

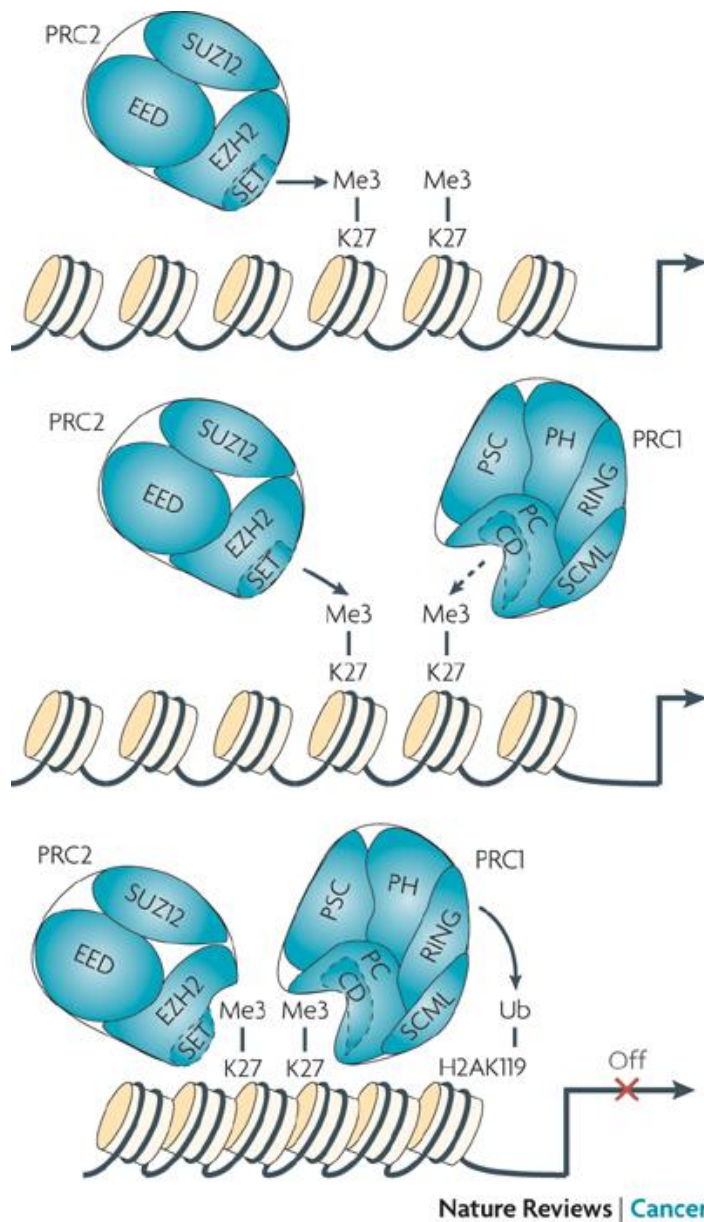
Polycomb group proteins: navigators of lineage pathways led astray in cancer

Adrian P. Bracken & Kristian Helin

Abstract

The Polycomb group (PcG) proteins are transcriptional repressors that regulate lineage choices during development and differentiation. Recent studies have advanced our understanding of how the PcG proteins regulate cell fate decisions and how their deregulation potentially contributes to cancer. In this Review we discuss the emerging roles of long non-coding RNAs (ncRNAs) and a subset of transcription factors, which we call cell fate transcription factors, in the regulation of PcG association with target genes. We also speculate about how their deregulation contributes to tumorigenesis.

Considerable attention is currently focused on identifying the events that lead to the development of so-called 'tumour-initiating cells', as understanding this might facilitate the design of more effective cancer therapies^{1, 2, 3}. It is becoming increasingly evident that, in addition to genetic alterations, tumour development involves the alteration of gene expression patterns owing to epigenetic changes⁴. Recent studies have implicated the Polycomb group proteins (PcG proteins) as key contributors to these changes^{5, 6, 7, 8, 9}. The PcG proteins form multiprotein repressive complexes, called Polycomb repressive complexes (PRCs), which repress transcription by a mechanism that probably involves the modification of chromatin (Fig. 1).



Nature Reviews | Cancer

Two major Polycomb repressive complexes (PRCs) have been described. The PRC2 complex contains the histone methyltransferase enhancer of zeste homologue 2 (EZH2), which together with embryonic ectoderm development (EED) and suppressor of zeste 12 homolog (SUZ12) catalyses the trimethylation of histone H3 at lysine K27 (H3K27me3). The EZH2 SET domain confers this activity. Multiple forms of the PRC1 complex exist and these contain combinations of at least four PC proteins (CBX2, CBX4, CBX7 and CBX8), six PSC proteins (BMI1, MEL18, MBLR, NSPC1, RNF159 and RNF3), two RING proteins (RNF1 and RNF2), three PH proteins (HPH1, HPH2 and HPH3) and two SCML proteins (SCML1 1 and SCML2). Some results have suggested that PRC1 complexes are recruited by the affinity of chromodomains in chromobox (Cbx) proteins to the H3K27me3 mark. PRC1 recruitment results in the RNF1 and RNF2-mediated ubiquitylation of histone H2A on lysine 119, which is thought to be important for transcriptional repression. PC, Polycomb; PSC, Posterior sex combs ; SCML, Sex combs on midleg .

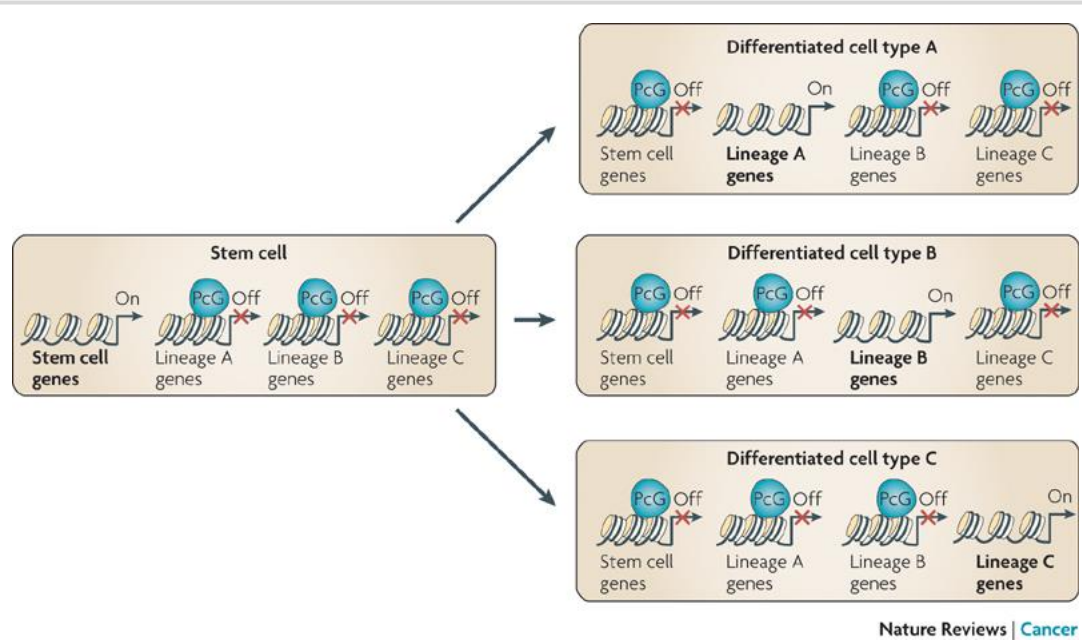
Figure 1 | Coordinated action of Polycomb repressive complexes.

Several genetic studies in different organisms have firmly established the vital and conserved roles for PcG proteins in embryonic development and adult somatic cell differentiation¹⁰. Moreover, recent studies have demonstrated that the PcG proteins are required for maintaining the correct identities of stem, progenitor and differentiated cells¹¹. The genome-wide mapping of PcG target

genes in mammalian cells has offered scientists the opportunity to start to unravel the molecular mechanisms of PcG protein action^{12, 13, 14}. The PcG proteins have been found to bind and repress the promoters of genes that encode proteins with key roles in cell fate determination in many different cellular lineages. Although these data support the large body of evidence that points to crucial roles for the PcG proteins in both development and adult homeostasis, we are only beginning to understand how the PcG proteins actually regulate their target genes.

Initial studies have established that the PcG proteins are displaced from certain target genes, for example the homeobox (Hox) genes, on their transcriptional activation during differentiation^{13, 14, 15}. However, subsequent studies demonstrated that the binding of PcG proteins is much more dynamic than anticipated, showing that the PcG proteins are also recruited to the promoters of certain genes in response to differentiation signals and, importantly, that this recruitment is required for their silencing during differentiation^{15, 16, 17, 18, 19}. On the basis of these results, we and others have proposed a model in which the PcG proteins function dynamically during development and differentiation to lock off the expression of alternative cell fate regulators in any particular lineage (Fig. 2). In this Review we propose that the deregulation of these mechanisms is central to tumour initiation.

Figure 2 | Dynamic recruitment and displacement of Polycomb group proteins during lineage specification.



The Polycomb group (PcG) proteins are displaced from the promoters of one set of target genes, while being recruited to the promoters of another set of target genes during lineage specification. This model depicts a progenitor or stem cell that has the potential to differentiate into three different lineages: A, B or C. As an undifferentiated cell it expresses 'stem cell genes' that are required to maintain its proliferative undifferentiated state. Before the signal to differentiate, the PcG proteins repress the transcription of lineage-specific differentiation genes. On stimulation to differentiate, the PcG proteins are recruited to the stem cell-specific genes, independently of the type of lineage specification signal. The PcG proteins are then displaced from lineage-specific gene promoters during differentiation depending on the lineage specification signal. The mechanisms by which the PcG proteins are displaced and recruited to target genes are not well understood. In differentiated cells, PcG proteins silence not only the expression of stem cell genes but also the expression of genes that encode regulators of alternative lineages. This mechanism of 'locking' cell fate is thought to be central to how cells maintain their identity through subsequent cell divisions. Importantly, these mechanisms are more dynamic and plastic than previously anticipated, and they are reversed during cellular reprogramming and are deregulated in cancer.

PcG proteins and cancer

The PcG proteins are essential for the maintenance of both normal and cancer stem cell populations^{20, 21, 22}. This is partly attributed to their ability to bind to and repress the CDKN2B and CDKN2A loci, which encode the tumour suppressors INK4B (encoded by CDKN2B), INK4A and ARF (both encoded by CDKN2A)^{21, 23, 24, 25, 26, 27, 28, 29, 30, 31}. INK4A and INK4B function upstream in the RB pathway, and ARF functions upstream in the p53 pathway^{25, 32}. In addition to frequent genetic alterations, this locus is often epigenetically silenced by DNA methylation in cancer, and the PcG proteins have been proposed to contribute to this²⁶. Many additional PcG target genes accumulate DNA methylation on their promoters in cancer, such as Wilm's tumour 1 (WT1), retinoic acid receptor-beta (RARβ), kruppel-like factor 4 (KLF4), inhibitor of DNA binding 4 (ID4), GATA binding protein 3 (GATA3) chromodomain helicase DNA

binding protein 5 (CHD5) and PU.1 (also known as SPI1)¹³. The reports that enhancer of zeste homologue 2 (EZH2)³³ and chromobox homologue 7 (CBX7)³⁴ can physically associate with DNA methyltransferases (DNMTs) suggest a mechanism whereby the PcG proteins directly contribute to the altered DNA methylation profiles that are observed in multiple cancer types. In fact, PcG target genes are as much as 12 times as likely to be aberrantly silenced by DNA methylation in cancer as non-PcG target genes^{7, 8, 9} and poorly differentiated and aggressive human tumours show preferential repression of PcG target genes⁶. Taken together, these results suggest a possible scenario in which PcG proteins and DNA methylating enzymes (such as DNMTs) cooperate to aberrantly silence pro-differentiation and anti-proliferative genes, which leads to the accumulation of a population of cells unable to respond to differentiation signals. It is thought that the consequent block of differentiation may allow these tumour-initiating cells to linger and accumulate the additional epigenetic and/or genetic alterations necessary to develop into a tumour.

However, a key question remains unanswered: what triggers the aberrant silencing of PcG target genes that is observed in many cancer types? One potential scenario is that PcG proteins, such as EZH2 and BMI1, become aberrantly upregulated, leading to the progressive recruitment of DNMTs to PcG target genes, a switch to a more permanent transcriptional silencing and the generation of tumour-initiating cells. Supporting evidence for this hypothesis includes the fact that several PcG proteins are highly expressed in cancer¹⁰. For example, BMI1 is amplified and overexpressed in B cell lymphoma and functions as an oncogene that cooperates with Myc in a mouse model of lymphoma^{35, 36, 37, 38}. Similarly, suppressor of zeste 12 homologue (SUZ12) is translocated in endometrial cancer³⁹, and EZH2 is amplified and highly expressed in many tumour types^{40, 41, 42, 43, 44, 45, 46, 47}. Potentially contributing to these increased EZH2 levels, the microRNA miR-101 has recently been reported to directly target EZH2 and is itself deleted in some cancers^{48, 49}. However, despite the functional evidence for a role of PcG proteins, particularly BMI1, in the development of cancer, the higher levels of these proteins frequently observed in tumours could partly be a consequence of the high proportion of proliferating and/or 'stem-like' cells in tumours. For example, BMI1 has been reported to be highly expressed in normal stem cells⁵⁰ and EZH2 expression correlates with proliferation rate as it is controlled by the RB-E2F pathway⁴¹. Therefore, in this Review we discuss an alternative and complementary hypothesis in which PcG proteins are led astray in cancer by the deregulation of factors that are required for their association to target genes. We propose that the deregulation of these factors directly contributes to the aberrant modulation of transcriptional programmes observed in many cancers.

PcG recruitment to target genes

PcG proteins do not have the ability to bind specific DNA motifs. Therefore, a key mechanistic question concerns how they are recruited to and displaced from their target genes during lineage specification. The answer to this question not

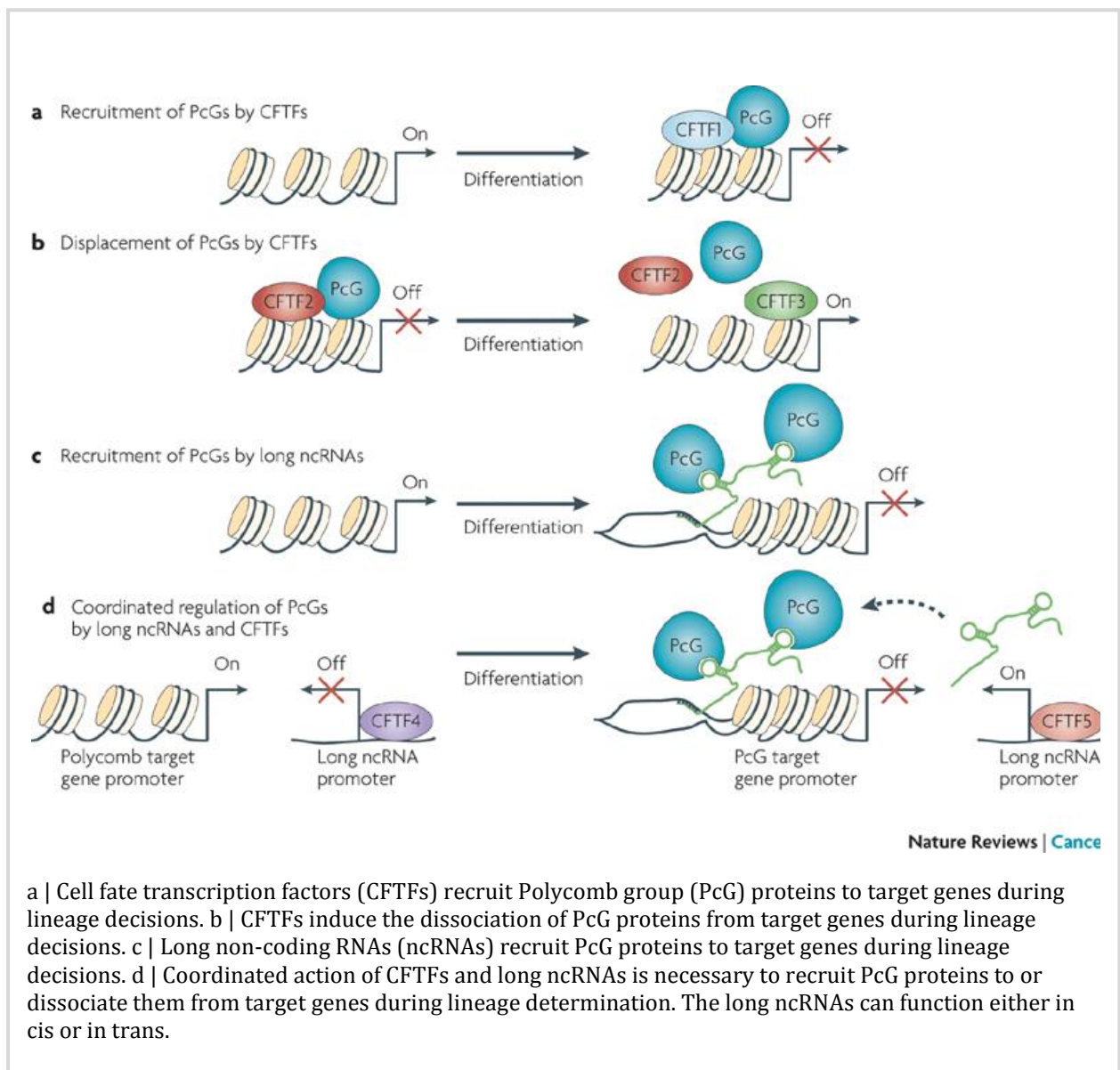
only has implications for our fundamental understanding of lineage choice during development and differentiation, but may also considerably contribute to our understanding of the initiating events in cancer.

Transcription factors recruit PcG proteins. In *Drosophila melanogaster*, several transcription factors are required to recruit PcG proteins to polycomb repressive elements (PREs) during development⁵¹. One such transcription factor, encoded by Yy1 (also known as Pho), has recently been shown to co-occupy most PREs with PRC1 and PRC2 components in *D. melanogaster* embryos and larvae^{19, 52}. The PRE in *D. melanogaster* is not an easily recognizable DNA sequence motif as it is not a single transcription factor binding site. Instead, it is a collection of binding sites, defined as an 'element' of several hundred base pairs. To date, PREs have not been defined in mammalian cells, despite the mapping of several thousand binding sites for the PcG proteins^{12, 13, 14}. This suggests that many different mammalian transcription factors contribute to the recruitment of the PcG proteins. In fact, if one looks at the target genes regulated by the PcG proteins in mammalian cells and considers how they are expressed in different cell types, it becomes difficult to imagine that only a few transcription factors are involved in PcG recruitment and displacement. It is likely that the requirement of multiple transcription factors confers a much greater flexibility of target gene regulation. On this basis, it will be essential to define these transcription factors, because their deregulation could be key to inducing cancer — that is, they could work as oncogenes or tumour suppressors.

So which transcription factors control the association of PcG proteins with their target genes? It has been estimated that the human genome encodes approximately 2,600 transcription factors^{53, 54}. We propose that 'cell fate' transcription factors (CFTFs) are strong candidates for the regulation of PcG protein recruitment to and dissociation from their target genes. We define CFTFs as all transcription factors that function to regulate cell fate decisions during either embryogenesis or adult cell differentiation. Interestingly, most — if not all — CFTFs are themselves PcG target genes^{12, 13, 14}. Some examples include the Hox, Sox, Runx, Fox, Pax and Gata transcription factor families. Functionally, they are known to regulate many key cell fate decisions, both in stem cells and during differentiation, by activating the target genes of specific fates and also by repressing alternative-fate genes^{55, 56, 57, 58}. Emerging data suggest that several well-characterized CFTFs exercise their roles in cell fate decisions through recruitment of PcG proteins to their target genes (Fig. 3a). Analogous to YY1 in *D. melanogaster*, the three embryonic stem cell (ES cell) 'core' transcription factors OCT4, SOX2 and NANOG co-occupy a subset of PcG target genes in human ES cells¹⁴. Several mammalian transcription factors physically associate with PcG proteins, such as YY1 (Ref. 59), RING1 and YY1-binding protein (RYBP)⁶⁰, PLZF^{61, 62}, GATA3 (Ref. 63) and E2F6 (Ref. 64). Furthermore, YY1, IKAROS and PLZF have been shown to be required for PcG association with target genes during mammalian development and differentiation^{59, 61, 65}. It is likely that other CFTFs displace PcG proteins from

their target genes during mammalian cellular differentiation (Fig. 3b). Supporting this possibility is the finding that tissue-specific TATA box-binding protein (TBP)-associated factors (TAFs) can displace PcG proteins from the promoters of key differentiation genes during terminal differentiation of *D. melanogaster* germ stem cells⁶⁶. Taken together, these data suggest that a large number of different CTFs could function in many different tissue types to regulate PcG function and therefore lineage choices.

Figure 3 | Potential mechanisms by which cell fate transcription factors and long non-coding RNAs function to regulate Polycomb group protein association with target genes during lineage choices and specification.



a | Cell fate transcription factors (CTFs) recruit Polycomb group (PcG) proteins to target genes during lineage decisions. b | CTFs induce the dissociation of PcG proteins from target genes during lineage decisions. c | Long non-coding RNAs (ncRNAs) recruit PcG proteins to target genes during lineage decisions. d | Coordinated action of CTFs and long ncRNAs is necessary to recruit PcG proteins to or dissociate them from target genes during lineage determination. The long ncRNAs can function either in cis or in trans.

Long non-coding RNAs (ncRNAs) recruit PcG proteins. Long ncRNAs are a subset of ncRNAs that are longer than 200 nucleotides and have diverse cellular functions (Box 1). Interestingly, they are also emerging as recruiters of PcG proteins to target genes⁶⁷. Recently, the long ncRNAs HOTAIR, KCNQ10T1 and REPA have been shown to recruit PcG proteins to chromatin through interaction with the PRC2 complex^{68, 69, 70, 71}. For example, Rinn et al. identified HOTAIR, which is expressed from within the HOXC locus, and demonstrated that it is required to repress transcription in trans across 40 kb of the HOXD locus⁶⁹. This repression is mediated through a direct interaction between HOTAIR and the PRC2 complex. The loss of the PRC2-mediated trimethylation of lysine 27 of histone H3 (H3K27me₃) on the HOXD locus in HOTAIR-depleted cells led the authors to speculate that HOTAIR recruits the PRC2 complex to this genomic region. Recently, several thousand additional long ncRNAs have been identified in both mouse and human cells, but little is known about their function^{69, 72, 73}. The fact that many have differential expression patterns during lineage specification suggests that they may have important roles during development and differentiation^{69, 72, 73}. Interestingly, one-fifth of all human long ncRNAs identified to date have been reported to physically associate with the PRC2 complex, suggesting that they may have a general role in recruiting PcG proteins to their target genes⁷³. Intriguingly, Khalil et al. carried out RNA fluorescence in situ hybridization on HOTAIR and four additional PcG-associated long ncRNAs and reported multiple 'speckles' in the cell nuclei, suggesting that they might function to recruit PcG proteins to many different target genes. The finding that RNA interference-mediated knockdown of long ncRNAs did not affect the expression levels of genes located in cis further supports the idea that long ncRNAs function in trans.

So, how do long ncRNAs recruit PcG proteins to their target genes? Current models (Fig. 3c) are based on the proposed abilities of long ncRNAs to bind genomic DNA in a sequence-specific manner at the promoters of target genes and to recruit PcG proteins through interaction with the SET domains and/or chromodomains that are present in several PcG proteins^{67, 74}. SET domains are characteristic of histone methyltransferases, such as EZH2 and the trithorax proteins MLL and Ash1, and have been shown to be general protein–nucleic acid interaction modules⁷⁵. Beyond a role in the recruitment of the PRC2 complex, long ncRNAs may also have a role in recruiting the PRC1 complex (which contains the chromodomain-containing Cbx proteins). This idea is supported by the observation that, similar to SET domains, chromodomains are protein–nucleic acid interaction modules⁷⁶. In fact, the mouse Cbx proteins have been shown to bind RNA⁷⁷ and the association of these and other chromodomain-containing proteins with chromatin seems to be RNA dependent^{76, 77, 78}. For example, Bernstein et al. demonstrated that the association of mouse CBX7 with the H3K27me₃ mark was attenuated following RNase treatment⁷⁷. This observation, coupled with the fact that the affinity of the Cbx protein chromodomains for H3K27me₃ is low, indicates that long ncRNAs may be required together with the trimethylation mark to stabilize the binding of PRC1 at target loci.

An additional potential mechanism for PcG recruitment may involve transcriptional control of long ncRNAs by CTFs (Fig. 3d). This hypothetical mechanism could represent a regulatory layer, which has received little attention so far, and would be consistent with the genetic evidence that CTFs have vital roles in lineage specification (Table 1). Evidence supporting this model was first provided by the groundbreaking chromatin immunoprecipitation (ChIP)-chip (Box 2) mapping of the binding sites of three transcription factors, p53, MYC and SP1, on the entire length of chromosomes 21 and 22 (Ref. 79). The key finding was that most transcription factor binding sites were in fact distal to the known transcription start sites of coding genes. The authors provided evidence that some of the binding sites were instead at the promoters of non-coding genes. Consistent with this, Guttman et al. recently reported that a subset of long ncRNAs is directly regulated by p53, nuclear factor-kappaB (NF-kappaB), SOX2, OCT4 and NANOG80. Using either whole-genome ChIP-chip or ChIP-sequencing approaches, four additional studies have substantiated these observations by demonstrating that the CTFs FOXA2, peroxisome proliferator-activated receptor-gamma (PPARgamma), CCAAT/enhancer-binding protein-alpha (C/EBPalpha) and FOXP3 also primarily bind away from the promoters of coding genes81, 82, 83, 84. The proportion of these distal binding sites that represent the promoters of long ncRNAs remains to be determined.

Table 1 | Cell fate transcription factors that are deregulated in human cancer

| Gene name | Role | Genetic alteration in human cancer | Evidence for role in cancer | Target genes |
|-----------|----------|---|--|--|
| ETV5 | Oncogene | Translocated in prostate cancer ¹²⁶ | <i>In vitro</i> model ¹²⁶ | |
| ETV7 | Oncogene | Overexpressed in lymphoma ¹²⁷ | <i>In vivo</i> model ¹²⁷ | |
| GATA6 | Oncogene | Amplified in pancreatic cancer ¹²⁸ | <i>In vitro</i> model ¹²⁸ | |
| HOXA9 | Oncogene | Translocated in myeloid leukaemia ¹⁰² | <i>In vivo</i> model ¹²⁹ | |
| LMO1 | Oncogene | Translocated in T cell leukaemia ¹³⁰ | <i>In vivo</i> model ¹³¹ | |
| LMO2 | Oncogene | Translocated in T cell leukaemia ¹³⁰ | <i>In vivo</i> model ¹³² | |
| MITF | Oncogene | Amplified in melanoma ¹³³ | <i>In vitro</i> model ¹³³ | CDKN2A ^{RNKA4A} (REF. 134) |
| MYB | Oncogene | Mutated in colon cancer and translocated in T-ALL ⁸⁵ | <i>In vivo</i> model ¹³⁵ | |
| MYCN | Oncogene | Amplified in neuroblastoma ¹³⁶ | <i>In vivo</i> model ¹³⁷ | |
| OTX2 | Oncogene | Amplified in medulloblastoma ^{138,139} | <i>In vitro</i> model ¹³⁹ | |
| PAX3 | Oncogene | Translocated in alveolar rhabdomyosarcoma ¹⁴⁰ | <i>In vivo</i> model ¹⁴¹ | |
| PLZF | Oncogene | Translocated in acute promyelocytic leukaemia ¹⁴² | <i>In vivo</i> model ¹⁴³ | |
| RUNX1 | Oncogene | Translocated in AML ¹⁴⁴ | <i>In vivo</i> model ¹⁴⁵ | NF1 (REF. 146) |
| TAL1 | Oncogene | Translocated or mutated in T-ALL ^{131,147} | <i>In vivo</i> model ^{131,147} | CD4 (REF. 147) |
| TBX2 | Oncogene | Amplified in breast cancer ¹⁴⁸ | <i>In vitro</i> model ¹⁴⁸ | CDKN2A ^{ARF} (REF. 148) |
| TTF1 | Oncogene | Amplified in lung cancer ^{149,150} | <i>In vitro</i> model ^{149,150} | |
| CDX2 | TS | Mutated in colon cancer ^{151,152} | <i>In vivo</i> model ¹⁵³ | CDKN1A (REF. 154) |
| CEBPA | TS | Mutated in AML ¹⁵⁵ | <i>In vivo</i> model ¹⁵⁶ | |
| FOXP3 | TS | Mutated and deleted in breast cancer ^{157,158} | <i>In vitro</i> and <i>in vivo</i> models ^{157,158} | ERBB2 (REF. 158) and SKP2 (REF. 157) |
| GATA3 | TS | Silenced and mutated in breast cancer ^{159,160} | <i>In vivo</i> model ¹⁵⁹ | FOXA1 (REF. 159) and CDKN2C ¹⁶¹ |
| HOXA5 | TS | Silenced in breast cancer ¹⁶² | <i>In vitro</i> model ¹⁶² | TP53 (REF. 162) |
| IKAROS | TS | Deleted in AML ^{90,107} | <i>In vitro</i> model ¹⁶⁵ | HES1 (REF. 65) |
| ING1 | TS | Mutated in squamous cell carcinoma ^{163,164} | <i>In vivo</i> model ¹⁶⁵ | |
| KLF6 | TS | Mutated in prostate cancer ¹⁶⁶ and deregulated in GBM ¹⁶⁷ | <i>In vivo</i> model ¹⁶⁷ | ATF3 (REF. 168) |
| PAX5 | TS | Mutated, deleted or fused in ALL ¹⁰⁷ | <i>In vitro</i> model ¹⁰⁷ | CD19 and CD72 (REF. 107) |
| PJ.1 | TS | Inactivated or mutated in AML ^{169,170} | <i>In vivo</i> model ^{171,172} | JUNB ¹⁷¹ |
| RUNX3 | TS | Mutated or silenced in gastric cancer ¹⁷³ | <i>In vivo</i> model ¹⁷⁴ | |
| SMAD4 | TS | Mutated in pancreatic cancer ¹⁷⁵ | <i>In vivo</i> model ¹⁷⁶ | |
| WT1 | Both | Mutated in hepatic cancer ^{177,178} | <i>In vitro</i> model ¹⁷⁹ | CDKN1A (REF. 179) |

ALL, acute lymphoblastic leukaemia; AML, acute myeloid leukaemia; ATF3, activating transcription factor 3; CEBPA, CCAAT/enhancer binding protein- α ; CDKN1A, cyclin-dependent kinase inhibitor 2A; CDKN2A, cyclin-dependent kinase 2A; CDKN2C, cyclin-dependent kinase inhibitor 2C; CDX2, caudal type homeobox 2; ETV, ets variant; Fox, forkhead box; GBM, glioblastoma multiforme; HES1, hairy and enhancer of split 1; Hox, homeobox; LMO, LIM domain only; ING1, inhibitor of growth family, member 1; KLF6, Kruppel-like factor 6; MITF, microphthalmia-associated transcription factor; NF1, neurofibromin 1; OTX2, orthodenticle homeobox 2; Pax, paired box; Runx, runt-related transcription factor; SKP2, S-phase kinase-associated protein 2; TAL1, T-cell acute lymphocytic leukaemia 1; T-ALL, T cell acute lymphoblastic leukaemia; TBX2, T-box 2; TS, tumour suppressor; WT1, Wilms tumor 1.

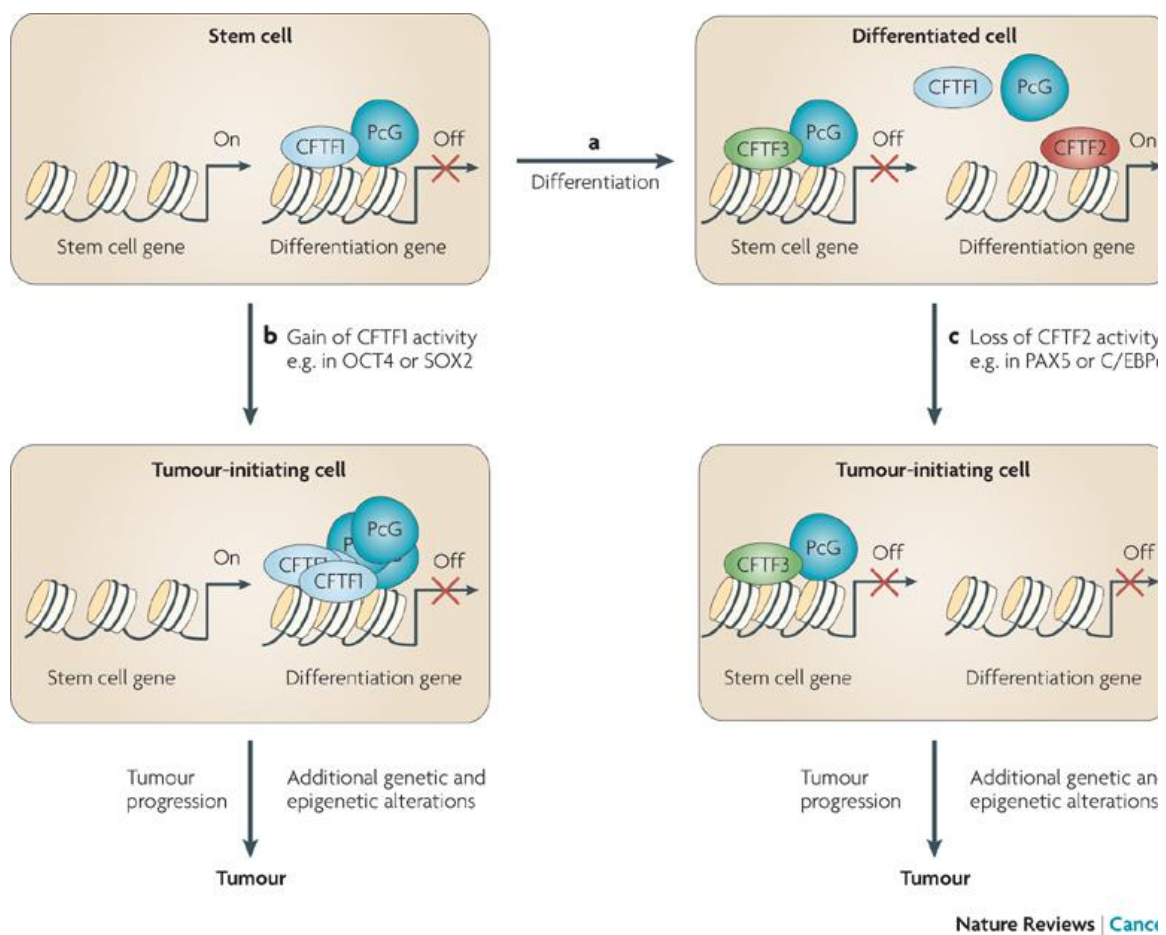
The potential role of CFTFs and ncRNAs in cancer

What is known about the role of CFTFs and long ncRNAs in cancer? The genetic evidence supporting a role for transcription factors in cancer is probably stronger than for any other functional group of proteins. For example, MYC is one of the best characterized human oncogenes, and TP53 (which encodes p53 in humans) and RB1 are the two most studied human tumour-suppressor genes. There is also strong evidence that at least 30 CFTFs are genetically altered and contribute to cancer in a tissue-specific manner (Table 1). The precise mechanisms of action are still poorly understood in many cases. However, the evidence suggests that their normal roles in the regulation of lineage-specific cell fate decisions become perturbed on their mutation, amplification, translocation and deletion in cancer. Some specific examples include the oncogenes MYB85, SOX2 (Ref. 180), MITF86 and GATA2 (Ref. 87), and the tumour suppressors GATA3 (Ref. 88), CEBP &agr;89, IKAROS and PAX5 (Ref. 90). Importantly, in

addition to being genetically altered in human cancers, mouse models have established the significance of CTFs as cancer-relevant genes (Table 1).

The model emerging is that CTFs can be subdivided into two classes on the basis of their normal function, and that the deregulation of both classes can potentially contribute to the formation of non-differentiated or tumour-initiating cells (Fig. 4). Oncogenes belong to the first class, as they are normally expressed in stem or progenitor cells, and the tumour suppressors are in the second class, as they are normally expressed during differentiation and are required for lineage specification. We propose that the deregulation of either class of CTFs leads to the accumulation of cells incapable of undergoing differentiation (Fig. 4). These pre-tumorigenic cells then have the potential to further progress to become tumours after the accumulation of additional genetic and/or epigenetic alterations. To illustrate this hypothesis we describe some examples of these two classes of CTFs and pay particular attention to those CTFs for which there is evidence of a functional interaction with PcG proteins.

Figure 4 | Gain and loss of cell fate transcription factors may lead to the formation of tumour-initiating cells.



This model illustrates how the loss or gain of function of two putative cell fate transcription factors (CFTFs) may lead to the formation of a tumour-initiating cell. a | Normal differentiation of a stem or progenitor cell. The levels of CFTF1 decrease during differentiation and the levels of CFTF2 and CFTF3 increase. The decrease of CFTF1 and the increase of CFTF2 lead to displacement of Polycomb group (PcG) proteins from the promoter of the differentiation gene. The increase in CFTF3 levels leads to the recruitment of PcG proteins to the promoter of the stem cell gene. b | Conversion of normal stem cells to tumour-initiating cells. In this case, the levels of CFTF1 become aberrantly high in the stem cells and as a consequence the differentiation gene remains repressed, rendering the cell insensitive to differentiation signals. c | The conversion of differentiated cells to tumour-initiating cells. In this scenario, CFTF2 function is lost in a differentiated cell (for example, by mutation or deletion of the gene) and therefore the cell reverts or de-differentiates to a more stem cell-like state or to a tumour-initiating cell in which the differentiation gene is aberrantly silenced. Notably, it is also possible that CFTF1 could be activated in a differentiated cell or that CFTF2 could be deleted or mutated in a stem cell or progenitor. In addition, the loss of CFTF3 activity could lead to an inability to repress the stem cell gene.

OCT4 is normally expressed in pluripotent cells of the early embryo and in ES cells, and it is required for maintaining these cells in an undifferentiated state⁹¹. A potential oncogenic activity of OCT4 was revealed when it was shown to be highly expressed in human germ cell tumours and was required for their continued growth⁹². In addition, the ectopic expression of OCT4 blocks progenitor cell differentiation and causes dysplasia in epithelial tissues⁹³. Importantly, OCT4 occupies several hundred PcG target genes in human ES cells and is thought to contribute to the sustained recruitment of PcG proteins to the promoters of differentiation genes¹⁴. Therefore, the upregulation of OCT4 in cancer might lead to the persistent or sustained PcG-mediated repression of differentiation genes and a consequent block of the ability of cells to respond to

differentiation cues (Fig. 4). Several other CTFs, such as MYB85, PLZF94, HOXA9 (Refs 95–97), PAX3 (Ref. 98) and PAX7 (Ref. 99), are known to function in tissue-specific stem and progenitor cells and have been found to have gain of function in cancer. PLZF recruits BMI1 and the associated PRC1 complex to repress the HOXD locus during mouse development⁶¹. PLZF is expressed in haematopoietic stem and progenitor cells and is an essential regulator of spermatogonial stem cell maintenance^{94, 100}. Importantly, the PLZF–retinoic acid receptor- α (RAR α) fusion protein, like the promyelocytic leukaemia (PML)–RAR α fusion protein, can aberrantly recruit PcG proteins to target genes during cancer development^{62, 101}. This raises the possibility that other CTFs form fusion proteins with this ability. For example, HOXA9 is expressed in haematopoietic stem cells and progenitors and is translocated in myeloid leukaemia¹⁰². Similarly, the PAX3 and PAX7 CTFs function during embryonic myogenesis (muscle development) and are translocated in alveolar rhabdomyosarcoma — a childhood cancer of skeletal muscle cells⁵⁸. Several other CTFs that are normally expressed in undifferentiated cells are deregulated in cancer without being expressed as fusion proteins (Table 1). For example, MYB is expressed in colon stem cells and progenitor cells; it is genetically disrupted in colon cancer by a mutation in an intron, leading to higher expression levels⁸⁵. SOX2, like OCT4, occupies a subset of PcG target genes in ES cells and is also expressed in tissue-specific stem and progenitor cells, including neural, lung and oesophageal cells^{103, 104, 105}. Interestingly, SOX2 is amplified in both lung and oesophageal squamous cell carcinomas, suggesting that its role in cancer is to maintain cells in a pre-terminally differentiated state¹⁸⁰. It will be interesting to discover whether MYB, PAX3, PAX7 and other CTFs are involved in regulating PcG target genes and/or whether they regulate PcG function.

We propose that the second group of CTFs promotes differentiation by recruiting PcG proteins to stem cell genes and/or by displacing PcG proteins from differentiation target genes (Fig. 4). When inactivated in cancer (for example by deletion or mutation) this would lead to cells being unable to repress stem cell genes and/or activate a programme of differentiation genes. Two potential examples, PAX5 and IKAROS, are required for B lymphocyte differentiation^{58, 106} and are frequently deleted in acute lymphoblastic leukaemias^{90, 107}. Interestingly, mice with loss of function of Ikaros or deletion of Pax5 have reduced H3K27me₃ at certain loci, suggesting that these CTFs functionally interact with PcG proteins^{65, 108}. Another example is GATA3, which is known to be required for luminal cell differentiation in breast epithelia⁸⁸ and is mutated in breast cancers¹⁰⁹. C/EBP α , however, is required for granulocytic differentiation of bipotent granulocyte-macrophage progenitor cells and is mutated in acute myeloid leukaemia⁸⁹. Most of these lineage-specific CTFs remain to be characterized in terms of their interactions with PcG proteins and epigenetic modifiers per se (Table 1).

The idea that the deregulation of CTFs can change cell fate and lead to tumour development has been highlighted by the recent interest in cellular reprogramming^{110, 111, 112} (Box 3). Adult somatic cells can be induced to trans-differentiate into cell types of other lineages or de-differentiate into embryonic stem-like cells called induced pluripotent stem cells. It is now clear that the controlled gain or loss of expression of specific sets of CTFs in different contexts has the power to reprogramme cell identity. This has been shown to lead to a resetting of the epigenetic landscape in these cells^{113, 114}, further supporting the hypothesis that the deregulation of even one CTF could potentially induce epigenetic reprogramming and contribute to tumour initiation. Importantly, it is also likely that oncogenic CTFs (such as OCT4 and SOX2) when activated or tumour suppressive CTFs (such as PAX5 and IKAROS) when inactivated lead to a block of the differentiation of immature cells or a de-differentiation of more mature cells (Fig. 4).

Do CTFs always have to be genetically altered as outlined in Table 1? The most likely answer is no. It is well established that cellular signalling pathways are commonly deregulated in human cancer¹¹⁵. Most — if not all — of these pathways regulate cell fate decisions by controlling the abundance and/or activity of downstream effector CTFs. So, for example, although OCT4 has not been shown to be genetically altered in germ cell cancers, it responds to signalling from fibroblast growth factors, the leukaemia inhibitory factor–signal transducer and activator of transcription 3 pathway, the transforming growth factor-beta–bone morphogenic protein pathway and the nodal and Wnt pathways, any of which could be deregulated¹¹⁶. It is therefore logical to assume that many additional CTFs are deregulated in cancer as a consequence of altered signalling pathways. Consistent with this, several CTFs are epigenetically, rather than genetically, deregulated in cancer. For example, PAX2 is upregulated in endometrial cancer¹¹⁷ and HOXA5 is downregulated in breast cancer¹⁰⁹.

The potential of long ncRNAs as drivers of tumour formation is also apparent. There is emerging, if so far limited, evidence that long ncRNA genes are indeed cancer relevant. The expression of HOTAIR is higher in metastatic breast cancer than in primary breast epithelial cells (R. Gupta and H.Y. Chang, personal communication). Moreover, these observations provide evidence that HOTAIR contributes to the metastatic phenotype, and that this correlated with the aberrant recruitment of PcG proteins to multiple target genes. In another recent study, Yu et al. searched for antisense transcripts associated with 21 well-known tumour suppressor genes¹¹⁸. They identified a 34.8 kb transcript (which they called p15AS) that was associated with the cyclin-dependent kinase inhibitor and PcG target gene CDKN2B, which is frequently silenced in leukaemia. The authors examined the expression of both CDKN2B and p15AS in leukaemic leukocytes and found that in most cases p15AS expression was increased with a concomitant decrease in INK4B expression. Ectopic expression of p15AS was shown to increase DNA methylation levels at the CDKN2B promoter. An

interesting and so far unexplored possibility is that the aberrantly high levels of this long ncRNA that are observed in cancer could lead to the permanent repression of the CDKN2B locus through PcG recruitment and subsequent accumulation of DNA methylation. Another study identified a 7 kb long ncRNA, named hcn, as a marker for mouse hepatocellular carcinoma (HCC)119. Expression of this long ncRNA was found to be eightfold higher in a mouse model of HCC compared with matched normal liver tissue. This high expression was observed in all stages of HCC, implicating it as a potential initiating lesion in the development in cancer. Furthermore, the authors identified a human orthologue of hcn, metastasis associated lung adenocarcinoma transcript 1 (MALAT1), which is highly expressed in human cancers. It will be interesting to determine the biological function of these long ncRNAs and in particular whether they function to recruit PcG proteins to target genes during lineage choices and whether their deregulation contributes to cancer. It is clear that long ncRNAs represent a promising candidate set of potential oncogenes and tumour suppressor genes.

Perspectives

Current research efforts are directed at understanding the mechanisms by which PcG proteins are recruited to and displaced from their target genes during lineage specification. Both CTFs and long ncRNAs are emerging as key regulators of these events. In the next few years, we will see a number of papers in which the target genes of cancer-relevant CTFs will be delineated. Analogous to the unravelling of the transcriptional networks controlling ES cells, we expect that similar efforts will establish the transcriptional networks of adult stem cells, progenitors and differentiated cells. Studies will also address the hypothesis that PcG recruitment is regulated by CTFs and long ncRNAs during lineage choice. It will be important to determine whether specific subsets of PcG target genes are activated or repressed by specific CTFs or long ncRNAs during lineage choices. Moreover, it will be interesting to determine to what extent CTFs regulate the expression of long ncRNAs, and whether long ncRNAs dictate the recruitment of the PcG proteins. These studies will provide important information regarding the molecular mechanisms that control normal cell fate decisions and also how the deregulation of key players (that is, CTFs and long ncRNAs) might lead to cancer. For instance, do the many alterations in CTFs that have been documented in various tumours (Table 1) contribute to the altered epigenetic profiles observed in these tumours? Currently, remarkably little is known about these specific transcriptional alterations and how they are influenced by the deregulation of the CTFs and/or long ncRNAs.

Over the next 5 years many functionally important long ncRNAs will probably be uncovered. It will be fascinating to learn how many of the long ncRNAs will emerge as bona fide oncogenes and tumour suppressor genes. Anticancer therapies targeting lineage-specific CTFs or long ncRNAs may have advantages over drugs directed at PcG proteins or DNA methylation enzymes, as they are more likely to be cell type specific. In conclusion, it is clear that a better understanding of the role of CTFs and long ncRNAs in modulating the

epigenetic activity of the PcG proteins will provide more targets for anticancer therapy, and therefore is promising for the tailor-made individualized treatment of cancer patients.

Acknowledgements

We thank G. Brien for help with Table 1, E. Bernstein for critical reading of the manuscript and members of the Bracken and Helin laboratories for very helpful discussions. Work in the Bracken laboratory is supported by the Smurfit Institute of Genetics, the Adelaide & Meath Hospital, including the National Children's Hospital, and the Trinity College Dublin start-up fund for new lecturers. Work in the Helin laboratory is supported by the Danish National Research Foundation, the Danish Cancer Society, the Novo Nordisk Foundation, the Danish Medical Research Council, the Danish Natural Science Research Council, the Lundbeck Foundation, the Association for International Cancer Research and the Excellence Programme of the University of Copenhagen.

References

Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* 414, 105–111 (2001).

Stingl, J. & Caldas, C. Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. *Nature Rev. Cancer* 7, 791–799 (2007).

Vescovi, A. L., Galli, R. & Reynolds, B. A. Brain tumour stem cells. *Nature Rev. Cancer* 6, 425–436 (2006).

Feinberg, A. P., Ohlsson, R. & Henikoff, S. The epigenetic progenitor origin of human cancer. *Nature Rev. Genet.* 7, 21–33 (2006).

The authors discuss the epigenetic progenitor model of cancer, in which tumour development is proposed to involve the alteration of gene expression patterns owing to epigenetic changes. This model is under active investigation.

Ohm, J. E. & Baylin, S. B. Stem cell chromatin patterns: an instructive mechanism for DNA hypermethylation? *Cell Cycle* 6, 1040–1043 (2007).

Ben-Porath, I. et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nature Genet.* 40, 499–507 (2008).

Ohm, J. E. et al. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nature Genet.* 39, 237–242 (2007).

Schlesinger, Y. et al. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nature Genet.* 39, 232–236 (2007).

Widschwendter, M. et al. Epigenetic stem cell signature in cancer. *Nature Genet.* 39, 157–158 (2007).

References 7–9 established that PcG target genes are frequently aberrantly hypermethylated by DNA methylation in cancer.

Sparmann, A. & van Lohuizen, M. Polycomb silencers control cell fate, development and cancer. *Nature Rev. Cancer* 6, 846–856 (2006).

Pietersen, A. M. & van Lohuizen, M. Stem cell regulation by polycomb repressors: postponing commitment. *Curr. Opin. Cell Biol.* 20, 201–207 (2008).

Boyer, L. A. et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441, 349–353 (2006).

Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H. & Helin, K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* 20, 1123–1136 (2006).

Lee, T. I. et al. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125, 301–313 (2006).

References 12–14 were the first studies to perform genome-wide mapping of target genes for the PcG proteins.

Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M. & Helin, K. The Polycomb Group protein Suz12 is required for embryonic stem cell differentiation. *Mol. Cell. Biol.* 27, 3769–3779 (2007).

Ezhkova, E. et al. Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. *Cell* 136, 1122–1135 (2009).

Mohn, F. & Schubeler, D. Genetics and epigenetics: stability and plasticity during cellular differentiation. *Trends Genet.* 25, 129–136 (2009).

Mohn, F. et al. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* 30, 755–766 (2008).

Oktaba, K. et al. Dynamic regulation by polycomb group protein complexes controls pattern formation and the cell cycle in *Drosophila*. *Dev. Cell* 15, 877–889 (2008).

Lessard, J. & Sauvageau, G. Bmi-1 determines the proliferative capacity of and leukaemic stem cells. *Nature* 423, 255–260 (2003).

Molofsky, A. V. et al. Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. *Nature* 425, 962–967 (2003).

Park, I. K. et al. Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* 423, 302–305 (2003).

Dietrich, N. et al. Bypass of senescence by the polycomb group protein CBX8 through direct binding to the INK4A-ARF locus. *EMBO J.* 26, 1637–48 (2007).

Gil, J., Bernard, D., Martinez, D. & Beach, D. Polycomb CBX7 has a unifying role in cellular lifespan. *Nature Cell Biol.* 6, 67–72 (2004).

Gil, J. & Peters, G. Regulation of the INK4b-ARF-INK4a tumour suppressor locus: all for one or one for all. *Nature Rev. Mol. Cell Biol.* 7, 667–677 (2006).

Bracken, A. P. et al. The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. *Genes Dev.* 21, 525–30 (2007).

Bruggeman, S. W. et al. Ink4a and Arf differentially affect cell proliferation and neural stem cell self-renewal in Bmi1-deficient mice. *Genes Dev.* 19, 1438–1443 (2005).

Chen, H. et al. Polycomb protein Ezh2 regulates pancreatic beta-cell Ink4a/Arf expression and regeneration in diabetes mellitus. *Genes Dev.* 23, 975–985 (2009).

Dhawan, S., Tschen, S. I. & Bhushan, A. Bmi-1 regulates the Ink4a/Arf locus to control pancreatic beta-cell proliferation. *Genes Dev.* 23, 906–911 (2009).

Leung, C. et al. Bmi1 is essential for cerebellar development and is overexpressed in human medulloblastomas. *Nature* 428, 337–341 (2004).

Molofsky, A. V., He, S., Bydon, M., Morrison, S. J. & Pardal, R. Bmi-1 promotes neural stem cell self-renewal and neural development but not mouse growth and survival by repressing the p16Ink4a and p19Arf senescence pathways. *Genes Dev.* 19, 1432–1437 (2005).

Lowe, S. W. & Sherr, C. J. Tumor suppression by Ink4a-Arf: progress and puzzles. *Curr. Opin. Genet. Dev.* 13, 77–83 (2003).

Vire, E. et al. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 439, 871–874 (2006).

Mohammad, H. P. et al. Polycomb CBX7 promotes initiation of heritable repression of genes frequently silenced with cancer-specific DNA hypermethylation. *Cancer Res.* 69, 6322–6330 (2009).

Bea, S. et al. BMI-1 gene amplification and overexpression in hematological malignancies occur mainly in mantle cell lymphomas. *Cancer Res.* 61, 2409–2412 (2001).

Rubio-Moscardo, F. et al. Mantle-cell lymphoma genotypes identified with CGH to BAC microarrays define a leukemic subgroup of disease and predict patient outcome. *Blood* 105, 4445–4454 (2005).

van Lohuizen, M. et al. Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. *Cell* 65, 737–752 (1991).

Jacobs, J. J. et al. Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via INK4a/ARF. *Genes Dev.* 13, 2678–2690 (1999).

Koontz, J. I. et al. Frequent fusion of the JAZF1 and JJAZ1 genes in endometrial stromal tumors. *Proc. Natl Acad. Sci. USA* 98, 6348–6353 (2001).

Bachmann, I. M. et al. EZH2 expression is associated with high proliferation rate and aggressive tumor subgroups in cutaneous melanoma and cancers of the endometrium, prostate, and breast. *J. Clin. Oncol.* 24, 268–273 (2006).

Bracken, A. P. et al. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J.* 22, 5323–35 (2003).

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).

Saramaki, O. R., Tammela, T. L., Martikainen, P. M., Vessella, R. L. & Visakorpi, T. The gene for polycomb group protein enhancer of zeste homolog 2 (EZH2) is amplified in late-stage prostate cancer. *Genes Chromosom. Cancer* 45, 639–645 (2006).

van Kemenade, F. J. et al. Coexpression of BMI-1 and EZH2 polycomb-group proteins is associated with cycling cells and degree of malignancy in B-cell non-Hodgkin lymphoma. *Blood* 97, 3896–3901 (2001).

Varambally, S. et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* 419, 624–629 (2002).

Vekony, H. et al. High expression of Polycomb group protein EZH2 predicts poor survival in salivary gland adenoid cystic carcinoma. *J. Clin. Pathol.* 61, 744–749 (2008).

Weikert, S. et al. Expression levels of the EZH2 polycomb transcriptional repressor correlate with aggressiveness and invasive potential of bladder carcinomas. *Int. J. Mol. Med.* 16, 349–353 (2005).

Friedman, J. M. et al. The putative tumor suppressor microRNA-101 modulates the cancer epigenome by repressing the polycomb group protein EZH2. *Cancer Res.* 69, 2623–2629 (2009).

Varambally, S. et al. Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. *Science* 322, 1695–1699 (2008).

Raaphorst, F. M., Otte, A. P. & Meijer, C. J. Polycomb-group genes as regulators of mammalian lymphopoiesis. *Trends Immunol.* 22, 682–690 (2001).

Ringrose, L. & Paro, R. Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development* 134, 223–232 (2007).

Kwong, C. et al. Stability and dynamics of polycomb target sites in *Drosophila* development. *PLoS Genet.* 4, e1000178 (2008).

Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* 25, 25–29 (2000).

Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M. & Teichmann, S. A. Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.* 14, 283–291 (2004).

Blyth, K., Cameron, E. R. & Neil, J. C. The RUNX genes: gain or loss of function in cancer. *Nature Rev. Cancer* 5, 376–387 (2005).

Myatt, S. S. & Lam, E. W. The emerging roles of forkhead box (Fox) proteins in cancer. *Nature Rev. Cancer* 7, 847–859 (2007).

Pearson, J. C., Lemons, D. & McGinnis, W. Modulating Hox gene functions during animal body patterning. *Nature Rev. Genet.* 6, 893–904 (2005).

Robson, E. J., He, S. J. & Eccles, M. R. A PANorama of PAX genes in cancer and development. *Nature Rev. Cancer* 6, 52–62 (2006).

Caretti, G., Di Padova, M., Micales, B., Lyons, G. E. & Sartorelli, V. The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. *Genes Dev.* 18, 2627–2638 (2004).

Garcia, E., Marcos-Gutierrez, C., del Mar Lorente, M., Moreno, J. C. & Vidal, M. RYBP, a new repressor protein that interacts with components of the mammalian Polycomb complex, and with the transcription factor YY1. *EMBO J.* 18, 3404–3418 (1999).

Barna, M. et al. Plzf mediates transcriptional repression of HoxD gene expression through chromatin remodeling. *Dev. Cell* 3, 499–510 (2002).

Boukarabila, H. et al. The PRC1 Polycomb group complex interacts with PLZF/RARA to mediate leukemic transformation. *Genes Dev.* 23, 1195–1206 (2009).

Hosokawa, H. et al. Regulation of Th2 cell development by Polycomb group gene bmi-1 through the stabilization of GATA3. *J. Immunol.* 177, 7656–7664 (2006).

Trimarchi, J. M., Fairchild, B., Wen, J. & Lees, J. A. The E2F6 transcription factor is a component of the mammalian Bmi1-containing polycomb complex. *Proc. Natl Acad. Sci. USA* 98, 1519–1524 (2001).

Kleinmann, E., Geimer Le Lay, A. S., Sellars, M., Kastner, P. & Chan, S. Ikaros represses the transcriptional response to Notch signaling in T-cell development. *Mol. Cell. Biol.* 28, 7465–7475 (2008).

Chen, X., Hiller, M., Sancak, Y. & Fuller, M. T. Tissue-specific TAFs counteract Polycomb to turn on terminal differentiation. *Science* 310, 869–872 (2005).

Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nature Rev. Genet.* 10, 155–159 (2009).

Pandey, R. R. et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* 32, 232–246 (2008).

Rinn, J. L. et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323 (2007).

This paper identified HOTAIR as a long ncRNA that is transcribed from the HOXC locus, which the authors showed associates with the PRC2 complex and is required for recruitment of PcGs to the HOXD locus.

Terranova, R. et al. Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev. Cell* 15, 668–679 (2008).

Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322, 750–756 (2008).

Dinger, M. E. et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* 18, 1433–1445 (2008).

Khalil, A. M. et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl Acad. Sci. USA* 106, 11667–11672 (2009).

Hogg, J. R. & Collins, K. Structured non-coding RNAs and the RNP Renaissance. *Curr. Opin. Chem. Biol.* 12, 684–689 (2008).

Krajewski, W. A., Nakamura, T., Mazo, A. & Canaani, E. A motif within SET-domain proteins binds single-stranded nucleic acids and transcribed and supercoiled DNAs and can interfere with assembly of nucleosomes. *Mol. Cell. Biol.* 25, 1891–1899 (2005).

Akhtar, A., Zink, D. & Becker, P. B. Chromodomains are protein-RNA interaction modules. *Nature* 407, 405–409 (2000).

Bernstein, E. et al. Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol. Cell. Biol.* 26, 2560–2569 (2006).

Muchardt, C. et al. Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1alpha. *EMBO Rep.* 3, 975–981 (2002).

Cawley, S. et al. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* 116, 499–509 (2004).

A landmark paper, which established that three transcription factors do not primarily bind at the promoters of known coding genes, as previously assumed.

Guttman, M. et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458, 223–227 (2009).

This paper identified 1,600 conserved long ncRNAs in humans, and showed the differential expression of some of these during embryonic development and in response to DNA damage.

Lefterova, M. I. et al. PPARgamma and C/EBP factors orchestrate adipocyte biology via adjacent binding on a genome-wide scale. *Genes Dev.* 22, 2941–2952 (2008).

Nielsen, R. et al. Genome-wide profiling of PPARgamma:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Genes Dev.* 22, 2953–2967 (2008).

Robertson, A. G. et al. Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding. *Genome Res.* 18, 1906–1917 (2008).

Zheng, Y. et al. Genome-wide analysis of Foxp3 target genes in developing and mature regulatory T cells. *Nature* 445, 936–940 (2007).

Ramsay, R. G. & Gonda, T. J. MYB function in normal and cancer cells. *Nature Rev. Cancer* 8, 523–534 (2008).

Garraway, L. A. & Sellers, W. R. Lineage dependency and lineage-survival oncogenes in human cancer. *Nature Rev. Cancer* 6, 593–602 (2006).

Zhang, S. J. et al. Gain-of-function mutation of GATA-2 in acute myeloid transformation of chronic myeloid leukemia. *Proc. Natl Acad. Sci. USA* 105, 2076–2081 (2008).

Kouros-Mehr, H., Kim, J. W., Bechis, S. K. & Werb, Z. GATA-3 and the regulation of the mammary luminal cell fate. *Curr. Opin. Cell Biol.* 20, 164–170 (2008).

Nerlov, C. The C/EBP family of transcription factors: a paradigm for interaction between gene expression and proliferation control. *Trends Cell Biol.* 17, 318–324 (2007).

Mullighan, C. G. et al. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* 453, 110–114 (2008).

Nichols, J. et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 95, 379–391 (1998).

Gidekel, S., Pizov, G., Bergman, Y. & Pikarsky, E. Oct-3/4 is a dose-dependent oncogenic fate determinant. *Cancer Cell* 4, 361–370 (2003).

Hochedlinger, K., Yamada, Y., Beard, C. & Jaenisch, R. Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. *Cell* 121, 465–477 (2005).

Reid, A. et al. Leukemia translocation gene, PLZF, is expressed with a speckled nuclear pattern in early hematopoietic progenitors. *Blood* 86, 4544–4552 (1995).

Ivanova, N. B. et al. A stem cell molecular signature. *Science* 298, 601–604 (2002).

Sauvageau, G. et al. Differential expression of homeobox genes in functionally distinct CD34⁺ subpopulations of human bone marrow cells. *Proc. Natl Acad. Sci. USA* 91, 12223–12227 (1994).

Thorsteinsdottir, U. et al. Overexpression of the myeloid leukemia-associated Hoxa9 gene in bone marrow cells induces stem cell expansion. *Blood* 99, 121–129 (2002).

Lang, D. et al. Pax3 functions at a nodal point in melanocyte stem cell differentiation. *Nature* 433, 884–887 (2005).

Olguin, H. C. & Olwin, B. B. Pax-7 up-regulation inhibits myogenesis and cell cycle progression in satellite cells: a potential mechanism for self-renewal. *Dev. Biol.* 275, 375–388 (2004).

Kotaja, N. & Sassone-Corsi, P. Plzf pushes stem cells. *Nature Genet.* 36, 551–553 (2004).

Villa, R. et al. Role of the polycomb repressive complex 2 in acute promyelocytic leukemia. *Cancer Cell* 11, 513–525 (2007).

Nakamura, T. et al. Fusion of the nucleoporin gene NUP98 to HOXA9 by the chromosome translocation t(7;11)(p15;p15) in human myeloid leukaemia. *Nature Genet.* 12, 154–158 (1996).

Que, J. et al. Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. *Development* 134, 2521–2531 (2007).

Que, J., Luo, X., Schwartz, R. J. & Hogan, B. L. Multiple roles for Sox2 in the developing and adult mouse trachea. *Development* 136, 1899–1907 (2009).

Chen, X. et al. Multilayered epithelium in a rat model and human Barrett's esophagus: similar expression patterns of transcription factors and differentiation markers. *BMC Gastroenterol.* 8, 1 (2008).

Georgopoulos, K. Haematopoietic cell-fate decisions, chromatin regulation and ikaros. *Nature Rev. Immunol.* 2, 162–174 (2002).

Mullighan, C. G. et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 446, 758–764 (2007).

Xu, C. R., Schaffer, L., Head, S. R. & Feeney, A. J. Reciprocal patterns of methylation of H3K36 and H3K27 on proximal vs. distal IgVH genes are modulated by IL-7 and Pax5. *Proc. Natl Acad. Sci. USA* 105, 8685–8690 (2008).

Tlsty, T. D. Luminal cells GATA have it. *Nature Cell Biol.* 9, 135–136 (2007).

Gurdon, J. B. & Melton, D. A. Nuclear reprogramming in cells. *Science* 322, 1811–1815 (2008).

Hochedlinger, K. & Plath, K. Epigenetic reprogramming and induced pluripotency. *Development* 136, 509–523 (2009).

Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676 (2006).

This is a seminal paper that describes how the expression of just four CTFs, OCT4, SOX2, KLF4 and MYC, could de-differentiate skin fibroblasts into embryonic stem cells. It inspired efforts to generate patient-specific tissues for regenerative medicine. From a cancer perspective, together with subsequent papers, it highlights the reprogramming power of CTFs.

Deng, J. et al. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nature Biotechnol.* 27, 353–360 (2009).

Maherali, N. et al. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* 1, 55–70 (2007).

Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* 100, 57–70 (2000).

Boiani, M. & Scholer, H. R. Regulatory networks in embryo-derived pluripotent stem cells. *Nature Rev. Mol. Cell Biol.* 6, 872–884 (2005).

Wu, H. et al. Hypomethylation-linked activation of PAX2 mediates tamoxifen-stimulated endometrial carcinogenesis. *Nature* 438, 981–987 (2005).

Yu, W. et al. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* 451, 202–206 (2008).

This paper identified a long ncRNA that represses CDKN2B.

Lin, R., Maeda, S., Liu, C., Karin, M. & Edgington, T. S. A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. *Oncogene* 26, 851–858 (2007).

Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Rev. Genet.* 10, 57–63 (2009).

Zhou, Q. & Melton, D. A. Extreme makeover: converting one cell into another. *Cell Stem Cell* 3, 382–388 (2008).

Nutt, S. L., Heavey, B., Rolink, A. G. & Busslinger, M. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* 401, 556–562 (1999).

Laiosa, C. V., Stadtfeld, M., Xie, H., de Andres-Aguayo, L. & Graf, T. Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBPalpha and PU.1 transcription factors. *Immunity* 25, 731–744 (2006).

Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise reprogramming of B cells into macrophages. *Cell* 117, 663–676 (2004).

Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. A. In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* 455, 627–632 (2008).

Helgeson, B. E. et al. Characterization of TMPRSS2:ETV5 and SLC45A3:ETV5 gene fusions in prostate cancer. *Cancer Res.* 68, 73–80 (2008).

Cardone, M. et al. The novel ETS factor TEL2 cooperates with Myc in B lymphomagenesis. *Mol. Cell. Biol.* 25, 2395–2405 (2005).

Kwei, K. A. et al. Genomic profiling identifies GATA6 as a candidate oncogene amplified in pancreaticobiliary cancer. *PLoS Genet.* 4, e1000081 (2008).

Kroon, E., Thorsteinsdottir, U., Mayotte, N., Nakamura, T. & Sauvageau, G. NUP98-HOXA9 expression in hemopoietic stem cells induces chronic and acute myeloid leukemias in mice. *EMBO J.* 20, 350–361 (2001).

Grutz, G. G. et al. The oncogenic T cell LIM-protein Lmo2 forms part of a DNA-binding complex specifically in immature T cells. *EMBO J.* 17, 4594–4605 (1998).

Chervinsky, D. S. et al. Disordered T-cell development and T-cell malignancies in SCL LMO1 double-transgenic mice: parallels with E2A-deficient mice. *Mol. Cell. Biol.* 19, 5025–5035 (1999).

Fisch, P. et al. T-cell acute lymphoblastic lymphoma induced in transgenic mice by the RBTN1 and RBTN2 LIM-domain genes. *Oncogene* 7, 2389–2397 (1992).

Garraway, L. A. et al. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* 436, 117–122 (2005).

Loercher, A. E., Tank, E. M., Delston, R. B. & Harbour, J. W. MITF links differentiation with cell cycle arrest in melanocytes by transcriptional activation of INK4A. *J. Cell Biol.* 168, 35–40 (2005).

Lidonnici, M. R., Corradini, F., Waldron, T., Bender, T. P. & Calabretta, B. Requirement of c-Myb for p210(BCR/ABL)-dependent transformation of hematopoietic progenitors and leukemogenesis. *Blood* 111, 4771–4779 (2008).

Brodeur, G. M., Seeger, R. C., Schwab, M., Varmus, H. E. & Bishop, J. M. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* 224, 1121–1124 (1984).

Weiss, W. A., Aldape, K., Mohapatra, G., Feuerstein, B. G. & Bishop, J. M. Targeted expression of MYCN causes neuroblastoma in transgenic mice. *EMBO J.* 16, 2985–2995 (1997).

Boon, K., Eberhart, C. G. & Riggins, G. J. Genomic amplification of orthodenticle homologue 2 in medulloblastomas. *Cancer Res.* 65, 703–707 (2005).

Di, C. et al. Identification of OTX2 as a medulloblastoma oncogene whose product can be targeted by all-trans retinoic acid. *Cancer Res.* 65, 919–924 (2005).

Barr, F. G. et al. Rearrangement of the PAX3 paired box gene in the paediatric solid tumour alveolar rhabdomyosarcoma. *Nature Genet.* 3, 113–117 (1993).

Keller, C. et al. Alveolar rhabdomyosarcomas in conditional Pax3:Fkhr mice: cooperativity of Ink4a/ARF and Trp53 loss of function. *Genes Dev.* 18, 2614–2626 (2004).

Chen, Z. et al. Fusion between a novel Kruppel-like zinc finger gene and the retinoic acid receptor-alpha locus due to a variant t(11;17) translocation associated with acute promyelocytic leukaemia. *EMBO J.* 12, 1161–1167 (1993).

He, L. Z. et al. Distinct interactions of PML-RARalpha and PLZF-RARalpha with co-repressors determine differential responses to RA in APL. *Nature Genet.* 18, 126–135 (1998).

Zhang, J., Kalkum, M., Yamamura, S., Chait, B. T. & Roeder, R. G. E protein silencing by the leukemogenic AML1-ETO fusion protein. *Science* 305, 1286–1289 (2004).

Yan, M. et al. A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. *Nature Med.* 12, 945–949 (2006).

Yang, G. et al. Transcriptional repression of the Neurofibromatosis-1 tumor suppressor by the t(8;21) fusion protein. *Mol. Cell. Biol.* 25, 5869–5879 (2005).

O'Neil, J., Shank, J., Cusson, N., Murre, C. & Kelliher, M. TAL1/SCL induces leukemia by inhibiting the transcriptional activity of E47/HEB. *Cancer Cell* 5, 587–596 (2004).

Jacobs, J. J. et al. Senescence bypass screen identifies TBX2, which represses Cdkn2a (p19ARF) and is amplified in a subset of human breast cancers. *Nature Genet.* 26, 291–299 (2000).

Kwei, K. A. et al. Genomic profiling identifies TITF1 as a lineage-specific oncogene amplified in lung cancer. *Oncogene* 27, 3635–3640 (2008).

Weir, B. A. et al. Characterizing the cancer genome in lung adenocarcinoma. *Nature* 450, 893–898 (2007).

Aoki, K., Tamai, Y., Horiike, S., Oshima, M. & Taketo, M. M. Colonic polyposis caused by mTOR-mediated chromosomal instability in *Apc*⁺/ Δ 716 *Cdx2*^{plusminus} compound mutant mice. *Nature Genet.* 35, 323–330 (2003).

Wicking, C. et al. CDX2, a human homologue of *Drosophila* caudal, is mutated in both alleles in a replication error positive colorectal cancer. *Oncogene* 17, 657–659 (1998).

Chawengsaksophak, K., James, R., Hammond, V. E., Kontgen, F. & Beck, F. Homeosis and intestinal tumours in *Cdx2* mutant mice. *Nature* 386, 84–87 (1997).

Bai, Y. Q., Miyake, S., Iwai, T. & Yuasa, Y. CDX2, a homeobox transcription factor, upregulates transcription of the p21/WAF1/CIP1 gene. *Oncogene* 22, 7942–7949 (2003).

Pabst, T. et al. Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. *Nature Genet.* 27, 263–270 (2001).

Kirstetter, P. et al. Modeling of C/EBPalpha mutant acute myeloid leukemia reveals a common expression signature of committed myeloid leukemia-initiating cells. *Cancer Cell* 13, 299–310 (2008).

Zuo, T. et al. FOXP3 is a novel transcriptional repressor for the breast cancer oncogene SKP2. *J. Clin. Invest.* 117, 3765–3773 (2007).

Zuo, T. et al. FOXP3 is an X-linked breast cancer suppressor gene and an important repressor of the HER-2/ErbB2 oncogene. *Cell* 129, 1275–1286 (2007).

Kouros-Mehr, H., Slorach, E. M., Sternlicht, M. D. & Werb, Z. GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* 127, 1041–1055 (2006).

Usary, J. et al. Mutation of GATA3 in human breast tumors. *Oncogene* 23, 7669–7678 (2004).

Pei, X. H. et al. CDK inhibitor p18INK4c is a downstream target of GATA3 and restrains mammary luminal progenitor cell proliferation and tumorigenesis. *Cancer Cell* 15, 389–401 (2009).

Raman, V. et al. Compromised HOXA5 function can limit p53 expression in human breast tumours. *Nature* 405, 974–978 (2000).

Garkavtsev, I. et al. The candidate tumour suppressor p33ING1 cooperates with p53 in cell growth control. *Nature* 391, 295–298 (1998).

Maestro, R. et al. Chromosome 13q deletion mapping in head and neck squamous cell carcinomas: identification of two distinct regions of preferential loss. *Cancer Res.* 56, 1146–1150 (1996).

Kichina, J. V. et al. Targeted disruption of the mouse *ing1* locus results in reduced body size, hypersensitivity to radiation and elevated incidence of lymphomas. *Oncogene* 25, 857–866 (2006).

Narla, G. et al. KLF6, a candidate tumor suppressor gene mutated in prostate cancer. *Science* 294, 2563–2566 (2001).

Kimmelman, A. C. et al. Suppression of glioblastoma tumorigenicity by the Kruppel-like transcription factor KLF6. *Oncogene* 23, 5077–5083 (2004).

Huang, X., Li, X. & Guo, B. KLF6 induces apoptosis in prostate cancer cells through up-regulation of ATF3. *J. Biol. Chem.* 283, 29795–29801 (2008).

Mueller, B. U. et al. Heterozygous PU.1 mutations are associated with acute myeloid leukemia. *Blood* 101, 2074 (2003).

Vangala, R. K. et al. The myeloid master regulator transcription factor PU.1 is inactivated by AML1-ETO in t(8;21) myeloid leukemia. *Blood* 101, 270–277 (2003).

Steidl, U. et al. Essential role of Jun family transcription factors in PU.1 knockdown-induced leukemic stem cells. *Nature Genet.* 38, 1269–1277 (2006).

Rosenbauer, F. et al. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nature Genet.* 36, 624–630 (2004).

Li, Q. L. et al. Causal relationship between the loss of RUNX3 expression and gastric cancer. *Cell* 109, 113–124 (2002).

Ito, K. et al. RUNX3 attenuates beta-catenin/T cell factors in intestinal tumorigenesis. *Cancer Cell* 14, 226–237 (2008).

Hahn, S. A. et al. DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* 271, 350–353 (1996).

Izeradjene, K. et al. KrasG12D and Smad4/Dpc4 haploinsufficiency cooperate to induce mucinous cystic neoplasms and invasive adenocarcinoma of the pancreas. *Cancer Cell* 11, 229–243 (2007).

Haber, D. A. et al. An internal deletion within an 11p13 zinc finger gene contributes to the development of Wilms' tumor. *Cell* 61, 1257–1269 (1990).

Yang, L., Han, Y., Suarez Saiz, F. & Minden, M. D. A tumor suppressor and oncogene: the WT1 story. *Leukemia* 21, 868–876 (2007).

Englert, C., Maheswaran, S., Garvin, A. J., Kreidberg, J. & Haber, D. A. Induction of p21 by the Wilms' tumor suppressor gene WT1. *Cancer Res.* 57, 1429–1434 (1997).

Bass, A. J. et al. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nature Genet.* 4 Oct 2009 (doi: 10.1038/ng.465).

Author affiliations

The Smurfit Institute of Genetics, Trinity College Dublin and The Adelaide & Meath Hospital, including the National Children's Hospital, Dublin, Ireland.

Email: adrian.bracken@tcd.ie

Biotech Research and Innovation Centre (BRIC) and Centre for Epigenetics, University of Copenhagen, Copenhagen 22009, Denmark.

Email: kristian.helin@bric.ku.dk