

REVIEW

Received 22 Jul 2016 | Accepted 29 Mar 2017 | Published 25 May 2017

DOI: 10.1038/ncomms15422

OPEN

# Modeling the process of human tumorigenesis

Sneha Balani<sup>1</sup>, Long V. Nguyen<sup>1</sup> & Connie J. Eaves<sup>1</sup>

Modelling the genesis of human cancers is at a scientific turning point. Starting from primary sources of normal human cells, it is now possible to reproducibly generate several types of malignant cell populations. Powerful methods for clonally tracking and manipulating their appearance and progression in serially transplanted immunodeficient mice are also in place. These developments circumvent historic drawbacks inherent in analyses of cancers produced in model organisms, established human malignant cell lines, or highly heterogeneous patient samples. In this review, we survey the advantages, contributions and limitations of current *de novo* human tumorigenesis strategies and note several exciting prospects on the horizon.

Identifying established malignancies in humans is usually not a difficult clinical issue as most human cancers do not become symptomatic until they are well advanced. Indeed, by the time they are detected, most will already consist of millions, if not trillions, of cells exhibiting many abnormal features<sup>1</sup>. These include evidence of invasive behaviour, deregulated growth, cells with an abnormal morphology and disorganized histology, and possession of a diversity of mutations<sup>2</sup>. How these can be usefully combined to generate more refined assessments of categories and stages of human cancer development has, however, challenged pathologists for decades. In addition, the molecular events involved in the early genesis of malignant human cell populations have been particularly elusive. This is because these stages are not usually detected in patients and, when they are, very little if any tissue is available for research studies. At the same time, there is expanding interest in the possibility that a better understanding of the initial changes that lead to an irreversibly transformed state and new ways to identify such changes might revolutionize early detection strategies as well as therapeutic success rates. Many approaches, both prospective and retrospective, have thus been developed to recreate and interrogate the process of tumorigenesis. All have particular advantages, but also significant caveats and shortcomings. What is new, are recent technological advances that are now enabling malignant populations of human cells to be generated *de novo* from primary tissue sources.

Here, we first review the background of information on which current understanding of the process of human oncogenesis has been founded. This is then followed by a review of newer developments and findings emanating from *de novo* tumorigenesis experiments that are driving new concepts relevant to this rapidly evolving topic. When coupled with unbiased DNA barcoding, reprogramming, and CRISPR/Cas9 technologies, these methods hold promise for obtaining further insights into the different stages of development of malignant human cell populations with unprecedented precision and clinical relevance.

<sup>1</sup>Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, British Columbia, Canada. Correspondence and requests for materials should be addressed to C.J.E. (email: ceaves@bccrc.ca).

### The Past—origin of current concepts

**Tumorigenesis viewed as an evolutionary process.** The concept that almost all human tumours represent aberrant clonal outgrowths is well established<sup>3</sup>. However, this simply means that the malignant population that ultimately appears represents the deregulated growth of a single cell. It does not imply biologic or even genetic identity among its progeny. Nevertheless, it does make it likely that all members of the clone will carry a trace of the earliest genetic or epigenetic changes that drove its deregulated expansion. It is also important to remember that, by the time a cancer is first detectable, it will already contain many millions or even billions of cells produced through dozens of amplifying divisions. During this process, additional genetic diversification and evolution occurs (Fig. 1). This is due in part just to normal rates of incorrect DNA replication (estimated as  $2.3 \times 10^{-8}$  mutations per nucleotide per cell generation in human cells)<sup>4</sup>. A decreased control of DNA stability is also a common feature of malignant cells. Other mechanisms that contribute to the variable diversification of malignant clones include the tissue and genotype of the cell in which the process of transformation is initiated, the age of the individual in which this process begins and numerous environmental factors<sup>5–9</sup>.

Modern sequencing data has now revealed the enormous heterogeneity that exists within the genomes of malignant populations<sup>10</sup>. This heterogeneity is often apparent within a single cancer assessed at a single moment and sampled from a single site. Additional heterogeneity may also be encountered when different sites are examined, or the same tumour is sampled at different times, or from different individuals with tumours that have arisen in the same tissue. The genetic diversity that characterizes many malignant populations has multiple causes. One is that the full transformation of human cells relies on the accumulation of multiple mutations through a stepwise process in which different rare mutations and/or epigenetic changes must be accrued in order for cells to acquire a deregulated ability for unlimited division as well as other abnormal phenotypic properties<sup>11,12</sup>. From this concept, it has been inferred that the process of spontaneous oncogenesis in humans would most often be slow, at least initially, spanning a period of several years. Such a model, in turn, implies that most cancers would originate in a cell type that is normally long-lived but still has sufficient proliferative capacity to accumulate the number of rare mutations

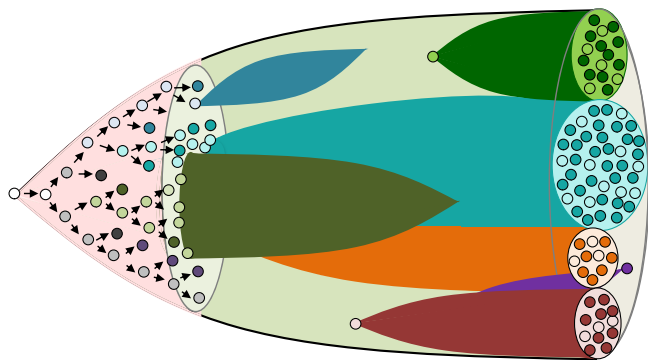
required to become transformed. Interestingly, it also implies the possible generation of neoplastic clones with mutations that predispose to further transformation but may never do so, as is being increasingly documented in many tissues<sup>13–15</sup>. However, accelerating the process, and a cause of more rapidly increasing genetic diversity, would be the acquisition of one or more mutations or epigenetic changes that compromise genomic stability<sup>16–19</sup>. At the same time, recent genomic analyses of some human cancers are suggesting alternative trajectories in which transforming populations early on may undergo cataclysmic bursts of genomic diversification<sup>20</sup>.

Most of our knowledge of the process of oncogenesis in human tissues has been derived historically from 3 different types of investigation: (i) genomic analyses of cells obtained from patients at different times during the development and evolution of their cancer with retrospective inferences drawn from the spectrum and prevalence of sequence changes (best determined at the single cell level)<sup>21–24</sup>; (ii) inferences derived from single cell or clonal transcriptome studies<sup>25,26</sup>; and (iii) studies of genetically engineered malignant populations in model organisms, and mice in particular<sup>27–29</sup>. The first two of these types of investigation have shown that most human cancers, by the time they become clinically evident, consist of cells that have already acquired multiple mutations<sup>2</sup>. However, the extent to which the time course or order in which the same mutations or other relevant (for example, epigenetically determined) changes are accrued in any given human cancer still remains poorly understood<sup>30</sup>.

Although, mutations do not occur randomly<sup>31–33</sup>, the likelihood of a given complement of mutations being accrued in the lifetime of an individual human cell or its progeny remains low. It is therefore not surprising that the more primitive cells in tissues of high cell turnover; that is, those cells responsible for sustaining the required cell outputs of the tissue, have commonly been assumed to serve as reservoirs of premalignant changes that accrue throughout life.

In the human hematopoietic system, this concept has been most strongly supported by genomic analyses of cells from patients with chronic myeloid leukemia (CML). This leukemia is unusual in that it is typically diagnosed at an early stage (referred to as chronic phase). Although the clone can already be huge (typically more than a trillion cells), the process of differentiation is minimally perturbed and hence most of the cells are actually highly differentiated, short-lived mature blood cells, indistinguishable from their normal counterparts. Cell production within the chronic phase clone, in fact, largely mirrors the normal process of adult human hematopoiesis and is likewise sustained by a phenotypically similar and rare subset of multi-potent self-renewing CML (*BCR-ABL1*+) cells as shown by their unique capacity for prolonged CML cell output both *in vitro* and *in vivo*<sup>34,35</sup>. Interestingly, effective treatment of CML patients with tyrosine kinase inhibitors (TKIs) selectively eliminates the bulk of the clone (>99.9%), but has less effect on the CML stem cells which can then rapidly regenerate an enlarged derivative population if TKI administration is transiently arrested<sup>36</sup>. Moreover, if the chronic phase clone is not adequately reduced, subclones of CML cells with additional mutations and biological properties eventually appear and produce an inevitably fatal acute lymphoid or myeloid leukemia<sup>37</sup>. Additional studies have suggested that at least some other forms of human acute myeloid leukemia (AML)<sup>23,38–42</sup> myelodysplastic syndromes (MDS)<sup>43,44</sup> fit a similar model of step-wise disease progression.

**The cancer stem cell concept.** The restriction of transplantable tumour-initiating ability (in immunodeficient mice) to a rare,



**Figure 1 | Schematic depiction of the subclonal evolution and diversification of cell types in developing malignant populations.** In this diagram, subclones identified by accumulating genetic changes are shown by different colours. Cells within each clone that have proliferative potential are shown as pale cells in contrast to some of their progeny that can no longer divide that are shown as dark cells (to illustrate the diversification of biological properties that occurs both within and between subclones), with some clones being transient, whereas others are persistent but variably expanding.

prospectively separable subset of malignant human cell populations was first demonstrated in human AML<sup>45</sup>. This was followed soon thereafter by analogous experimental evidence of tumour-initiating ability confined to distinct and prospectively separable subsets of cells in primary samples of human breast<sup>46</sup>, brain<sup>47</sup> and colon cancers<sup>48</sup>. Together, these findings established the concept of cancer stem cells thus identified by their transplantable tumour-initiating property. Phenotypic and transcriptional similarities between the subsets of malignant human cells and the stem cells of the tissue of origin have reinforced the association of this 'cancer stem cell' functionality with the retention or reactivation of molecular mechanisms essential to the maintenance of normal tissue stem cells. Of particular interest in this regard, has been an association of quiescence with many types of both normal and cancer stem cells<sup>49–53</sup>, thus accounting for the innate resistance of the cancer stem cell population to standard chemotherapeutic agents that primarily target dividing cells. The relative insensitivity of these cells to treatment also makes them dangerous reservoirs for the accumulation of additional mutations that make their elimination even more problematic.

However, in other human cancers, like melanoma<sup>54</sup>, a majority of the cells with self-sustaining tumorigenic properties is often already apparent at diagnosis. In addition, there is increasing evidence of the reversibility of the properties that confer transplantable tumour-initiating activity in a growing number of human tumours. This phenomenon, referred to as phenotypic instability or plasticity, is thought to reflect changes triggered by cues from the environment or stochastically regulated intrinsic parameters<sup>55–58</sup>, and may add significantly to the heterogeneous dynamic evolution of tumours caused by a stepwise and branched process of mutation accrual noted already.

**Transgenic mouse models of *de novo* tumorigenesis.** Mouse models have been valuable for analysing early oncogenic events because of the ease with which their genomes can be genetically manipulated in defined manners and times. This has made it possible to interrogate many specific molecular, cellular and tissue-specific characteristics associated with inherited and somatically acquired mutations in humans that have been inferred to play a role in their transformation. There are, however, numerous physiological differences between mice and humans that limit the utility of this approach to provide insights into mechanisms operative in humans. Of particular note in this regard is the apparent ease with which mouse cells can be transformed, in contrast to their human counterparts<sup>59</sup>.

Mouse models have been particularly helpful in tracing the clonal origin of normal and transformed cells in various tissues. This type of experiment typically takes advantage of a known lineage-specific gene promoter to transgenically engineer the expression of an oncogene exclusively in the cell types in which that promoter is active. This strategy can thus be used to track subsequent changes in predefined cell types before the formation of a detectable tumour. In addition, this strategy has been useful for identifying the origin of any malignant populations ultimately obtained and demonstrating different properties in the tumours thus generated<sup>27,60–63</sup>.

For example, in the adult mammary gland, the specificity of expression of different types of cytokeratins associated with either basal or luminal mammary epithelial cells has enabled tumours derived from these different cell types to be produced and compared. These include the production of estrogen receptor-positive and progesterone receptor-positive (ER<sup>+</sup>PR<sup>+</sup>) and ER<sup>-</sup>PR<sup>-</sup> tumours from mouse basal and luminal cells

engineered separately to express *PIK3CA*(H1047R), thereby creating models that appear to mirror analogous phenotypes of human breast cancer<sup>64,65</sup>. However, it is also noteworthy that potent stimulation of normal adult mouse luminal mammary cells, even in the absence of a mutant gene, can induce their production of both basal and luminal progeny<sup>66</sup>, despite evidence of their more restricted differentiation behaviour *in vitro* and *in vivo*<sup>67,68</sup>. Similar transgenic models have been used to study concomitant *PIK3CA* expression and *p53* deletion<sup>64,65,69</sup>, as well as concomitant *BRCA1/PTEN/p53* depletion in mouse mammary cells<sup>70</sup>.

Familial adenomatous polyposis, a human condition with a recognized step-wise process of transformation, is another interesting example of a tumour that has been modelled in mice. In this case, the inducible loss of only one *APC* allele in *Lrig1*-expressing cells resulted first in their acquisition of pre-neoplastic changes and then their generation of multiple colonic tumours within 50 days<sup>71</sup>. In the epidermis, *Sox2* expression has been found to initiate the *de novo* growth of a squamous cell carcinoma<sup>72</sup>, and glioblastoma in the brain<sup>73</sup>. On the other hand, forced expression of *Sox9* in the epidermis was sufficient to induce the *de novo* formation of basal cell carcinomas<sup>74</sup>. Mouse models have been similarly used to study the ability of *Wnt*-activation<sup>75</sup>, oncogenic *Ras*<sup>76</sup>, *Myb*<sup>77</sup>, and *Pgc-1b*<sup>78</sup> to initiate *de novo* tumorigenesis in intestinal epithelial cells. Studies in the hematopoietic system of mice have also reported analogous strategies for the *de novo* generation of AML<sup>79</sup>, plasma cell neoplasms<sup>80</sup>, and lymphoid malignancies<sup>81</sup>.

A key advantage of such transgenic mouse models of oncogenesis is their ability to mimic the effect of acquiring endogenous oncogenic mutations under homeostatic conditions. They also enable events critical to the initiation of tumour formation within the normal tissue architecture to be examined. However, it is also possible that the level of oncogene expression driven by promoters of lineage-specific (tracer) genes would be different from that characteristic of tumours caused by mutations in the corresponding endogenous genes. The different epigenetic mechanisms regulating their expression in the two different sites would be expected to contribute to the types, or at least speed, of the phenotypic changes induced. Furthermore, interactions of spontaneously arising mutant human cells within their native microenvironments, as well as intrinsic aspects of the process of normal tissue control, may not be well modelled in the mouse. On the other hand, transgenic mouse models offer the advantage that they can now be readily engineered to regulate not only the lineage, but also the level, timing and duration of expression of genes of interest that have been introduced into the germ line.

Mouse knock-in (KI) protocols now enable some of the shortcomings of historic transgenic models to be addressed. These involve the use of homologous recombination to replace endogenous genes in mouse embryonic stem cells with candidate oncogenic mutant alleles and subsequent introduction of the selected cells into the mouse germ line. KI mice have thus been generated to explore the function of *Kit* mutations in gastrointestinal stromal tumours, the role of *TP53* in Li-Fraumeni syndrome tumours and for recapitulating familial cancers. This technique has also been further adapted to induce spontaneous expression of the oncogenic *KRAS*(*G12D*) gene enabling the genesis of lung cancer, thymic lymphoma, and aberrant intestinal crypt foci to be created and studied<sup>82</sup>. CRISPR/Cas9 technology to introduce oncogenic mutations into the endogenous locus of the mouse germ line (with or without controlled expression) is now offering even greater precision in the types of genetic manipulations that can be deployed to analyse downstream biological effects<sup>83–85</sup>.

**Transformation of human cell lines.** Much of the process of malignant transformation of human cells has historically focused on immortalized non-tumorigenic human cell lines as the starting material. MCF-10A is an example of a human mammary epithelial cell line of this type that has been used for this purpose. Forced over-expression of *MYC*<sup>86</sup>, *EGFR*<sup>87</sup>, *Nicastrin*<sup>88</sup>, *ERBB2* (ref. 89), and *RANK*<sup>90</sup> in MCF-10A cells produced several phenotypic changes associated with transformation. These included changes in their growth properties in three-dimensional (3D) cultures<sup>91</sup>, as well as in xenotransplant models<sup>89</sup>. However, a study based on a non-tumorigenic human prostate cell line, EPT2-D5, showed that metastatic prostate tumours could be generated in xenografted mice simply by first culturing the cells in a protein-free media to generate 'spheres'<sup>92</sup>. Moreover, recent epigenomic and transcriptional analysis of MCF-10A and two other immortalized (184-hTERT) non-tumorigenic human mammary cell lines has revealed that all three have acquired significant differences from primary sources of human mammary cells<sup>93</sup>. It thus remains unclear how well the transformation of such immortalized cell lines will reflect those relevant to the spontaneous genesis of cancers from primary sources of human cells.

### The present—utility of *de novo* human models

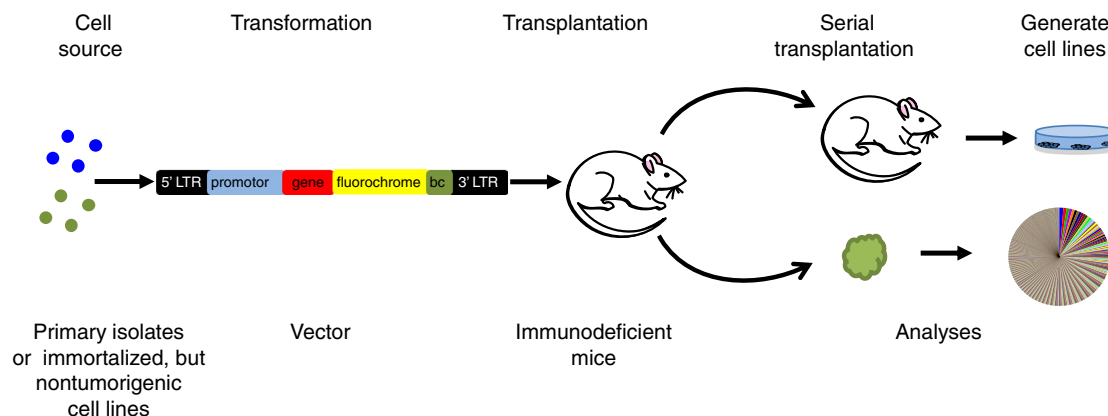
**External factors.** It is also now clear that most malignant cells are not capable of autonomous growth, although the acquisition of autocrine properties and other intrinsic mechanisms for enhanced auto-stimulation and survival are common. Tumours, like normal tissues, also require a blood supply and frequently contain numerous other non-clonally derived cell types, even to the point where these latter cell types may become predominant<sup>94–96</sup>. These cells may include many different stromal cell types as well as the components of vessel walls and extravasated blood and lymphoid cells. External cues emanating from these different sources both positively and negatively regulate the local rate of growth of neighbouring malignant cells as well as their invasive behaviour<sup>97,98</sup>, thereby contributing to the overall dynamic evolution of the genomic, biologic and geographic properties of the entire population of malignant cells<sup>10</sup>.

**Persistence of tissue-specific programs.** In addition to these multiple sources of cellular heterogeneity, malignant populations

also frequently retain many of the distinguishing but heterogeneous features of the tissue from which they have arisen even after they have metastasized. In fact, almost by definition, a stepwise model of transformation would imply that this would have to be the case unless it completely froze the state of the first affected cell resulting in the generation of a clone of identical daughter phenotypes. If not, then it would be predicted that the composition of a malignant clone would also reflect some of the phenotypic heterogeneity determined by persisting elements of processes that normally sustain the maintenance, organization and differentiation of cells in that tissue<sup>99</sup>. Thus, the phenotypic heterogeneity inherent in many bulk malignant populations would be anticipated to also reflect some of the complex diversity of persisting epigenetic and transcriptional elements characteristic of the formation and maintenance of the tissue from which the cells arose. In this regard, it is noteworthy that different tissues make different utilization of multiple, albeit overlapping basic mechanisms that ultimately co-ordinate their proliferative capacity, survival and epigenetic control of lineage restriction to ultimately establish non-dividing or nonviable cell states. Within the same tissue, these may also change during development and aging as well as under different conditions of homeostasis, wounding, inflammation and infection. Even at a single moment in time, there may be multiple pathways by which a single mature cell type can be generated. Moreover, the properties of 'self-renewal' and long-term quiescence in normal cells are now known, at least as suggested by lineage tracing studies in mice, to be more broadly distributed and less rigidly associated with the capacity to regenerate the entire tissue<sup>100–102</sup>.

These issues are important because they have necessitated a modification to the historic assumption as to the singular state of 'stem cells' even within a given tissue. Hence a greater diversity of cell types can now be envisaged as potential initial cells in which the genetic and epigenetic changes, required to generate irreversibly expanding malignant clones, might originate. As detailed below, these considerations also undermine assumptions underpinning the historic concept of a cancer stem cell, without negating it altogether—a difficult development that the field has now had to accommodate<sup>10</sup>.

***De novo* human models of tumorigenesis.** Experimental models of *de novo* tumorigenesis starting from cells isolated directly from normal human tissues are attractive because they circumvent the



**Figure 2 | *De novo* generation of tumours from 'normal' human cells.** Most examples of successful transformation of primary sources of normal human cells (or non-tumorigenic human cell lines) have used retro- or lenti-viruses encoding one or more oncogenes and a fluorochrome (for example, GFP) to enable malignant cells to be later isolated and characterized. The transduced cells are then transplanted into a receptive site in immunodeficient mice. When a tumour forms, the cells can then be removed for morphological, immunohistochemical, flow cytometric and/or various molecular and clonal analyses. When this method is efficient, polyclonal tumours may be generated (as illustrated by the pie chart). Retrieved viable cells can also be further transplanted or may be used to generate cell lines.

**Table 1 | Examples of *de novo* tumour models from primary human cell sources.**

Human Tissue	Gene	Finding	Reference
Prostate	AKT, ERG, AR	Tumour formation in basal cells but not in luminal cells.	103
Colon	APC, SMAD4, TP53, KRAS, PIK3CA	Organoids engineered to express all five mutations grew independently of niche factors <i>in vitro</i> , and could be transplanted to form tumours in mice.	104
Mammary	p53(R175H), CCND1, PIK3CA, and KRAS(G12V)	Cells with BRCA1-mutation form tumours and showed increased basal differentiation compared to cells from non-carrier tissues. EpCAM + CD10 – luminal cells from both BRCA1+/+ and BRCA1mut/+ tissues were enriched for tumour-forming ability.	105
	SV40 and KRAS(G12V)	Transformation of EpCAM + cells resulted in the development of common forms of breast cancer, including ER + and ER – tumours with luminal and basal-like characteristics, respectively. Transformation of CD10 + (basal) cells resulted in the development of rare metaplastic tumours similar to the claudin-low subtype.	106
	KRAS(G12D)	Both basal and luminal cells generated polyclonal serial transplantable tumours containing a mixture of phenotypes and clones with variable growth dynamics revealed in serial transplants.	107
Blood	MLL-AF9	Some leukemia stem cells (LSCs) were multi-potent and could be lineage directed by altering either the growth factors or the recipient strain of mouse, highlighting the importance of microenvironment. Other LSCs were strictly lineage committed, demonstrating the heterogeneity of the stem cell compartment in the MLL diseases produced.	108,109
	MLL-AF4	Generation of a model of t(4;11) pro-B ALL that fully recapitulated the immunophenotypic and molecular features of the disease that appears in patients.	110
	MN1 and NUP98HOXD13	Co-transduction of an activated HOX gene (NUP98HOXD13) with MN1 induced a serially transplantable AML.	111
	BCR-ABL1 and dnIKAROS (IK6)	An aggressive AML with disseminated myeloid sarcomas developed within 4 weeks following transplantation of cord blood cells transduced with both genes.	112
	MYC and BCL2	Production of a model of lymphoma that recapitulates the histopathological and clinical aspects of steroid-, chemotherapy- and rituximab-resistant human 'double-hit' MYC-BCL2 lymphoma.	113
	DEK-NUP214	Development of a human cell AML with phenotypic characteristics of a t(6;9) disease and CD45 + CD13 + CD34 + CD38 + immunophenotype.	165
ZMYM2-FGFR1	Development of myeloproliferative disease that progresses to AML. Mice show hepatosplenomegaly, hypercellular bone marrow and a CD45 + CD34 + CD13 + immunophenotype.	166	

concerns inherent in extrapolating from immortalized cell line data or species differences (Fig. 2). However, the frequency of success has thus far been much more limited, perhaps due to a historic lag in the development of appropriate methods to isolate the relevant target cells in viable form, and/or to transduce them at an adequate efficiency. Indeed, where these issues have been carefully addressed, some models have been generated (Table 1). For example, normal human basal prostate cells transduced with a combination of vectors encoding cDNAs for *AKT*, *ERG* and the androgen receptor have been found to produce tumours in transplanted immunodeficient mice<sup>103</sup>. In contrast, in the same study, transplants of co-isolated luminal prostate cells transduced with the same vectors did not yield tumours. This result is interesting because the gene expression profile of prostate cancers appears closer to that of the luminal cells of the normal prostate. One explanation is that prostate cancers in which these genes are characteristically altered actually originate in basal cells that then generate progeny that have luminal features<sup>103</sup>. Alternatively, it may be argued that a malignant phenotype can originate *in vivo* directly in luminal cells but the conditions used to date simply fail to support this process experimentally.

*De novo* models of tumorigenesis starting from primary human cells have also now been reported for colon and mammary cells. For example, a recent study demonstrated the formation of tumours in immunodeficient mice transplanted with organoids expanded *in vitro* from colon cells in which CRISPR/Cas9 was used to generate suppressive mutations in *APC*, *SMAD4* and *TP53* and activating mutations in *KRAS* and *PIK3CA*<sup>104</sup>. *De novo* genesis of human breast tumours has also now been achieved in

immunodeficient mice transplanted with primary isolates of normal cells transduced with *p53(R175H)*, *CCND1*, *PIK3CA* and *KRAS(G12V)*<sup>105</sup>, *SV40* plus *KRAS(G12V)*<sup>106</sup>, and more recently with just *KRAS(G12D)* alone<sup>107</sup>. However, in contrast to the results described for the prostate model, immunohistological analyses of these human breast tumours has indicated the presence of a mixture of phenotypes, possibly related to the polyclonal composition of the tumours generated<sup>107</sup>. The robustness of these models and speed of the tumorigenic process in at least some cases should make them useful for future elucidation of the minimal cellular and extrinsic factors required for their genesis.

Examples of *de novo* leukaemogenesis using primitive (CD34<sup>+</sup>) subsets of hematopoietic cells isolated from human cord blood is also accruing. In this case, examples of genes whose vector-mediated forced overexpression in normal cells have produced overt leukaemic populations in immunodeficient mice include cDNAs for *MLL-AF9* (refs 108,109), *MLL-AF4* (ref. 110), *MN1* plus *NUP98HOXD13* (ref. 111), *BCR-ABL* plus a dominant-negative form of *IKAROS*<sup>112</sup>, and *MYC* plus *BCL*<sup>113</sup>. The mixed lineage leukaemia (*MLL*) gene is rearranged and fuses with multiple partner genes in both spontaneously arising human AML and acute lymphoid leukemia (ALL), but the *MLL-AF9* fusion oncogene is associated almost exclusively with AML in humans<sup>108</sup>. Interestingly, overexpression of *MLL-AF9* in normal cord blood cells produced ALL in transplanted Non-obese diabetic-Scid (NOD-SCID) hosts, but AML in NOD-SCID mice that had been engineered to express three human growth factors. This result illustrates the ability of external factors to dictate the

phenotype of the malignant cells produced and further underscores the fallacy in assuming the predominant cell type necessarily reflects the cell of origin or the specific oncogene driving the tumorigenic state. The *MNI-NUP98HOXD13* model also serves to illustrate the difference between effects obtained in analogous mouse and human target cells. In mouse cells, *MNI* alone was sufficient to induce a myeloid leukaemia, whereas in human cord blood cells it induced only a transient myeloproliferation. Only when an activated *HOX* gene (that is, *NUP98HOXD13*), was also introduced was a serially transplantable AML obtained in the human cells.

These studies demonstrate some of the unique aspects of tumorigenesis in primary human cells that are not well predicted by mouse models. They also provide examples of experimentally generated *de novo* human tumours whose characteristics are heavily influenced by factors beyond the cell of origin, or the genetic alterations used.

**Functional diversity of human tumour subclones.** Most studies attempting to analyse the subclonal diversity of malignant human populations that arise in patients have relied on retrospective inferences derived from analyses of the genomic or phenotypic characterization of primary samples and/or changes incurred in serial transplants of these cells in immunodeficient mouse hosts<sup>23,114–118</sup>. However, these approaches generally do not allow the frequency of clonogenic cells to be quantified, nor the definition of their genomic or phenotypic properties. These are important parameters because the genomic instability of many established neoplastic human populations can produce genetic alterations that are irrelevant to the continued growth of the tumour. Examples of such mutations have been well documented in genetic analyses of the leukemic cells from patients with chronic phase CML<sup>119</sup>.

Limiting dilution analysis (LDA) and vector-based tracking of clonogenic cells offer powerful approaches to quantify malignant cells with proliferative potential *in vivo*. The LDA approach has been extensively applied to a number of primary human tumour types that can engraft immunodeficient mice. These include malignant populations that arise in the human brain<sup>47,120</sup>, colon<sup>48,121</sup>, prostate<sup>122</sup>, breast<sup>46</sup>, ovary<sup>123</sup>, skin/melanoma<sup>54,124</sup> and the hematopoietic system<sup>45,125–128</sup>. Coupling these approaches to prospectively isolated phenotypes of cells within the transplanted populations has been the basis of identifying the subset(s) of cells actually possessing the tumorigenic activity detected in the recipients. However, to exploit this approach to characterize the clonal growth properties of *single* tumorigenic cells, the transplants must be initiated with cell numbers that produce tumours at a frequency of less than one in three mice. The large numbers of mice required to obtain clonal data can thus become rapidly prohibitively expensive, if not impractical, particularly when clonal outputs may vary quantitatively as well as phenotypically and dynamically within and between different primary tumours being assessed.

The use of massively parallel DNA sequencing has improved the sensitivity and efficiency of vector insert analysis to identify clones based on their semi-random integration sites<sup>129</sup>, thus circumventing many of the limitations of LDA approaches. Nevertheless, ~50% of insertion sites are still elusive to detection using the latest versions of this methodology<sup>129</sup>. A more recent alternative has been the creation and use of libraries of DNA barcoded lentiviral vectors<sup>130</sup>. When such libraries are used at a concentration that results in the transduction of less than one in three cells in the target population being interrogated, and the diversity of the library is greater than the number of cells being investigated, the likelihood that each of the initially marked cells

contains a unique barcode is high. The frequency of each barcode in the DNA extracted from populations derived from barcoded cells can then be used to infer the number of cells present in different clones. For both vector strategies, phenotypic purification of the cells to be analysed before extracting the DNA for insert site or barcode analysis allows the phenotypic composition of each clone to also be characterized<sup>109,131–137</sup>.

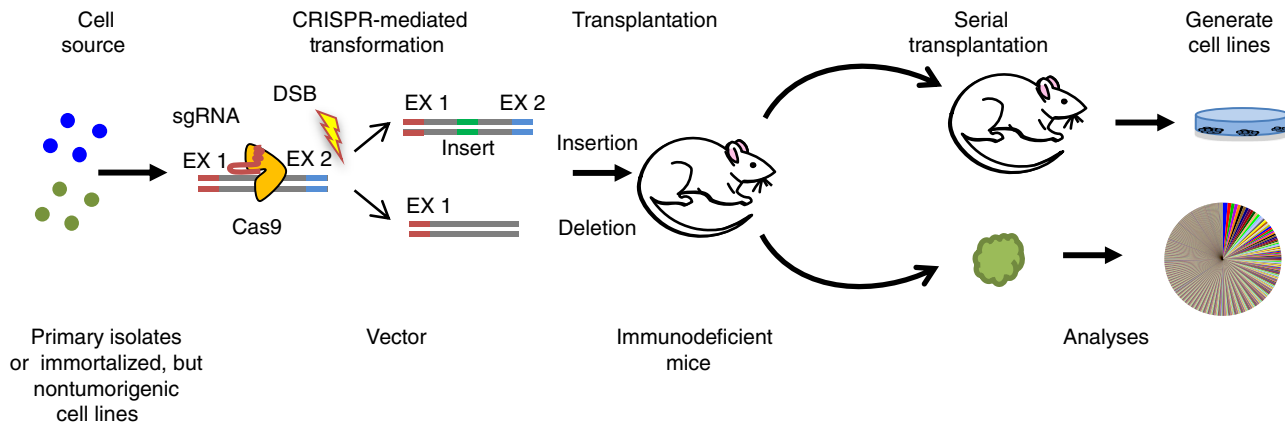
Application of these protocols has revealed complex clonal dynamics in human tumours derived from serially passaged cells<sup>131</sup>. Interestingly, this exercise has documented highly diverse growth behaviours of single cells from cell lines when large numbers of these are transplanted, even though single-cell transplant experiments indicate more than one third can generate a tumour. Thus, to relate the growth activity of the clones with their genotype, it is necessary to isolate and sequence the individual clones generated, as has been done in human glioblastoma<sup>138</sup>.

### The future—new opportunities

Technological advances are well known to be the drivers of new information. Nowhere is this more relevant than to current issues of interest in the general arena of biomedicine. In the field of human cell oncogenesis, three technologies are of particular relevance: organoid cultures, reprogramming and CRISPR/Cas9. Organoid culture systems are allowing more faithful tissue development to be generated from primitive normal cells and are showing promise for supporting the expansion *in vitro* of cells derived directly from human tumours<sup>139–141</sup>. Reprogramming allows permanent clonal lines of induced pluripotent stem cells (iPSCs) to be generated from individual malignant cells. CRISPR/Cas9 enables the use of precise gene editing to create new models of human tumours and to test the role of specific genetic alterations in establishing or maintaining the transformed status of patient-derived malignant cells.

**Organoids.** This term refers to 3D structures generated in suspension cultures under conditions where 2D growth of attached cells is blocked or prevented (either by suspending the cells in a semi-solid medium-like matrigel or the use of a culture container that prevents cell attachment). The generation of 3D organoids to engineer tissues that are histologically and functionally more similar to their *in vivo* counterparts will facilitate the modelling of human disorders, performing drug screens, and the creation *in vitro* of replacement tissues and/or organs. Methods combining directed differentiation of cells in culture systems that promote organoid formation are being developed for many complex tissues such as the liver, kidney, intestine, eye, and brain, to name a few<sup>142</sup>. The advantages of human organoid cultures include better control of the cellular milieu than is associated with tumorigenesis *in vivo*, better spatial organization of cell types than is achieved in 2D *in vitro* systems and improved mimicry of *in vivo* cell behaviour, potential for larger scale testing of variables and drug effects, and reduced ethical concerns and costs associated with animal xenograft experiments.

**Induced pluripotent stem cells.** iPSC technology offers a novel method of capturing the genomic state of single cells from almost any source, including transformed cells. However, this comes at the cost of losing the malignant phenotype of the original cells from which each iPSC clone is derived. Nevertheless, from the limited experience where it has been applied to the study of cancer to date, it appears that at least some properties of the malignant cells of origin may be reactivated by stimulating the iPSCs to differentiate back into the lineage or tissue from which the malignant cells arose and may then also be used for drug screening<sup>143,144</sup>.



**Figure 3 | Use of CRISPR/Cas9 gene editing to examine the tumorigenic consequences of modifying specific genes in human cells.** In this methodology, the test cells are exposed to CRISPR/Cas9 reagents and then transplanted into immunodeficient mice as in Fig. 2.

iPSCs have been derived from human chronic phase CML cells with documented retention and expression of the signature *BCR-ABL1* fusion gene<sup>145–147</sup>. Additional examples of human hematopoietic cells from which iPSCs (or iPSC-like cells) have been obtained include cells from patients with MDS<sup>148</sup>, polycythemia vera<sup>149</sup> and juvenile myelomonocytic leukemia<sup>150</sup>, as well as originally normal human hematopoietic cells forced to overexpress *MLL*<sup>151</sup>. iPSCs have also been generated from human pancreatic ductal adenocarcinomas<sup>152</sup>, Multiple endocrine neoplasia type 2A (MEN2A)<sup>153</sup>, bladder cancer cell lines<sup>154</sup> and cells from an individual with Li Fraumeni syndrome, a congenital cancer predisposition genotype in which one allele of the *p53* gene is mutated<sup>155</sup>. In this case, the iPSCs were used to study the epigenetic states of tumour formation<sup>156</sup> and the bidirectional reversibility of epigenetic processes associated with iPSC reprogramming<sup>157</sup>.

**CRISPR/Cas9.** The CRISPR/Cas9 technology originates from type II CRISPR-Cas systems that evolved to provide bacteria with a form of adaptive immunity to viruses and plasmids. The CRISPR-associated Cas9 protein is an endonuclease that introduces a site-specific double-strand break (DSB) into DNA. The ability of CRISPR/Cas9 to introduce DSBs at defined positions using a guide RNA can also be used to generate human cell lines and primary cells bearing chromosomal translocations resembling those described in human cancers (Fig. 3). Examples to date include lung cancer<sup>158</sup>, AML<sup>159</sup> and Ewing's sarcoma<sup>160</sup>. An improved method to generate models of liver cancer or myeloid malignancy in mice using CRISPR/Cas9 was also recently reported<sup>161,162</sup>. Combining CRISPR/Cas9 with barcoding has been used to track the clonal dynamics of lung cancer cells during their acquired resistance to EGFR inhibitors and subsequent response to combined drug therapies<sup>163,164</sup>. Combining the generation of iPSCs with CRISPR/Cas9 holds the future possibility of both creating new models of human cancer and of testing the ultimate therapeutic potential of gene-targeting approaches.

### Summary

Recreating the process of human tumorigenesis in naive human cells has been a challenging frontier not anticipated by the historically contrasting ease of achieving this outcome in mouse cells. However, several major developments are now changing this picture. One has been the generation of long-lived mice in which multiple components of the innate and acquired immune system have been suppressed. This advance now allows transformed human cells to grow and evolve into fully malignant populations in an *in vivo* setting. Together with improvements in

lentiviral design and production, and methods to isolate and purify normal viable human cells, these advances are now enabling a growing number of *de novo* models of human tumorigenesis to be developed. The use of DNA barcoding of the initial cells is now also providing a tractable method to analyse the clonal composition and dynamics of the malignant cells generated and should provide a powerful approach to examine therapeutic effects in the future. These approaches coupled with emerging methods to modify and regulate cell behaviour even more precisely herald a new era of improved understanding of the complex processes of human tumorigenesis.

### References

- Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation *Cell* **144**, 646–674 (2011).
  - Atlas, T. C. G. TCGA Research Network Publications. Available at: <https://cancergenome.nih.gov/publications>.
  - Friedman, J. M. & Fialkow, P. J. Cell marker studies of human tumorigenesis. *Transplant. Rev.* **28**, 17–33 (1976).
  - Nachman, M. W. & Crowell, S. L. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297–304 (2000).
  - Whiteman, D. C. & Wilson, L. F. The fractions of cancer attributable to modifiable factors: A global review. *Cancer Epidemiol.* **44**, 203–221 (2016).
  - Sitas, F. Twenty five years since the first prospective study by Forman *et al.* (1991) on *Helicobacter pylori* and stomach cancer risk. *Cancer Epidemiol.* **41**, 159–164 (2016).
  - Grulich, A. E., van Leeuwen, M. T., Falster, M. O. & Vajdic, C. M. Incidence of cancers in people with HIV/AIDS compared with immunosuppressed transplant recipients: a meta-analysis. *Lancet (London)* **370**, 59–67 (2007).
  - McMahon, B. *et al.* Hepatitis-associated liver cancer: gaps and opportunities to improve care. *J. Natl Cancer. Inst.* **108**, 4. doi:10.1093/jnci/djv359 (2016).
  - Brock, A., Krause, S. & Ingber, D. E. Control of cancer formation by intrinsic genetic noise and microenvironmental cues. *Nat. Rev. Cancer* **15**, 499–509 (2015).
  - Rosenbloom, D. I., Camara, P. G., Chu, T. & Rabadan, R. Evolutionary scalpels for dissecting tumor ecosystems. *Biochim. Biophys. Acta.* doi:10.1016/j.bbcan.2016.11.005 (2016).
  - Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976).
  - Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).
  - Steensma, D. P. *et al.* Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* **126**, 9–16 (2015).
- This report summarizes data from very recent studies documenting an increasing age-associated appearance of oncogenic mutations in amplified clones of normally differentiating cells without any evidence of immediate or necessary progression to malignancy. These findings established the concept of clonal evolution preceding acquisition of a fully malignant state in the blood-forming system where abnormal cell distributions and arrangements in situ are difficult to discern.*
- Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).

15. Blokzijl, F. *et al.* Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**, 260–264 (2016).
16. Negri, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability [mdash] an evolving hallmark of cancer. *Nat. Rev. Mol. Cell. Biol.* **11**, 220–228 (2010).
17. Choi, J. D. & Lee, J. S. Interplay between Epigenetics and Genetics in cancer. *Genomics Inform.* **11**, 164–173 (2013).
18. Avgustinova, A. & Benitah, S. A. The epigenetics of tumour initiation: cancer stem cells and their chromatin. *Curr. Opin. Genet. Dev.* **36**, 8–15 (2016).
19. Lee, J. K., Choi, Y. L., Kwon, M. & Park, P. J. Mechanisms and consequences of cancer genome instability: lessons from genome sequencing studies. *Annu. Rev. Pathol.* **11**, 283–312 (2016).
20. Notta, F. *et al.* A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* **538**, 378–382 (2016)  
**This study was the first to challenge the generally accepted model of how a solid tumour develops. Through a detailed genomic analysis of patients' samples the authors obtained evidence of a sudden burst of important mutational events rather than a slow stepwise accumulation of driving mutations.**
21. Aparicio, S. & Caldas, C. The implications of clonal genome evolution for cancer medicine. *N. Engl. J. Med.* **368**, 842–851 (2013).
22. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
23. Shlush, L. I. *et al.* Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* **506**, 328–333 (2014).
24. Qian, M., Wang, D. C., Chen, H. & Cheng, Y. Detection of single cell heterogeneity in cancer. *Semin. Cell Dev. Biol.* **64**, 143–149 (2016).
25. Sandberg, R. Entering the era of single-cell transcriptomics in biology and medicine. *Nat. Methods* **11**, 22–24 (2014).
26. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
27. Blanpain, C. Tracing the cellular origin of cancer. *Nat. Cell Biol.* **15**, 126–134 (2013).
28. Kucherlapati, R. Genetically modified mouse models for biomarker discovery and preclinical drug testing. *Clin. Cancer Res.* **18**, 625–630 (2012).
29. Becher, O. J. & Holland, E. C. Genetically engineered models have advantages over xenografts for preclinical studies. *Cancer Res.* **66**, 3355–3358, discussion 3358–3359 (2006).
30. Ashworth, A., Lord, C. J. & Reis-Filho, J. S. Genetic interactions in cancer progression and treatment. *Cell* **145**, 30–38 (2011).
31. Shi, Y. *et al.* Chromatin accessibility contributes to simultaneous mutations of cancer genes. *Sci. Rep.* **6**, 35270 (2016).
32. Narayan, S., Bader, G. D. & Reimand, J. Frequent mutations in acetylation and ubiquitination sites suggest novel driver mechanisms of cancer. *Genome Med.* **8**, 55 (2016).
33. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
34. Petzer, A. *et al.* Characterization of primitive subpopulations of normal and leukemic cells present in the blood of patients with newly diagnosed as well as established chronic myeloid leukemia. *Blood* **88**, 2162–2171 (1996).
35. Eisterer, W. *et al.* Different subsets of primary chronic myeloid leukemia stem cells engraft immunodeficient mice and produce a model of the human disease. *Leukemia* **19**, 435–441 (2005).
36. Chu, S. *et al.* Persistence of leukemia stem cells in chronic myelogenous leukemia patients in prolonged remission with imatinib treatment. *Blood* **118**, 5565–5572 (2011).
37. Pellicano, F., Mukherjee, L. & Holyoake, T. L. Concise review: cancer cells escape from oncogene addiction: understanding the mechanisms behind treatment failure for more effective targeting. *Stem Cells* **32**, 1373–1379 (2014).
38. Fearon, E. R., Burke, P. J., Schiffer, C. A., Zehnbauser, B. A. & Vogelstein, B. Differentiation of leukemia cells to polymorphonuclear leukocytes in patients with acute nonlymphocytic leukemia. *N. Engl. J. Med.* **315**, 15–24 (1986).
39. Eppert, K. *et al.* Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat. Med.* **17**, 1086–1093 (2011).
40. Jan, M. *et al.* Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci. Transl. Med.* **4**, 149ra118 (2012).
41. Corces-Zimmerman, M. R., Hong, W. J., Weissman, I. L., Medeiros, B. C. & Majeti, R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proc. Natl Acad. Sci. USA* **111**, 2548–2553 (2014).
42. Smith, F. O., Raskind, W. H., Fialkow, P. J. & Bernstein, I. D. Cellular biology of acute myelogenous leukemia. *J. Pediatr. Hematol. Oncol.* **17**, 113–122 (1995).
43. Prchal, J. T. *et al.* A common progenitor for human myeloid and lymphoid cells. *Nature* **274**, 590–591 (1978).
44. Raskind, W. H., Tirumali, N., Jacobson, R., Singer, J. & Fialkow, P. J. Evidence for a multistep pathogenesis of a myelodysplastic syndrome. *Blood* **63**, 1318–1323 (1984).
45. Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat. Med.* **3**, 730–737 (1997)  
**This study was the first to demonstrate that human cells able to regenerate leukemic populations in vivo (in a transplanted, sublethally irradiated immunodeficient mouse) are restricted to the CD34+ subset and that includes cells with serially transplantable leukemia propagating activity indicative of leukemic stem cell activity.**
46. Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl Acad. Sci. USA* **100**, 3983–3988 (2003).
47. Singh, S. K. *et al.* Identification of human brain tumour initiating cells. *Nature* **432**, 396–401 (2004).
48. O'Brien, C. A., Pollett, A., Gallinger, S. & Dick, J. E. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* **445**, 106–110 (2007).
49. Holyoake, T., Jiang, X., Eaves, C. & Eaves, A. Isolation of a highly quiescent subpopulation of primitive leukemic cells in chronic myeloid leukemia. *Blood* **94**, 2056–2064 (1999)  
**This study was the first to document the presence of a quiescent population of malignant cells setting the stage for understanding why anti-proliferative agents are not effective longterm in eradicating many human cancers.**
50. Gao, M. Q., Choi, Y. P., Kang, S., Youn, J. H. & Cho, N. H. CD24+ cells from hierarchically organized ovarian cancer are enriched in cancer stem cells. *Oncogene* **29**, 2672–2680 (2010).
51. Pece, S. *et al.* Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. *Cell* **140**, 62–73 (2010).
52. Roesch, A. *et al.* A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. *Cell* **141**, 583–594 (2010).
53. Dembinski, J. L. & Krauss, S. Characterization and functional analysis of a slow cycling stem cell-like subpopulation in pancreas adenocarcinoma. *Clin. Exp. Metastasis* **26**, 611 (2009).
54. Quintana, E. *et al.* Efficient tumour formation by single human melanoma cells. *Nature* **456**, 593–598 (2008)  
**This study was the first to show that in many human melanomas almost every cell has the innate ability to regenerate a tumour, thus providing an important example of a human tumour that does not fit a stem cell model.**
55. Liu, S. *et al.* Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Rep.* **2**, 78–91 (2014).
56. Friedl, P. & Alexander, S. Cancer invasion and the microenvironment: plasticity and reciprocity. *Cell* **147**, 992–1009 (2011).
57. Chaffer, C. L., San Juan, B. P., Lim, E. & Weinberg, R. A. EMT, cell plasticity and metastasis. *Cancer Metastasis Rev.* **35**, 645–654 (2016).
58. Lehuède, C., Dupuy, F., Rabinovitch, R., Jones, R. G. & Siegel, P. M. Metabolic plasticity as a determinant of tumor growth and metastasis. *Cancer Res.* **76**, 5201–5208 (2016).
59. Beer, P. A. & Eaves, C. J. Modeling normal and disordered human hematopoiesis. *Trends in Cancer* **1**, 199–210 (2016).
60. Liu, X. *et al.* Low CD38 identifies progenitor-like inflammation-associated luminal cells that can initiate human prostate cancer and predict poor outcome. *Cell Rep.* **17**, 2596–2606 (2016).
61. Yoo, Y. A. *et al.* Bmi1 marks distinct castration-resistant luminal progenitor cells competent for prostate regeneration and tumour initiation. *Nat. Commun.* **7**, 12943 (2016).
62. Shin, S. *et al.* Genetic lineage tracing analysis of the cell of origin of hepatotoxin-induced liver tumors in mice. *Hepatology (Baltimore)* **64**, 1163–1177 (2016).
63. Leeman, K. T., Fillmore, C. M. & Kim, C. F. Lung stem and progenitor cells in tissue homeostasis and disease. *Curr. Top. Dev. Biol.* **107**, 207–233 (2014).
64. Van Keymeulen, A. *et al.* Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nature* **525**, 119–123 (2015).
65. Koren, S. *et al.* PIK3CA(H1047R) induces multipotency and multi-lineage mammary tumours. *Nature* **525**, 114–118 (2015).
66. Makarewicz, M. *et al.* Developmental changes in the in vitro activated regenerative activity of primitive mammary epithelial cells. *PLoS Biol.* **11**, e1001630 (2013).
67. Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem cells. *Nature* **439**, 993–997 (2006).
68. Shackleton, M. *et al.* Generation of a functional mammary gland from a single stem cell. *Nature* **439**, 84–88 (2006).
69. Meyer, D. S. *et al.* Luminal expression of PIK3CA mutant H1047R in the mammary gland induces heterogeneous tumors. *Cancer Res.* **71**, 4344–4351 (2011).
70. Melchor, L. *et al.* Identification of cellular and genetic drivers of breast cancer heterogeneity in genetically engineered mouse tumour models. *J. Pathol.* **233**, 124–137 (2014).



71. Powell, A. E. *et al.* Inducible loss of one Apc allele in Lrig1-expressing progenitor cells results in multiple distal colonic tumors with features of familial adenomatous polyposis. *Am. J. Physiol. Gastrointest. Liver Physiol.* **307**, G16–G23 (2014).
72. Boumahdi, S. *et al.* SOX2 controls tumour initiation and cancer stem-cell functions in squamous-cell carcinoma. *Nature* **511**, 246–250 (2014).
73. Vanner, R. J. *et al.* Quiescent sox2(+) cells drive hierarchical growth and relapse in sonic hedgehog subgroup medulloblastoma. *Cancer Cell* **26**, 33–47 (2014).
74. Larsimont, J. C. *et al.* Sox9 Controls Self-Renewal of Oncogene Targeted Cells and Links Tumor Initiation and Invasion. *Cell Stem Cell* **17**, 60–73 (2015).
75. Schwitalla, S. *et al.* Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties. *Cell* **152**, 25–38 (2013).
76. Moller, Y. *et al.* Oncogenic Ras triggers hyperproliferation and impairs polarized colonic morphogenesis by autocrine ErbB3 signaling. *Oncotarget* **7**, 53526–53539 (2016).
77. Malaterre, J. *et al.* Intestinal-specific activatable Myb initiates colon tumorigenesis in mice. *Oncogene* **35**, 2475–2484 (2016).
78. Bellafante, E. *et al.* PGC-1 $\beta$  promotes enterocyte lifespan and tumorigenesis in the intestine. *Proc. Natl Acad. Sci. USA* **111**, E4523–E4531 (2014).
79. Pineault, N. *et al.* Induction of acute myeloid leukemia in mice by the human leukemia-specific fusion gene NUP98-HOXD13 in concert with Meis1. *Blood* **101**, 4529–4538 (2003).
80. Asai, T. *et al.* Generation of a novel, multi-stage, progressive, and transplantable model of plasma cell neoplasms. *Sci. Rep.* **6**, 22760 (2016).
81. Scourzac, L. *et al.* DNMT3A(R882H) mutant and Tet2 inactivation cooperate in the deregulation of DNA methylation control to induce lymphoid malignancies in mice. *Leukemia* **30**, 1388–1398 (2016).
82. Rappaport, A. & Johnson, L. Genetically engineered knock-in and conditional knock-in mouse models of cancer. *Cold Spring Harb. Protoc.* **2014**, 897–911 (2014).
83. Yi, L. & Li, J. CRISPR-Cas9 therapeutics in cancer: promising strategies and present challenges. *Biochim. Biophys. Acta* **1866**, 197–207 (2016).
84. Kannan, R. & Ventura, A. The CRISPR revolution and its impact on cancer research. *Swiss Med. Wkly* **145**, w14230 (2015).
85. Annunziato, S. *et al.* Modeling invasive lobular breast carcinoma by CRISPR/Cas9-mediated somatic genome editing of the mammary gland. *Genes Dev.* **30**, 1470–1480 (2016).
86. Wasylishen, A. R. *et al.* New model systems provide insights into Myc-induced transformation. *Oncogene* **30**, 3727–3734 (2011).
87. Bessette, D. C. *et al.* Using the MCF10A/MCF10CA1a breast cancer progression cell line model to investigate the effect of active, mutant forms of EGFR in breast cancer development and treatment using gefitinib. *PLoS ONE* **10**, e0125232 (2015).
88. Lombardo, Y. *et al.* Nicastrin regulates breast cancer stem cell properties and tumor growth in vitro and in vivo. *Proc. Natl Acad. Sci. USA* **109**, 16558–16563 (2012).
89. Ward, T. M. *et al.* Truncated p110 ERBB2 induces mammary epithelial cell migration, invasion and orthotopic xenograft formation, and is associated with loss of phosphorylated STAT5. *Oncogene* **32**, 2463–2474 (2013).
90. Palafox, M. *et al.* RANK induces epithelial-mesenchymal transition and stemness in human mammary epithelial cells and promotes tumorigenesis and metastasis. *Cancer Res.* **72**, 2879–2888 (2012).
91. Mullins, S. R. *et al.* Three-dimensional cultures modeling premalignant progression of human breast epithelial cells: role of cysteine cathepsins. *Biol. Chem.* **393**, 1405–1416 (2012).
92. Qu, Y. *et al.* Generation of prostate tumor-initiating cells is associated with elevation of reactive oxygen species and IL-6/STAT3 signaling. *Cancer Res.* **73**, 7090–7100 (2013).
93. Pellacani, D. *et al.* Analysis of normal human mammary epigenomes reveals cell-specific active enhancer states and associated transcription factor networks. *Cell Rep.* **17**, 2060–2074 (2016).
94. Finak, G. *et al.* Stromal gene expression predicts clinical outcome in breast cancer. *Nat. Med.* **14**, 518–527 (2008).
95. Wu, J., Liang, C., Chen, M. & Su, W. Association between tumor-stroma ratio and prognosis in solid tumor patients: a systematic review and meta-analysis. *Oncotarget* **7**, 68954–68965 (2016).
96. Lee, D. *et al.* Intratumor stromal proportion predicts aggressive phenotype of gastric signet ring cell carcinomas. *Gastric Cancer*, 1–11. doi:10.1007/s10120-016-0669-2 (2016).
97. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* **19**, 1423–1437 (2013).
98. McAllister, S. S. & Weinberg, R. A. The tumour-induced systemic environment as a critical regulator of cancer progression and metastasis. *Nat. Cell Biol.* **16**, 717–727 (2014).
99. Valent, P. *et al.* Cancer stem cell definitions and terminology: the devil is in the details. *Nat. Rev. Cancer* **12**, 767–775 (2012).
100. Wuidart, A. *et al.* Quantitative lineage tracing strategies to resolve multipotency in tissue-specific stem cells. *Genes Dev.* **30**, 1261–1277 (2016).
101. Beumer, J. & Clevers, H. Regulation and plasticity of intestinal stem cells during homeostasis and regeneration. *Development (Cambridge)* **143**, 3639–3649 (2016).
102. Chen, F. & Fine, A. Stem Cells in Lung Injury and Repair. *Am. J. Pathol.* **186**, 2544–2550 (2016).
103. Goldstein, A. S., Huang, J., Guo, C., Garraway, I. P. & Witte, O. N. Identification of a cell of origin for human prostate cancer. *Science* **329**, 568–571 (2010).  
**This study was the first to produce human prostate cancers de novo and show that the histological characterization of cancers does not necessarily correlate with the cellular origin of the malignant clone.**
104. Matano, M. *et al.* Modeling colorectal cancer using CRISPR-Cas9-mediated engineering of human intestinal organoids. *Nat. Med.* **21**, 256–262 (2015).  
**This study was the first study to combine the use of CRISPR-Cas9 and organoid technologies to engineer the generation of colorectal cancer from normal human intestinal epithelium.**
105. Proia, T. A. *et al.* Genetic predisposition directs breast cancer phenotype by dictating progenitor cell fate. *Cell Stem Cell* **8**, 149–163 (2011).
106. Keller, P. J. *et al.* Defining the cellular precursors to human breast cancer. *Proc. Natl Acad. Sci. USA* **109**, 2772–2777 (2012).
107. Nguyen, L. V. *et al.* Barcoding reveals complex clonal dynamics of de novo transformed human mammary cells. *Nature* **528**, 267–271 (2015).  
**This study presented the first evidence that a single oncogene (KRAS-G12D) when introduced into either normal human mammary basal or luminal cells could force them to produce invasive tumours in transplanted immunodeficient mice. This study also revealed that serially transplantable clones were not predominant in the primary tumours revealing previously unknown cellular dynamics in the process of carcinogenesis.**
108. Wei, J. *et al.* Microenvironment determines lineage fate in a human model of MLL-AF9 leukemia. *Cancer Cell* **13**, 483–495 (2008).
109. Barabe, F., Kennedy, J. A., Hope, K. J. & Dick, J. E. Modeling the initiation and progression of human acute leukemia in mice. *Science* **316**, 600–604 (2007).
110. Lin, S. *et al.* Instructive role of MLL-fusion proteins revealed by a model of t(4;11) pro-B acute lymphoblastic leukemia. *Cancer Cell* **30**, 737–749 (2016).
111. Imren, S. *et al.* Modeling de novo leukemogenesis from human cord blood with MN1 and NUP98HOXD13. *Blood* **124**, 3608–3612 (2014).
112. Theocharides, A. P. *et al.* Dominant-negative Ikaros cooperates with BCR-ABL1 to induce human acute myeloid leukemia in xenografts. *Leukemia* **29**, 177–187 (2015).
113. Leskov, I. *et al.* Rapid generation of human B-cell lymphomas via combined expression of Myc and Bcl2 and their use as a preclinical model for biological therapies. *Oncogene* **32**, 1066–1072 (2013).
114. Eirew, P. *et al.* Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* **518**, 422–426 (2015).
115. Bruna, A. *et al.* A Biobank of breast cancer explants with preserved intratumor heterogeneity to screen anticancer compounds. *Cell* **167**, 260–274.e222 (2016).
116. Navin, N. *et al.* Inferring tumor progression from genomic heterogeneity. *Genome Res.* **20**, 68–80 (2010).
117. Bashashati, A. *et al.* Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *J. Pathol.* **231**, 21–34 (2013).
118. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
119. Jiang, X., Saw, K. M., Eaves, A. & Eaves, C. Instability of BCR-ABL gene in primary and cultured chronic myeloid leukemia stem cells. *J. Natl Cancer. Inst.* **99**, 680–693 (2007).
120. Bao, S. *et al.* Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* **444**, 756–760 (2006).
121. Ricci-Vitiani, L. *et al.* Identification and expansion of human colon-cancer-initiating cells. *Nature* **445**, 111–115 (2007).
122. Collins, A. T., Berry, P. A., Hyde, C., Stower, M. J. & Maitland, N. J. Prospective identification of tumorigenic prostate cancer stem cells. *Cancer Res.* **65**, 10946–10951 (2005).
123. Suzuki, S. *et al.* Identification and characterization of cancer stem cells in ovarian yolk sac tumors. *Cancer Sci.* **101**, 2179–2185 (2010).
124. Schatton, T. *et al.* Identification of cells initiating human melanomas. *Nature* **451**, 345–U311 (2008).
125. Lapidot, T. *et al.* A cell initiating human acute myeloid-leukemia after transplantation into Scid mice. *Nature* **367**, 645–648 (1994).
126. Sarry, J. E. *et al.* Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2R gamma c-deficient mice. *J. Clin. Invest.* **121**, 384–395 (2011).

127. Notta, F. *et al.* Evolution of human BCR-ABL1 lymphoblastic leukaemia-initiating cells. *Nature* **469**, 362–367 (2011).
128. Ebinger, S. *et al.* Characterization of rare, dormant, and therapy-resistant cells in acute lymphoblastic leukemia. *Cancer Cell* **30**, 849–862 (2016).
129. Bystrykh, L. V., Verovskaya, E., Zwart, E., Broekhuis, M. & de Haan, G. Counting stem cells: methodological constraints. *Nat. Methods* **9**, 567–574 (2012).
130. Bystrykh, L. V. & Belderbos, M. E. Clonal analysis of cells with cellular barcoding: when numbers and sizes matter. *Methods Mol. Biol.* **1516**, 57–89 (2016).
131. Nguyen, L. V. *et al.* DNA barcoding reveals diverse growth kinetics of human breast tumour subclones in serially passaged xenografts. *Nat. Commun.* **5**, 5871 (2014).
132. Bhang, H. E. *et al.* Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat. Med.* **21**, 440–448 (2015).
133. Klauke, K. *et al.* Tracing dynamics and clonal heterogeneity of Cbx7-induced leukemic stem cells by cellular barcoding. *Stem Cell Rep.* **4**, 74–89 (2015).
134. Wu, C. *et al.* Clonal tracking of rhesus macaque hematopoiesis highlights a distinct lineage origin for natural killer cells. *Cell Stem Cell* **14**, 486–499 (2014).
135. Cheung, A. M. *et al.* Analysis of the clonal growth and differentiation dynamics of primitive barcoded human cord blood cells in NSG mice. *Blood* **122**, 3129–3137 (2013).
136. Verovskaya, E. *et al.* Heterogeneity of young and aged murine hematopoietic stem cells revealed by quantitative clonal analysis using cellular barcoding. *Blood* **122**, 523–532 (2013).
137. Naik, S. H. *et al.* Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature* **496**, 229–232 (2013).
138. Meyer, M. *et al.* Single cell-derived clonal analysis of human glioblastoma links functional and genomic heterogeneity. *Proc. Natl Acad. Sci. USA* **112**, 851–856 (2015).
139. Clevers, H. Modeling development and disease with organoids. *Cell* **165**, 1586–1597 (2016).
140. Weeber, F. *et al.* Preserved genetic diversity in organoids cultured from biopsies of human colorectal cancer metastases. *Proc. Natl Acad. Sci. USA* **112**, 13308–13311 (2015).
141. Baker, L. A., Tiriac, H., Clevers, H. & Tuveson, D. A. Modeling pancreatic cancer with organoids. *Trends Cancer* **2**, 176–190 (2016).
142. Lancaster, M. A. & Knoblich, J. A. Organogenesis in a dish: modeling development and disease using organoid technologies. *Science* **345**, 1247125 (2014).
143. Papapetrou, E. P. Patient-derived induced pluripotent stem cells in cancer research and precision oncology. *Nat. Med.* **22**, 1392–1401 (2016).
144. Lim, K. L. *et al.* Reprogramming cancer cells: overview & current progress. *Expert Opin. Biol. Ther.* **16**, 941–951 (2016).
145. Kumano, K. *et al.* Generation of induced pluripotent stem cells from primary chronic myelogenous leukemia patient samples. *Blood* **119**, 6234–6242 (2012).
146. Cayette, J. E. *et al.* Generation of iPSCs from cultured human malignant cells. *Blood* **115**, 4039–4042 (2010).
147. Hu, K. & Slukvin, I. in *Pluripotent Stem Cells: Methods and Protocols* (eds Lakshminarayanan, U. & Vemuri, M. C.) 163–176 (Humana Press, 2013).
148. Kotini, A., Dolezal, E. K., Nimer, S. & Papapetrou, E. P. An iPSC-based model Of MDS for phenotype-driven gene and drug discovery. *Blood* **122**, 859–859 (2013).
149. Ye, Z. *et al.* Human-induced pluripotent stem cells from blood cells of healthy donors and patients with acquired blood disorders. *Blood* **114**, 5473–5480 (2009).
150. Gandre-Babbe, S. *et al.* Patient-derived induced pluripotent stem cells recapitulate hematopoietic abnormalities of juvenile myelomonocytic leukemia. *Blood* **121**, 4925–4929 (2013).
151. Muñoz-López, A. *et al.* Development refractoriness of MLL-rearranged human B cell acute leukemias to reprogramming into pluripotency. *Stem Cell Rep.* **7**, 602–618 (2016).
152. Kim, J. *et al.* An iPSC line from human pancreatic ductal adenocarcinoma undergoes early to invasive stages of pancreatic cancer progression. *Cell Rep.* **3**, 2088–2099 (2013).
153. Hadoux, J. *et al.* Generation of an induced pluripotent stem cell line from a patient with hereditary multiple endocrine neoplasia 2A (MEN2A) syndrome with RET mutation. *Stem Cell Res.* **17**, 154–157 (2016).
154. Iskender, B., Izgi, K. & Canatan, H. Reprogramming bladder cancer cells for studying cancer initiation and progression. *Tumour Biol.* **37**, 13237–13245 (2016).
155. Lee, D. F. *et al.* Modeling familial cancer with induced pluripotent stem cells. *Cell* **161**, 240–254 (2015).
156. Ron-Bigger, S. *et al.* Aberrant epigenetic silencing of tumor suppressor genes is reversed by direct reprogramming. *Stem Cells* **28**, 1349–1354 (2010).
157. Ohnishi, K. *et al.* Premature termination of reprogramming *in vivo* leads to cancer development through altered epigenetic regulation. *Cell* **156**, 663–677 (2014).
158. Choi, P. S. & Meyerson, M. Targeted genomic rearrangements using CRISPR/Cas technology. *Nat. Commun.* **5**, 3728 (2014).
159. Chen, C. *et al.* MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia. *Cancer Cell* **25**, 652–665 (2014).
160. Torres, R. *et al.* Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. *Nat. Commun.* **5**, 3964 (2014).
161. Heckl, D. *et al.* Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nat. Biotechnol.* **32**, 941–946 (2014).
162. Xue, W. *et al.* CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature* **514**, 380–384 (2014).
163. Guernet, A. *et al.* CRISPR-Barcoding for intratumor genetic heterogeneity modeling and functional analysis of oncogenic driver mutations. *Mol. Cell* **63**, 526–538 (2016).
164. Frieda, K. L. *et al.* Synthetic recording and *in situ* readout of lineage information in single cells. *Nature* **541**, 107–111 (2017).
165. Qin, H., Malek, S., Cowell, J. K. & Ren, M. Transformation of human CD34+ hematopoietic progenitor cells with DEK-NUP214 induces AML in an immunocompromised mouse model. *Oncogene* **35**, 5686–5691 (2016).
166. Ren, M. *et al.* Development of ZMYM2-FGFR1 driven AML in human CD34+ cells in immunocompromised mice. *Int. J. Cancer* **139**, 836–840 (2016).

### Acknowledgements

This work is supported by grants from the Canadian Cancer Society Research Institute and a University of British Columbia Graduate Studentship to S.B. We also thank all members of the Eaves lab for suggestions and comments.

### Author contributions

S.B. and C.J.E. developed the concept and structure of the review. S.B., L.V.N. and C.J.E. wrote and edited the manuscript.

### Additional information

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Balani, S. *et al.* Modeling the process of human tumorigenesis. *Nat. Commun.* **8**, 15422 doi: 10.1038/ncomms15422 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017