.20	0.22	0.33

Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Array CGH analysis of primary tumour–metastasis pairs. a, Number of aberrations detected by array comparative genomic hybridization in primary tumours and matched lung metastases. Dot plot with median; statistical analysis by Mann–Whitney *U*-test. **b**, Heatmap of copy number states for the 28 primary tumours and 44 matched metastases across chromosomes 1–19 and X. Light, medium and dark yellow or blue colours indicate weak, intermediate and strong amplification (yellow) or deletion (blue) amplitudes, respectively (thresholds at $\pm 0.1, \pm 0.2, \pm 0.3$). **c**, Prototype aberrations (top) constructed from segmented array CGH profiles (bottom) of the primary tumour (PT) and the matched metastases (Met 1–3) of mouse 3769 (phylogenetic tree and phylogenetic paths displayed in Fig. 4h, i). Prototypes (top) are organized in stacked rows per chromosome and numbered according to chromosome and positional order of their first change point, for example, 1.2 denotes the second prototype of chromosome 1. These prototype aberrations are then used to construct the phylogenetic paths (for example, Fig. 4i) and trees. For better visibility, small focal aberrations were enlarged to have a minimal extension of 300 probes. Yellow, amplification (+1); blue, deletion (−1). Corresponding segmentation profiles of the normalized and wavy-pattern-corrected array CGH data (grey dots) are indicated by red lines (bottom). For segmentation and prototype construction, see Methods. **d**, Table for calculating the relative time points of dissemination (Fig. 4j). PT ab, number of aberrations in the primary tumour; M_k ab, number of aberrations in the matched metastases ($k = 1, 2, 3$); PT–M_k cab, number of common aberrations between primary tumours and metastases; PT–M_k pcab, proportion of common aberrations relative to the primary tumour, that is, pcab = cab/PT ab.



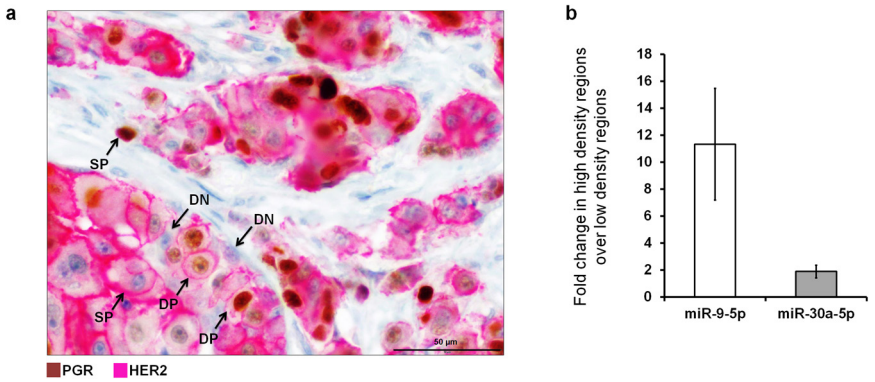
Extended Data Figure 8 | Phylogenetic analysis of metastasis (early divergence). Phylogenetic trees of the top 13 out of 28 primary tumours and matched metastasis samples listed according to earliest time point of dissemination (for details see tables in Extended Data Figs 7, 9). A1–7, inferred common ancestors (intermediates); M, metastases (M1–3); N, normal cell; P, primary tumour. (see Methods). The ordinate indicates the number of aberrations per profile (on a square root scale). For the first three matched samples, the phylogenetic tree paths (middle), prototype aberrations (top right) and segmented array CGH profiles (bottom right) are also shown in addition to the phylogenetic tree (left). Aberration profiles along phylogenetic paths run from N via A1–7 to P or M1–3. Aberration prototypes are named according to chromosome and positional order of their first change point, for example, 2.2 denotes the second aberration prototype of chromosome 2 (see Methods and Extended Data Fig. 7c).



b Position of phylogenetic trees in Extended Data Figures 8 and 9 based on ranking according to relative time point of dissemination

mouse ID	3689	3743	3622	3644	3620	3620	3794	3794	3045	3770	3770	3770	3769	3703	3769	3620	3315	3315	5423	5423	3522	3206	3635	3145	5419	3700	3639	5423	3524	3525	3569	3769	3890	3627	3644	3012	3511	3676	3525	3525	3120	3631	3315	3145
met	M1	M1	M2	M2	M2	M3	M1	M2	M1	M1	M2	M3	M2	M1	M3	M1	M2	M3	M1	M3	M1	M1	M1	M2	M1	M1	M1	M2	M1	M3	M1	M1	M1	M1	M1	M1	M1	M1	M2	M1	M1	M1	M1	M1
pos EDF8	1	2	3	4	5	5	6	6	7	8	8	8		9		5			10	10	3	11	12		13			10							4										
pos EDF9																	2	2						1		14	13		12	5	11		10	9		8	7	6	5	5	4	3	2	1
rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44
PT-M pcab	0.00	0.00	0.04	0.04	0.04	0.06	0.06	0.06	0.06	0.09	0.10	0.10	0.10	0.15	0.17	0.19	0.19	0.20	0.20	0.22	0.22	0.24	0.27	0.28	0.29	0.29	0.29	0.32	0.33	0.33	0.33	0.41	0.41	0.41	0.42	0.48	0.50	0.50	0.52	0.52	0.52	0.53	0.63	0.70	0.79

Extended Data Figure 9 | Phylogenetic analysis of metastasis (late divergence). a, Phylogenetic trees of the top 14 out of 28 primary tumour and matched metastasis samples listed according to latest time point of dissemination. For the first three mice the phylogenetic paths (mid), prototype aberrations (top right) and segmented array CGH profiles (bottom right) are also shown next to the phylogenetic tree (left). See Methods and Extended Data Figs 7, 8. **b**, Summary table of all phylogenetic analyses indicates the position of the corresponding mouse phylogenetic tree in Extended Data Figs 8 and 9 (EDF8 and EDF9) for each primary tumour–metastasis pair. The two bottom rows indicate the rank and the corresponding relative time point of dissemination as measured by the proportion of aberrations shared between primary tumour and metastasis (PT–M pcab; see Extended Data Fig. 7 and Fig. 4j). Note that only metastases ranked on position 36–44 diverged late. The phylogenetic tree and phylogenetic paths for mouse 3769 are displayed in Fig. 4h, i. In the pos-EDF8 and pos-EDF9 rows the darker colours are samples of which all data, including phylogenetic paths, prototype aberrations, segmented array CGH profiles and phylogenetic trees are shown. Faint colours cells in pos-EDF8 and pos-EDF9 are samples where only phylogenetic trees are shown.



Extended Data Figure 10 | PGR, HER2 signalling and dissemination in breast cancer patients. a, Double staining of a HER2^{high}/PGR^{high} human breast cancer sample (PGR, brown, nucleus; HER2, red/pink, membrane). Cells with varying expression levels of HER2 and PGR, as well as negative, single- or double-positive cells can be seen. Scale bar, 100 µm. Arrows indicate, double-positive (DP), double-negative (DN) and single-positive (SP) cells. **b**, Lack of PGR expression in high-density areas of HER2^{high}/PGR^{high}-classified tumour samples (see Fig. 5c) is directly linked to high miR-9-5p and miR-30a-5p expression. Data are mean ± s.d.