Genetic determinants of cancer metastasis

Don X. Nguyen and Joan Massagué

Abstract | Metastasis can be viewed as an evolutionary process, culminating in the prevalence of rare tumour cells that overcame stringent physiological barriers as they separated from their original environment and developmental fate. This phenomenon brings into focus long-standing questions about the stage at which cancer cells acquire metastatic abilities, the relationship of metastatic cells to their tumour of origin, the basis for metastatic tissue tropism, the nature of metastasis predisposition factors and, importantly, the identity of genes that mediate these processes. With knowledge cemented in decades of research into tumour-initiating events, current experimental and conceptual models are beginning to address the genetic basis for cancer colonization of distant organs.

Primary tumour
Cancer that arises from the malignant conversion of cells from an initial organ site.

Histological grade
Morphologically identifiable steps of tumour progression that are used to classify disease stage.

Organ tropism
A predilection of a primary tumour to spread to specific secondary organs.

Angiogenesis
The pathological growth of new blood vessels to support tumour growth.

Intravasation
Entry of tumour cells into the bloodstream.

Cancer Biology and Genetics Program, and Howard Hughes Medical Institute, Box 116, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York 10021, USA.
Correspondence to J.M.
e-mail: massaguj@mskcc.org
doi:10.1038/nrg2101

As they abandon their original niche, most tumour cells die from various stresses caused by the departure from their original microenvironment, or the unfavourable surroundings that they encounter in distant organs. Our organs and tissues are highly evolved systems that, once their development is complete, do not support the outgrowth of solitary cells that wander in from elsewhere (bone-marrow-derived progenitor cells and immune cells being exceptions). However, because of their genetic instability and heterogeneity, tumours have a finite chance of generating cells with phenotypic attributes that are sufficient to withstand and exploit otherwise incompatible microenvironments.

Metastasis causes 90% of deaths from solid tumours and displays a remarkably diverse set of clinical manifestations. In breast cancer, metastases can be undetectable and can remain latent for many years following primary tumour removal, only to emerge as incurable lesions that are triggered by unknown causes. By contrast, pancreatic cancer and small-cell lung cancer often present with widespread metastases at the time of initial diagnosis. Although lethal, glioblastoma remains locally invasive, seldom metastasizing outside the CNS. In paediatric neuroblastomas, metastasis can spontaneously regress, for mysterious reasons. The risk of metastatic recurrence can sometimes be predicted from certain features of the primary tumour, as in the case of breast cancer, for which tumour size, histological grade, and gene-expression pattern are indicative of the risk of relapse[1]. In other instances, metastatic disease presents itself with an unknown primary tumour source, rendering such correlations impossible to make[2].

One of the most intriguing biological aspects of metastasis is the pattern of organ dissemination. The spread of tumour cells frequently includes invasion of local lymph nodes through the lymphatic system. However, aggressive tumour cells typically enter the bloodstream and reach distant tissues (BOX 1; FIG. 1). This dissemination has stereotypical patterns of organ tropism that reflects the heterogeneity of tumour cells and depends on the cancer type (TABLE 1). In addition, patients can harbour simultaneous metastases in different organs, and for each patient it is unknown if such lesions have arisen from the same primary tumour (synchronous seeding), or whether they result from a sequence of seeding events in which some metastases themselves subsequently colonize other organs (metachronous seeding)[3]. Moreover, disseminated tumour cells could theoretically also return to their original site (self-seeding), thereby accounting for the progressive accumulation of aggressive cells in primary tumours and local recurrences[4].

Work over the past three decades has identified many genes for which gain or loss of function confers autonomous proliferative activity, resistance to cell death cues, angiogenesis, altered cell adhesion and motility[5]. In other words, genetic alterations that mediate the initiation and local progression of tumours, and that collectively confer the prerequisites for metastasis, have been identified. However, we are only beginning to learn about the genetic determinants of metastasis proper; that is, those that mediate tumour cell invasion, intravasation, survival in circulation, scattering to distant tissues, extravasation into parenchyma, and colonization of vital organs (BOX 1; FIG. 1). But unlocking the nature and function of these

Extravasation
Exit of tumour cells out of capillary beds into the parenchyma of an organ.

Parenchyma
The main functional portion of an organ.

Emboli
Clumped tumour cells that typically lodge in blood vessels.

Haematogenous dissemination
The spread of cancer cells through the bloodstream.

determinants is an endeavour with clear clinical ramifications, as the roles of these genes will dictate whether lethal metastasis will ensue. Recent reviews have comprehensively analysed the biological traits and steps of cancer metastasis[6–10]. On the basis of this conceptual framework, we here provide a synthesis and perspective on the ongoing search for its genetic determinants.

Metastasis theories: seeds, soils and signatures

Already noted by the Egyptians of 1500 BC, malignant disease was described with remarkable coherence by Greek physicians in the time of Hippocrates (fifth century BC), who saw axillary swelling as a sign that breast tumours would be followed by secondary growths of fatal outcome[11]. However, such lesions were considered as independent malignancies that arose from the spread of toxic humours. The humoral theory prevailed until the eighteenth century and beyond, when the discovery of the cell as the basic unit in live tissue stimulated new debate, with some viewing metastases as degenerate cell components arising from a primary tumour[3].

Congenial soils. By the end of the nineteenth century, the cellular origin of cancer attained credence, and the unique patterns of organ colonization by different tumour types now caught attention. Fuchs and others noted that the striking preference of various tumours to metastasize to certain organs must imply a hospitable environment in those tissues[12]. Building on this, Paget proposed that tumour cells ('seeds') must be predisposed to arrest and proliferate only in those anatomical sites ('soil') that provide a congenial ground — a basic tenet of his seed-and-soil hypothesis[12]. Ewing later challenged this concept, postulating that the direction of circulation alone influences the distribution of metastasis[13]. The importance of vasculature was further supported by mechanical distinctions between lymphatic and haematogenous dissemination[14]. Eventually, these disparate views were reconciled in the proposal that circulatory routes influence the distribution of cancer cells, but the eventual outgrowth of macrometastases depends on the seeding of compatible tissues[7].

Darwinian concepts. Inspired by evolutionary theory, neoplastic development started to be viewed as a process of natural selection, with genomic instability producing variant cell populations from which aggressive clones emerge[15,16]. Consistent with this view, experiments using melanoma cells that were repeatedly inoculated into mice intravenously and the resulting metastatic colonies harvested from the lungs produced cell variants with different grades of metastatic proclivity[17]. Notably, pre-existing clones from the initial cell line already displayed marked variability for metastasis *in vivo*, suggesting that metastatic progression might require the acquisition of additional genetic alterations that confer a selective advantage to rare clones in the tumour cell population[18].

Over the past three decades, identification of oncogenes and tumour suppressor genes, as well as analysis of their mutation frequency during tumour development, have given rise to a model whereby sequentially acquired abnormalities mediate the transitions between progressive tumour stages[19]. By contrast, the identity of genetic alterations that would specifically mediate metastasis has remained rather elusive[20]. Indeed, questions have been raised as to the extent to which metastasis depends on the acquisition of further genetic abnormalities beyond those that turn an incipient tumour into a locally aggressive malignancy[21].

Pre-determining gene-expression signatures. Technological advances derived from the sequencing of the human genome have recently enabled the genome-wide analysis of tumours, providing ways to scrutinize the complex process of metastatic progression. One major advance has been the application of gene-expression profiling to tumour classification and prognosis, initially

Box 1 | The metastasis process and its requirements

Prerequisites for metastasis

To metastasize, malignant cells must fulfil certain tumorigenic functions that become prerequisites for metastasis. These include: unlimited proliferation, evasion of cell-intrinsic and environmental constraints, attraction of a blood supply, and the capacity to detach and move away from the original location. Tumour-initiating mutations, alterations that are secondary to genomic instability and epigenetic changes, underlie the acquisition of these functions in developing tumours. As tumours grow, they must selectively conquer environmental pressures including cytotoxic immunity, low oxygen tension and an acidic environment. Many prerequisite tumorigenic properties have been discussed elsewhere[5,6,8,117]. Such functions are acquired during tumour initiation and local development, but must remain active throughout malignant progression, as they favour the emergence of metastasis-prone tumour cells in the context of local and distant microenvironments.

Metastatic initiation and dissemination

Dissemination starts in earnest when aggressive tumour cells become invasive and readily enter the bloodstream through the vasculature that they have attracted. The fenestrated, pathologically leaky vasculature that is typical of many tumours can facilitate the entry of cells into the circulation. Intravasation is also enhanced by an epithelial-to-mesenchymal transition that endows carcinoma cells with embryonic plasticity and added motility[33,34]. The rate of malignant cell shedding generally increases with tumour size, but dissemination can occur in the early stages of the primary tumour, to the point that some metastases have no known primary source. Surviving the stresses of the bloodstream requires tumour-intrinsic functions for the evasion of detachment-triggered cell death (anoikis), as well as co-option of extrinsic accomplices such as platelet aggregates[118].

Metastatic colonization

The dissemination of tumour cells to various organs is influenced by circulation patterns and the mechanical lodging of tumour cells in capillary beds. Adhesion molecules that mediate attachment to capillary walls and chemokine sources that stimulate survival and proliferation of tumour cells that express suitable receptors also have important roles. Metastatic cells enter the parenchyma of a target organ by breaching the capillaries in which they are embedded, either by vascular-remodelling events that allow transmigration across the capillary wall, or as a result of mechanical disruption of capillaries by expanding tumour emboli. On entry, tumour cells from one organ are confronted by a different microenvironment, which they must survive and eventually overtake. Recruited bone-marrow-derived progenitor cells might jump-start this process by providing a permissive niche for metastasis. Active colonization can be achieved by the co-option of organ-specific components of the tumour microenvironments, such as the activation of bone-resorbing osteoclasts by breast cancer cells during osteolytic metastasis[9]. Full metastatic colonization can occur by immediate growth of tumour cells upon their extravasation, or after a prolonged period of micrometastatic dormancy. The survival of dormant lesions, their eventual activation and outgrowth requires additional tumour-intrinsic or tumour-extrinsic factors. As each organ presents a highly specialized microenvironment, distinct sets of functions are thought to be required for colonizing different tissues[8,10].

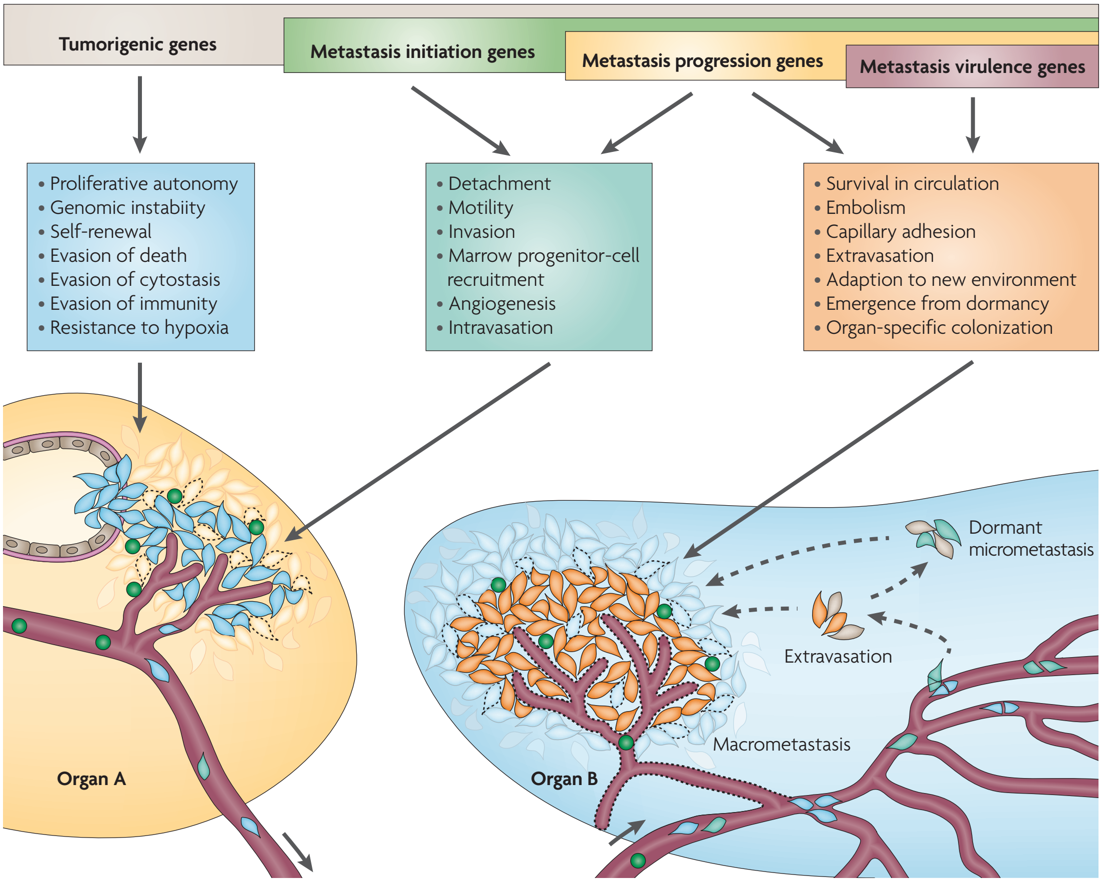

Figure 1 | **Classes of genes participating in the metastasis process.** On the basis of their level of participation in the metastatic process, we distinguish three general classes of metastasis genes. 'Metastasis initiation' genes are those that provide an advantage in primary tumours, paving the way for tumour cells to enter the circulation. 'Metastasis progression' genes are those that fulfil certain rate-limiting functions in primary tumour growth, and other specific functions in metastatic colonization. 'Metastasis virulence' genes are those that provide a selective advantage in secondary sites but not in the primary tumour, thus participating in metastatic colonization but not in primary tumour development. These genes act to promote metastasis after tumorigenic genes have given rise to a transformed, locally aggressive tumour cell population. Metastasis genes might act from the tumour cells in the primary site (blue cells) or their metastatic derivatives in a distant organ (orange cells). Metastasis genes might also act from the tumours microenvironment in the primary site or the metastatic site (light-coloured cells), some of which (dotted-contour cells) are co-opted by the growing tumour. Tumorigenic and metastatic functions also include the recruitment of distant accomplices, such as bone-marrow-derived progenitor cells, endothelial cells and macrophages (collectively represented by green cells) that support tumour growth and metastasis development. See BOX 1 for further details.

in haematological malignancies[22,23] and more recently in solid tumours[24–27]. Bioinformatic analysis of such data sets from primary tumour samples has uncovered complex gene-expression patterns, or 'signatures', that can predict the risk of metastatic recurrence (BOX 2). The detection of such traits in primary tumours would be consistent with the hypothesis of clonal dominance, according to which rare malignant subpopulations must become overgrown in the primary tumour if they are to reach ample cell numbers for metastasis[28]. Nonetheless, these findings have also been interpreted as evidence that the metastatic predisposition of tumours might already be fulfilled by the traits of a primary malignancy. According to this model, metastasis would be largely pre-determined by the combined action of mutations that give rise to a locally aggressive primary tumour[21]. In this case, the acquisition of further pro-metastatic changes by rare variant clones would have only a complementary role.

Towards an integrative view. A model of metastatic progression that is solely based on pre-determined traits would not be readily compatible with the continuous nature of somatic evolution, nor would it account for additional events whereby disseminated tumour cells that lay dormant for years eventually give rise to overt metastases. Genes that are conspicuously expressed in metastatic cells but not in primary tumours, and that mediate metastasis but not primary tumorigenicity, have been identified, particularly in bone metastases[9]. Moreover, various candidate metastasis suppressor genes have been identified that impede metastatic progression without affecting primary malignancy[29].

Melanoma
Skin cancer that is initiated by the transformation of melanocytes.

Metastasis suppressor gene
A gene in which loss of function specifically enhances metastasis without affecting primary tumour growth.

Table 1 | **Stereotypical patterns of tumour metastasis to distant organs.**

Cancer type	Site of metastasis
Breast cancers	Primarily bone and lung; less frequently liver and brain. ER-positive tumours preferentially spread to bone; ER-negative tumours metastasize more aggressively to visceral organs. This might be related to the cell of origin: ER-positive tumours can originate from luminal progenitor cells, whereas many ER-negative tumours emerge from cells in basal layers.
Lung cancers	The two most common types of lung cancer have different aetiologies. Small-cell lung carcinomas (SCLC) disseminate rapidly to many organs. Non-small-cell lung carcinomas (NSCLC) often spread to the contra-lateral lung, the brain and also to the adrenal glands, liver and bones.
Prostate cancer	Almost exclusively to bone, forming osteoblastic lesions and filling the marrow cavity with mineralized bone matrix, unlike the osteolytic metastasis that is caused by breast cancer.
Pancreatic cancer	Aggressive spread to the liver and the surrounding viscera.
Colon cancer	The mesenteric circulation pattern favours dissemination to the liver and peritoneal cavity, but metastasis also occurs in the lungs.
Ovarian cancer	Local spread in the peritoneal cavity; rarely affects other organs, as inadvertently demonstrated by the absence of distant metastases in patients who received palliative shunts to divert ascites to the circulation for abdominal pain relief [116].
Sarcomas	Various types of sarcoma (tumours of mesenchymal origin) mainly metastasize to the lungs.
Myeloma	Haematological malignancy of the bone marrow that causes osteolytic bone lesions, sometimes spreading to other organs.
Glioblastoma	These brain tumours show little propensity for distance organ metastasis, despite aggressively invading the CNS.
Neuroblastoma	Paediatric tumours that arise from nervous tissue of the adrenal gland. Forms bone, liver and lung metastases, which in some cases spontaneously regress.

ER – oestrogen receptor.

Another remarkable observation comes from the cytogenetic analysis of solitary tumour cells that have been isolated from the bone marrow of breast cancer patients. Tumour cells found in the marrow during early disease progression share few chromosomal aberrations with their corresponding primary tumours[30]. Because genetic instability enables clonal evolution, this suggests that some cells acquire mutations and disseminate early on, after which they continue to evolve independently of primary tumour cells[31]. Whether overt metastasis eventually emerges from a progenitor pool that is present in such early seeding, or from the later dissemination of cells that acquired their aggressiveness in the primary tumour, remains an open question.

These diverse lines of evidence can be assembled into an integrative model (FIG. 2) in which the proclivity to disseminate can be acquired as a developing tumour becomes locally aggressive, whereas the outgrowth of distal colonizing cells necessitates further selection from subsequent genetic heterogeneity. The cell type of origin and the circulation pattern represent other important factors. Underlying each of these aspects are genetic determinants, the identity of which, along with their impact in the metastasis process, must be discerned.

Classes of metastasis genes

Metastasis encompasses several unique biological processes that, as a whole, enable the colonization of distant organs. Despite this, the terms 'metastasis' or 'metastatic spread' have been rather liberally used in the literature to describe individual features of an aggressive tumour. Genes that mediate the initial formation and progression of tumours, by cell-autonomous or local microenvironment -modifying functions, can confer tumorigenic properties, but might not specifically mediate the steps of metastasis. Tumorigenic functions, including an unrestricted proliferative ability, resistance to cell death signals or immune evasion, can be considered as prerequisites for metastasis. Genetic alterations that drive these functions remain indispensable in metastatic cells, as shown by the clinical success of drugs such as trastuzumab (Herceptin), which suppresses breast cancer metastatic disease by antagonizing the *HER2* oncogene[32]. However, such genes are not considered here as metastasis genes *per se*, because they do not directly mediate the tumour dissemination and colonization of distant organs (BOX 1; FIG. 1).

With imminent breakthroughs in the study of metastasis, it would seem timely to define metastasis genes more strictly as those in which gain or loss of function specifically enables tumour cells to circulate, home to, penetrate or colonize distant organs (BOX 1; FIG. 1). Several classes of genes can be distinguished that fulfil these criteria. One class, which could be called metastasis initiation genes, includes genes that provide an advantage in a primary tumour and, in so doing, pave the way for tumour cells to escape into the circulation. Most genes that underlie tumour cell motility, invasion or angiogenesis would fall into this class. Invasiveness provides a local advantage by enabling tumour cells to seek vasculature and other favourable elements of the microenvironment. At the same time, this allows tumour cells to enter the circulation and disseminate to distant organs. Examples include several genes that confer a mesenchymal, invasive phenotype[33,34], and genes such as caspase 8 (*CASP8*), the loss of function of which favours dissemination by protecting tumour cells from programmed death upon release of integrin-mediated anchoring at the invasive front[35]. Furthermore, malignant cells growing in tumours that are angiogenic and densely vascularized can readily intravasate into circulation. Invasiveness and local angiogenesis are therefore

Mesenteric circulation
Blood flow from the intestines.

Peritoneal cavity
The space within the abdomen that contains the intestines, the stomach and the liver.

Palliative shunt
Diversion of malignant fluid to alleviate symptoms that arise from its accumulation in the body cavity.

Ascites
A detectable accumulation of free fluid in the peritoneal cavity.

Metastasis initiation genes
A gene that is engaged in the invasion and intravasation of metastatic cells.

Box 2 | Gene signatures that predict metastasis

Cancer-associated gene-expression signatures serve as new tools to classify tumour subtypes, discern disease aetiology, predict risk of relapse, identify genes that mediate disease progression, and select optimal therapeutic options. Different transcriptomic signatures have been described, their biological significance depending on the manner in which they were delineated. Among solid tumour types, breast cancer has been a prominent target in this type of analysis. The heterogeneous nature of this disease, the variability in patient survival rates that is associated with the disease subtype, the prolonged period during which relapse is possible but uncertain, the existence of different therapeutic options to choose from, and the availability of clinical samples from large patients cohorts have all contributed to making breast cancer a hotbed for prognostic gene-expression signature research. Here are nine prominent signatures that predict outcome.

The 'intrinsic subtypes' breast cancer gene-expression signature was obtained by unsupervised clustering of breast tumours on the basis of their gene-expression profile[27]. The expression pattern classifies breast cancers into five subtypes: a normal-tissue-like subtype, two luminal-cell-derived subtypes (A and B) that express oestrogen receptor, a subtype that is derived from the basal epithelial cell layer and does not express oestrogen receptor, and a subtype that overexpresses the human epidermal growth factor receptor-2 (HER2 or ErbB2). Each subtype is characterized by a distinct histopathology and survival rate[119].

A second group of gene-expression signatures were obtained from groups of tumours with known clinical outcome. The 70-gene 'poor-prognosis' signature[25,26] and the 76-gene poor-prognosis signature[98] were obtained from comparisons between early-stage breast tumours for which the clinicial outcome was either favourable or poor. A 21-gene 'recurrence score' was delineated on the basis of metastatic recurrence in breast cancer patients whose lymph nodes were metastasis-free and who were treated with tamoxifen[120].

Other gene-expression signatures were derived from functions that are associated with tumour progression. Aggressive tumours with poor prognosis, which have long been considered wounds that do not heal, express a 512-gene 'wound-signature', which reflects the presence of growth-factor-stimulated fibroblasts in the tumour stroma[121]. Aggressive tumours can also express a 'hypoxia' signature, which reflects the presence of tumour cells that manage to thrive in a hypoxic microenvironment because of their ability to express survival and pro-angiogenic functions[122]. A 186-gene 'invasion' signature was derived by comparing normal breast epithelial cells with breast cancer samples that were enriched for tumorigenic cells[102].

A fourth group of signatures was generated by seeking genes that mediate tumour progression and metastasis. Gene-expression patterns that reflect the presence of oncogenic RAS, SRC, WNT or E2F stimuli — the 'oncogenic pathways' signature — correlate with clinical outcome and point at the specific pathways that are involved[123]. An 18-gene 'lung metastasis' signature, derived from *in vivo* selection of breast cancer cells that selectively metastasize to the lungs, correlates with a risk of pulmonary metastasis and poor overall survival[37]. Being candidate mediators of metastasis, the genes in these signatures are, in principle, susceptible to therapeutic intervention.

Although obtained from breast cancer studies, some of these gene-expression signatures also correlate with poor outcome in other tumour types. For example, the joint deregulation of the WNT and SRC pathways in the 'oncogenic pathways' signature is indicative of poor prognosis in breast, lung and ovarian cancers, whereas the 'invasion' signature correlates with poor outcome in breast, medulloblastoma, lung and prostate cancers.

important events for metastasis, although they are not in themselves sufficient. Gliomas are extremely invasive and adenomas can be highly vascularized, yet neither tumour type metastasizes to other organs.

Formation of distant metastases depends on whether some circulating tumour cells — usually a tiny minority of them — manage to enter and eventually overtake the parenchyma of a target organ. Consequently, metastasis genes become engaged as a result of the selective advantage that they provide against pressures that bar circulating tumour cells from different tissues. Genes that mediate the local progression of a primary tumour are not selected in response to the pressures that exist in distant organs, and so they might not be expected to satisfy demands that are unique to particular organ microenvironments. Nonetheless, certain genes that contribute to primary tumorigenesis might fulfil an additional function that happens to be advantageous at the site of metastasis. We refer to this class as metastasis progression genes, which we define as genes that fulfil certain rate-limiting functions in primary tumour growth and other specific functions in metastatic colonization. Metastasis progression genes are thus distinguished from oncogenes that carry out the same cell-autonomous transforming function throughout the course of a malignant disease. The specific advantage that is conferred by a metastasis progression gene might be restricted to one particular target organ, in which case metastasis progression genes would mechanistically couple primary tumour progression and tissue-specific spread. As such, these genes can be found within gene-expression signatures that correlate certain primary tumours with risk of organ-specific dissemination (BOX 1; FIG. 1).

A third class, the 'metastasis-virulence' genes, are defined as those that provide a selective advantage in secondary sites but not in the primary tumour. These genes participate in metastatic colonization but not in primary tumour development. Consequently, metastasis virulence genes add to the aggressiveness of metastatic tumour cells in a secondary site. Their altered expression or activity becomes discernable in tumours at distant organs. Such genes would rarely be present in primary tumour 'poor-prognosis' gene-expression signatures, as their function would not provide a selective advantage in primary tumours.

Recent work on organ-specific metastasis provides evidence for the existence of metastasis progression genes and metastasis virulence genes. Of 54 genes that were found to be associated with the lung metastatic activity of a breast cancer cell-line model, only 18 are expressed in primary tumours of breast cancer patients. Despite this, genes from both groups functionally mediate lung colonization from circulation[36,37]. The 18 genes that are expressed in primary tumours stimulate primary tumorigenesis, and clinically correlate with a larger primary breast tumour size at diagnosis[37,38]. Three of these genes, namely, *EREG*, which encodes the EGF receptor ligand epiregulin, *COX2* (also known as *PTGS2*), which encodes cyclooxygenase 2, and the matrix metalloproteinase *MMP1*, constitute a vascular-remodelling programme that facilitates pathological angiogenesis and intravasation in mammary tumours, while additionally mediating extravasation of circulating tumour cells from the lung capillaries[36]. These properties classify such mediators as metastasis progression genes. Conversely, various metastasis virulence genes mediate lung colonization in this model, but fail to support primary tumour growth, and their expression in primary tumours is not associated with risk of pulmonary recurrences[36,37]. The general distinction between metastasis progression and virulence genes awaits further characterization from other model systems and in other metastatic cancers.

Metastasis progression gene
A gene that has dual functions in mediating primary tumorigenesis and metastatic colonization.

Metastasis virulence gene
A gene that is exclusively involved in distant organ colonization.

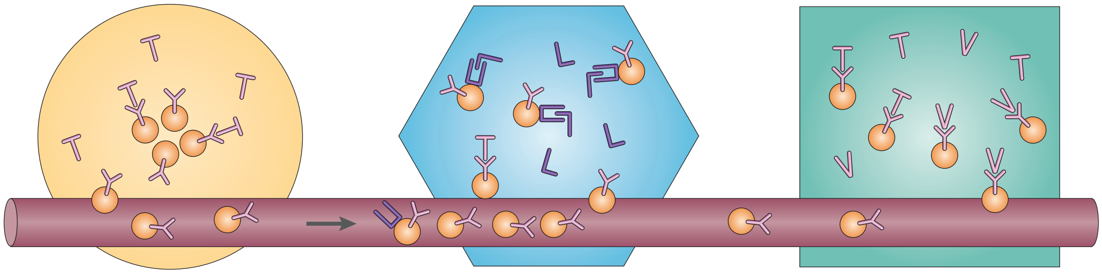

Figure 2 | **A model for the integration of four metastasis theories.** Four views on how metastasis develops can be incorporated in one single model. The seed-and-soil hypothesis[12] posits that certain organs (soils) are congenial for certain tumour cells (seeds). In abstract terms, the microenvironment of an organ X might provide a factor L that supports the survival of rare U-expressing tumour cells, whereas organ Y provides factor V for the survival of Y-expressing tumour cells. The hypothesis of dominant circulation patterns[13] postulates that cells that are released from a tumour preferentially target the first highly vascularized organ that they encounter. Accordingly, the circulation pattern from the tumour source would favour metastasis to organ X over organ Y. Evolution models[18] propose that metastasis is a rare event because it results from the selection of rare clones that emerge from a genetically heterogenous and unstable population, becoming selected in specific organs by virtue of having randomly acquired key pro-metastatic mutations. Thus, a genomically unstable tumour cell population might give rise to rare U-positive clones that become selected in organ X by the presence of factor L. Pre-determination models[21] argue that poor-prognosis genes that are expressed in primary tumours predict metastasis, because these genes can support distant metastasis in addition to primary tumour growth. Accordingly, Y-positive cancer cells would become dominant in a primary tumour (and gene Y would be detectable in a poor-prognosis signature) because Y allows these cells to benefit from factor T during tumorigenesis, and later drives metastasis to other organs that express T or T-like factors (for example, factor V).

Genetic determinants of metastasis

The precise nature of the alterations that turn certain genes into mediators or suppressors of metastatic initiation, progression and virulence is gradually being revealed. Various genetic and epigenetic mechanisms, many somatic but some inherited, could be at play as determinants of cancer metastasis.

Genetic versus epigenetic alterations. If metastasis is driven by natural selection of the fittest among genetically unstable malignant cells, any anomaly that results in the activation of pro-metastatic genes or the repression of genes that interfere with metastasis could be involved, as long as it yields a selective advantage. By this logic, activation of metastatic functions could entail genetic alterations such as mutations, gene deletions, copy-number aberrations or chromosomal rearrangements, as well as epigenetic changes such as covalent modification of DNA or chromatin, deregulated transcription factors or microRNAs, or altered translational and post-translational mechanisms. For most of these alterations, the number of clinically validated examples to date is scarce.

Few mutations, genomic alterations or allelic imbalances are currently known to distinctively endow tumour cells with metastatic functions. The tumour suppressor gene *CDH1* has been linked to both tumorigenic and metastatic processes. Although *CDH1* suppression occurs epigenetically or post-translationally, inactivating mutations have been found in gastric cancer[39]. The product of *CDH1* is E-cadherin, a cell-adhesion receptor, the loss of function of which is characteristic of epithelial-to-mesenchymal transition (EMT), a developmental phenotype that is akin to the acquisition of invasive behaviour by carcinoma cells[33,34,40]. The candidate metastasis suppressor gene *NM23* encodes a histidine kinase[29], and loss of heterozygozity for *NM23* might occur more specifically upon metastatic progression in colon cancer[41]. In another example, *NEDD9*, which encodes an adaptor protein that enhances focal contact formation and invasion, is amplified in an *HRAS*-driven mouse melanoma model and in metastatic human melanomas[42]. It is unclear whether the invasive activities that are conferred by NEDD9 upregulation or E-cadherin inactivation fulfil specific organ colonization functions beyond satisfying general prerequisites for tumour dissemination. Moreover, although the tumorigenic role of E-cadherin loss is well documented, the clinical manifestation of EMT during metastasis remains somewhat controversial[43,44]. Nevertheless, these examples suggest that genomic alterations could indeed directly or indirectly influence the metastatic competency of a tumour.

Many of the candidate metastasis genes identified to date are engaged in this process through changes in gene-expression levels. One reason for this could be that recent genome-wide searches for metastasis-associated events have focused more on gene-expression changes than on mutations or gene copy-number alterations. Although the molecular basis for perturbations in expression of metastatic genes remains largely unknown, it can sometimes be traced to tumour-initiating mutations. For instance, the von Hippel–Lindau tumour suppressor, pVHL, restricts the expression of hypoxia-inducible genes by targeting the hypoxic-responsive transcription factor HIF1α for ubiquitination and

Focal contacts
Dynamic cell-adhesion structures.

degradation[45]. During renal carcinoma development, inactivating mutations in *VHL* leave HIF1α free to hyperactivate many target genes, including that encoding the chemokine receptor CXCR4 (REF. 46). CXCR4 is a mediator of tumour cell survival and metastasis in microenvironments such as the bone marrow and the lung parenchyma that are rich in its ligand, SDF1, also known as CXCL12 (REFS 47,48). Consequently, a side benefit to mutant *VHL* tumour cells that mount a hypoxia response is the expression of a chemokine receptor that will support their ability to colonize certain chemokine-rich tissues.

Numerous epigenetic alterations in cancer have now been characterized. On the one hand, perturbations in chromatin structure, DNA methylation and genomic imprinting can be considered as surrogates for gene mutations, eventually contributing to clonal tumour progression. On the other hand, it has been postulated that epigenetic modulators of cell plasticity can initiate the malignant phenotype by conferring a stem or progenitor state[49]. If the capacity for metastatic dissemination is acquired early on, then genes that epigenetically regulate progenitor cells can also imprint metastasis. One example is provided by the EZH2 component of the Polycomb complexes, which catalyse histone modifications and promote DNA methylation in mammalian stem cells[50]: EZH2 overexpression is predictive of invasive growth in metastatic prostate and breast cancers[51,52]. Another instance is the activation of the WNT/β-catenin pathway, which leads to the expansion of stem- or progenitor-cell populations in various tissue types. In prostate cancer cells, it has been reported that a β-catenin–reptin chromatin-remodelling complex silences the expression of *KAI1* (also known as *CD82*)[53]. Repression of *KAI1* enables malignant prostate cells to dislodge from the receptor DARC, an endothelial membrane protein that normally binds to *KAI1*-expressing cells, causing them to arrest in circulation[54]. Thus, epigenetic determinants can re-programme the expression of genes that affect metastatic outcome.

Inherited predispositions. Population-genetic studies reveal clearly inherited tendencies to develop cancer[55]. In some cases, these can be traced to penetrant germline mutations, as shown by the association of *BRCA1* recessive mutations with familial breast and ovarian cancers. Moreover, certain somatically altered genes that indirectly mediate metastasis can also be found as germline mutations: inherited mutations in *CDH1* result in dominant hereditary diffuse gastric cancer (HDGC)[39] and the von Hippel–Lindau syndrome is caused by inherited loss-of-function mutations in *VHL*[45].

Are there alleles that specifically confer a heritable increase in the probability of metastatic progression? Evidence of this in humans is currently limited to circumstantial observations. For instance, there is a noticeable propensity among young African women to develop aggressive forms of breast cancer. It is difficult to predict whether the genetic determinants that underlie this predisposition work by modifying the tumour-initiating events or by directly affecting metastatic functions. Moreover, it is estimated that the majority of breast cancer risk is attributed to low-penetrant polygenic effects, rendering the identification of individual susceptibility genes difficult[56]. Nonetheless, a study combining gene-expression analysis and population genetics has revealed that tumours from pre-menopausal African-American patients express a gene signature that typifies the basal subtype of breast cancer (BOX 2) with a higher prevalence than those from Caucasian patients[57]. As this polygenic signature correlates with aggressive metastatic relapse[27], the existence of alleles that predispose breast cancer metastasis seems plausible.

Host hereditary factors have recently been experimentally characterized. In the polyoma middle-T transgenic mouse model of mammary carcinoma, the efficiency of pulmonary metastasis depends on the genetic background[58]. By applying QTL analysis, a metastasis-efficiency locus was mapped to chromosome 19 (REF. 59). A subsequent multiple-cross mapping and candidate-based strategy identified a SNP in *Sipa1* that determined metastasis susceptibility[60]. *Sipa1* encodes a GTPase-activating protein (GAP) that inhibits the activity of RAP1 and RAP2, and the identified SNP results in an amino-acid substitution that hinders RAP–GAP function, attenuating pulmonary metastasis. *SIPA1* polymorphisms have been reported in human breast-cancer samples and correlate with poor prognosis[61].

The existence of metastasis-promoting germline mutations could provide yet another source of shared transcriptomic profiles between metastatic and primary tumours. Thus, directly or indirectly, the genetic background can determine metastasis susceptibility in breast cancer.

Solitary mediators versus multi-gene programmes. Metastasis genes can be selected during tumour progression as the sole mediators of a particular metastatic step. Such solitary mediators might be expressed in the tumour cells themselves. Contributions by factors that are secreted from the metastatic microenvironment are also possible[62].

Although solitary mediators are likely to be important determinants of metastasis, a growing body of evidence points to the role of complex sets of genes that are engaged by various mechanisms to mediate metastasis. Lineage-specific factors might represent a form of malignant and metastatic programming. uring melanoma progression, the melanocyte lineage-specification programme potentiates metastasis through the activity of slug, a regulator of neural crest cell migration[63]. Furthermore, rhabdomyosarcoma metastasis requires the activity of the homeobox transcription factor SIX1, which normally controls developmental myogenesis[64]. In a lobular breast cancer model, overexpression of twist and snail, master transcriptional regulators of morphogenesis, promotes EMT and intravasation[65]. The pleiotropic effects of these developmental and lineage-determining factors might contribute to multiple steps, from the aggressiveness of local tumours cells to their dissemination and onset of distant organ metastasis.

Chemokine
A group of small signalling proteins (cytokines) that are usually secreted by immune cells.

Neural crest cell
Highly motile cells that originate from the ectoderm during development.

Rhabdomyosarcoma
Paediatric malignancy that arises from skeletal muscle cells.

Some transcription factors are not associated with differentiation programmes, but are rather effectors of physiological pathways that are triggered by changes in the tumour microenvironment. Although the biological function of stromal interactions in metastasis has been reviewed elsewhere[8], it is worth mentioning one prominent example involving the pro-tumorigenic role of inflammation[66] and the NFκB transcription factor. NFκB lies downstream of most inflammatory pathways and regulates a plethora of genes that have been implicated in stromal activation[67]. The transcriptional response towards NFκB activation during metastasis seems to be context dependent. In breast cancer metastatic cells, NFκB mediates expression of granulocyte–macrophage colony-stimulating factor (GM-CSF), which induces osteoclastogenesis during bone colonization[68]. In a mouse model of prostate cancer metastasis to the liver, cytokine-mediated activation of the NFκB pathway through the kinase IKKα leads to transcriptional inhibition of maspin, a metastasis suppressor gene[69].

Pleiotropic regulators can also contribute to metastasis in a non-cell-autonomous manner, by directly impinging on the microenvironment to enable metastatic colonization. Exemplifying this is transforming growth factor-β (TGFB), which is released by tumour cells and bone matrix alike, and in mouse models can induce the expression of osteoclast-activating genes (such as the genes encoding parathyroid-hormone-related protein (PTHrP) and interleukin 11 (IL11)) that fuel a vicious cycle of osteolytic bone metastasis[9,47,70]. The SMAD transcription factors as well as SMAD-independent pathways that lie downstream of the TGFB receptor have been implicated in these pro-metastatic effects. In another example, the vascular endothelial growth factor (VEGF) that is released by several cell types binds to its receptors VEGFR1 and VEGFR2, present on bone-marrow-derived progenitor and mature endothelial cells, respectively. In addition to its angiogenic properties[71], supra-physiological secretion of VEGF from cancer cells can mediate pulmonary extravasation by increasing the permeability of lung capillaries[72]. Moreover, VEGF promotes the recruitment of VEGFR1-positive bone-marrow-derived progenitors to metastatic niches, thereby priming the colonization of distant organs[73]. Because of their various distinctive functions during tumorigenesis, TGFB[74,75] and VEGF[76,77] could be classified as metastasis progression factors.

Searching for metastasis genes

The goal of identifying metastasis genes is primarily driven by the need to address the problem of cancer. Therefore, a premium must be placed on the identification of metastasis genes that have demonstrable clinical relevance. *In vitro* systems have been used to great advantage for the analysis of pre-metastatic activities, such as a tumour cell motility, anchorage-free survival and basement-membrane invasion. A new generation of tractable methods uses three-dimensional cultures to more accurately model some of these functions[78]. Nevertheless, *in vitro* assays do not fully replicate the microenvironment of metastasis target organs. Consequently, most efforts to discover metastasis genes and functions rely on *in vivo* systems and human tissue-based approaches.

Genetically engineered mouse models. Sophisticated genetically engineered mouse models are being designed to approximate the characteristics of human tumour development[79]. Some models of the more aggressive breast[80] and prostate[81] cancer do show metastases, but their reliance on viral oncogenes generates malignancies of uncertain clinical relevance. For instance, the TRAMP (transgenic adenocarcinoma of mouse prostate) model, which is based on prostate-specific expression of the SV40 T-antigen oncogene, does not yield the osteoblastic metastases that typify human prostatic cancer[81]. Although such transgenic animals are useful to dissect oncogene suppressor and tumour suppressor pathways, stromal interactions and pre-clinical strategies, relatively few of these have hitherto managed to faithfully reproduce the clinical features of metastatic disease (TABLE 1). More recent mouse models have been engineered to sporadically acquire various human tumour mutations, including lung adenocarcinomas caused by combined *KRAS* mutations and *p53* inactivation[82,83]. It will be interesting to see how closely the metastatic disease in these animals mimics the development of brain metastasis and other lesions that are typical of lung adenocarcinoma in humans[84].

Functional genomic screens. Genetic screens represent a functional and unbiased strategy for identifying genes that can mediate a given function. With the genome sequences of many organisms now complete, the potential for identifying genes that cause experimental metastasis by genome-wide screens is clear. However, the complexities and costs that are associated with carrying out such screens in mammals have prompted investigators to use surrogate *in vitro* assays[85], or 'simpler' models such as *Drosophila melanogaster*[86]. Several screens in *D. melanogaster* have identified metastasis-specific genes by starting with tumours that had been initiated by defined mutations. For instance, in a background of tumours arising from deletion of *lethal giant larvae (l(2)gl)*, insertional mutagenesis identified *semaphorin 5c* as a mediator of tumour invasion[87]. In non-invasive eye tumours that were initiated by the *Ras^{V12}* oncogene, tissue-restricted gene disruption by homologous recombination led to the identification of the cell-polarity gene *scribble*, which cooperates with *Ras^{V12}* to induce basement-membrane degradation and invasion[88].

These examples notwithstanding, there are obvious limitations to invertebrate models. Such organisms have an open circulatory system that obviates the need for angiogenesis, a cardinal feature of mammalian cancers[5], and precludes the study of intra- and extravasation. Moreover, despite some genetic homology, the vast anatomical differences between mammals and invertebrates are not conducive to the study of the unique barriers that are encountered by tumour cells that colonize organs such as the liver, lung and bones.

Stromal activation
Stimulation and mobilization of host cells in the microenvironment that surrounds a tumour.

Osteoclastogenesis
The differentiation and activation of osteoclasts that mediates bone resorption.

Recent technological advances have rendered the design of genome-wide screens more amenable to the identification of metastasis genes in mice. Tumour progression genes are being sought by adopting transposable elements in murine somatic tissues[89,90]. Alternatively, RNAi libraries can be exploited for tumour suppressor gene identification in transplantable mouse cancers[91]. Although most of these technologies have been applied to models of haematological malignancy, we can anticipate their adaptation for solid tumour metastasis. As candidate genes are identified in these systems, it will be essential to determine how these findings relate to metastasis in human.

Profiling human tissue. Microarray platforms coupled with bioinformatics tools can detect with remarkable resolution transcriptional signatures (BOX 2), gene copy-number abnormalities[92], SNPs[93], epigenetic changes[94], microRNA levels[95], proteomic alterations[96] and somatic mutations[97]. With an increasing appreciation for their potential clinical usefulness, these tools are regularly applied to interrogate tumour samples from large patient cohorts, and retrospectively derive indicators of disease outcome.

Different breast cancer gene-expression signatures that predict disease outcome have been reported (BOX 2). Surprisingly, these genetic profiles show almost no overlap with each other, raising questions about the relationship and biological significance of the genes in question. A striking example is provided by two poor-prognosis signatures that were obtained from similar early-stage breast tumours: of the 86 genes that could be compared, only 3 genes are shared by these two signatures[26,98]. Different patient cohorts, gene microarray platforms (Agilent versus Affymetrix) and bioinformatics tools might underlie this pronounced divergence. And yet, largely non-overlapping signatures can identify the same patients in a given cohort, as recently demonstrated in a comparative analysis of the intrinsic-subtype, poor-prognosis, wound-response, and recurrence-score signatures[99].

Although these signatures comprise largely distinct sets of genes, they encompass common segments of 'prognostic space', defined as the entire range of features that affect the clinical outcome of a tumour type[100]. Each of these transcriptional profiles might represent one of the biological features that is acquired by tumour cell populations as they become aggressive. That is, each signature might represent a different picture of the same aggressive entity. Nevertheless, certain gene-expression signatures become prognostic when combined with others, as shown by combining the wound-response signature with the poor-prognosis signature[101], the molecular-subtypes signature[101] or the invasion signature[102]. Protocols that incorporate the sequential or combined analysis of gene-expression data sets could improve prognostic guidance in the clinical management of individual patients.

An obvious drawback to any kind of tissue profiling is that the resulting gene list bears unknown significance as mediators of metastasis. To extract biologically relevant information from molecular signatures, integrative *in silico* methods are being developed[103], but functional validation in model systems is inevitable when the intent is to identify genes that mediate, and not just mark, the ability of tumours to develop metastasis.

Mining for pre-existing metastatic entities. Of the millions of cancer cells entering the bloodstream, only a minute proportion manages to survive in distant organs[104,105]. In xenograft models, thousands of tumour cells that are inoculated into mice, either at the corresponding (orthotopic) sites of the primary neoplasm or directly into the circulation, typically result in a limited number of metastatic colonies[106]. The fact that a few cells from a heterogeneous population can metastasize at all suggests that relevant metastatic entities can be separated from the bulk of indolent cells by selection *in vivo*[17]. The selection that is imposed by this process incorporates multiple barriers of the metastatic cascade, especially when the xenograft is carried out from the orthotopic site, and it enables comparisons of phenotypically diverse cell populations that originated from the same tumour.

Microarray profiling of *in vivo*-selected melanoma cell lines identified the small-cell motility GTPase RhoC as a mediator of lung metastasis in mice[107]. In a model of paediatric rhabdomyosarcoma, lung metastatic variants show increased expression of SIX1, which induces expression of the cytoskeletal signalling factor ezrin[64]. Notably, ezrin upregulation is also observed in metastatic osteosarcoma[108], suggesting a common selective pressure for ezrin expression in paediatric sarcoma metastases.

In vivo selection has also shed light on the genetic determinants of metastatic organ tropism. In studies using a cell line derived from a patient with advanced metastatic breast cancer, cell progenies that show predilections for distinct tissues were isolated following injection into the arterial or venous circulation of immunodefficient animals[47]. Consistent with the selectivity of metastatic progression, these different cell variants were found to already exist in the parental, unselected population[47,109]. Furthermore, subpopulations that were metastatic to bone, adrenal gland or lung tissue were associated with distinct gene-expression profiles. Bone metastatic cells, for instance, express the survival chemokine receptor CXCR4 and the osteoclast–activating cytokine IL11 (REF. 47), whereas lung metastatic cells from the same source express different mediators of pulmonary colonization[37]. These distinctive gene sets are therefore selected for interactions with unique microenvironments, providing a genetic basis for metastatic tropism.

An often-mentioned drawback to xenografts is that they do not model the early stages of cancer progression within endogenous tissues, and rely on cell lines that have alterations of unknown relevance. Further caution is required when submitting unstable tumour cells to multiple rounds of selection in mice, which could yield deviant entities with properties that bear little resemblance to the authentic metastatic cells from the original population. Therefore, as with the other experimental approaches, validation of clinical relevance must accompany the evidence of functional significance.

Xenograft
Implantation of human tumour cells into an immunocompromized animal.

Orthotopic site
Transplant of tumour cells into the anatomical location of an animal that best recapitulates the original source of primary tumorigenesis.

Validating metastasis genes

Bona fide metastasis genes should be defined as those with an activity that enables locally invasive tumour cells to enter the circulation, to home, extravasate and colonize distant tissues in a clinically relevant manner. These can be efficiently identified by combining functional testing in an animal model with validation in relevant clinical samples. Several examples of combined approaches have been reported in recent years.

Filtering human transcriptomic data with in vivo selection. *In vivo* sorting of metastatic cell populations can isolate pre-existing cell subpopulations with distinct metastatic organ tropism[109]. In this way, a gene-expression profile was derived from human breast cancer cells showing a proclivity for pulmonary colonization. By combining bioinformatic analysis of transcriptional data sets with functional validation in a xenograft mouse model, this gene list was culled to obtain an 18-gene lung metastasis signature (LMS) that is differentially expressed in human primary tumours[37]. The LMS was shown to predict risk of lung metastatic recurrence in an extended cohort of more than 700 primary breast tumours with known clinical outcome[38]. Functional assays also demonstrate the ability of these genes to mediate pulmonary metastasis[36,37]. Consequently, the biological filter that is afforded by *in vivo* selection and the clinical filter that is provided by large gene-expression data sets can be exploited to identify genetic determinants of metastasis that are relevant to human disease.

Linking signatures with chromosomal abnormalities. Analogous to the use of microarray platforms for gene expression, array comparative genomic hybridization (aCGH) can report perturbations in DNA copy number[110]. But uncovering a direct association between genomic alterations and metastatic outcome is a complex undertaking, as aCGH has revealed numerous and sometimes expansive regions of allelic imbalances in human neoplasms. The remaining challenge is to distinguish the true mediators from the bystander genes that map to the same loci. Taking the wound-healing signature (BOX 2) as an adequate marker of metastatic phenotype, a method was recently reported to link the probability of wound-healing signature expression with alterations in gene copy number[111]. In this manner, recurrent amplifications in a region of chromosome 8q were found to correlate with the wound-healing programme, and the list of candidates in this chromosome region was narrowed down to *MYC* and *CSN5*. *MYC* is a known oncogene that participates in the wound-healing response, whereas CSN5 can ubiquitinate MYC, causing increased MYC degradation, target-promoter licensing and cooperative activation of almost all 512 wound-healing genes. This unexpected finding suggests that there might be numerous incidental markers of metastasis within this signature, and further shows how genetic determinants of metastasis can be masked by dynamic gene-expression events. Similar methods have linked chromosomal aberrations to the transcriptomic profiles of distinct breast cancer subtypes[112].

Comparative oncogenomics. In a recent study that used an inducible Ha-RAS^{V12G} model of murine melanoma, oncogene-independent metastatic tumours were generated and characterized for regions of chromosomal gains and losses[42]. By comparing this information to human aCGH data, a minimal common region was mapped that contains a single gene, *NEDD9*. Amplifications and overexpression of *NEDD9* is seen in both human and mouse melanoma samples. An important feature of this approach is the use of a mouse model as biological filter. Thus, NEDD9 function could be validated *in vivo*, and was subsequently shown to regulate invasion.

Comparative oncogenomics has also been applied to identify oncogenes in liver cancer[113], and to sort relevant transcriptomic patterns between human and murine prostatic and lung adenocarcinoma[114,115]. However, this methodology is not specifically designed for the identification of metastasis genes, and requires previous knowledge of the initiating mutations in a specialized transgenic model, as shown by the fact that NEDD9 seems to mediate murine melanoma invasion only following oncogenic RAS/RAF activation[42].

Conclusions

In the light of recent advances, we have attempted to synthesize the current knowledge of the genetic basis for cancer metastasis. Insights from genome-wide analyses of human tumour tissue and innovative technologies for the exploitation of laboratory models are helping us reassess classical concepts about the evolving nature of genetically unstable metastatic cells, the interplay between circulating cancer cells and permissive tissue microenvironments, as well as the pressures that select for pro-metastatic traits at different stages of tumour progression. We have emphasized the distinction between tumorigenic events that are prerequisite for metastasis (such as transformation and immune evasion) and events that specifically satisfy the unique requirements of metastatic invasiveness, intravasation, extravasation and colonization of distant organs. It is the second set of functions that bears the greatest relevance as we face the challenge of explaining how tumours manage to metastasize. Conceptual clarity might further help in the design of clinical interventions within optimized therapeutic windows.

Opposing views on metastasis as either a predetermined or an evolving process have dominated the field. Our distinction between metastasis initiation genes, progression genes and virulence genes could be useful in reconciling such differences. The available examples for various types of metastatic determinants — genetic and epigenetic, somatic and inherited — scarce as they still might be, serve as precedents for the future identification of more genes that are engaged in metastasis. In this search, there must be an attempt to reach standards of clinical validation. Sooner or later, these anticipated discoveries must explain the many paradoxical properties of metastasis as a clinical reality, and answer important conceptual quandaries. Why does metastasis aggressively target multiple organs in some tumour types, why is it restricted to just one organ in others, and why does it not occur at all in certain invasive cancers? Does metastasis of one tumour type to different organs involve common as well as non-overlapping sets of genes? Does metastasis of different tumour types to the same organ involve common or different genes? How do these genes determine the organ-selective nature of metastasis? And finally, what renders micrometastases dormant and indolent for years or even decades, only to awake with lethal consequences? Answering these questions should help us better understand the biology of normal tissues that are faced with renegade cells, while providing invaluable help for a better prevention and management of metastatic diseases.

1. Weigelt, B., Peterse, J. L. & van 't Veer, L. J. Breast cancer metastasis: markers and models. *Nature Rev. Cancer* **5**, 591–602 (2005).
2. van de Wouw, A. J., Jansen, R. L., Speel, E. J. & Hillen, H. F. The unknown biology of the unknown primary tumour: a literature review. *Ann. Oncol.* **14**, 191–6 (2003).
3. Weiss, L. Metastasis of cancer: a conceptual history from antiquity to the 1990s. *Cancer Metastasis Rev.* **19**, 193–383 (2000).
4. Norton, L. & Massagué, J. Is cancer a disease of self-seeding? *Nature Med.* **12**, 875–878 (2006).
5. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
6. Christofori, G. New signals from the invasive front. *Nature* **441**, 444–450 (2006).
7. Fidler, I. J. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nature Rev. Cancer* **3**, 453–458 (2003).
8. Gupta, G. P. & Massagué, J. Cancer metastasis: building a framework. *Cell* **127**, 679–695 (2006).
9. Mundy, G. R. Metastasis to bone: causes, consequences and therapeutic opportunities. *Nature Rev. Cancer* **2**, 584–593 (2002).
10. Steeg, P. S. Tumor metastasis: mechanistic insights and clinical challenges. *Nature Med.* **12**, 895–904 (2006).
11. Capasso, L. L. Antiquity of cancer. *Int. J. Cancer* **113**, 2–13 (2005).
12. Paget, S. The distribution of secondary growths in cancer of the breast. 1889. *Cancer Metastasis Rev.* **8**, 98–101 (1989).
13. Ewing, J. *Neoplastic Diseases* edn 6 (Saunders, Philadelphia, 1928).
14. Fisher, B. & Fisher, E. R. The interrelationship of hematogenous and lymphatic tumor cell dissemination. *Surg. Gynecol. Obstet.* **122**, 791–798 (1966).
15. Crespi, B. & Summers, K. Evolutionary biology of cancer. *Trends Ecol. Evol.* **20**, 545–552 (2005).
16. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976).
17. Fidler, I. J. Selection of successive tumour lines for metastasis. *Nature New Biol.* **242**, 148–149 (1973).
18. Fidler, I. J. & Kripke, M. L. Metastasis results from preexisting variant cells within a malignant tumor. *Science* **197**, 893–895 (1977).
19. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).
20. Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nature Med.* **10**, 789–799 (2004).
21. Bernards, R. & Weinberg, R. A. A progression puzzle. *Nature* **418**, 823 (2002).
22. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
23. Golub, T. R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537 (1999).
24. Ramaswamy, S., Ross, K. N., Lander, E. S. & Golub, T. R. A molecular signature of metastasis in primary solid tumors. *Nature Genet.* **33**, 49–54 (2003).
25. van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
26. van 't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
27. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
28. Kerbel, R. S., Waghorne, C., Korczak, B., Lagarde, A. & Breitman, M. L. Clonal dominance of primary tumours by metastatic cells: genetic analysis and biological implications. *Cancer Surv.* **7**, 597–629 (1988).
29. Steeg, P. S. Metastasis suppressors alter the signal transduction of cancer cells. *Nature Rev. Cancer* **3**, 55–63 (2003).
30. 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).
A cytogenetic analysis of single tumour cells from the bone marrow of breast cancer patients, leading to the suggestion that metastatic cells disseminate early and evolve independently of their primary tumour.
31. Schardt, J. A. *et al.* Genomic analysis of single cytokeratin-positive cells from bone marrow reveals early mutational events in breast cancer. *Cancer Cell* **8**, 227–239 (2005).
32. Slamon, D. J. *et al.* Use of chemotherapy plus a monoclonal antibody against *HER2* for metastatic breast cancer that overexpresses *HER2*. *N. Engl. J. Med.* **344**, 783–792 (2001).
33. Huber, M. A., Kraut, N. & Beug, H. Molecular requirements for epithelial–mesenchymal transition during tumor progression. *Curr. Opin. Cell Biol.* **17**, 548–558 (2005).
34. Thiery, J. P. Epithelial–mesenchymal transitions in tumour progression. *Nature Rev. Cancer* **2**, 442–454 (2002).
35. Stupack, D. G. *et al.* Potentiation of neuroblastoma metastasis by loss of caspase-8. *Nature* **439**, 95–99 (2006).
36. Gupta, G. P. *et al.* Mediators of vascular remodelling co-opted for metastatic extravasation. *Nature* **446**, 765–770 (2007).
37. Minn, A. J. *et al.* Genes that mediate breast cancer metastasis to lung. *Nature* **436**, 518–524 (2005).
References 36 and 37 integrate *in vivo* selection with clinical validation to identify mediators of lung-specific metastasis, linking aggressive primary tumorigenesis to organ-specific colonization.
38. Minn, A. J. *et al.* Lung metastasis genes couple breast tumor size and metastatic spread. *Proc. Natl Acad. Sci. USA* **104**, 6740–6745 (2007).
39. Richards, F. M. *et al.* Germline E-cadherin gene (CDH1) mutations predispose to familial gastric cancer and colorectal cancer. *Hum. Mol. Genet.* **8**, 607–610 (1999).
This report links the inactivation of a developmentally regulated cell-adhesion gene with predisposition to cancer progression.
40. Cavallaro, U. & Christofori, G. Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. *Nature Rev. Cancer* **4**, 118–132 (2004).
41. Kapitanovic, S. *et al.* nm23-H1 expression and loss of heterozygosity in colon adenocarcinoma. *J. Clin. Pathol.* **57**, 1312–1318 (2004).
42. Kim, M. *et al.* Comparative oncogenomics identifies *NEDD9* as a melanoma metastasis gene. *Cell* **125**, 1269–1281 (2006).
An integrative approach that uses a mouse model to filter human aCGH data and characterize chromosomal aberrations that are associated with melanoma metastasis.
43. Thompson, E. W. & Newgreen, D. F. Carcinoma invasion and metastasis: a role for epithelial–mesenchymal transition? *Cancer Res.* **65**, 5991–5995 (2005).
44. Tarin, D. The fallacy of epithelial mesenchymal transition in neoplasia. *Cancer Res.* **65**, 5996–6000 (2005).
45. Kaelin, W. G. The von Hippel–Lindau tumor suppressor protein: roles in cancer and oxygen sensing. *Cold Spring Harb. Symp. Quant. Biol.* **70**, 159–166 (2005).
46. Staller, P. *et al.* Chemokine receptor CXCR4 downregulated by von Hippel–Lindau tumour suppressor pVHL. *Nature* **425**, 307–311 (2003).
An example of how somatic mutations that are acquired during tumour progression can affect the expression of a metastasis-specific gene.
47. Kang, Y. *et al.* A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3**, 537–549 (2003).
48. Muller, A. *et al.* Involvement of chemokine receptors in breast cancer metastasis. *Nature* **410**, 50–56 (2001).
This paper describes how non-immunological tumour cells can express the chemokine receptor CXCR4 and respond to a chemokine source to settle in certain organs.
49. Feinberg, A. P., Ohlsson, R. & Henikoff, S. The epigenetic progenitor origin of human cancer. *Nature Rev. Genet.* **7**, 21–33 (2006).
50. Baylin, S. B. & Ohm, J. E. Epigenetic gene silencing in cancer — a mechanism for early oncogenic pathway addiction? *Nature Rev. Cancer* **6**, 107–116 (2006).
51. Kleer, C. G. *et al.* EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc. Natl Acad. Sci. USA* **100**, 11606–11611 (2003).
52. Varambally, S. *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624–629 (2002).
This paper reports the deregulation of stem-cell epigenetic regulators during metastatic progression.
53. Kim, J. H. *et al.* Transcriptional regulation of a metastasis suppressor gene by Tip60 and β-catenin complexes. *Nature* **434**, 921–926 (2005).
A study that links the progenitor WNT/β-catenin pathway to the transcriptional repression of a metastasis suppressor gene.
54. Bandyopadhyay, S. *et al.* Interaction of KAI1 on tumor cells with DARC on vascular endothelium leads to metastasis suppression. *Nature Med.* **12**, 933–938 (2006).
55. Frank, S. A. Genetic predisposition to cancer — insights from population genetics. *Nature Rev. Genet.* **5**, 764–772 (2004).
56. Pharoah, P. D. *et al.* Polygenic susceptibility to breast cancer and implications for prevention. *Nature Genet.* **31**, 33–36 (2002).
57. Carey, L. A. *et al.* Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* **295**, 2492–2502 (2006).
58. Lifsted, T. *et al.* Identification of inbred mouse strains harboring genetic modifiers of mammary tumor age of onset and metastatic progression. *Int. J. Cancer* **77**, 640–644 (1998).
59. Park, Y. G. *et al.* Comparative sequence analysis in eight inbred strains of the metastasis modifier QTL candidate gene *Brms1*. *Mamm. Genome* **13**, 289–292 (2002).
60. Park, Y. G. *et al.* *SIPA1* is a candidate for underlying the metastasis efficiency modifier locus *MTES1*. *Nature Genet.* **37**, 1055–1062 (2005).
The first experimental identification of a polymorphism that affects metastatic potential in mice.
61. Crawford, N. P. *et al.* Germline polymorphisms in *SIPA1* are associated with metastasis and other indicators of poor prognosis in breast cancer. *Breast Cancer Res.* **8**, R16 (2006).
62. Hiratsuka, S., Watanabe, A., Aburatani, H. & Maru, Y. Tumour-mediated upregulation of chemoattractants and recruitment of myeloid cells predetermines lung metastasis. *Nature Cell Biol.* **8**, 1369–1375 (2006).
63. Gupta, P. B. *et al.* The melanocyte differentiation program predisposes to metastasis after neoplastic transformation. *Nature Genet.* **37**, 1047–1054 (2005).
64. Yu, Y. *et al.* Expression profiling identifies the cytoskeletal organizer ezrin and the developmental homeoprotein SIX-1 as key metastatic regulators. *Nature Med.* **10**, 175–181 (2004).
65. Yang, J. *et al.* Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117**, 927–939 (2004).
A gene-expression analysis that identified a role for a transcriptional regulator of embryo development during mouse mammary tumour intravasation.

66. de Visser, K. E., Eichten, A. & Coussens, L. M. Paradoxical roles of the immune system during cancer development. *Nature Rev. Cancer* **6**, 24–37 (2006).
67. Karin, M. Nuclear factor-κB in cancer development and progression. *Nature* **441**, 431–436 (2006).
68. Park, B. K. *et al.* NF-κB in breast cancer cells promotes osteolytic bone metastasis by inducing osteoclastogenesis via GM-CSF. *Nature Med.* **13**, 62–69 (2007).
69. Luo, J. L. *et al.* Nuclear cytokine-activated IKKα controls prostate cancer metastasis by repressing Maspin. *Nature* 18 March 2007 (doi:10.1038/nature05656).
70. Kang, Y. *et al.* Breast cancer bone metastasis mediated by the Smad tumor suppressor pathway. *Proc. Natl Acad. Sci. USA* **102**, 13909–13914 (2005).
71. Olsson, A. K., Dimberg, A., Kreuger, J. & Claesson-Welsh, L. VEGF receptor signalling — in control of vascular function. *Nature Rev. Mol. Cell Biol.* **7**, 359–371 (2006).
72. Weis, S., Cui, J., Barnes, L. & Cheresh, D. Endothelial barrier disruption by VEGF-mediated Src activity potentiates tumor cell extravasation and metastasis. *J. Cell Biol.* **167**, 223–229 (2004).
73. Kaplan, R. N. *et al.* VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. *Nature* **438**, 820–827 (2005).
74. Bierie, B. & Moses, H. L. Tumour microenvironment: TGFβ: the molecular Jekyll and Hyde of cancer. *Nature Rev. Cancer* **6**, 506–520 (2006).
75. Siegel, P. M. & Massagué, J. Cytostatic and apoptotic actions of TGFβ in homeostasis and cancer. *Nature Rev. Cancer* **3**, 807–821 (2003).
76. Carmeliet, P. Angiogenesis in life, disease and medicine. *Nature* **438**, 932–936 (2005).
77. Carmeliet, P. & Jain, R. K. Angiogenesis in cancer and other diseases. *Nature* **407**, 249–257 (2000).
78. Debnath, J. & Brugge, J. S. Modelling glandular epithelial cancers in three-dimensional cultures. *Nature Rev. Cancer* **5**, 675–688 (2005).
79. Van Dyke, T. & Jacks, T. Cancer modeling in the modern era: progress and challenges. *Cell* **108**, 135–144 (2002).
80. Guy, C. T., Cardiff, R. D. & Muller, W. J. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol. Cell. Biol.* **12**, 954–961 (1992).
81. Gingrich, J. R. *et al.* Metastatic prostate cancer in a transgenic mouse. *Cancer Res.* **56**, 4096–4102 (1996).
82. Jackson, E. L. *et al.* The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res.* **65**, 10280–10288 (2005).
83. Kim, C. F. *et al.* Mouse models of human non-small-cell lung cancer: raising the bar. *Cold Spring Harb. Symp. Quant. Biol.* **70**, 241–250 (2005).
84. Nathoo, N., Toms, S. A. & Barnett, G. H. Metastases to the brain: current management perspectives. *Expert Rev. Neurother.* **4**, 633–640 (2004).
85. Douma, S. *et al.* Suppression of anoikis and induction of metastasis by the neurotrophic receptor TrkB. *Nature* **430**, 1034–1039 (2004).
86. Brumby, A. M. & Richardson, H. E. Using *Drosophila melanogaster* to map human cancer pathways. *Nature Rev. Cancer* **5**, 626–639 (2005).
87. Woodhouse, E. C. *et al.* *Drosophila* screening model for metastasis: Semaphorin 5c is required for l(2)gl cancer phenotype. *Proc. Natl Acad. Sci. USA* **100**, 11463–11468 (2003).
88. Pagliarini, R. A. & Xu, T. A genetic screen in *Drosophila* for metastatic behavior. *Science* **302**, 1227–1231 (2003).
89. Dupuy, A. J., Akagi, K., Largaespada, D. A., Copeland, N. G. & Jenkins, N. A. Mammalian mutagenesis using a highly mobile somatic *Sleeping Beauty* transposon system. *Nature* **436**, 221–226 (2005).
90. Collier, L. S., Carlson, C. M., Ravimohan, S., Dupuy, A. J. & Largaespada, D. A. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature* **436**, 272–276 (2005).
91. Dickins, R. A. *et al.* Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. *Nature Genet.* **37**, 1289–1295 (2005).
92. Ellsworth, R. E. *et al.* Allelic imbalance in primary breast carcinomas and metastatic tumors of the axillary lymph nodes. *Mol. Cancer Res.* **3**, 71–77 (2005).
93. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
94. Laird, P. W. The power and the promise of DNA methylation markers. *Nature Rev. Cancer* **3**, 253–266 (2003).
95. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–838 (2005).
96. Varambally, S. *et al.* Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell* **8**, 393–406 (2005).
97. Sjoblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268–274 (2006).
98. Wang, Y. *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365**, 671–679 (2005).
99. Fan, C. *et al.* Concordance among gene-expression-based predictors for breast cancer. *N. Engl. J. Med.* **355**, 560–569 (2006).
An analysis of five prominent breast cancer gene signatures that shows that, despite little overlap in gene identity, these signatures can classify similar subsets of patient who are at risk for metastatic relapse.
100. Massagué, J. Sorting out breast-cancer gene signatures. *N. Engl. J. Med.* **356**, 294–297 (2007).
101. Chang, H. Y. *et al.* Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc. Natl Acad. Sci. USA* **102**, 3738–3743 (2005).
102. Liu, R. *et al.* The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N. Engl. J. Med.* **356**, 217–226 (2007).
103. Rhodes, D. R. & Chinnaiyan, A. M. Integrative analysis of the cancer transcriptome. *Nature Genet.* **37**, S31–S37 (2005).
104. Luzzi, K. J. *et al.* Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am. J. Pathol.* **153**, 865–873 (1998).
105. Fidler, I. J. & Nicolson, G. L. Fate of recirculating B16 melanoma metastatic variant cells in parabiotic syngeneic recipients. *J. Natl Cancer Inst.* **58**, 1867–1872 (1977).
106. Khanna, C. & Hunter, K. Modeling metastasis *in vivo*. *Carcinogenesis* **26**, 513–523 (2005).
107. Clark, E. A., Golub, T. R., Lander, E. S. & Hynes, R. O. Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* **406**, 532–535 (2000).
The first study to combine genomic profiling and *in vivo* selection for the identification of metastasis genes.
108. Khanna, C. *et al.* The membrane–cytoskeleton linker ezrin is necessary for osteosarcoma metastasis. *Nature Med.* **10**, 182–186 (2004).
109. Minn, A. J. *et al.* Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. *J. Clin. Invest.* **115**, 44–55 (2005).
110. Pinkel, D. & Albertson, D. G. Comparative genomic hybridization. *Annu. Rev. Genomics Hum. Genet.* **6**, 331–354 (2005).
111. Adler, A. S. *et al.* Genetic regulators of large-scale transcriptional signatures in cancer. *Nature Genet.* **38**, 421–430 (2006).
112. Chin, K. *et al.* Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* **10**, 529–541 (2006).
113. Zender, L. *et al.* Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. *Cell* **125**, 1253–1267 (2006).
114. Sweet-Cordero, A. *et al.* An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nature Genet.* **37**, 48–55 (2005).
115. Ellwood-Yen, K. *et al.* Myc-driven murine prostate cancer shares molecular features with human prostate tumors. *Cancer Cell* **4**, 223–238 (2003).
116. Tarin, D., Vass, A. C., Kettlewell, M. G. & Price, J. E. Absence of metastatic sequelae during long-term treatment of malignant ascites by peritoneo-venous shunting. A clinico-pathological report. *Invasion Metastasis* **4**, 1–12 (1984).
117. Joyce, J. A. Therapeutic targeting of the tumor microenvironment. *Cancer Cell* **7**, 513–520 (2005).
118. Nierodzik, M. L. & Karpatkin, S. Thrombin induces tumor growth, metastasis, and angiogenesis: evidence for a thrombin-regulated dormant tumor phenotype. *Cancer Cell* **10**, 355–362 (2006).
119. Sorlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA* **98**, 10869–10874 (2001).
120. Paik, S. *et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351**, 2817–2826 (2004).
121. Chang, H. Y. *et al.* Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol.* **2**, e7 (2004).
122. Chi, J. T. *et al.* Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. *PLoS Med.* **3**, e47 (2006).
123. Bild, A. H. *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353–357 (2006).

Acknowledgements
We apologize for omitting primary references and the work of those we could not cite owing to space limitations. We would like to thank A. Incassati, A. Chiang, P. Bos, D. Padua, G. Gupta and S.Tavazoie for insightful discussions. J.M. was funded by the National Institutes of Health, USA, grant P01-94060, and by a grant of the Keck Foundation, USA. D.X.N. is a Berlex postdoctoral fellow of the Damon Runyon Cancer Research Foundation, USA. J.M. is an Investigator of the Howard Hughes Medical Institute, USA.

Competing interests statement
The authors declare no competing financial interests.

DATABASES
The following terms in this article are linked online to:
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
BRCA1 | *CASP8* | *CDH1* | *CD82* | *CSN5* | *EREG* | *HER2* | *HRAS* | *KRAS* | *lethal giant larvae* | *MMP1* | maspin | *MYC* | *NM23* | *NEDD9* | *PTGS2* | *p53* | *semaphorin 5c* | *Sipa1* | *SIPA1* | *VHL*
UniProtKB: http://ca.expasy.org/sprot
CXCR4 | DARC | E-cadherin | EZH2 | ezrin | GM-CSF | HIF1α | NFκB | pVHL | RhoC | SDF1 | SIX1 | slug | twist | TGFB | VEGF | VEGFR1 | VEGFR2

FURTHER INFORMATION
The Massagué laboratory at HHMI: http://www.hhmi.org/research/investigators/massague.html
The Massagué laboratory at SKI: http://www.ski.edu/lab_homepage.cfm?lab=133
Access to this links box is available online.