Emerging themes in cohesin cancer biology

Todd Waldman

Abstract | Mutations of the cohesin complex in human cancer were first discovered ~10 years ago. Since then, researchers worldwide have demonstrated that cohesin is among the most commonly mutated protein complexes in cancer. Inactivating mutations in genes encoding cohesin subunits are common in bladder cancers, paediatric sarcomas, leukaemias, brain tumours and other cancer types. Also in those 10 years, the prevailing view of the functions of cohesin in cell biology has undergone a revolutionary transformation. Initially, the predominant view of cohesin was as a ring that encircled and cohered replicated chromosomes until its cleavage triggered the metaphase-to-anaphase transition. As such, early studies focused on the role of tumour-derived cohesin mutations in the fidelity of chromosome segregation and aneuploidy. However, over the past 5 years the cohesin field has shifted dramatically, and research now focuses on the primary role of cohesin in generating, maintaining and regulating the intra-chromosomal DNA looping events that modulate 3D genome organization and gene expression. This Review focuses on recent discoveries in the cohesin field that provide insight into the role of cohesin inactivation in cancer pathogenesis, and opportunities for exploiting these findings for the clinical benefit of patients with cohesin-mutant cancers.

Even as the structure and function of sister chromatids was elucidated during the late twentieth century, the composition of the 'molecular glue' responsible for cohering them remained unknown. In 1997, this mystery was finally solved when several groups identified this molecular glue as a protein complex comprising components of the structural maintenance of chromosomes (SMC) family of chromatin proteins[1–3]. Yeast cells harbouring mutations in genes encoding these proteins displayed loss of cohesion, visualized as an increased physical distance between sister chromatids in metaphase (FIG. 1a). The discovery of this highly conserved protein complex, now universally referred to as 'cohesin', was a significant moment in modern cell biology because it provided substantial mechanistic insight into a key aspect of mitosis. Subsequent work by laboratories worldwide established what is now known as the 'canonical' model of cohesin structure and function — that cohesin is a multiprotein ring complex (FIG. 1b) that encircles replicated chromosomes until cleavage of its double-strand-break repair protein rad21 homologue (RAD21) subunit by the protease Separase opens the ring, releasing the cohered sister chromatids and triggering the metaphase-to-anaphase transition (reviewed in REFS[4,5]) (FIG. 1c).

Subsequently, in a series of discoveries from 2008 to 2011, researchers found that somatic inactivating mutations of genes encoding cohesin subunits were a major cause of human cancer. The first cohesin mutations to be discovered were in the genes encoding SMC1A, SMC3 and nipped-B-like protein (NIPBL)[6]. However, because these early mutations were heterozygous missense mutations of unknown functional significance, it was unclear at the time whether they were cancer-causing mutations or passenger mutations. This uncertainty was resolved when homozygous deletions of cohesin genes were discovered in human leukaemias[7], and when we discovered truncating mutations of the gene encoding the cohesin subunit SA-2 (STAG2) in a wide range of cancers[8]. Because the *STAG2* gene is on the X chromosome, there was no ambiguity about the functional significance of the *STAG2* mutations, as most resulted in truncation of a protein encoded by a functionally single-copy gene. Subsequent cancer genome sequencing efforts demonstrated that mutations of genes encoding cohesin subunits (most commonly truncating mutations in *STAG2*, but also heterozygous missense mutations in *RAD21*, *SMC1A*, *SMC3* and *NIPBL*) were present in bladder cancer, Ewing sarcoma, myeloid leukaemia, brain tumours, endometrial cancers and other tumour types (TABLE 1 and references therein). Of note, the missense mutations in *RAD21*, *SMC1A*, *SMC3* and *NIPBL* were mutually exclusive with truncating mutations in *STAG2*. Despite the fact that *STAG2* and *SMC1A*

Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University School of Medicine, Washington, DC, USA.

e-mail: waldmant@georgetown.edu

https://doi.org/10.1038/s41568-020-0270-1

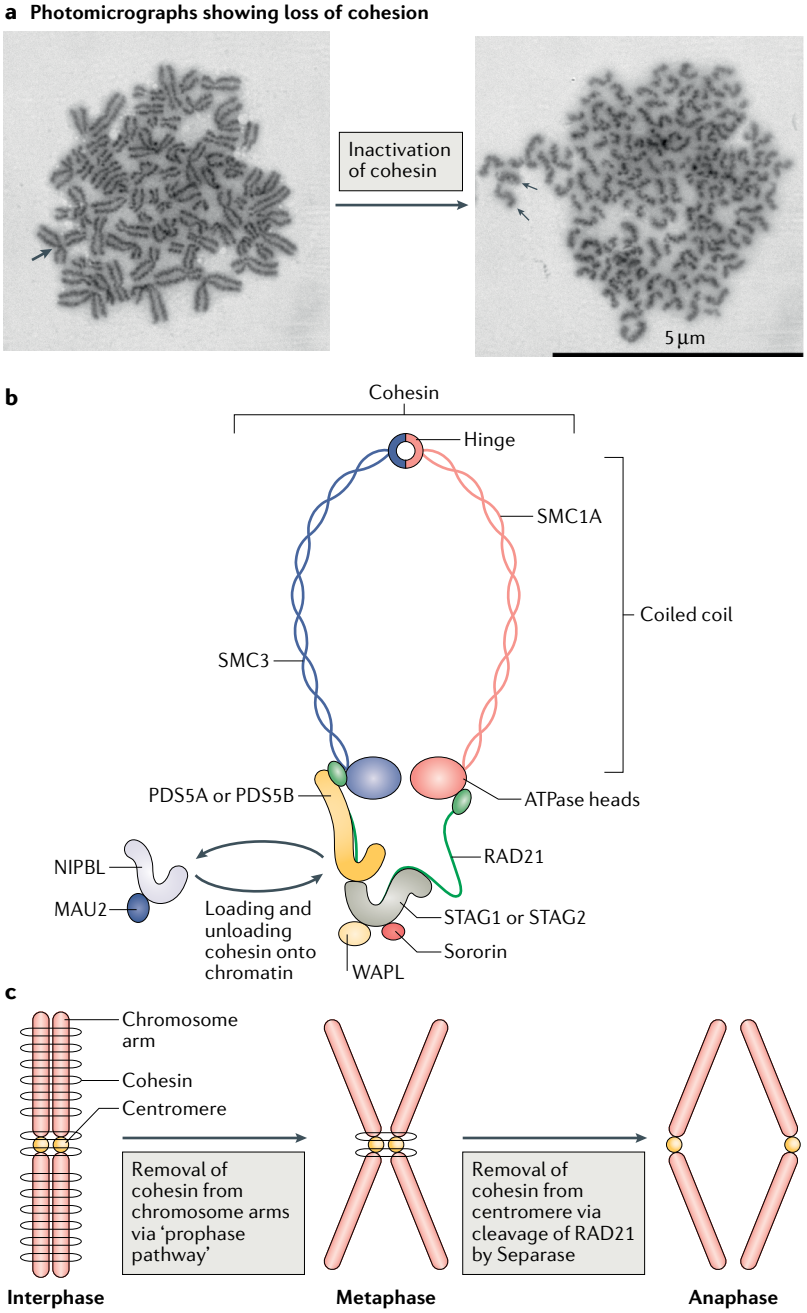

Fig. 1 | **Canonical model of cohesin action. a** | Inactivation of cohesin results in the inability of cells to appropriately cohere replicated DNA throughout the cell cycle. This is most prominent and easily visualized in metaphase. **b** | In vertebrate somatic cells, the cohesin complex consists of four core subunits: structural maintenance of chromosomes protein 1A (SMC1A), SMC3, double-strand-break repair protein rad21 homologue (RAD21) and either cohesin subunit SA-1 (STAG1) or SA-2 (STAG2). Several additional components serve primarily to regulate the core cohesin complex, including nipped-B-like protein (NIPBL) and MAU2 chromatid cohesion factor homologue (MAU2), which are required for loading cohesin onto chromatin; wings apart-like protein homologue (WAPL), sister chromatid cohesion protein PDS homologue A (PDS5A) and PDS5B, which are required for unloading of cohesin from chromatin; and sororin, which is involved in the establishment of sister chromatid cohesion. **c** | Cohesin is initially loaded onto chromatin in late G1 phase, and coheres replicated chromosomes until it is removed from chromosomal arms in prophase via polo-like kinase 1 (PLK1)-dependent phosphorylation of the STAG1 or STAG2 subunit, giving metaphase chromosomes their classic X-shape. Centromeric cohesin is then cleaved by Separase, allowing sister chromatids to snap apart at the metaphase-to-anaphase transition, resulting in separation of chromatids to daughter cells. Part **a** adapted with permission from REF.[8], AAAS. Part **b** adapted with permission from REF.[104], Elsevier.

are on the X chromosome, there is no known gender bias to inactivating mutations in these genes, as the X chromosome is single copy in males and functionally single copy in females due to X-inactivation.

This discovery that cohesin genes were tumour suppressors was unexpected and counter-intuitive, for at least two reasons. First, cohesin was well established to be a key component of the basic cellular machinery that controls mitosis[1–3]. Inactivation of such a key component of the cell cycle machinery was expected to be incompatible with cellular proliferation. This may help explain why many tumour-derived mutations of cohesin subunit genes are heterozygous missense mutations that inactivate some, but not all, functions of the encoded protein. Second, when the first somatic mutations of genes encoding cohesin subunits were discovered in human cancers, it was already known that inherited heterozygous mutations of the *NIPBL* gene were responsible for an inherited neurodevelopmental syndrome known as Cornelia de Lange syndrome[9,10], which is not known to be a cancer predisposition syndrome. This phenotypic difference between inherited and somatic mutations in genes encoding cohesin subunits was unexpected because it disobeyed Knudsen's two-hit hypothesis, which predicts that inherited mutations of a tumour suppressor gene will cause a cancer predisposition syndrome. However, we now know that inherited mutations of many other tumour suppressor genes encoding chromatin proteins similarly disobey Knudsen's hypothesis (BOX 1).

In this Review, I focus on emerging themes in cohesin cell and cancer biology that help explain the seemingly paradoxical existence of these mutations in cancer cells. I also specifically emphasize recent data that provide insights into which of the known functions of cohesin appear to be inactivated by tumour-derived mutations in cohesin genes. Finally, I review recent studies identifying potential areas of translational relevance for cancer diagnosis and treatment.

Cohesin gene mutations and aneuploidy

The canonical model of cohesin action predicts that inactivating cohesin subunits will cause aberrant segregation of sister chromatids, chromosomal instability and aneuploidy. Prior to the discovery of cohesin mutations in cancer, this prediction had been tested and confirmed by inactivating several cohesin subunits (for example, STAG1, wings apart-like protein homologue (WAPL)) in model organisms, including yeast, mice, *Arabidopsis* and others[11–15]. Furthermore, knockout of the *Smc1b* gene had been shown to cause aneuploidy in mouse oocytes[16,17]. Therefore, the initial discovery of mutations in genes encoding cohesin subunits was particularly exciting because it suggested a possible cause of aneuploidy in human cancer.

Early functional studies supported a role for mutations in cohesin genes in causing aneuploidy in human tumours. We showed that correction of endogenous mutant alleles of *STAG2* in human brain cancer cells led to restoration of sister chromatid cohesion and reduction in chromosome counts[8]. Additionally, others showed that *STAG2* depletion in HeLa cells directly undermined the fidelity of chromosome segregation, leading

Chromosomal instability
The condition when cells have an abnormally high rate of mis-segregation of replicated chromosomes to daughter cells in mitosis. Results in aneuploidy.

Aneuploidy
The presence of an abnormal number of chromosomes in a cell, generally due to aberrant segregation of replicated chromosomes to daughter cells in mitosis.

Chromatin immunoprecipitation-sequencing
A technique that combines chromatin immunoprecipitation (ChIP) with next-generation DNA sequencing to comprehensively identify, in an unbiased way, all of the genomic binding sites of chromatin-associated proteins. Often referred to as ChIP-seq.

Haematopoietic stem and progenitor cells
(HSPCs). Multipotent, self-renewing adult stem cells that give rise to all types of differentiated blood cells in the lymphoid and myeloid lineages. HSPCs are found primarily in the bone marrow of adults, but are also found in umbilical cord blood and in peripheral blood.

to whole-chromosome instability[18]. Furthermore, depletion of *STAG2* led to aneuploidy in karyotypically normal SV40-transformed human ureteral epithelial cells[19]. Together, these data suggested that mutations of *STAG2* caused aneuploidy in human cancer cells.

However, as genome sequencing studies identified even more frequent mutations in cohesin subunits in a wide range of cancers (details in TABLE 1), the link to aneuploidy grew increasingly tenuous. For example, most non-muscle invasive bladder cancers (NMIBCs) harbouring mutations of *STAG2* were euploid[20]. Similarly, mutations in the genes encoding cohesin subunits STAG2, RAD21, SMC1A and SMC3 were common in myelodysplastic syndrome, a group of cancers that are among the most benign myeloid neoplasms and are therefore mostly euploid[21]. Also, many myeloid leukaemias harbouring mutations of cohesin genes were euploid, and there was no correlation between the presence of a cohesin gene mutation and aneuploidy. However, the studies also showed that myeloid leukaemias in patients with Down syndrome, a constitutional aneuploidy, harbour the most frequent cohesin gene mutations of any cancer type[22]. Other studies demonstrated that Ewing sarcomas harbouring mutations of *STAG2* were often euploid, and that there was no correlation between the presence of a *STAG2* mutation and aneuploidy[23–25]. Finally, there is little evidence for alterations in sister chromatid cohesion, chromosomal instability or aneuploidy in cells from patients with Cornelia de Lange syndrome harbouring inherited mutations of *NIPBL* and other cohesin subunits[26].

In an effort to reconcile these findings with our earlier work implicating *STAG2* mutations in aneuploidy, we used gene editing to create isogenic sets of human cells harbouring nine different tumour-derived mutations of *STAG2* (REF.[27]). Only one of the nine mutations tested induced overt alterations in chromosome counts. Perhaps the most surprising finding was that tumour-derived *STAG2* missense mutations retained their ability to enforce sister chromatid cohesion. This demonstrated that tumour-derived mutations in *STAG2* do not uniformly inactivate cohesin's ability to enforce sister chromatid cohesion. Taken together, these studies indicated that neither loss of sister chromatid cohesion nor the initiation of chromosomal instability and aneuploidy is likely to be the primary phenotypic outcome of cohesin gene mutations in human cancer.

Table 1 | **Frequency of cohesin gene mutations in human cancer**

Cancer type	Cohesin mutation frequency (%)	Clinical correlations	Most commonly mutated subunits	Refs
Bladder cancer	15–40[a]	*STAG2*-mutant non-muscle invasive bladder cancers have a favourable prognosis	~50% are in *STAG2*; remaining are distributed among many other cohesin subunits	20,72–76
Ewing sarcoma	16–22	*STAG2*-mutant tumours have worse clinical outcomes	All mutations are in *STAG2*	23–25
Myeloid leukaemia	5–53[b]	Unresolved	~50% are in *STAG2*; remaining are distributed among many other cohesin subunits	21,22, 45,77–79, 105–109
Endometrial cancer	19	None	~50% *STAG2*, ~50% *NIPBL*	110
Glioblastoma	7	None	~50% are in *STAG2*; remaining are distributed among other many other cohesin subunits	111,112

NIPBL, nipped-B-like protein; STAG2. cohesin subunit SA-2. [a]In non-muscle invasive bladder cancer the mutation frequency is 35–40%, whereas in muscle invasive bladder cancer the mutation frequency is ~10%. [b]In most types of myeloid leukaemia, the mutation frequency is ~5–20%. However, in Down syndrome-associated leukaemia, the mutation frequency is >50%.

Chromatin structure and gene expression

Cohesin gene mutations and the interaction of cohesin with chromatin. In addition to its well-defined role in sister chromatid cohesion, cohesin plays a vital role in regulating the structure and function of chromatin. Exhaustive chromatin immunoprecipitation-sequencing analysis performed by the ENCODE (Encyclopedia of DNA Elements) Consortium and other groups worldwide has demonstrated that cohesin binds to chromatin at thousands of distinct sites whose locations depend in part on the cell type and the differentiation state[28,29]. This discovery that cohesin binds to chromatin at discrete sites was one of the first clues that cohesin played roles in cell biology other than enforcing sister chromatid cohesion, because, as pointed out previously, there would be no obvious need for cohesin to interact with chromatin at discrete, regulated sites if its role was simply to cohere replicated chromosomes[30]. Cohesin is recruited to these chromatin binding sites by CCCTC-binding factor (CTCF) and by the presence of active transcription[31].

This underlying biology raised the fundamental question of whether tumour-derived mutations of individual cohesin subunits alter the interaction of the entire cohesin complex with chromatin. The first evidence for this showed that levels of cohesin proteins were substantially reduced in bulk chromatin fractions from human leukaemia cells harbouring mutations in the genes encoding cohesin subunits STAG2, RAD21, SMC3, NIPBL and STAG1 (REF.[21]). This finding suggested that tumour-derived mutations of cohesin genes reduced the ability of cohesin to bind to chromatin. Several more recent chromatin immunoprecipitation-sequencing studies further addressed this issue at a more granular level. Chromatin immunoprecipitation-sequencing performed in haematopoietic stem and progenitor cells (HSPCs) derived from wild-type and myeloid-specific *Stag2* knockout mice showed that the presence or absence of STAG2 had no discernible effect on the interaction of cohesin or CTCF with chromatin[32]. In contrast, STAG2 was required for cohesin to be recruited to a subset of its chromatin binding sites in human colon cancer cells that had been engineered with a STAG2 auxin-inducible degron (AID)[33]. These differences are potentially explained by differences in the cell systems studied (pluripotent mouse HSPCs versus non-differentiating human cancer cells) and by differences in the kinetics and efficiency of STAG2 inactivation by knockout and AID. Given these conflicting data, whether STAG2 is required for the interaction of cohesin with chromatin is, at present, unresolved.

Box 1 | **Inherited mutations in genes encoding cohesin cause neurodevelopmental disorders, not cancer**

Knudsen's two-hit hypothesis predicts that inherited heterozygous mutations of a tumour suppressor gene will cause a cancer predisposition syndrome. However, cohesin tumour suppressor genes violate Knudsen's hypothesis — instead of causing cancer predisposition syndromes, inherited mutations of cohesin genes cause a range of phenotypically related paediatric neurodevelopmental disorders. The first of these to be discovered was Cornelia de Lange syndrome, which is most often caused by inherited mutations of the gene encoding the nipped-B-like protein (NIPBL) subunit of cohesin[9,10]. Soon thereafter, other developmental disorders phenotypically similar to Cornelia de Lange syndrome were found to be caused by inherited mutations of genes encoding other components of cohesin, including structural maintenance of chromosomes protein 1A (SMC1A), SMC3, double-strand-break repair protein rad21 homologue (RAD21) and others[113–116]. These phenotypically similar disorders are now collectively referred to as 'cohesinopathies'[117].

It has recently become increasingly clear that other tumour suppressor genes encoding chromatin proteins similarly violate Knudsen's hypothesis, in that somatic mutations cause cancer but inherited mutations cause neurodevelopmental disorders[118]. One of the most striking examples is the *ARID1A* (AT-rich interaction domain 1A) gene, which encodes a component of the SWI/SNF chromatin remodelling complex. *ARID1A* is one of the most commonly mutated tumour suppressor genes in human cancer[119]. However, inherited mutations of *ARID1A* do not cause a cancer predisposition syndrome; instead, they cause Coffin–Siris syndrome, a paediatric neurodevelopmental disorder[120]. As whole-genome sequencing has been increasingly applied to rare developmental disorders, the number of chromatin proteins for which somatic mutations cause cancer and inherited mutations cause neurodevelopmental disorders has exploded[121]. In some cases, the identical mutation has been shown to cause cancer when somatically mutated and a neurodevelopmental disorder when mutated in the germline[122]. This discrepancy suggests that the role of cohesin and other epigenetic chromatin regulatory proteins in cancer pathogenesis is related to a role in cellular differentiation and organismal development. It is tempting to speculate that as the genetics of cohesin mutations so closely resembles that of the SWI/SNF chromatin remodelling complex, these two complexes may have even more biochemical and functional similarities than are currently appreciated[123]. The basis of these intriguing departures from Knudsen's two-hit hypothesis is likely to provide important future insight into the function of cohesin and the role of cohesin mutations in the pathogenesis of cancer.

Effect of cohesin gene mutations on 3D genome organization. The study of the 3D structure of chromosomes and their relative positioning in the nucleus has recently been revolutionized by the development of a next-generation sequencing technique known as Hi-C (high-throughput chromosome conformation capture; described in BOX 2). In 2017, Hi-C studies revealed that cohesin controls key aspects of 3D genome organization[30,34–36]. Most prominently, inactivation of cohesin subunits RAD21 and NIPBL led to the immediate collapse of topologically associating domains (TADs)[30,34–36]. This role for cohesin in maintaining TAD structure is likely due to its recently demonstrated biochemical function as a catalytic engine for DNA loop extrusion[37,38]. Inactivation of RAD21 and NIPBL also strengthened A/B compartment structure, in that the 'A' and 'B' compartments could be bioinformatically detected and defined more easily in cohesin-inactivated cells than in wild-type cells.

However, these discoveries were made after inactivating components of cohesin that were comparatively infrequently mutated in cancer. To test whether these alterations in 3D genome organization might drive neoplastic transformation, several groups tested whether inactivation of STAG2 led to alterations in TADs and/or compartment structure. Hi-C on transformed human breast epithelial cells with small interfering RNA (siRNA)-mediated *STAG2* depletion showed that TADs did not collapse[39]. Instead, the number of TADs decreased by ~20% and their borders became less well defined. There was no effect of *STAG2* depletion on compartment structure. In a second study, Hi-C on mouse embryonic stem cells with siRNA-mediated knockdown of *Stag2* showed that neither TADs nor A/B compartment structure was substantially altered by depleting *STAG2* (REF.[40]). When STAG2 was inactivated using an AID, most TADs were unaffected and there were no overt changes in compartment structure[33]. Of the minority of TADs that did change, most did not disappear completely but, instead, had alterations in their boundaries. There were no overt changes in compartment structure after acute inactivation of STAG2. Finally, Viny et al. performed Hi-C on *STAG2* knockout mouse HSPCs and found no differences in TADs or compartment structure after deleting *STAG2* (REF.[32]). These four studies demonstrated that STAG2 inactivation had either no effect or a relatively minor effect on TADs and compartments in mammalian cells, depending on the cell type. These findings raised the question of why STAG2 is apparently dispensable for maintenance of TADs, whereas other components of cohesin (for example, RAD21, NIPBL, WAPL) are required. The most likely explanation is that, unlike all other components of cohesin, STAG2 has a homologue (STAG1) that can functionally compensate for its role in generating and maintaining TAD structure.

In contrast to its dispensability for maintenance of TADs and compartments, emerging evidence suggests that inactivation of STAG2 can alter the structure of individual chromatin loops (FIG. 2). For example, siRNA-mediated depletion of *Stag2* in mouse embryonic stem cells reduced the short-range chromatin interactions that likely correspond to enhancer–promoter contacts[40]. Moreover, siRNA-mediated depletion of *STAG2* in cultured human cells resulted in the generation of new, longer, chromatin loops, suggesting that STAG1-containing cohesin complexes (which substitute for STAG2-cohesin in STAG2-inactivated cells) were more efficient at chromatin loop extrusion than STAG2-containing cohesin complexes[41]. Other studies have shown that cohesin modulates the number of Polycomb group (PcG)-marked chromatin loops (for additional details, see 'Emerging future directions'). Of note, identifying and studying individual chromatin loops requires extremely deep sequencing of Hi-C libraries, as well as advanced bioinformatics approaches

Auxin-inducible degron
A 68-amino-acid tag that, when added to an endogenous protein via gene editing, makes it possible to rapidly and completely degrade the tagged protein by adding auxin to the culture medium.

Hi-C
(High-throughput chromosome conformation capture). A next-generation DNA sequencing-based technique that makes it possible to comprehensively identify, in an unbiased way, regions of the genome that tend to co-localize in the 3D space comprising the interior of the nucleus.

3D genome organization
The 3D structure of chromosomes and their relative positioning in the nucleus. Sometimes also referred to as nuclear organization.

that are just now being developed (for example, see preliminary results in REF.[42]). Therefore, the effect of tumour-derived cohesin mutations on the formation, maintenance and regulation of individual chromatin loops is a particularly active area of current investigation.

Finally, studies focusing on CTCF have further implicated 3D genome organization as a mechanism of tumour suppression. CTCF is a sequence-specific DNA binding protein that interacts with cohesin[43] to create and define TADs (reviewed in REF.[44]) (BOX 2). *CTCF* is a tumour suppressor gene, mutated in myeloid leukaemias, lymphoblastic leukaemias, endometrial cancers and other tumour types[22,45,46]. Consistent with its role as a tumour suppressor, *Ctcf* knockout mice are predisposed to a wide range of cancers[47]. In addition, bioinformatics analyses of cancer genomes have revealed that CTCF chromatin binding sites are altered by mutation more frequently than would be expected by chance alone, and that these mutations can result in alterations to TAD boundaries[48–50]. Together, these data further implicate regulation of 3D genome organization as a key to cohesin's tumour-suppressing activities.

Effect of cohesin gene mutations on global gene expression. The interaction of cohesin with its chromatin binding sites is thought to play an important role in regulating gene expression by generating TADs, which increase the local concentration of contiguous stretches of chromatin and help bring enhancers into proximity to the promoters they regulate[51]. If cohesin controls regulated gene expression in this way, it might make sense that tumour-derived mutations of genes encoding cohesin components would lead to alterations in gene expression that drive tumorigenesis.

To test this hypothesis, several groups have performed gene expression analyses on isogenic sets of cells that differ only in the presence or absence of STAG2 (REFS[8,21,32,33,39,40]). These experiments were performed in human leukaemia and brain, breast and colon cancer cells, mouse haematopoietic precursor cells and mouse embryonic stem cells using ectopic expression, gene editing, siRNA and AID. Considering the widely accepted view that cohesin is a direct regulator of gene expression, surprisingly few STAG2-regulated genes were identified in these studies. Furthermore, the magnitude of the reported gene expression changes in response to STAG2 inactivation was low, in most cases being less than twofold.

These relatively modest changes raised the fundamental question of whether these expression changes are attributable to a direct role for STAG2 in transcriptional regulation, or are instead an indirect effect of alterations in cell biology (for example, cellular differentiation) caused by STAG2 inactivation. The generally low magnitude of these expression changes tends to argue against a direct role for STAG2 in transcriptional regulation, as important transcriptionally regulated effector genes would be expected to be robustly regulated by STAG2. Furthermore, if STAG2-regulated genes were direct effectors of STAG2 tumour suppression, a subset of the most important transcriptionally regulated genes would be expected to occur in common between the different experimental systems studied. However, this has not so far been the case. As one example, no STAG2-regulated genes were found in common between two distinct human glioblastoma cancer cell lines harbouring endogenous mutations of *STAG2* (REF.[8]). As such, at present, it appears most likely that the surprisingly modest changes in gene expression caused by loss of STAG2 are indirect effects of STAG2 inactivation rather than a primary, causative mechanism for cohesin-mediated tumour suppression. However, if genes are ultimately identified that are robustly regulated by STAG2 and are shared between experimental systems, this conclusion could change.

Cohesin, stemness and differentiation

Since the initial discovery of cohesin gene mutations in cancer, researchers have struggled to identify phenotypes of cohesin inactivation in cancer-relevant experimental systems. Early data suggesting a role for cohesin inactivation in aneuploidy were convincingly challenged

Box 2 | **Principles of 3D genome organization**

Chromosomes are packed into the mammalian nucleus in a non-random, multilayered and highly dynamic fashion. At the molecular level, chromosomes are wrapped around histones to form nucleosomes. At a higher level of genomic organization, whole chromosomes are known to reside in specific regions of the nucleus known as 'chromosome territories'. However, defining additional types of genomic organization that exist between nucleosomes and chromosome territories had been limited by a lack of technology.

This all changed in 2009 with the development of a technique known as Hi-C (high-throughput chromosome conformation capture), which made it possible to comprehensively identify, in an unbiased way, all of the regions of DNA that tend to co-localize in the 3D space comprising the interior of the nucleus[124–126]. To perform Hi-C, cells are first treated with formaldehyde to form covalent cross-links between non-contiguous segments of DNA that are in proximity to each other in 3D space. Next, a series of enzymatic manipulations are performed to create a next-generation sequencing library in which these nearby segments of DNA are ligated together. In this way, a single sequencing read can identify two non-contiguous segments of DNA that were in proximity to each other in the 3D space of the nucleus prior to the addition of cross-linker. After sequencing the library to a depth of hundreds of millions (or billions) of reads, advanced bioinformatics is performed to construct a 3D model of the location of all segments of DNA in the nucleus in relationship to each other.

This new Hi-C approach has led to the discovery of two new types of genomic organization (reviewed in REF.[127]). Topologically associating domains (TADs) are regions of contiguous chromatin of ~100 kb–2 Mb that self-interact more frequently than they interact with other regions of chromatin on the same chromosome[128]. TAD boundaries are generally defined by the presence of CCCTC-binding factor (CTCF) and cohesin, and are thought to be largely conserved between cell types and to be stable during cellular differentiation. Much research is currently underway to determine the mechanism for generation of TADs. Currently, the most popular model is the 'loop extrusion model', which states that cohesin encircles chromatin in *cis*, creating a loop that is then actively extruded through the cohesin ring until it is halted by CTCF barriers, which then define the base of the TAD (reviewed in REF.[129]). TADs appear to play a major role in transcriptional regulation by providing a spatially restricted domain for bringing enhancers into proximity to the promoters they regulate.

Chromosome A/B compartments are self-interacting domains that are much larger than TADs (>2 Mb). 'A' compartments tend to comprise open, transcriptionally active chromatin, whereas 'B' compartments tend to comprise closed, transcriptionally inactive chromatin. Compartments are thought to be more dynamic than TADs; one study reported that 36% of the genome undergoes compartment switching during differentiation of embryonic stem cells[130]. Cohesin appears to be involved in maintaining the pliability of A/B compartment structure; cells lacking cohesin have more well-defined compartments than cells with intact cohesin[30,34–36]. The development of Hi-C and its use to define TADs and A/B compartments is one of the major breakthroughs in the biological sciences in the past decade.

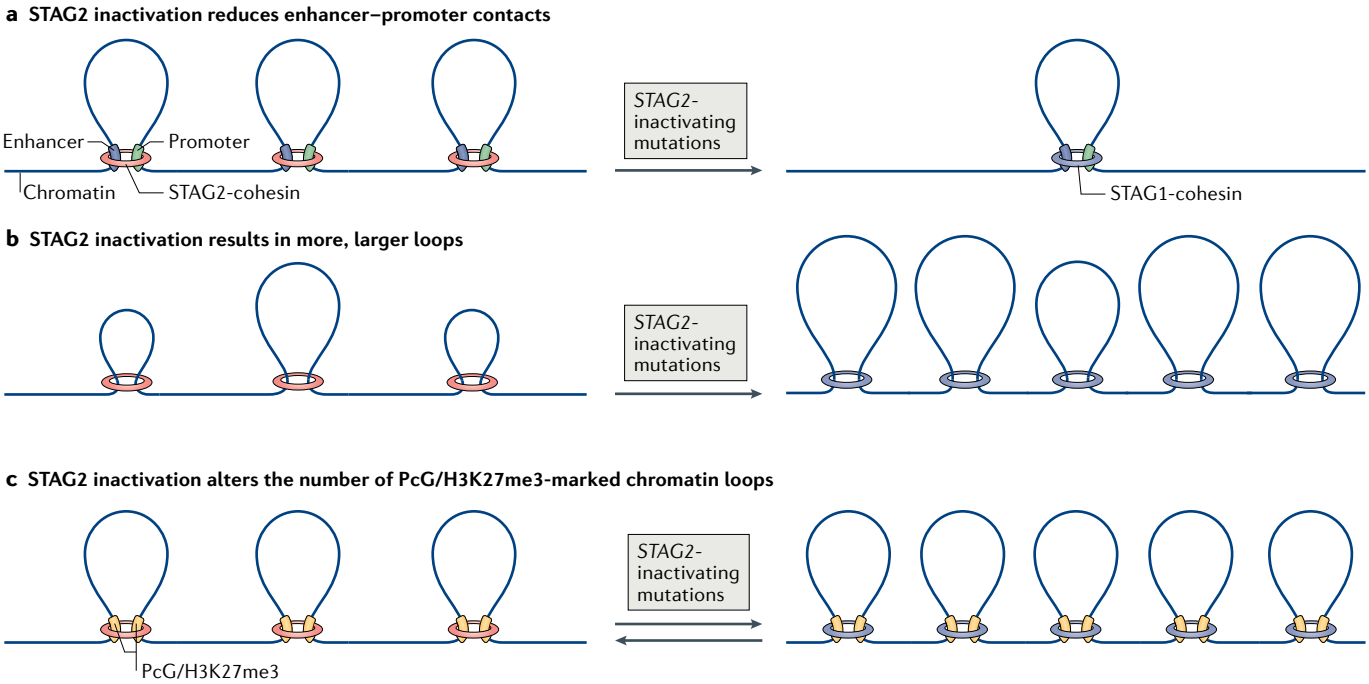

Fig. 2 | **Current models of the effects of STAG2 inactivation on individual chromatin loops.** Inactivation of many individual cohesin subunits results in the complete collapse of topologically associating domain structure. However, inactivation of cohesin subunit SA-2 (STAG2) has much more subtle effects on 3D genome organization, likely because STAG2 has a functionally compensating homologue (STAG1). These subtle effects of STAG2 inactivation appear to be primarily at the level of individual chromatin loops. Recent studies have suggested several different models for the effects of STAG2 inactivation on chromatin loops. **a** | Cuadrado et al. have shown that inactivation of STAG2 can result in a reduction in short-range chromatin interactions that correspond to enhancer–promoter contacts[40]. **b** | Wutz et al. have demonstrated that human cells with inactivation of STAG2 have more chromatin loops, and larger chromatin loops[41]. **c** | Several groups have demonstrated that cohesin inactivation can lead to alterations in Polycomb group (PcG)-marked chromatin loops, which are defined by the histone H3 K27 trimethylation (H3K27me3) chromatin mark. However, there is disagreement in the literature as to whether cohesin inactivation results in an increase[101,102] or decrease[40,103] in H3K27me3-marked loops.

by subsequent studies, as described herein. Furthermore, inactivating mutations or depletion of STAG2 in cultured untransformed epithelial cells and in cancer cells did not produce a proliferation advantage; instead, it led to slowed proliferation and cell cycle arrest, the opposite of what would be expected for inactivation of a wild-type tumour suppressor gene[27,52].

However, beginning in 2015, work from several laboratories began to define a role for mutated cohesin subunits in stemness and differentiation in HSPCs[53–55]. In particular, these groups showed that inactivation of STAG2, RAD21, SMC1A and SMC3 in human and mouse HSPCs led to dramatically enhanced replating in vitro, a key measure of stem cell self-renewal (FIG. 3). This finding was particularly important because it provided the first incontrovertible evidence for a proliferation advantage in cells harbouring inactivating mutations of cohesin subunits. The studies also generated various tissue-specific *STAG2*, *RAD21*, *SMC1A* or *SMC3* knockout and knockdown mice. They showed that inactivation of cohesin subunits in vivo led to alterations in myeloid and erythroid differentiation (FIG. 3), leading to myeloproliferative disorders that mimicked early human neoplasia[53,55]. When *Smc3* knockout mice were crossed with mice harbouring a second mutated cancer gene (*Flt3*), cohesin inactivation led to aggressive leukaemia[55].

More recent studies have confirmed these early findings and extended them in important ways. Using an unbiased genome-wide short hairpin RNA screen, Galeev et al. also demonstrated a role for cohesin subunits STAG2, RAD21, STAG1 and SMC3 in stemness and differentiation in human HSPCs[56]. Because they used an unbiased screen, the authors were able to further conclude that cohesin was among the most important of all protein complexes in regulating stemness and differentiation in HSPCs. Tothova et al. used CRISPR to introduce inactivating mutations into the genes encoding multiple components of cohesin in human HSPCs[57]. They showed that whereas heterozygous inactivation of cohesin resulted in cellular transformation, homozygous inactivation was lethal, providing an explanation for the haploinsufficiency of most cohesin mutations in human cancer. Finally, Cuartero et al. have demonstrated that cohesin is required for induced, tissue-specific gene expression programmes in myeloid-derived cells[58].

These findings raised the question of what the mechanisms are for these effects of cohesin inactivation on stemness and differentiation. In many of these studies, cohesin inactivation was accompanied by alterations in chromatin accessibility, 3D genome organization and gene expression[53–57,59]. However, it has been difficult to determine causation — that is, which (if any)

Haploinsufficiency
One of the two alleles of a gene is inactivated by mutation, resulting in a pathology such as cancer. Generally used to refer to tumour suppressor genes in which inactivation of one allele produces cancer, whereas inactivation of both alleles is lethal to the cell.

Biomarkers
Measurable biological substances, such as DNA, RNA or protein, that provide predictive information about a patient's likely clinical outcome.

Synthetic lethality
The simultaneous inactivation of two gene products results in cell death, whereas inactivation of only one of the gene products does not. When applied to anticancer drug discovery, one of the inactivated proteins is encoded by a mutated tumour suppressor gene and the other protein is inactivated via pharmacological inhibition.

of these are causes of the stemness and differentiation phenotype and which are effects. One intriguing recent study demonstrated that ectopic re-expression of a potentially important STAG2-regulated gene (*Ebf1*) rescued the alterations in stemness and differentiation caused by deletion of *Stag2* (REF.[32]). However, whether cohesin-mediated changes in 3D genome organization, chromatin accessibility and/or gene expression are required for the demonstrated alterations in stemness and differentiation remains fundamentally unresolved.

Several other hypotheses have been put forth to explain the mechanism through which cohesin regulates stemness and differentiation. For example, the role of cohesin in differentiation may be secondary to a more primary effect on DNA replication, suggesting a mechanistic connection between DNA replication, stemness and differentiation[60]. Others have suggested that cohesin is required for activating key inflammation-related gene expression programmes during HSPC differentiation[58,61], and that cohesin inactivation drives the epithelial-to-mesenchymal transition[62]. An important area for future research will be to determine the effect of cohesin inactivation on well-defined differentiation programmes in cultured embryonic stem cells and/or induced pluripotent stem cells. It will also be informative to determine whether cohesin inactivation alters the stemness and differentiation profiles of the progenitor cells for relevant solid tumours, such as bladder cancer, Ewing sarcoma and glioblastoma.

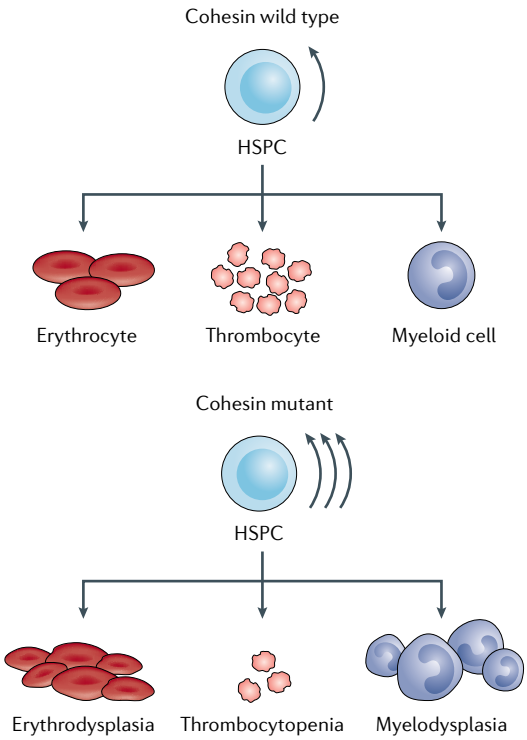

Fig. 3 | **Functions of cohesin in haematopoietic stem cell self-renewal and differentiation.** After it became clear that mutations of cohesin subunits were important drivers of myeloid leukaemias, several groups tested whether inactivation of cohesin subunits in haematopoietic stem cells (often referred to as haematopoietic stem and precursor cells (HSPCs)) would result in alterations to their self-renewal and/or differentiation potential. Despite the fact that these studies were performed using both mouse and human HSPCs and different approaches for inactivation, and targeted various components of cohesin (structural maintenance of chromosomes protein 1A (SMC1A), SMC3, cohesin subunit SA-1 (STAG1), STAG2, double-strand-break repair protein rad21 homologue (RAD21)), the conclusions were remarkably consistent. Inactivation of cohesin subunits resulted in dramatically enhanced self-renewal of HSPCs that preserved their immature state, leading to myeloproliferative disorders in mice. Inactivation of cohesin in HSPCs also resulted in alterations in the differentiation programmes of both red and white blood cells, resulting in phenotypes including red cell dysplasia, thrombocytopenia and myelodysplasia.

Clinical utility of cohesin mutations

In addition to understanding the mechanisms of cohesin tumour suppression, a major focus of cohesin cancer research has been to translate knowledge of these mutations into clinical benefit for patients with cohesin-mutant cancers. These studies have fallen into two main categories: identifying targets for therapeutic inhibition in cohesin-mutant cancers; and developing prognostic biomarkers based on the cohesin gene mutation status.

Synthetic lethality of PARP and STAG1 inhibition with STAG2-inactivating mutations. The discovery of cancer genes is useful not only for understanding the molecular basis of cancer pathogenesis but also for developing therapeutics that specifically target cancer cells with specific mutant genes. However, in the case of mutated tumour suppressor genes, it can be challenging to develop such therapeutic strategies, because tumour suppressor genes represent a loss of function whereas most drug development strategies are based on inhibition of a gain of function. Therefore, in the case of mutant tumour suppressor genes, the classical approach is to identify gene products whose inhibition is selectively cytotoxic or cytostatic to cells harbouring mutations in the tumour suppressor gene — known as a synthetic lethality strategy[63]. Two such synthetic lethalities have thus far been identified for tumours harbouring mutations of *STAG2* — inactivation of STAG1 and inactivation of poly(ADP-ribose) polymerase (PARP) (FIG. 4).

STAG1 is a STAG2 homologue that has distinct functions in cohesin biology, but that can partially compensate for STAG2 when *STAG2* is mutated in cancer[64]. The presence of this compensation pathway suggested that inactivation of STAG1 might be preferentially cytotoxic to *STAG2*-mutant cancers. Several groups have recently provided evidence to support this hypothesis, demonstrating that STAG1 inactivation is a potent synthetic lethality in *STAG2*-mutant cancer cells in vitro and in vivo[65–67]. At first, it was assumed that this synthetic lethality was based on a requirement for either STAG1 or STAG2 in maintaining sister chromatin cohesion. However, it now seems more likely that STAG1-cohesin complexes are required for the maintenance of TAD structure (and therefore for cellular viability) in *STAG2*-mutant cells. Regardless of the mechanism, these synthetic lethality data indicate that pharmacological

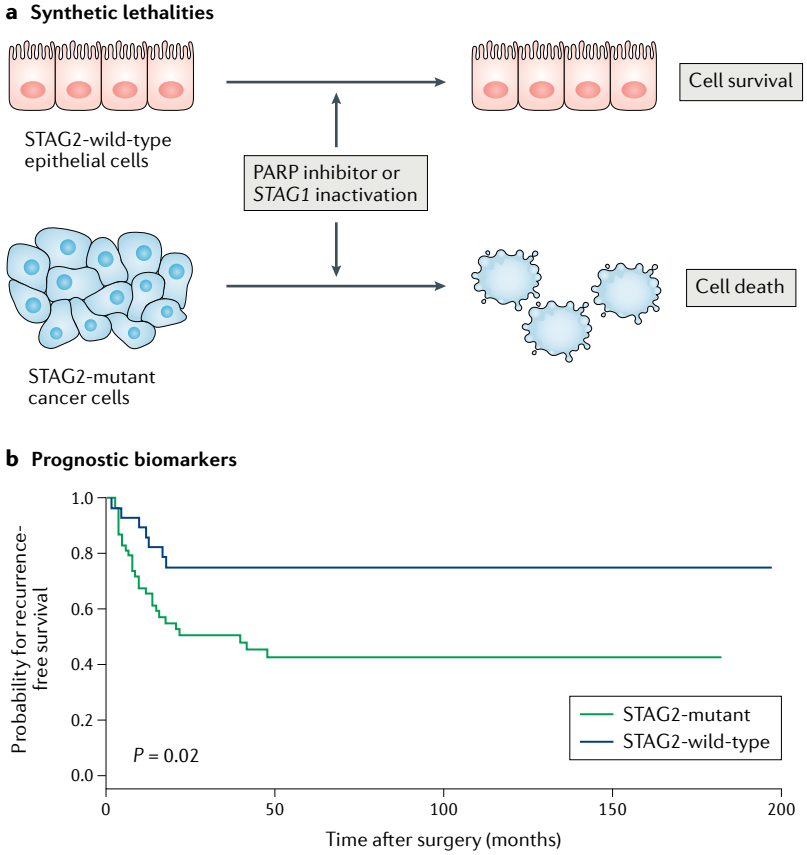

Fig. 4 | **Translational potential of cohesin mutations in cancer. a** | Two synthetic lethalities have been identified for cells harbouring mutations of cohesin subunit SA-2 (*STAG2*). Glioblastoma cells harbouring inactivating mutations of *STAG2* are specifically sensitized to inhibitors of poly(ADP-ribose) polymerase (PARP), which regulates DNA replication. Similarly, bladder cancer cells harbouring inactivating mutations of *STAG2* are sensitive to genetic inactivation of its homologue *STAG1*. **b** | *STAG2* mutations define a previously unrecognized subtype of non-muscle invasive bladder cancer with a favourable prognosis. *STAG2*-mutant tumours are less likely to recur and invade the muscle that surrounds the bladder than tumours with the wild-type *STAG2* gene. Part **b** adapted with permission from the American Association for Cancer Research: Lelo, A. et al. STAG2 is a biomarker for prediction of recurrence and progression in papillary non-muscle-invasive bladder cancer. *Clin. Cancer Res.* **24**, 4145–4153 https://doi.org/10.1158/1078-0432.CCR-17-3244 (2018)[75].

inhibition of STAG1 is a potential therapeutic strategy for treating *STAG2*-mutant cancers. However, an important caveat is that because STAG1 does not have enzymatic activity, it is not possible to employ routine approaches to identify pharmacological inhibitors. Furthermore, it is not currently technically feasible to perform therapeutic genetic inactivation of *STAG1* in patients (for example, via CRISPR). Therefore, the challenge now facing researchers is how to develop novel therapeutics that work through non-traditional mechanisms to inhibit STAG1 function in *STAG2*-mutant tumours.

PARPs are a family of proteins that detect and initiate repair of single-strand DNA breaks[68]. Prior to the initial discovery of cohesin mutations in cancer, Phil Hieter's group had demonstrated that cohesin mutations were synthetically lethal with mutations in genes encoding components of the DNA replication fork[69]. Therefore, they tested whether pharmacological inhibitors of PARP — known to play a role in replication fork processivity — demonstrated synthetic lethality with cohesin inactivation. They found that PARP inhibitors were, in fact, selectively cytotoxic to *STAG2*-mutant human cancer cells, including cells harbouring natural tumour-derived mutations in *STAG2* (REF.[70]). This observation has recently been confirmed in vivo[67]. A recent study has further demonstrated that this synthetic lethality may be attributable to a newly discovered role for STAG2 in maintaining the physical interactions between cohesin and components of the replication fork[52]. These findings have potentially immediate clinical relevance because potent pharmacological inhibitors of PARP (for example, olaparib, talazoparib) are already FDA-approved for the treatment of ovarian and breast cancers. Based on these types of preclinical data, a clinical trial has recently been initiated to determine the efficacy of talazoparib in patients with cohesin-mutant myeloid leukaemias[71].

Development of cohesin as a cancer biomarker. Mutated cancer genes are often useful as prognostic biomarkers because they can define a previously unrecognized tumour subtype with a distinct prognosis. Recent studies have demonstrated that *STAG2* mutations may be a clinically useful biomarker in bladder cancer.

In 2013, we and others reported that inactivating mutations of *STAG2* were present in ~35% of NMIBCs, but in only ~10% of muscle invasive bladder cancer[20,72–74]. This discrepancy suggested that *STAG2*-mutant NMIBCs represented a previously unrecognized, less-aggressive subtype of NMIBC that was less likely to progress to muscle invasion. To test this suggestion, we and others developed an immunohistochemistry assay for determining the *STAG2* mutation status of NMIBCs and correlated the mutational status with clinical outcomes[75,76]. NMIBC tumours with wild-type *STAG2* genes were twice as likely to recur and progress to muscle invasion compared with tumours with mutant *STAG2* (REF.[75]) (FIG. 4), suggesting that this biomarker assay may be useful for identifying NMIBCs that are most likely to require adjuvant therapy and intensive postoperative surveillance.

Similar biomarker validation studies have also been performed in other tumour types harbouring frequent cohesin gene mutations. In acute myelogenous leukaemia, the prognostic value of cohesin gene mutations is unclear, as published studies have suggested that patients with cohesin-mutant acute myelogenous leukaemia have either a more favourable prognosis[77] or a more unfavourable prognosis[78], or that cohesin gene mutations have no prognostic value at all[79]. In Ewing sarcoma, we and others have found there is a weak but statistically significant correlation between *STAG2* mutations and clinical outcomes[23–25]. However, this information is unlikely to be practically useful because otherwise healthy children with potentially lethal tumours are generally treated aggressively regardless of their biomarker status. In addition to these prognostic studies, Shen et al. have demonstrated that depletion of *STAG2* can confer therapeutic resistance to BRAF inhibitors in melanoma[80].

Emerging future directions

As described in this Review, over the past decade much progress has been made in evaluating the role of tumour-derived mutations in genes encoding cohesin subunits in gene expression, 3D genome organization, cellular differentiation and clinical translation. In addition to these well-established research findings, numerous additional but less well-developed avenues of investigation have the potential to drive the field into important, novel directions for the future.

Intersections of cohesin with the splicing machinery. In 2014, four independent groups reported that inactivating key components of the RNA splicing machinery resulted in loss of sister chromatid cohesin identical to inactivating cohesin itself[81–84]. Then, in 2019, we reported that cohesin physically interacts with a panoply of splicing factors and RNA binding proteins, including several of the splicing factors most commonly mutated in cancer[85]. This discovery of a physical interaction between cohesin and the general splicing machinery built on and generalized an earlier finding that cohesin interacted robustly with the SRRM1 splicing factor[86]. Taken together, these findings suggested that there was a novel intersection between cohesin and splicing factor biology. Work is currently ongoing in several laboratories to explain these findings and determine their potential relevance to cancer pathogenesis.

Functional assessment of tumour-derived missense mutations of cohesin. The fact that 85% of tumour-derived *STAG2* mutations are truncating mutations is a double-edged sword. On one hand, the prevalence of truncating mutations unambiguously identifies *STAG2* as a bona-fide tumour suppressor gene. On the other hand, it makes it difficult to identify the specific tumour-suppressing function(s) of STAG2 because truncating mutations (coupled with nonsense-mediated decay) simultaneously inactivate all functions of a protein. Therefore, tumour-derived missense mutations of *STAG2* and other cohesin subunits are particularly valuable because such mutations have the potential to inactivate the tumour-suppressing functions of cohesin while leaving other functions intact. We have shown that *STAG2* genes harbouring tumour-derived missense mutations are functionally wild type in sister chromatid cohesion assays[27]. However, Jessica Downs' laboratory has shown that tumour-derived missense mutations abrogate the ability of STAG2 to repress transcription at double-strand breaks, implicating this activity as being important for STAG2 tumour suppression[87]. Further study of tumour-derived missense mutations of cohesin subunits should provide insight into the specific functions of cohesin that drive neoplastic transformation.

Cohesin gene mutations and DNA replication. An arguably underemphasized aspect of cohesin function has been its robust, direct role in DNA replication. Early studies performed in *Xenopus* egg extracts demonstrated that recruitment of cohesin to chromatin required multiple components of the DNA pre-replication complex, and was dependent on replication licensing[88,89]. Recent studies demonstrated that this was due to a robust physical interaction between cohesin and the DNA replication machinery[90,91]. Acetylation of cohesin was a central determinant of fork processivity[92], and deleting cohesin slowed DNA replication[93]. David Solomon's laboratory has taken the first step in applying these lessons to understand the mechanisms and consequences of tumour-derived mutations in cohesin subunits. They showed that depletion of *STAG2* often stalls replication forks owing to a generalized disruption of the interaction between cohesin and the DNA replication machinery[52]. Most recently, Job de Lange's laboratory has proposed that active removal of cohesin from chromatin allows cancer cells to overcome DNA replication stress[94]. Further work on these compelling intersections between cohesin and DNA replication promises to provide further insight into the mechanisms and effects of cohesin mutations in cancer.

Cohesin and PcG-mediated epigenetic silencing. The PcG family of chromatin remodelling proteins enforces epigenetic silencing of gene expression by modulating histone methylation and 3D genome organization[95]. PcG proteins are well known to drive cancer pathogenesis[96,97], and pharmacological inhibitors of PcG proteins are in active development as anticancer therapeutics[98]. In 2012, work performed in *Drosophila* embryos suggested that PcG and cohesin complexes had interacting, yet opposing, functions in the transcriptional regulation of developmentally relevant genes[99,100]. Since then, studies performed in mouse embryonic stem cells and oocytes have further emphasized the important mechanistic intersection between cohesin and PcG in epigenetic silencing; however, whether inactivation of cohesin results in enhanced repression[101,102] or enhanced derepression[40,103] of PcG-regulated genes is unresolved. Together, these findings suggest that cohesin-mediated tumour suppression may involve modulation of PcG-mediated epigenetic silencing.

Conclusions

The genes encoding components of the cohesin complex together comprise a commonly mutated tumour suppressor pathway. STAG2 is the most commonly mutated subunit, and the only one that is completely inactivated by mutations. In contrast, the other components of cohesin do not have functionally compensating homologues (STAG1 in the case of STAG2) and are therefore targeted by more subtle heterozygous missense mutations, and are, as such, considered haploinsufficient tumour suppressor genes. Cohesin mutations are most commonly found in bladder cancer, myeloid leukaemia, Ewing sarcoma, glioblastoma and endometrial cancer. These mutations have potential translational relevance, as both drug targets and biomarkers, although these studies remain in the early stages.

The mechanisms and phenotypes of cohesin tumour suppression remain mysterious. Initially, it appeared that cohesin inactivation was a major cause of aneuploidy in cancer, but subsequent studies convincingly challenged this hypothesis. Instead, the best current data indicate that cohesin controls cellular stemness

and differentiation. The biochemical mechanism for this almost certainly relates to the known role for cohesin as a chromatin regulator; however, the specific mechanism remains unknown. Various laboratories have attempted to determine whether, and to what extent, mutations of individual cohesin subunits alter the interaction of cohesin with chromatin, gene expression and 3D genome organization. However, at present, these findings have not definitively pointed to any of these potential mechanisms as being critical for cohesin tumour suppression.

As such, the major affirmative findings of the first 10 years of cohesin cancer research were to determine the tumour types in which cohesin mutations were most common and to show that inactivating cohesin tumour suppression alters cellular differentiation and stemness. Research directions for the second decade of cohesin cancer research are clear: identify the specific biochemical and cell biological mechanism(s) through which cohesin mutations drive neoplastic transformation, and further develop approaches for translating these mutations into clinical benefit for patients with cohesin-mutant cancers.

Published online 8 June 2020

1. Michaelis, C., Ciosk, R. & Nasmyth, K. Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* **91**, 35–45 (1997).
2. Guacci, V., Koshland, D. & Strunnikov, A. A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in *S. cerevisiae*. *Cell* **91**, 47–57 (1997).
Together with Michaelis et al. (1997), this paper reports the initial discovery of the cohesin complex in yeast.
3. Losada, A., Hirano, M. & Hirano, T. Identification of *Xenopus* SMC protein complexes required for sister chromatid cohesion. *Genes. Dev.* **12**, 1986–1997 (1998).
4. Peters, J. M. & Nishiyama, T. Sister chromatid cohesion. *Cold Spring Harb. Perspect. Biol* **4**, a011130 (2012).
5. Haarhuis, J. H., Elbatsh, A. M. & Rowland, B. D. Cohesin and its regulation: on the logic of X-shaped chromosomes. *Dev. Cell* **31**, 7–18 (2014).
6. Barber, T. D. et al. Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. *Proc. Natl Acad. Sci. USA* **105**, 3443–3448 (2008).
7. Rocquain, J. et al. Alteration of cohesin genes in myeloid diseases. *Am. J. Hematol.* **85**, 717–719 (2010).
8. Solomon, D. A. et al. Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science* **333**, 1039–1043 (2011).
This paper is the first to report *STAG2* mutations in human cancer.
9. Krantz, I. D. et al. Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of *Drosophila melanogaster* Nipped-B. *Nat. Genet.* **36**, 631–635 (2004).
10. Tonkin, E. T., Wang, T. J., Lisgo, S., Bamshad, M. J. & Strachan, T. NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nat. Genet.* **36**, 636–641 (2004).
11. Nasmyth, K. Segregating sister genomes: the molecular biology of chromosome separation. *Science* **297**, 559–565 (2002).
12. Zhang, N. et al. Overexpression of Separase induces aneuploidy and mammary tumorigenesis. *Proc. Natl Acad. Sci. USA* **105**, 13033–13038 (2008).
13. Remeseiro, S. et al. Cohesin-SA1 deficiency drives aneuploidy and tumourigenesis in mice due to impaired replication of telomeres. *EMBO J.* **31**, 2076–2089 (2012).
14. Schubert, V. et al. Cohesin gene defects may impair sister chromatid alignment and genome stability in *Arabidopsis thaliana*. *Chromosoma* **118**, 591–605 (2009).
15. Covo, S., Puccia, C. M., Argueso, J. L., Gordenin, D. A. & Resnick, M. A. The sister chromatid cohesion pathway suppresses multiple chromosome gain and chromosome amplification. *Genetics* **196**, 373–384 (2014).
16. Hodges, C. A., Revenkova, E., Jessberger, R., Hassold, T. J. & Hunt, P. A. SMC1β-deficient female mice provide evidence that cohesins are a missing link in age-related nondisjunction. *Nat. Genet.* **37**, 1351–1355 (2005).
17. Chiang, T., Duncan, F. E., Schindler, K., Schultz, R. M. & Lampson, M. A. Evidence that weakened centromere cohesion is a leading cause of age-related aneuploidy in oocytes. *Curr. Biol.* **20**, 1522–1528 (2010).
18. Kleyman, M., Kabeche, L. & Compton, D. A. STAG2 promotes error correction in mitosis by regulating kinetochore-microtubule attachments. *J. Cell Sci.* **127**, 4225–4233 (2014).
19. Li, X. et al. Loss of STAG2 causes aneuploidy in normal human bladder cells. *Genet. Mol. Res.* **14**, 2638–2646 (2015).
20. Balbás-Martínez, C. et al. Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. *Nat. Genet.* **45**, 1464–1469 (2013).
21. Kon, A. et al. Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. *Nat. Genet.* **45**, 1232–1237 (2013).
This paper is the first to demonstrate that genes encoding cohesin subunits are commonly inactivated by mutations in myeloid leukaemia.
22. Yoshida, K. et al. The landscape of somatic mutations in Down syndrome-related myeloid disorders. *Nat. Genet.* **45**, 1293–1299 (2013).
23. Crompton, B. D. et al. The genomic landscape of pediatric Ewing sarcoma. *Cancer Discov.* **4**, 1326–1341 (2014).
24. Brohl, A. S. et al. The genomic landscape of the Ewing sarcoma family of tumors reveals recurrent STAG2 mutation. *PLOS Genet.* **10**, e1004475 (2014).
25. Tirode, F. et al. Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. *Cancer Discov.* **4**, 1342–1353 (2014).
Together with Crompton et al. (2014) and Brohl et al. (2014), this paper demonstrates that *STAG2* is the most commonly mutated gene in Ewing sarcoma other than the tumour-defining EWS–FLI translocation.
26. Castronovo, P. et al. Premature chromatid separation is not a useful diagnostic marker for Cornelia de Lange syndrome. *Chromosome Res.* **17**, 763–771 (2009).
27. Kim, J. S. et al. Intact cohesion, anaphase, and chromosome segregation in human cells harboring tumor-derived mutations in STAG2. *PLOS Genet.* **12**, e1005865 (2016).
28. DeMare, L. E. et al. The genomic landscape of cohesin-associated chromatin interactions. *Genome Res.* **23**, 1224–1234 (2013).
29. Lee, B. K. & Iyer, V. R. Genome-wide studies of CCCTC-binding factor (CTCF) and cohesin provide insight into chromatin structure and regulation. *J. Biol. Chem.* **287**, 30906–30913 (2012).
30. Wutz, G. et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**, 3573–3599 (2017).
31. Busslinger, G. A. et al. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**, 503–507 (2017).
32. Viny, A. D. et al. Cohesin members Stag1 and Stag2 display distinct roles in chromatin accessibility and topological control of HSC self-renewal and differentiation. *Cell Stem Cell* **25**, 682–696 (2019).
33. Casa, V. et al. Redundant and specific roles of cohesin STAG subunits in chromatin looping and transcription control. *Genome Res.* **30**, 515–527 (2020).
34. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 (2017).
This paper is among the first to demonstrate that cohesin is required to demarcate and anchor the boundaries of self-interacting contiguous regions of chromosomes known as TADs.
35. Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
36. Haarhuis, J. H. I. et al. The cohesin release factor WAPL restricts chromatin loop extension. *Cell* **169**, 693–707 (2017).
37. Kim, Y., Shi, Z., Zhang, H., Finkelstein, I. J. & Yu, H. Human cohesin compacts DNA by loop extrusion. *Science* **366**, 1345–1349 (2019).
38. Davidson, I. F. et al. DNA loop extrusion by human cohesin. *Science* **366**, 1338–1345 (2019).
Together with Kim et al. (*Science*, 2019), this paper is the first to demonstrate that purified human cohesin functions as a catalytic engine for loop extrusion in vitro.
39. Kojic, A. et al. Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. *Nat. Struct. Mol. Biol.* **25**, 496–504 (2018).
40. Cuadrado, A. et al. Specific contributions of cohesin-SA1 and cohesin-SA2 to TADs and polycomb domains in embryonic stem cells. *Cell Rep.* **27**, 3500–3510 (2019).
41. Wutz, G. et al. ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesin^{STAG1} from WAPL. *eLife* **9**, e52091 (2020).
42. Lu, L., et al. (2019) Robust Hi-C chromatin loop maps in human neurogenesis and brain tissues at high-resolution. *bioRxiv* https://doi.org/10.1101/744540 (2019).
43. Xiao, T., Wallace, J. & Felsenfeld, G. Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol. Cell. Biol.* **31**, 2174–2183 (2011).
44. Ghirlando, R. & Felsenfeld, G. CTCF: making the right connections. *Genes. Dev.* **30**, 881–891 (2016).
45. Ding, L. W. et al. Mutational landscape of pediatric acute lymphoblastic leukemia. *Cancer Res.* **77**, 390–400 (2017).
46. Marshall, A. D. et al. CTCF genetic alterations in endometrial carcinoma are pro-tumorigenic. *Oncogene* **36**, 4100–4110 (2017).
47. Kemp, C. J. et al. CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. *Cell Rep.* **7**, 1020–1029 (2014).
48. Katainen, R. et al. CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat. Genet.* **47**, 818–821 (2015).
This paper is the first to demonstrate that cohesin/CTCF DNA binding sites are mutated in cancer cells more frequently than would be expected by chance alone, emphasizing that alterations in cohesin-mediated 3D genome organization may be a fundamental feature of diverse cancer types.
49. Guo, Y. A. et al. Mutation hotspots at CTCF binding sites coupled to chromosomal instability in gastrointestinal cancers. *Nat. Commun.* **9**, 1520 (2018).
50. Liu, E. M. et al. Identification of cancer drivers at CTCF insulators in 1,962 whole genomes. *Cell Syst.* **8**, 446–455 (2019).
51. Kagey, M. H. et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
52. Mondal, G., Stevers, M., Goode, B., Ashworth, A. & Solomon, D. A. A requirement for STAG2 in replication fork progression creates a targetable synthetic lethality in cohesin-mutant cancers. *Nat. Commun.* **10**, 1686 (2019).

53. Mullenders, J. et al. Cohesin loss alters adult hematopoietic stem cell homeostasis, leading to myeloproliferative neoplasms. *J. Exp. Med.* **212**, 1833–1850 (2015).
54. Mazumdar, C. et al. Leukemia-associated cohesin mutants dominantly enforce stem cell programs and impair human hematopoietic progenitor differentiation. *Cell Stem Cell* **17**, 675–688 (2015).
55. Viny, A. D. et al. Dose-dependent role of the cohesin complex in normal and malignant hematopoiesis. *J. Exp. Med.* **212**, 1819–1832 (2015).
Together with Mullenders et al. (2015) and Mazumdar et al. (2015), this paper is the first to demonstrate that inactivation of cohesin dramatically alters the stemness and pluripotentiality of stem cells, identifying a primary role for cohesin in the regulation of cellular differentiation.
56. Galeev, R. et al. Genome-wide RNAi screen identifies cohesin genes as modifiers of renewal and differentiation in human HSCs. *Cell Rep.* **14**, 2988–3000 (2016).
57. Tothova, Z. et al. Multiplex CRISPR/Cas9-based genome editing in human hematopoietic stem cells models clonal hematopoiesis and myeloid neoplasia. *Cell Stem Cell* **21**, 547–555 (2017).
58. Cuartero, S. et al. Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation. *Nat. Immunol.* **19**, 932–941 (2018).
59. Sasca, D. et al. Cohesin-dependent regulation of gene expression during differentiation is lost in cohesin-mutated myeloid malignancies. *Blood* **134**, 2195–2208 (2019).
60. Lavagnolli, T. et al. Initiation and maintenance of pluripotency gene expression in the absence of cohesin. *Genes. Dev.* **29**, 23–38 (2015).
61. Chen, Z. et al. Cohesin-mediated NF-κB signaling limits hematopoietic stem cell self-renewal in aging and inflammation. *J. Exp. Med.* **216**, 152–175 (2019).
62. Yun, J. et al. Dynamic cohesin-mediated chromatin architecture controls epithelial–mesenchymal plasticity in cancer. *EMBO Rep.* **17**, 1343–1359 (2016).
63. O'Neil, N. J., Bailey, M. L. & Hieter, P. Synthetic lethality and cancer. *Nat. Rev. Genet.* **18**, 613–623 (2017).
64. Canudas, S. & Smith, S. Differential regulation of telomere and centromere cohesion by the Scc3 homologues SA1 and SA2, respectively, in human cells. *J. Cell Biol.* **187**, 165–173 (2009).
65. van der Lelij, P. et al. Synthetic lethality between the cohesin subunits STAG1 and STAG2 in diverse cancer contexts. *eLife* **6**, e26980 (2017).
66. Benedetti, L., Cereda, M., Monteverde, L., Desai, N. & Ciccarelli, F. D. Synthetic lethal interaction between the tumour suppressor STAG2 and its paralog STAG1. *Oncotarget* **8**, 37619–37632 (2017).
Together with van der Lelij et al. (2017), this paper is the first to demonstrate that inactivation of *STAG2* generates an absolute cellular dependency on its orthologue STAG1.
67. Liu, Y. et al. Somatic mutation of the cohesin complex subunit confers therapeutic vulnerabilities in cancer. *J. Clin. Invest.* **128**, 2951–2965 (2018).
68. Bai, P. Biology of poly(ADP-ribose) polymerases: the factotums of cell maintenance. *Mol. Cell.* **58**, 947–958 (2015).
69. McLellan, J. L. et al. Synthetic lethality of cohesins with PARPs and replication fork mediators. *PLOS Genet.* **8**, e1002574 (2012).
70. Bailey, M. L. et al. Glioblastoma cells containing mutations in the cohesin component STAG2 are sensitive to PARP inhibition. *Mol. Cancer Ther.* **13**, 724–732 (2014).
71. US National Library of Medicine. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT03974217 (2019).
72. Solomon, D. A. et al. Frequent truncating mutations of STAG2 in bladder cancer. *Nat. Genet.* **45**, 1428–1430 (2013).
73. Guo, G. et al. Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. *Nat. Genet.* **45**, 1459–1463 (2013).
Together with Balbás-Martínez et al. (2013) and Solomon et al. (2013), this paper demonstrates that cohesin gene mutations are among the most common genetic events in the pathogenesis of early-stage bladder cancers.
74. Taylor, C. F., Platt, F. M., Hurst, C. D., Thygesen, H. H. & Knowles, M. A. Frequent inactivating mutations of STAG2 in bladder cancer are associated with low tumour grade and stage and inversely related to chromosomal copy number changes. *Hum. Mol. Genet.* **23**, 1964–1974 (2014).
75. Lelo, A. et al. STAG2 Is a biomarker for prediction of recurrence and progression in papillary non-muscle-invasive bladder cancer. *Clin. Cancer Res.* **24**, 4145–4153 (2018).
76. Qiao, Y., Zhu, X., Li, A., Yang, S. & Zhang, J. Complete loss of STAG2 expression is an indicator of good prognosis in patients with bladder cancer. *Tumour Biol.* **37**, 10279–10286 (2016).
77. Tsai, C. H. et al. Prognostic impacts and dynamic changes of cohesin complex gene mutations in de novo acute myeloid leukemia. *Blood Cancer J.* **7**, 663 (2017).
78. Papaemmanuil, E. et al. Genomic classification and prognosis in acute myeloid leukemia. *N. Engl. J. Med.* **374**, 2209–2221 (2016).
79. Thol, F. et al. Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications. *Blood* **123**, 914–920 (2014).
80. Shen, C. H. et al. Loss of cohesin complex components STAG2 or STAG3 confers resistance to BRAF inhibition in melanoma. *Nat. Med.* **22**, 1056–1061 (2016).
81. Sundaramoorthy, S., Vázquez-Novelle, M. D., Lekomtsev, S., Howell, M. & Petronczki, M. Functional genomics identifies a requirement of pre-mRNA splicing factors for sister chromatid cohesion. *EMBO J.* **33**, 2623–2642 (2014).
82. van der Lelij, P. et al. SNW1 enables sister chromatid cohesion by mediating the splicing of sororin and APC2 pre-mRNAs. *EMBO J.* **33**, 2643–2658 (2014).
83. Watrin, E., Demidova, M., Watrin, T., Hu, Z. & Prigent, C. Sororin pre-mRNA splicing is required for proper sister chromatid cohesion in human cells. *EMBO Rep.* **15**, 948–955 (2014).
84. Oka, Y. et al. UBL5 is essential for pre-mRNA splicing and sister chromatid cohesion in human cells. *EMBO Rep.* **15**, 956–964 (2014).
85. Kim, J. S. et al. Systematic proteomics of endogenous human cohesin reveals an interaction with diverse splicing factors and RNA-binding proteins required for mitotic progression. *J. Biol. Chem.* **294**, 8760–8772 (2019).
86. McCracken, S. et al. Proteomic analysis of SRm160-containing complexes reveals a conserved association with cohesin. *J. Biol. Chem.* **280**, 42227–42236 (2005).
87. Meisenberg, C. et al. Repression of transcription at DNA breaks requires cohesin throughout interphase and prevents genome instability. *Mol. Cell.* **73**, 212–223 (2019).
88. Takahashi, T. S., Yiu, P., Chou, M. F., Gygi, S. & Walter, J. C. Recruitment of *Xenopus* Scc2 and cohesin to chromatin requires the pre-replication complex. *Nat. Cell Biol.* **6**, 991–996 (2004).
89. Gillespie, P. J. & Hirano, T. Scc2 couples replication licensing to sister chromatid cohesion in *Xenopus* egg extracts. *Curr. Biol.* **14**, 1598–1603 (2004).
90. Zheng, G., Kanchwala, M., Xing, C. & Yu, H. MCM2-7-dependent cohesin loading during S phase promotes sister-chromatid cohesion. *eLife* **7**, e33920 (2018).
91. Ivanov, M. P. et al. The replicative helicase MCM recruits cohesin acetyltransferase ESCO2 to mediate centromeric sister chromatid cohesion. *EMBO J.* **37**, e97150 (2018).
92. Terret, M. E., Sherwood, R., Rahman, S., Qin, J. & Jallepalli, P. V. Cohesin acetylation speeds the replication fork. *Nature* **462**, 231–234 (2009).
93. Guillou, E. et al. Cohesin organizes chromatin loops at DNA replication factories. *Genes. Dev.* **24**, 2812–2822 (2010).
94. Benedict, B. et al. WAPL-dependent repair of damaged DNA replication forks underlies oncogene-induced loss of sister chromatid cohesion. *Dev. Cell* **52**, 683–698. e7 (2020).
95. Loubiere, V., Martinez, A. M. & Cavalli, G. Cell fate and developmental regulation dynamics by polycomb proteins and 3D genome architecture. *Bioessays* **41**, e1800222 (2019).
96. Morin, R. D. et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat. Genet.* **42**, 181–185 (2010).
97. Jain, P. & Di Croce, L. Mutations and deletions of PRC2 in prostate cancer. *Bioessays* **38**, 446–454 (2016).
98. Kim, K. H. & Roberts, C. W. Targeting EZH2 in cancer. *Nat. Med.* **22**, 128–134 (2016).
99. Cunningham, M. D. et al. Wapl antagonizes cohesin binding and promotes Polycomb-group silencing in *Drosophila*. *Development* **139**, 4172–4179 (2012).
100. Schaaf, C. A. et al. Cohesin and polycomb proteins functionally interact to control transcription at silenced and active genes. *PLOS Genet.* **9**, e1003560 (2013).
101. Rhodes, J. D. P. et al. Cohesin disrupts polycomb-dependent chromosome interactions in embryonic stem cells. *Cell Rep.* **30**, 820–835.e10 (2020).
102. Du, Z. et al. Polycomb group proteins regulate chromatin architecture in mouse oocytes and early embryos. *Mol. Cell* **77**, 825–839.e7 (2020).
103. Stelloh, C. et al. The cohesin-associated protein Wapal is required for proper polycomb-mediated gene silencing. *Epigenetics Chromatin* **9**, 14 (2016).
104. Sedeño Cacciatore, Á. & Rowland, B. D. Loop formation by SMC complexes: turning heads, bending elbows, and fixed anchors. *Curr. Opin. Genet. Dev.* **55**, 11–18 (2019).
105. Walter, M. J. et al. Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* **366**, 1090–1098 (2012).
106. Welch, J. S. et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 (2012).
107. Cancer Genome Atlas Research Network. et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
108. Walter, M. J. et al. Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. *Leukemia* **27**, 1275–1282 (2013).
109. Thota, S. et al. Genetic alterations of the cohesin complex genes in myeloid malignancies. *Blood* **124**, 1790–1798 (2014).
110. Cancer Genome Atlas Research Network. et al. Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73 (2013).
111. Brennan, C. W. et al. The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).
112. Ceccarelli, M. et al. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* **164**, 550–563 (2016).
113. Musio, A. et al. X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. *Nat. Genet.* **38**, 528–530 (2006).
114. Deardorff, M. A. et al. Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. *Am. J. Hum. Genet.* **80**, 485–494 (2007).
115. Deardorff, M. A. et al. RAD21 mutations cause a human cohesinopathy. *Am. J. Hum. Genet.* **90**, 1014–1027 (2012).
116. Yuan, B. et al. Clinical exome sequencing reveals locus heterogeneity and phenotypic variability of cohesinopathies. *Genet. Med.* **21**, 663–675 (2019).
117. Zakari, M., Yuen, K. & Gerton, J. L. Etiology and pathogenesis of the cohesinopathies. *Wiley Interdiscip. Rev. Dev. Biol.* **4**, 489–504 (2015).
118. Ronan, J. L., Wu, W. & Crabtree, G. R. From neural development to cognition: unexpected roles for chromatin. *Nat. Rev. Genet.* **14**, 347–359 (2013).
119. Kadoch, C. et al. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* **45**, 592–601 (2013).
120. Tsurusaki, Y. et al. Mutations affecting components of the SWI/SNF complex cause Coffin–Siris syndrome. *Nat. Genet.* **44**, 376–378 (2012).
121. Bögershausen, N. & Wollnik, B. Mutational landscapes and phenotypic spectrum of SWI/SNF-related intellectual disability disorders. *Front. Mol. Neurosci.* **11**, 252 (2018).
122. Hoischen, A., Krumm, N. & Eichler, E. E. Prioritization of neurodevelopmental disease genes by discovery of new mutations. *Nat. Neurosci.* **17**, 764–772 (2014).
123. Muñoz, S., Minamino, M., Casas-Delucchi, C. S., Patel, H. & Uhlmann, F. A role for chromatin remodeling in cohesin loading onto chromosomes. *Mol. Cell* **74**, 664–673.e5 (2019).
124. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).

125. Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat. Rev. Genet.* **14**, 390–403 (2013).
126. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19**, 789–800 (2018).
127. Mirny, L. A., Imakaev, M. & Abdennur, N. Two major mechanisms of chromosome organization. *Curr. Opin. Cell. Biol.* **58**, 142–152 (2019).
128. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci. Adv.* **5**, eaaw1668 (2019).
129. Hassler, M., Shaltiel, I. A. & Haering, C. H. Towards a unified model of SMC complex function. *Curr. Biol.* **28**, R1266–R1281 (2018).
130. Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).

Acknowledgements
Cohesin cancer biology research in the T.W. laboratory is funded by NIH/National Cancer Institute (NCI) grant R01CA169345 and the Hyundai Hope on Wheels Foundation. The Lombardi Comprehensive Cancer Center is funded by NIH/NCI Cancer Center Support Grant P30CA051008.

Competing interests
The author declares no competing interests.

Peer review information
Nature Reviews Cancer thanks F. X. Real, A. Viny and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.