

RESEARCH

Open Access



# Chronic loss of STAG2 leads to altered chromatin structure contributing to de-regulated transcription in AML

James S. Smith<sup>1</sup>, Katrina M. Lappin<sup>1</sup>, Stephanie G. Craig<sup>1</sup>, Fabio G. Liberante<sup>1,2</sup>, Clare M. Crean<sup>1</sup>, Simon S. McDade<sup>1</sup>, Alexander Thompson<sup>1,3</sup>, Ken I. Mills<sup>1\*†</sup> and Kienan I. Savage<sup>1\*†</sup>

## Abstract

**Background:** The cohesin complex plays a major role in folding the human genome into 3D structural domains. Mutations in members of the cohesin complex are known early drivers of myelodysplastic syndromes (MDS) and acute myeloid leukaemia (AML), with *STAG2* the most frequently mutated complex member.

**Methods:** Here we use functional genomics (RNA-seq, ChIP-seq and HiChIP) to investigate the impact of chronic *STAG2* loss on three-dimensional genome structure and transcriptional programming in a clinically relevant model of chronic *STAG2* loss.

**Results:** The chronic loss of *STAG2* led to loss of smaller loop domains and the maintenance/formation of large domains that, in turn, led to altered genome compartmentalisation. These changes in genome structure resulted in altered gene expression, including deregulation of the *HOXA* locus and the MAPK signalling pathway, resulting in increased sensitivity to MEK inhibition.

**Conclusions:** The altered genomic architecture driven by the chronic loss of *STAG2* results in altered gene expression that may contribute to leukaemogenesis and may be therapeutically targeted.

**Keywords:** *STAG2*, Cohesion, Acute myeloid leukaemia, RNA-seq, ChIP-seq, HiChIP, *HOX*, MAPK pathway

## Background

Acute Myeloid Leukaemia (AML) is a highly clonal disease characterised by the rapid expansion of differentiation-blocked myeloid precursor cells, resulting in defective haematopoiesis and eventually, bone marrow failure [1]. In recent years, large-scale sequencing studies have identified a plethora of mutations within patient derived AML cells, expanding our understanding of the genomic landscape and highlighting the complexity of the varying subtypes of this disease [2]. One of the emerging

genomic disease subgroups involves mutations within chromatin remodelling and splicing related genes, including members of the cohesin complex. Approximately 11% of patients diagnosed with a myeloid malignancy, including AML, Myelodysplastic syndrome, (MDS) and Myeloproliferative neoplasm (MPN), have been shown to harbour a mutation within a member of the cohesin complex, with many more showing significantly reduced expression of complex members [3–5]. Mutations within cohesin complex genes have also been identified in clinically normal and aging individuals following sequencing analysis of large populations, indicating that these mutations are early events in leukaemogenesis, leading to the identification of clonal haematopoiesis of indeterminate potential (CHIP) [6].

\*Correspondence: k.mills@qub.ac.uk; k.savage@qub.ac.uk

†Ken I. Mills and Kienan I. Savage are joint senior authors

<sup>1</sup> Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast BT9 7AE, Northern Ireland, UK

Full list of author information is available at the end of the article



© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

The cohesin complex is an evolutionarily conserved multimeric protein complex consisting of SMC1A, SMC3, RAD21 and either STAG1 or STAG2. The complex plays a pivotal role in mitosis through sister chromatid cohesion, as well as in homologous recombination mediated DNA double strand break repair [7] and has been suggested to be involved in long-range interactions between *cis* regulatory elements of the genome [8].

The most commonly mutated gene within the cohesin complex is *STAG2*, with ~90% of *STAG2* mutations resulting in the introduction of premature stop codons likely to lead to loss of protein function [5]. The impact of loss of function *STAG2* mutations on cohesin function has yet to be fully elucidated. Cohesin and the CCCTC binding factor (CTCF) have been described as master weavers of the genome [9], with a key role in regulating the 3D architecture of the human genome. CTCF and cohesin are co-localised heavily throughout the genome, separating regions of active and repressive chromatin marks regulating gene expression [9, 10].

This study investigated the impact of a clinically relevant *STAG2* mutation on 3D genome architecture, global and local gene expression and therapeutic potential.

## Materials and methods

### OCI-AML3 and OCI-AML3<sup>ΔSTAG2</sup> cells

The male OCI-AML3 cell line (ACC-582) was sourced from DSMZ (Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Germany). Cells were authenticated using STR profiling at the Genomics Core Technology Unit, Belfast City Hospital prior to model generation. Cells were cultured in RPMI-1640 supplemented with 10% FBS and 100 U/mL penicillin and 100 µg/mL streptomycin at 37 °C and 5% CO<sub>2</sub> atmosphere.

Δ*STAG2* cells were generated using lentiviral CRISPR with sgRNA targeting Exon 20 of *STAG2*. Lentivirus was generated using the packaging and envelope vectors psPAX2 and pMD2.G and lentiCRISPR V2 plasmid containing the sgRNA using 293 T cells. Viral supernatant was collected, filtered and the target cells transduced by spinoculation at 500 × RCF for 30 min. Cells were selected in 1.5 µg/mL puromycin for 7 days before single cell dilution cloning and expansion to isolate clones.

### Western blotting

1 × 10<sup>6</sup> cells were collected during log phase of growth; washed in PBS and lysed using RIPA supplemented with protease inhibitors on ice for 30 min. Collected lysates were quantified using BCA assay and 30 µg protein was loaded onto SDS-PAGE gels. Gels were transferred to Polyvinylidene fluoride or polyvinylidene difluoride (PVDF) membranes and blocked in 5% w/v powdered

milk before overnight incubation with the antibody of choice. The following day, membranes were washed and incubated with secondary antibodies before chemiluminescent detection was performed.

### Sanger sequencing

1 × 10<sup>6</sup> cells were collected during log phase of growth and lysed using the Zymo quick-gDNA kit as per the manufacturer's instructions. PCR primers targeting a region spanning the sgRNA target site were used to PCR amplify the edited region in the cell types. PCR products were purified and sent to the Genomics Core Technology Unit at Belfast City Hospital.

### Drug screening

Cells were plated at 5 × 10<sup>5</sup> cells per mL at a volume of 100 µL in either sterile white (CellTitre Glo) or sterile black (CellTox Green) 96 well culture plates. Cells were exposed to either DMSO vehicle control or increasing concentrations of the experimental compound. DMSO final concentrations did not exceed 0.1%. Following incubation, either 100 µL of CellTitre Glo reagent or 100 µL CellTox Green was added and readout performed as per the manufacturer's instructions.

### RNA library preparation

Total RNA was isolated from 1 × 10<sup>6</sup> cells per biological replicate using the RNeasy kit (Qiagen), with 3 biological replicates performed per cell line. Quality was verified using the Agilent RNA 6000 Nano kit on a Bioanalyzer 2100 before moving to library preparation as per the manufacturer's instructions using the KAPA Stranded RNA-Seq Kit with RiboErase. Libraries were size selected to include fragments 150–700 bp in length and sequenced on an Illumina NextSeq 500 using 2 × 75 bp reads.

### ChIP library preparation

Between 1 × 10<sup>6</sup> and 2 × 10<sup>6</sup> cells were collected during log phase of growth, washed with PBS and fixed first in 1.5 mM EGS for 18 min followed by the addition of methanol-free formaldehyde to a final concentration of 1% for 8 min. Glycine was used to quench before cellular lysis. Chromatin was sonicated using a Bioruptor UCD-500 to an average length of 200–500 bp before overnight incubation with the antibody of choice bound to protein G beads (input DNA was also collected for sequencing to identify ChIP enriched reads). Chromatin bound beads were collected, washed in increasing concentration salt buffer and eluted overnight. Proteinase K was added and samples incubated for 2 h at 55 °C before RNase incubation at 37 °C for 1 h. DNA was isolated using phenol–chloroform–isoamylalcohol purification and quantified

using a Qubit 3.0. Libraries were generated using the Diagenode Microplex V2 library preparation kit as per the manufacturer's instructions, libraries were size selected to include fragments 150–700 bp in length and sequenced on an Illumina NextSeq 500 using  $1 \times 75$  bp read. A single ChIP experiment was performed for each target protein, for each condition, in each cell line.

#### HiChIP library preparation

Duplicate biological replicate experiments were carried out. HiChIP was performed as per the published protocol within [11]. In brief: Approximately  $1 \times 10^6$  cells were collected during log phase growth and washed in PBS prior to crosslinking with methanol-free formaldehyde at a final concentration of 1% for 10 min. Cells were then lysed in preparation for *in-situ* contact generation. Isolated nuclei were permeabilised, restriction digested with MboI and restriction sites filled with dNTP's using Biotin-14-dATP. Filled in ends were ligated together at room temperature for 4 h with rotation before nuclei were lysed, sonicated and the target proteins immunoprecipitated overnight using protein G Dynabeads. ChIP DNA was collected, washed and crosslinks reversed overnight using Proteinase K. ChIP DNA was eluted and samples purified using the DNA Clean and Concentrator columns using double elution steps. The DNA was quantified using a Qubit 3.0 before biotin ligation junction capture using streptavidin C-1 beads. Samples were washed and taken forward for Tn5 Tagmentation. Tagmentation was performed as per the manufacturer's instructions followed by PCR amplification for 10 cycles. Libraries were size selected to 200–700 bp and sequenced to a shallow depth on the Miseq using  $2 \times 100$  bp reads. Libraries of sufficient quality were then deep sequenced on the Nextseq using  $2 \times 80$  or  $2 \times 43$  bp.

#### Data availability

The HiChIP, ChIP-Seq and RNA-seq data files are accessible at GEO Series record GSE111537.

#### RNA-SEQ data processing

FASTQ files (three biological replicates for each cell line) were downloaded from BaseSpace sequencing hub and checked for quality using FASTQC. Reads were aligned to hg19 using STAR and Ensemble gene annotation. Gene count data generated by STAR was used as input to DESeq2 to generate differential expression data as well as normalised count tables.

#### ChIP-SEQ data processing

FASTQ files (input and ChIP in each cell line for: STAG1, STAG2, CTCF, H3K27ac & H3K27me3) were generated with Illumina pipeline software (OLB v1.8 and CASAVA

1.7.0 using the default chastity base call thresholds) and subsequently filtered to remove suboptimal reads ( $> 5\%$  N bases and/or  $> 95\%$  same base) and PCR duplicates. High-quality reads were then mapped to the hg19 (Ensembl 56) genome using Bowtie2 using the "very-sensitive" option. Following removal of inferior alignments (reads partially aligned or with more than two indels or more than three mismatches to the reference sequence), files were converted from SAM to BAM using SAMTOOLS and indexed. BAM files were converted to BigWig files using the DeepTools package bamCoverage. Reads were normalised during conversion using the "Reads per genome content" option and all downstream analysis was performed on the normalised BigWigs. MACS2 was used to analyse the resulting alignments and to assess the distribution of aligned reads from target protein ChIP libraries with matched input libraries. Peak lists generated were used to generate Heatmaps using the DeepTools package plotHeatmap and density plots using the plotProfile also in DeepTools.

#### HiChIP data processing

HiChIP data was processed using the Juicer pipeline [12] following sequencing. The pipeline takes read pairs and aligns to the genome using BWA-MEM. The contact matrixes generated as an output were KR normalised and used in post-processing stages of the pipeline.

#### Loops and domains

Loops were detected using HiCCUPS using default settings at resolutions of 5, 10 and 25 kb creating a merged consensus loop list in both cell types. Domains were called using Arrowhead using default settings at resolutions of 5, 10 and 25 KB with a consensus domain list created. For the  $\Delta$ STAG2 cells we found it necessary to scan the contact matrixes manually, overlaying the domain calls from the OCI-AML3 wild type cells and retaining domains observed in both. However, we only retained the domain location and not the score for these manually added domains.

#### Compartmentalisation

Compartmentalisation was assessed using the Eigenvector tool in the Juicer pipeline [12]. We generated 50 kb tracks and correlated the A/B compartments with H3K27ac and H3K27me3 ChIP-SEQ data for each cell type using the H3K27ac mark as an indicator of A compartments and the H3K27me3 mark as an indicator of B compartments.

#### Changes in transcription

Differences in transcription were assessed by filtering the normalised gene counts generated by DESeq2. Gene

counts were converted to Log<sub>2</sub> values and any with a count of 0 in both cell lines were removed. A consensus list of genes with expression levels of greater than Log<sub>2</sub> count 4 in one cell type was created. This allowed the determination of the core set of transcriptionally active genes in the cells, as well as detecting genes that had dropped below the cut off implemented to define transcription.

## Results

### Generation of an isogenic $\Delta$ STAG2 cell model

Given that the most common mutations within *STAG2* result in the introduction of a premature stop codon early in the *STAG2* coding region (including a mixture of frame shifting indels and stop gains)[3], we utilized a CRISPR/Cas9 system to target the *STAG2* coding region within the OCI-AML3 cell line. We selected the OCI-AML3 line, as it was derived from a normal karyotype (NK) patient with *NPM1c+*(Type A) and *DNTM3A* mutations, which have been shown to co-occur with *STAG2* mutations in patient samples [13, 14]. We targeted the commonly mutated exon 20 of *STAG2* (ensuring all possible isoforms were affected), resulting in a hemizygous (*STAG2* is encoded within the X-chromosome and is thus hemizygous in this male cell line) 11 bp deletion (c.1838\_1848del) and a single base insertion at the same site (c.1838-1848insG) resulting in a frameshift mutation and the introduction of a premature stop codon (p. L613CfsX5) 11 bp downstream of the mutation site (Fig. 1a).

Analysis of *STAG2* mRNA levels in the edited cell line ( $\Delta$ STAG2), highlighted a  $\sim$ sevenfold reduction in expression compared to the *STAG2*-WT counterpart, indicating that the mutant mRNA is degraded through non-sense mediated decay (Fig. 1b). When examining the mRNA expression of *STAG1* it was interesting to observe an increase of  $\sim$ 1.5 fold relative to the *STAG2*-WT counterpart (Fig. 1c). The other members of the cohesin complex (*RAD21*, *SMC1A* & *SMC3*) exhibited very similar patterns of expression at the mRNA level in both the *STAG2*-WT and  $\Delta$ STAG2 cell lines (Fig. 1c). In keeping with this, *STAG2* protein expression was completely abolished in this model (Fig. 1d). Additionally, when examining the expression of other members of the cohesin complex, including *STAG1*, similar protein levels were observed in both the  $\Delta$ STAG2 and *STAG2*-WT cells (Fig. 1d).

### STAG1 compensates for loss of *stag2* chromatin binding

Chromatin immunoprecipitation for *STAG1*, *STAG2* and CTCF followed by deep sequencing (ChIP-Seq) was carried out to confirm the loss of *STAG2* chromatin binding and to assess the chromatin binding of other

core members of the cohesin complex in the absence of *STAG2* (Fig. 2a). Binding peaks were called using MACS2 and Deeptools used to generate summary binding density plots to visualise binding across the genome (Fig. 2b). The genomic distribution of *STAG1* and *STAG2* in *STAG2*-WT cells was similar to that previously described in other cell lines [15] (Additional file 1: Figure S1 and Additional file 2: Table S1).

As expected, *STAG2* binding was completely abrogated in the  $\Delta$ STAG2 cells, whilst *STAG1* appeared to have a slightly stronger binding profile in the  $\Delta$ STAG2 cells compared to *STAG2*-WT cells (Fig. 2b). Conversely, the CTCF binding profile was slightly weaker in the  $\Delta$ STAG2 cells (Fig. 2b).

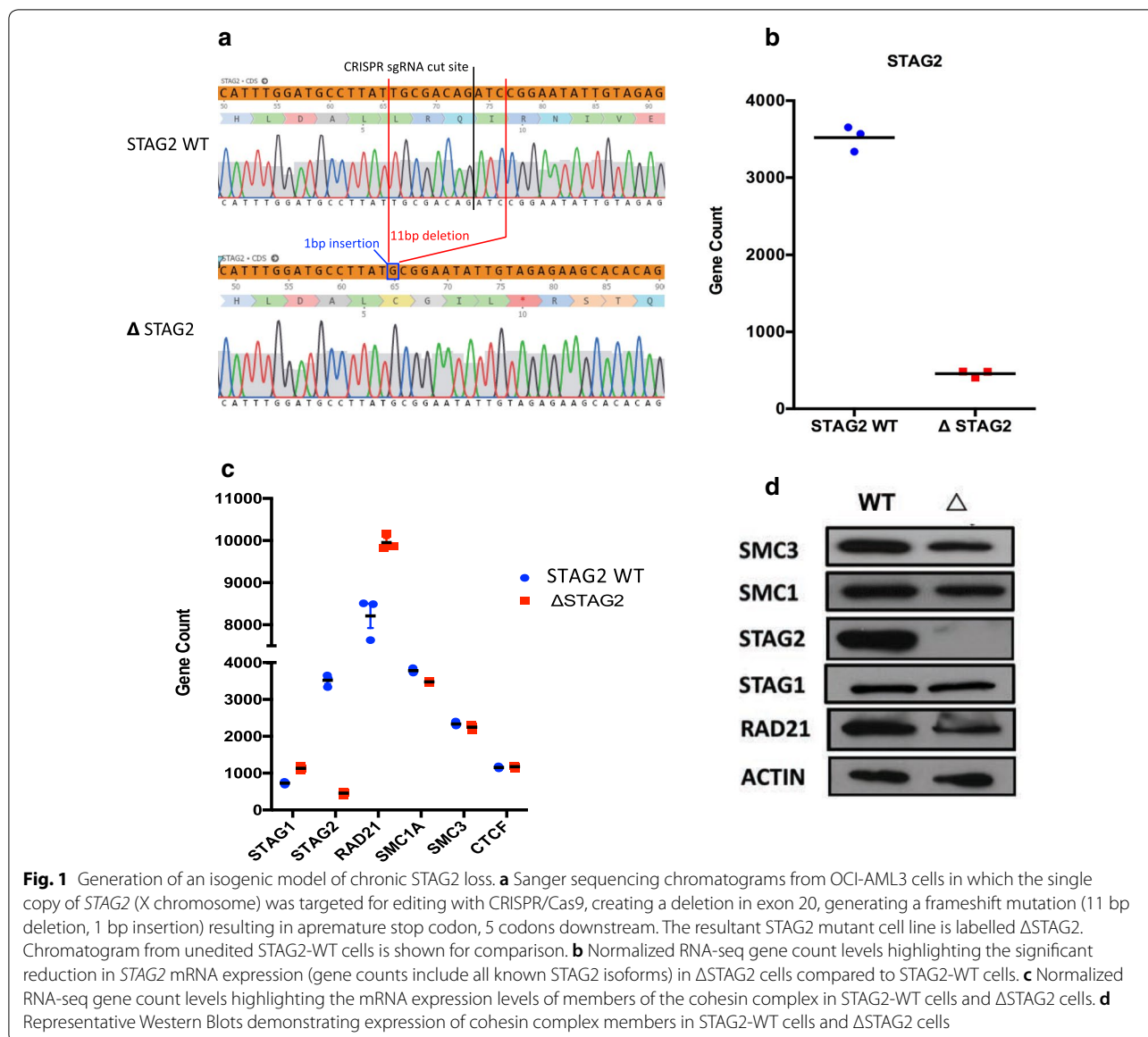
To assess this at an individual binding region level, genome-wide signal density plots for *STAG2*, *STAG1* and CTCF were generated using peak summit locations called in the *STAG2*-WT cells for each protein (Fig. 2c). A total of 34,670 *STAG2* binding peaks were identified in *STAG2*-WT cells, with none identified in the  $\Delta$ STAG2 cells, again confirming the complete loss of functional *STAG2* in these cells. Analysis of *STAG1* binding revealed a  $\sim$ 35% increase in the number of *STAG1* binding peaks (from 15,164 to 22,955) in the *STAG2*-WT and  $\Delta$ STAG2 cells respectively. Many of these additional *STAG1* binding sites corresponded to canonical *STAG2* binding sites suggesting a degree of compensation by *STAG1* in the absence of *STAG2*. To confirm this, *STAG1* peak calls from the *STAG2*-WT cells were intersected with *STAG1* binding sites from the  $\Delta$ STAG2 cells. This showed that  $\sim$ 95% (14,375/15,164) of the canonical *STAG1* binding peaks were maintained in the  $\Delta$ STAG2 cells but also identified an additional 8,580 new *STAG1* binding sites, only present in the  $\Delta$ STAG2 cells (Fig. 2d). To test if these new *STAG1* binding peaks represented regions previously occupied by *STAG2*, we intersected the locations of these binding peaks with the *STAG2* binding peaks called in *STAG2*-WT cells. Of the 8,580 new *STAG1* sites,  $\sim$ 86% (7,392) overlapped with identified sites in *STAG2*-WT cells; these also coincided with 8046 (94%) and 7947 (92%) CTCF sites in the *STAG2*-WT and  $\Delta$ STAG2 cells respectively.

This suggests that the increased *STAG1* binding observed in the  $\Delta$ STAG2 cells has partially, but not completely, compensated for loss of *STAG2* and was predominantly associated with CTCF binding sites.

### Chronic loss of *Stag2* leads to altered 3D genome architecture

3D chromatin structure is organised into loops and domains, which are anchored at convergent CTCF binding sites brought together through loop extrusion facilitated by cohesin [16]. Therefore, to examine 3D





architecture, HiChIP analysis [11] targeting CTCF to identify the CTCF/cohesin bound chromatin interaction sites present in both cell populations was undertaken.

Topologically associated domains (TAD's) were identified using the Juicer Tools Arrowhead algorithm at both 10 and 25 kb resolutions, retaining unique domain calls from both resolutions to generate a consensus domain list [12]. This identified 5,014 and 622 domains in *STAG2*-WT and  $\Delta$ STAG2 cells respectively. However, despite the reduction in domains called in the  $\Delta$ STAG2 cells, the characteristic square appearance of a significant number of TAD's was apparent when visually assessing contact maps generated from the  $\Delta$ STAG2 cells, but with significantly reduced signal (Fig. 3a). This indicates that

some signal was present and visually observable, but was below the call detection threshold of the Juicer Arrowhead detection software. Nonetheless, far fewer TADs were detected in the  $\Delta$ STAG2 cells, suggesting that this may occur due to fewer contacts between CTCF anchor points.

The Juicer Arrowhead software [12] generates scores that serve as a measure of likelihood that the boundaries of the domains called are actual domain corners, with interactions likely to occur within the domain. We found that domains called in the *STAG2*-WT cells scored on average 1.14 (range 0.11–1.88) whilst in the  $\Delta$ STAG2 cells the score was reduced to an average of 0.96 (range 0.19–1.69) (Fig. 3b). The decrease in average score

reflected the decreased frequency of domain forming interactions observed in the  $\Delta$ STAG2 cells.

Additionally, TADs in the STAG2-WT cells, ranged in size from 120 kb to 5.8 Mb (median 425 kb) with 89% of interactions less than 1 Mb, whilst in  $\Delta$ STAG2 cells, the domains ranged from 160 kb to 6.4 Mb (median 575 kb), with 60% of them less than 1 Mb in the  $\Delta$ STAG2 cells (Fig. 3c). This represents a shift in the domain size, with decreased numbers of smaller domains (<250 kb), maintenance of intermediate sized domains (250–800 kb) and formation of larger domains (>800 kb) associated with STAG2 loss, which is also visible when comparing the contact maps for STAG2-WT versus  $\Delta$ STAG2 HiChIP data (Fig. 3a & Additional file 1: Figure S2A,C).

Using virtual 4C (v4C) [17] plotting enabled changes in chromatin interactions with defined anchor regions between the STAG2-WT and  $\Delta$ STAG2 cells to be visualised. We plotted regions where we visually observed altered domain structure between the STAG2-WT and  $\Delta$ STAG2 cells in interaction maps and observed large differences in interaction frequency, with the  $\Delta$ STAG2 cells displaying greatly reduced interaction frequencies on normalised plots (Fig. 3d and Additional file 1: Figure S2B, D). The loss of STAG2 not only results in a reduced number of domains present, but the domains that are able to form with only STAG1 present predominantly result in the loss of smaller domains and larger domain formation.

### DNA loop signal intensity is reduced with chronic STAG2 loss

The altered TAD formation observed in  $\Delta$ STAG2 cells suggested that loop formation might be affected in these cells. Therefore, the presence/formation of loops was determined using the Hi-C Computational Unbiased Peak Search (HiCCUPS) tool [18]. The appearance of focal peaks in proximity ligation data is indicative of a “loop” between two DNA locations. The analysis was run at 5, 10 and 25 kb resolutions, with the final output being a merged consensus loop list from all 3 resolutions

for each replicate library. The final loop call list contained 2,859 loops in the STAG2-WT cells compared to 994 in the  $\Delta$ STAG2 cells. Surprisingly, the vast majority of the loops were of similar size range (Fig. 3e), albeit with some loss of smaller loops (<100 kb) in the  $\Delta$ STAG2 cells. Although significantly fewer loops were called in the  $\Delta$ STAG2 cells, this is not caused by the significant loss of loops of any specific size. Importantly, although much weaker loop peak signals were present in the  $\Delta$ STAG2 cells, the formation of new loops was not seen. Aggregate Peak Analysis (APA) of the consensus loop lists showed that despite the low loop peak signal within the  $\Delta$ STAG2 cells, clear peak foci were visible, albeit with a reduced pixel signal intensity, when compared to STAG2-WT cells (Fig. 3F).

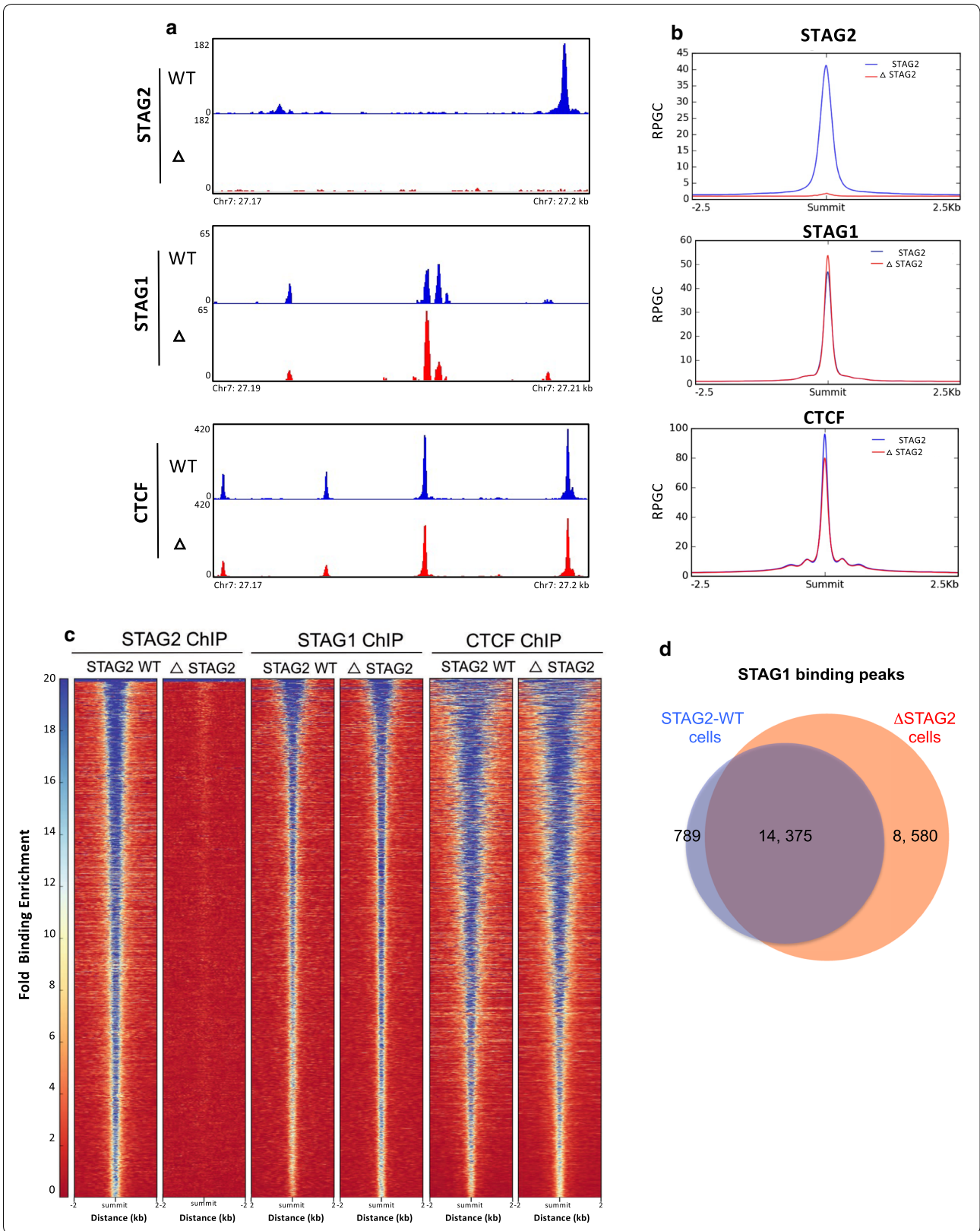
### Genome compartmentalisation analysis highlights compartment switching associated with chronic loss of STAG2

Transcriptionally active and repressed compartments within the genome are observable within the HiChIP data using Pearson's correlation matrices of observed over expected read densities creating a plaid pattern at a resolution of 50 kb [19]. These patterns can be loosely interpreted as compartments defined as A or B, which correlate with euchromatin and heterochromatin respectively. Eigenvector calculation at the same resolution was used to determine if any differences in compartmental shape and edges occur in the  $\Delta$ STAG2 cells compared to their WT counterparts (Fig. 4a). Positive eigenvector values were assigned by performing H3K27ac and H3K27me3 ChIP-Seq in both the WT and  $\Delta$ STAG2 cells and aligning with the eigenvector output.

Eigenvector scoring of the  $\Delta$ STAG2 HiChIP data showed that ~10% of compartments scored as type B in the WT cells had switched to type A compartments, whilst ~8% of compartments scored as A were type B compartments in the  $\Delta$ STAG2. This “switch” in compartment status is greater than would have been expected by chance (~5%) (Fig. 4b).

(See figure on next page.)

**Fig. 2** Cohesin binding patterns in the absence of STAG2 highlight STAG1 compensation. **a** Normalised BigWig binding profiles generated from STAG2, STAG1 and CTCF ChIP-Seq, in STAG2-WT and  $\Delta$ STAG2 cells viewed in IGV. Each region shown, represents the typical binding pattern observed in these cells, for each protein throughout the genome and shows the complete loss of functional STAG2, slightly increased binding of STAG1 and slightly decreased binding of CTCF in the  $\Delta$ STAG2 cells. **b** Consensus binding peak lists were generated for STAG2, STAG1 and CTCF from ChIP-Seq carried out in STAG2-WT cells. These lists were then used to plot the genome-wide average binding profiles of STAG2, STAG1 and CTCF at these consensus sites (–2.5–2.5 kb surrounding peak summits) in STAG2-WT and  $\Delta$ STAG2 cells. Average binding profiles are normalised to total reads per genome content, centred on peak summits. **c** Heatmap profiles of binding patterns for STAG2, STAG1 and CTCF across the genome in both STAG2-WT and  $\Delta$ STAG2 cells. Binding profiles are plotted from –2–2 Kb either side of the individual binding peak summits for each protein called in the STAG2—WT cells. Colour scale depicts fold binding enrichment at these sites. **d** Venn diagram of the number of STAG1 binding peaks called in STAG2-WT and  $\Delta$ STAG2 cells, with 14,375 peaks present in both STAG2-WT and  $\Delta$ STAG2 cells, but significantly more STAG1 binding peaks in  $\Delta$ STAG2 cells compared to STAG2-WT cells



Furthermore, when comparing the compartmental plaid patterns,  $\Delta$ STAG2 compartments appeared to “blend” at edges by switching from A–B or B–A at earlier or later points than seen in the WT counterparts, suggesting a degree of “slippage” at compartment edges (Fig. 4c).

### Chronic loss of STAG2 alters transcriptional programmes

Previous studies assessing the impact of acute or short-term depletion of members of the cohesin complex showed a modest impact on transcription [20]. To examine the impact of chronic loss of STAG2, a clinically relevant scenario, in a relevant genetic background, RNA-seq was performed on the matched isogenic cell pair and the resultant gene expression profiles compared (Fig. 5a). The chronic loss of STAG2 had a relatively large effect on overall gene expression with 24.5% of the active transcriptome (3149 genes) exhibiting some degree of significant change in expression; 1667 (53%) genes decreased and 1482 (47%) genes increased in expression (Fig. 5b & Additional file 3: Table S2). To identify genes whose expression was more definitively deregulated, a >twofold-change cut-off identified 225 (32.6%) upregulated and 466 (67.4%) genes downregulated under chronic loss of STAG2 (Additional file 3: Table S2). Increasing the cut-off level to >fivefold resulted in the identification of 13 (12.9%) up-regulated and 88 (87.1%) down-regulated genes (Additional file 3: Table S2).

To identify specific biological processes affected by chronic STAG2 loss, Ingenuity Pathway Analysis (IPA) of the 691 differentially expressed genes identified using >twofold-change cut-off, identified key networks that were relevant to “hematopoietic development and function” “cell death” and “cell morphology” (Additional file 4: Table S3), with the highest ranked altered disease network linked to haematological development and function (Fig. 5C & Additional file 4: Table S3). This suggests that whilst the majority of genes and pathways have not

become highly deregulated, the chronic loss of STAG2 function has resulted in the transcriptional deregulation of key leukaemogenic pathways, which may contribute to disease development/phenotype.

The disruption in transcription in  $\Delta$ STAG2 cells was unexpected, as previously published data had suggested that STAG1 was predominantly responsible for cohesin binding near gene promoter regions and thereby regulating transcription, whilst STAG2 may be more involved in the maintenance of global 3D chromatin structure [21]. To assess this, the ChIP-Seq data was used to re-analyse STAG1 and STAG2 binding surrounding transcription start sites (TSSs). Average signal density plots confirmed that, on a genome wide level in the STAG2-WT cells, STAG1 was predominantly bound at TSSs in comparison to STAG2 (Fig. 5d). Intriguingly, STAG1 binding at TSSs was nearly identical in the  $\Delta$ STAG2 cells when compared to the STAG2-WT, suggesting that STAG1 based cohesin complexes contribute to the vast majority of transcription-associated interactions in the genome.

Intriguingly, in wild type cells, STAG2 was predominantly bound near the TSSs or promoter regions of the 691 differentially expressed genes in comparison to STAG1 (Fig. 5e). Moreover, STAG1 binding at the TSS did not change in the setting of chronic STAG2 loss, indicating that at these sites, STAG1 was unable to compensate for STAG2. The presence of STAG2, and the lack of STAG1 compensation, at/near the TSSs of these genes further suggests that STAG2 may play a role in linking transcription regulation regions with promoters and enhancers and/or protecting certain promoters from interactions with enhancers in 3-dimensional space.

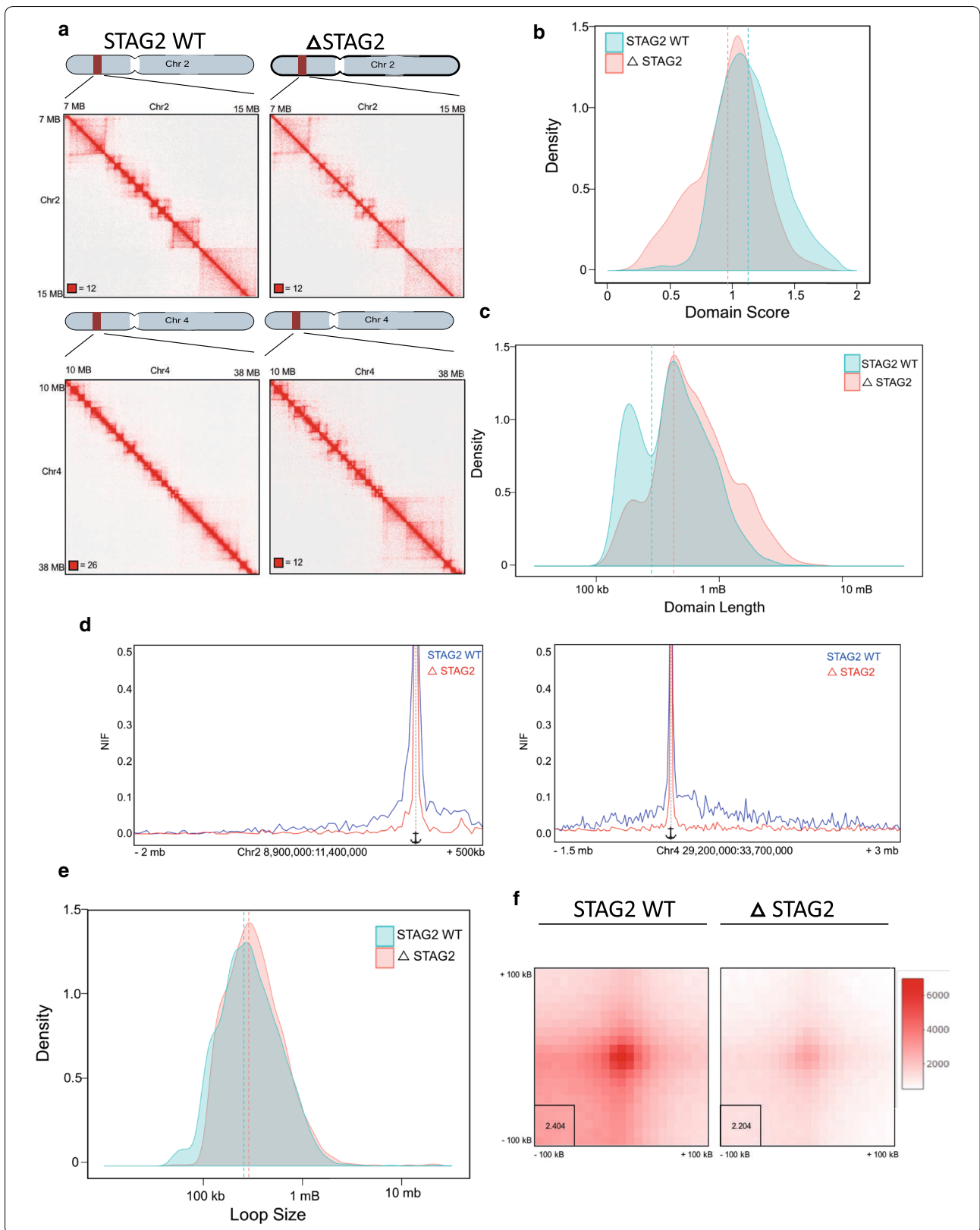
### Gene expression changes mediated by loss of STAG2 are linked with altered domain structure

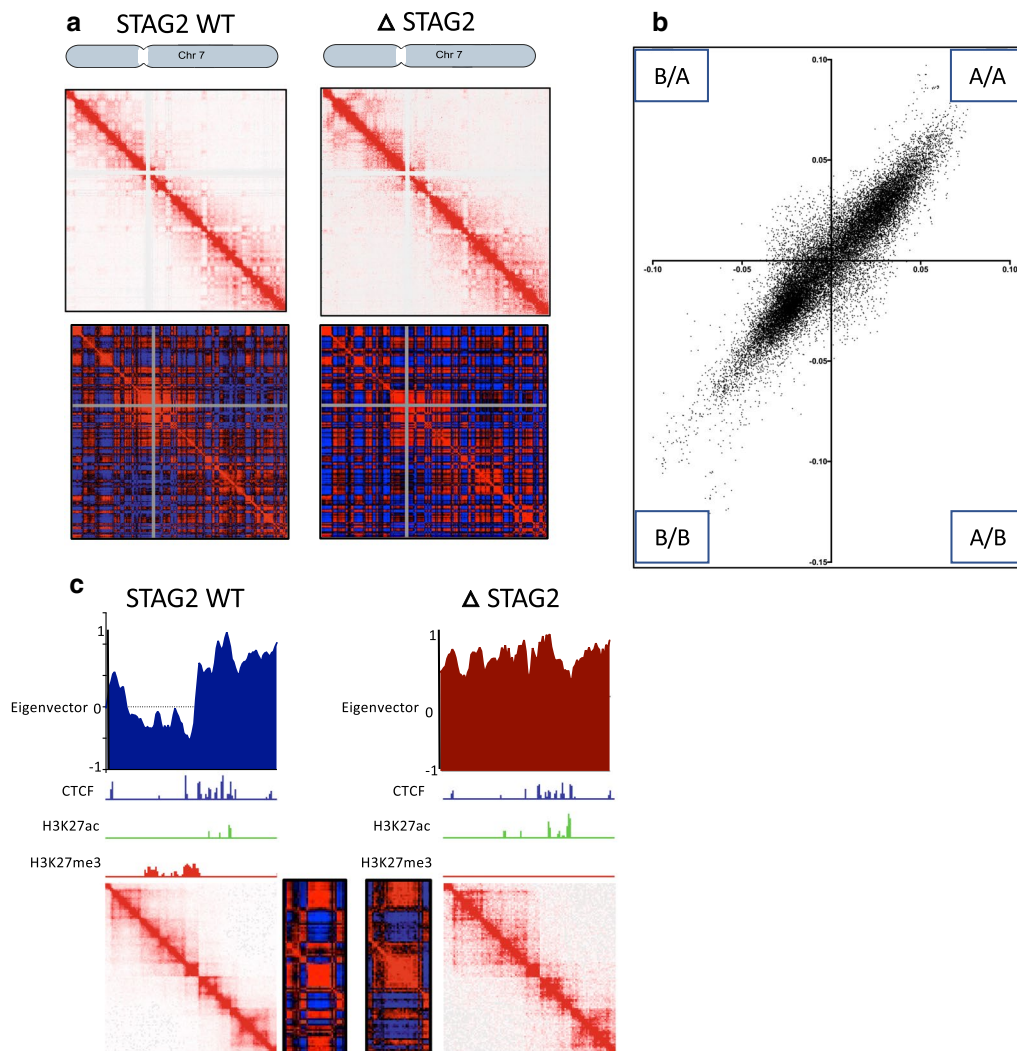
To examine the relationship between STAG2 mediated changes in genome structure and altered transcription, we specifically focused on the *HOXA* cluster, as multiple

(See figure on next page.)

**Fig. 3** Loss of STAG2 alters domain architecture and interaction frequency of the 3D genome. **a** Representative HiChIP contact matrixes displaying interactions covering an 8 Mb region of chromosome 2 and a 28 Mb region of chromosome 4, highlighting the reduced interaction frequency of smaller loop domains, maintenance of intermediate sized domains and formation of larger domains in  $\Delta$ STAG2 cells compared to STAG2-WT cells. **b** Density plot showing the distribution of domain scores generated using the Arrowhead algorithm on CTCF HiChIP data generated from STAG2-WT and  $\Delta$ STAG2 cells indicating reduced domain interaction frequency in  $\Delta$ STAG2 cells, compared to STAG2-WT cells. Dotted lines represent the median domain scores. **c** Density plot showing distribution of domain lengths called in the STAG2-WT and  $\Delta$ STAG2 cells demonstrating loss of smaller domain interactions, maintenance of intermediate size interactions and increased frequency of larger domain interactions in  $\Delta$ STAG2 cells compared to STAG2-WT cells. Dotted lines represent the median domain lengths. **d** Virtual 4C (V4c) plot displaying interactions anchored at a domain boundary region 11 Mb into chromosome 2 and 30.7 Mb into chromosome 4 demonstrating significantly reduced interaction frequency with this anchor point in  $\Delta$ STAG2 cells compared to STAG2-WT cells. **e** Density plot showing distribution of loop lengths called in the STAG2-WT and  $\Delta$  STAG2 cells, indicating the formation of similar sized loops in both STAG2-WT and  $\Delta$ STAG2 cells. Dotted lines represent the median loop sizes. **f** Aggregate Peak Analysis plots using loops called in the STAG2-WT cells as inputs, scaled internally indicating significantly reduced loop signal intensity in  $\Delta$  STAG2 cells in comparison to STAG2-WT cells



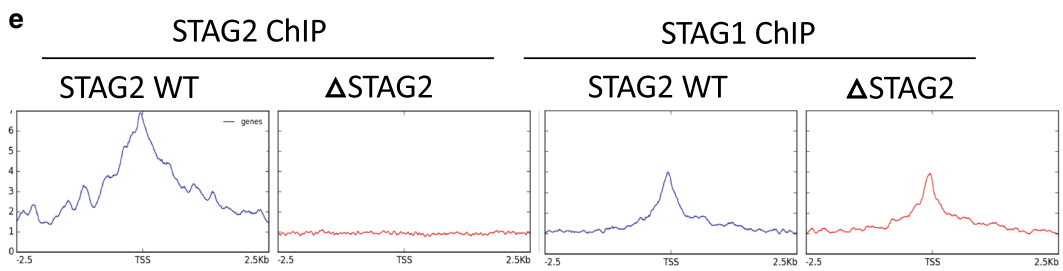
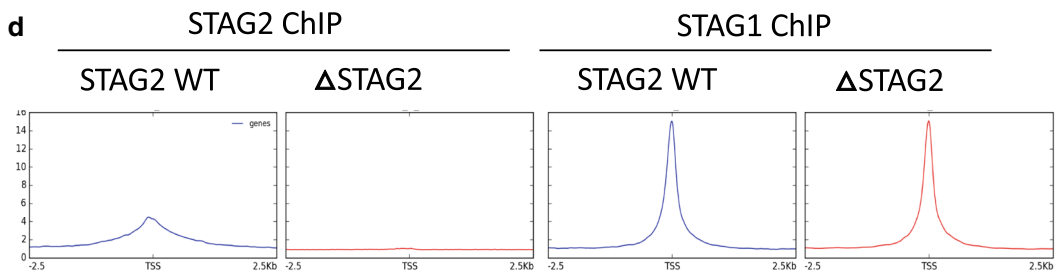
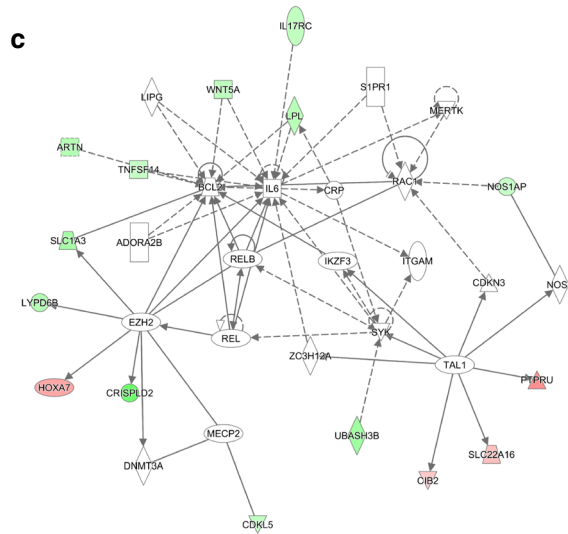
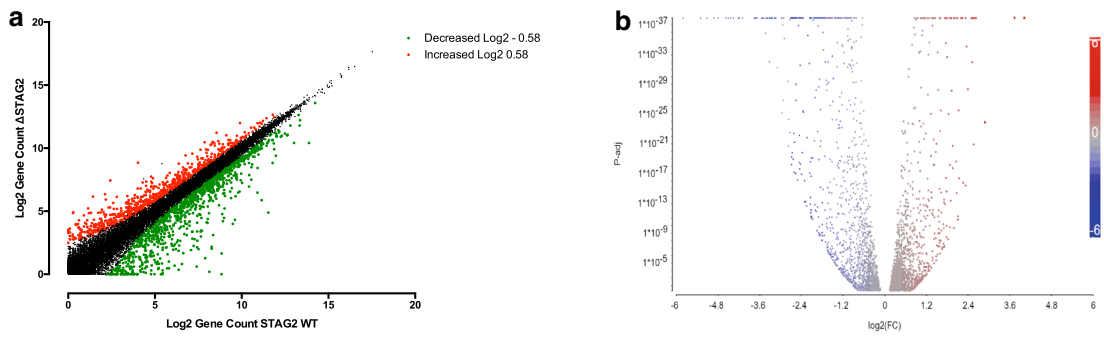




**Fig. 4** Genome compartmentalization highlights compartment switching. **a** Representative HiChIP Contact and Pearson's matrices of chromosome 7 demonstrating the altered compartmentalisation patterns observed in  $\Delta$ STAG2 cells in comparison to STAG2-WT cells. **b** Eigenvector compartment scores at 50 kb bin sizes highlighting the differences in A/B compartments between WT and  $\Delta$ STAG2 cells. **c** Example region covering Chr7:20-25mb in WT and  $\Delta$ STAG2 cells including; eigenvector scores, HiChIP and Pearson's correlation matrices and CTCF, H3K27ac and H3K27me3 ChIP-Seq tracks, highlighting the differences in compartment status in STAG2-WT and  $\Delta$ STAG2 cells

(See figure on next page.)

**Fig. 5** Chronic loss of STAG2 subtly alters global transcriptional programmes. **a** RNA-SEQ density plots of Log<sub>2</sub> gene counts in the  $\Delta$  STAG2 cells, compared to the STAG2-WT cells. Changes greater than or less than 1.5-fold ( $-/+ \text{Log}_2 0.585$ ) are indicated in green (decreased) and red (increased). **b** Volcano plot displaying the 3149 significantly altered ( $\text{adj } p < 0.05$ ) genes between the WT and  $\Delta$ STAG2 cells. **c** Deregulated Haematological developmental processes gene network identified through Ingenuity Pathway Analysis (IPA) as altered in  $\Delta$ STAG2 cells compared to STAG2-WT cells. Genes whose expression was increased or decreased in  $\Delta$  STAG2 cells are highlighted in red and green respectively. **d** ChIP-Seq derived average signal density plots depicting binding profiles for STAG1 and STAG2 centred on transcription start sites (TSSs) ( $-2.5$ – $2.5$  kb) of all coding genes. **e** ChIP-Seq derived average signal density plots depicting binding profiles for STAG1 and STAG2 centred on transcription start sites (TSSs) ( $-2.5$  –  $2.5$  kb) of genes whose expression was deregulated more than 1.5 fold in the  $\Delta$ STAG2 cells



*HOXA* genes were de-regulated upon chronic *STAG2* loss, and deregulation of these genes is highly associated with leukaemogenesis [22]. Interestingly, the region containing the *HOXA* cluster features extensive structural control, with many CTCF/cohesin sites positioned throughout the locus.

From the ChIP-Seq data, seven strong CTCF sites were identified that were associated with the *HOXA* locus (Fig. 6a). These sites linked the 3' region of a TAD (300 kb), to an upstream anchor within the *SKAP2* gene. The major boundary edge of this TAD (termed C7/9 boundary (C=CTCF)) was located between the *HOXA7* and *HOXA9* sites with *HOXA9*, *HOXA10*, *HOXA11*, *HOXA13* and *HOTTIP* lying downstream in a region of high H3K27ac (Fig. 6a). The expression of these genes remained relatively unchanged in the  $\Delta$ *STAG2* cells compared to WT (Fig. 6b). In contrast, the genes lying upstream of the C7/9 boundary (*HOXA2*, *HOXA3*, *HOXA4*, *HOXA5*, *HOXA6* and *HOXA7*) were all significantly up-regulated in the  $\Delta$ *STAG2* cells (2.2 to 3.41-fold higher than the *STAG2*-WT cells) (Fig. 6b). The upregulation of the genes upstream of the C7/9 boundary was also associated with increased H3K27ac and decreased H3K27me3, suggesting an association with altered compartmentalisation. However, the resolution of the Pearson's correlation and eigenvector calculations used to visualise genomic compartmentalisation is not sufficient to allow us to examine this finite region of the genome. Nonetheless, changes in compartmentalisation are generally driven by altered local chromatin structure, therefore v4C was used to visualise the domain region spanning the 300 kb TAD region using the major CTCF boundary site within *SKAP2* as the anchor point (Fig. 6c).

Significantly reduced interaction frequency and novel points of interaction between the *SKAP2* anchor point and with different CTCF sites within the *HOXA* cluster were identified. The major interaction in *STAG2*-WT

cells occurred between the *SKAP2* anchor point and the C7/9 boundary. The frequency of this interaction was significantly reduced in the  $\Delta$ *STAG2* cells, with a high-frequency interaction now occurring downstream at the C10/11 boundary, resulting in a larger sub-TAD domain (Fig. 5c).

The change in local structure within this TAD may contribute to altered compartmentalisation in this region, as indicated by the altered H3K27ac and H3K27me3 signals across the C7/9 boundary. This would lead to transcriptional activation of the genes within the TAD. Specifically, the genes upstream of the C7/9 boundary, which would include the early *HOX* genes (*HOXA2-A7*), would now lie within the transcriptionally active TAD in which the more highly expressed late *HOX* genes (*HOXA9-13*) were normally located.

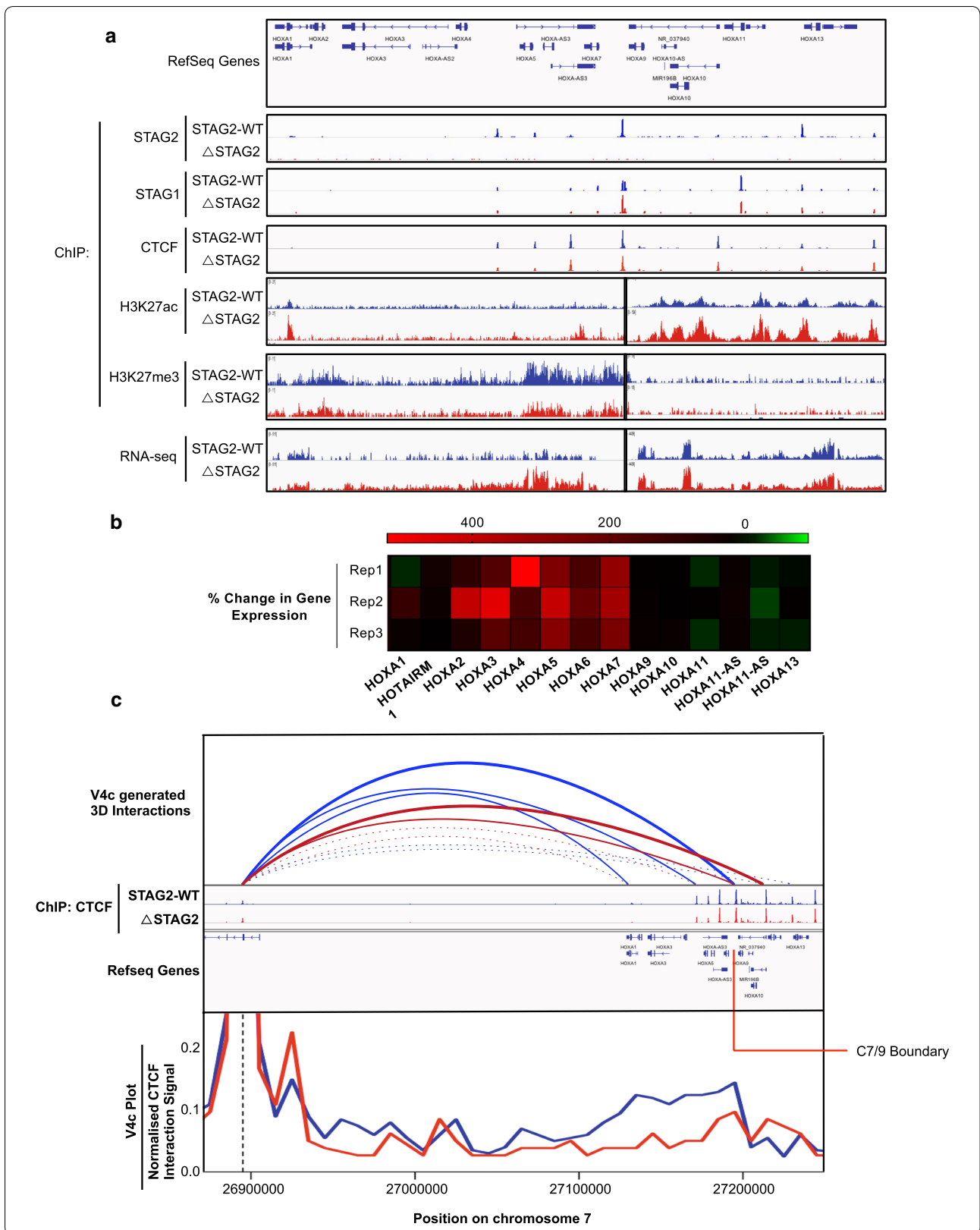
To assess whether the up-regulation of early *HOX* genes occurs in patients with *STAG2* mutant disease, *HOXA* cluster gene expression was examined in two different public datasets held within the Gene Expression Omnibus (TCGA AML dataset GSE68833 and CD34+ cells from MDS patients, GSE58831) [23]. Although *STAG2* mutant disease is relatively rare in these datasets and the fraction of *STAG2* mutant cells in either dataset is somewhat diluted due to the presence of non-leukemic lymphoid cells in each sample (the average blast cell count in AML samples is < 62%, whilst the average pre-isolated blast cell count is < 19%). However, both datasets showed significant up-regulation of many of the early *HOXA* genes; A5 and A7 in AML patients and A3, A4 and A7 in MDS patients (Additional file 1: Figures S3, S4).

The formation of a larger domain spanning the *HOXA* cluster led to upregulation of genes normally held within a repressive domain. However, the transcriptional consequences of the loss of smaller domain structures associated with chronic loss of *STAG2* were not clear. Interestingly, when examining biologically important

(See figure on next page.)

**Fig. 6** Gene expression changes mediated by loss of *STAG2* are linked with altered domain structure. **a** ChIP-Seq generated binding profile tracks for *STAG2*, *STAG1* and CTCF across the *HOXA* cluster region, displaying consistent *STAG1* and CTCF binding profiles in the absence of *STAG2*. ChIP-Seq generated H3K27ac and H3K273me tracks show increased H3K27ac across the *HoxA* locus, with decreased H3K273me across the region encompassing *HOXA1-7* in  $\Delta$ *STAG2* cells, compared to *STAG2*-WT cells. In keeping with this, the normalised RNA-seq read count track demonstrates increased transcription/gene expression across the *HOXA1-7* region in the  $\Delta$  *STAG2* cells compared to *STAG2*-WT cells (different scales were used to display H3K27ac, H3K273me and RNA-Seq read count tracks across the *HOXA1-7* and *HOXA9-13* regions in order to highlight the differences in these regions between *STAG2*-WT and  $\Delta$ *STAG2* cells). **b** Heatmap showing the % change in gene expression between *STAG2*-WT and  $\Delta$ *STAG2* cells of all genes located within the *HOXA* cluster represented above, in each replicate (Rep) RNA-Seq experiment carried out. **c** Virtual 4C plot displaying interactions anchored at the CTCF site located within the *SKAP2* gene located upstream of the *HOXA* cluster. CTCF binding tracks along with RefSeq gene tracks have been included for comparison. Additionally, v4c generated interaction loops across the region have been drawn, indicating the reduced interaction frequency between the *SKAP2* and the C7/9 CTCF sites, and the increased interaction frequency between the *SKAP2* and the C10/11 CTCF sites in  $\Delta$ *STAG2* cells compared to the *STAG2*-WT cells. This demonstrates altered 3D structure within the *HOXA* cluster in  $\Delta$ *STAG2* cells and indicates the inclusion of early *HOX* genes (*HOXA1-A7*) within the transcriptionally active TAD controlling expression of the late *HOX* genes (*HOXA9-13*) in these cells





pathways from our IPA analysis, a large number of down-regulated genes were associated with the MAPK signaling pathway (Fig. 7a, b). These genes mapped to genomic locations that were associated with loss of small domain structures, which may function to link promoter and enhancer regions (Fig. 7c, d & Additional file 1: Figure S5A, B). One gene of interest was *DUSP4*, which has been linked with sensitivity to MEK inhibition [24]. Analysis of *DUSP4* mRNA expression in the same publicly available datasets described earlier, showed that *DUSP4* mRNA expression was lower, but not significantly, in *STAG2* mutant samples compared to *STAG2* wild-type samples (Figure S5C-D). This was likely due to the limited number of *STAG2* mutated samples, the reduced blast cell fraction in these samples and the more subtle changes in expression of these genes observed in our  $\Delta$ *STAG2* model.

To determine if the downregulation of MAPK signaling genes in  $\Delta$ *STAG2* cells functionally impacts the MAPK signalling pathway, we assessed the sensitivity of the isogenic cell lines to MEK inhibition. As expected, the  $\Delta$ *STAG2* cells were significantly more sensitive to treatment with the dual MEK1/2 inhibitors Selumetinib and Trametinib, than the *STAG2*-WT cells (Fig. 7e, f). Consistent with this,  $\Delta$ *STAG2* cells exhibited lower dose inhibition of MEK1/2 (assessed by ERK1/2 phosphorylation) and extremely low dose induction of apoptosis (assessed by PARP and Caspase 3 cleavage) in comparison to *STAG2*-WT cells following MEK inhibition (Fig. 7g & Additional file 1: Figure S6a–c).

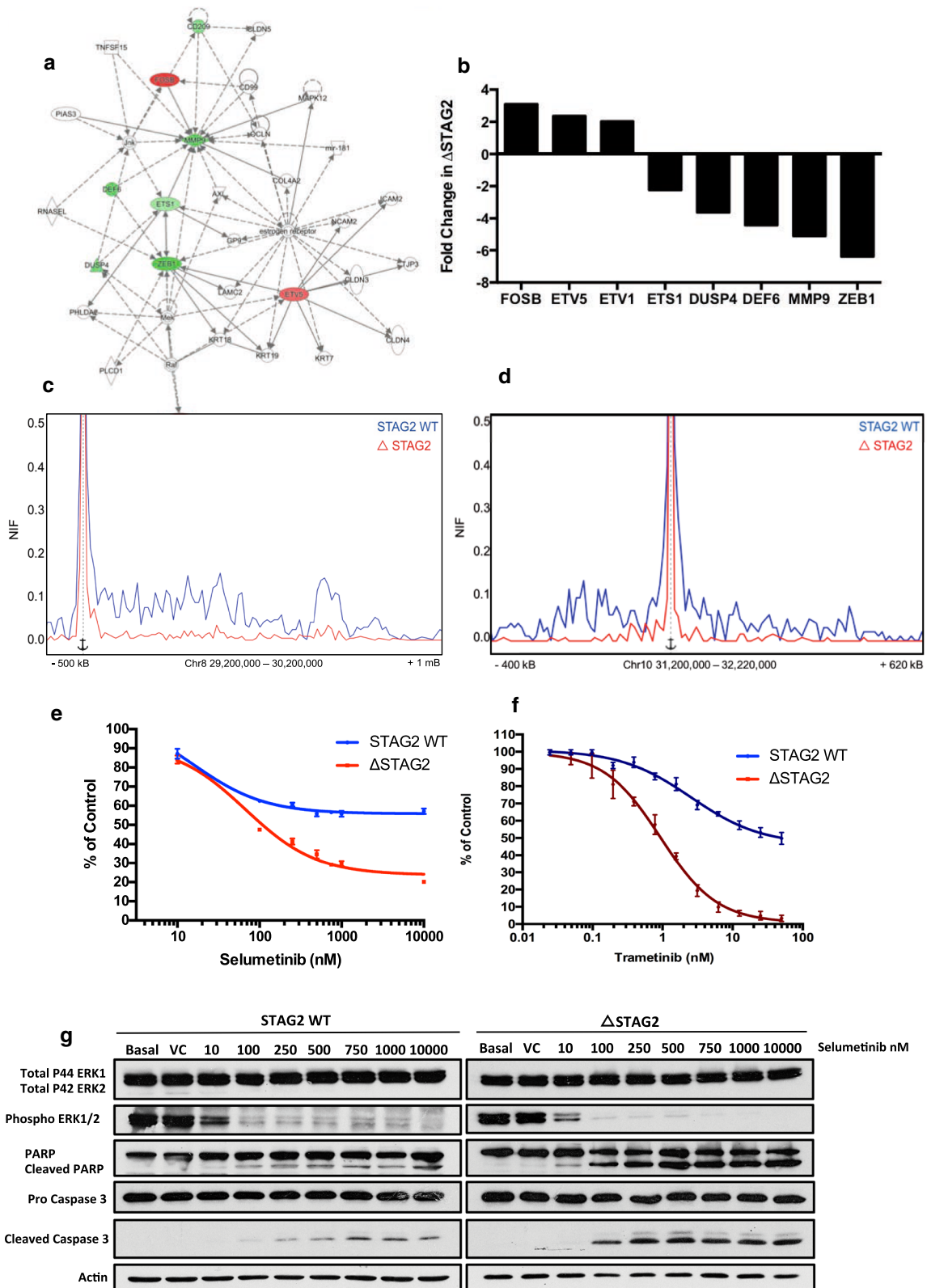
## Discussion

How the cohesin complex regulates 3D genome structure has been a focus of intensive study in recent years. However, much of this focus has been on examining the functional role of cohesin in the regulation of 3D chromatin structure through the depletion of loading and release factors, such as NIPBL and WAPL and via transient depletion of the core cohesin component RAD21. Here, we have focused on studying the impact of a clinically

relevant mutation in *STAG2*, a key component of the cohesin complex frequently mutated in MDS and AML. To date, this is the first study examining the function of cohesin following the chronic loss of a single member of the *STAG* proteins (*STAG2*) in a clinically relevant genetic background, therefore assessing the potential impact of this mutation on disease pathogenesis. Nonetheless, a number of recent studies have shed light on the functional role of cohesin in the regulation of 3D genome structure. Indeed, Rao et al. [25] recently examined the acute loss of cohesin complex function on chromatin structure using high resolution Hi-C following the degron mediated destruction and restoration of RAD21. This revealed that acute loss of cohesin leads to complete loss of all loop domains across the genome, leading to increased compartmentalisation (due to loss of restrictive loop boundaries), but limited changes to transcription [25]. Similarly, conditional deletion of the Cohesin loading factor NIPBL in mouse liver cells, leads to loss of all loop domain structures across the genome, supporting the important role of cohesin in the formation and maintenance of 3D chromatin structure [26]. In contrast, the deletion of the cohesin release factor WAPL, leads to loss of smaller loop domains and the maintenance and/or formation of larger loop domains [27]. This observation is consistent with the loop extrusion model of loop domain formation, in which cohesin slides along DNA, with DNA loops extruded through the cohesin ring structure until meeting DNA bound CTCF (bound at specific CTCF binding sites), where extrusion is paused/arrested forming a stable loop domain [27]. This loop domain remains stable, until cohesin is released; leading to loss of the loop, or CTCF is released, leading to further extrusion and the formation of a larger loop domain [27]. In the absence of WAPL mediated cohesin release, cohesin residence time on DNA exceeds that of CTCF, resulting in further extrusion and increased loop domain size [27]. Adding to this, a study using single molecule imaging and tracking, demonstrated that CTCF has a much shorter chromatin residence time (~1–2 min) than cohesin

(See figure on next page.)

**Fig. 7** Pathway analysis identifies MAPK signalling as a potential therapeutic target in  $\Delta$ *STAG2* cells. **a** Ingenuity Pathway Analysis (IPA) network diagram displaying altered expression of MAPK signaling related genes in  $\Delta$ *STAG2* cells. Genes whose expression was increased or decreased in  $\Delta$ *STAG2* cells are highlighted in red and green respectively. **b** Fold change in expression of genes specifically identified in the IPA MAPK signaling network generated from normalized RNA-Seq read counts in *STAG2*-WT and  $\Delta$ *STAG2* cells. **c** Virtual 4C plot displaying interactions over a 600 kb region of chromosome 8 encompassing the *DUSP4* gene. The V4c plot is anchored upstream of the *DUSP4* gene. **d** Virtual 4C plot displaying interactions over a 600 kb region of chromosome 8 encompassing the *ZEB1* gene. The V4c plot is anchored upstream of the *ZEB1* gene. **e, f** Dose response assay displaying increased sensitivity of  $\Delta$ *STAG2* cells compared to *STAG2*-WT cells to the MEKi Selumetinib (**e**) and Trametinib (**f**). Cell death was assessed 24-h following treatment using the cell titre-glow assay as a readout. Cellular survival is plotted as a percentage signal normalised to control/vehicle treated cells. **g** Representative Western Blot showing lower dose Selumetinib mediated inhibition of ERK1/2 phosphorylation and induction of apoptosis (indicated by PARP and Caspase 3 cleavage) in  $\Delta$ *STAG2* cells compared to the *STAG2*-WT cells. Densitometry analysis of this blot is shown in Additional file 1: Figure S4D–F



(~22 min), suggesting that cohesin may indeed bypass CTCF sites upon CTCF release/exchange and continue extrusion to form larger loop domain structures [28].

The ChIP-Seq data in our study has demonstrated that in the sustained/chronic absence of STAG2, STAG1 is able to partially compensate for loss of STAG2 mediated chromatin binding, by redistributing to sites previously and exclusively bound by STAG2. However, the compensatory binding by STAG1 at vacant STAG2 sites was unable to maintain the “normal” 3D structure of STAG2-WT cells and instead led to the loss of smaller domain structures and the maintenance and formation of larger domain structures, a phenotype strikingly similar to that observed upon deletion of WAPL. We propose two possible scenarios to explain this. One possibility is that by chronically depleting STAG2, we have depleted the pool of free functional cohesin able to be loaded onto chromatin and form loops, the continuous formation and extrusion of which forms smaller domains and sub-domains. Indeed, we did not observe an appreciable increase in STAG1 protein expression in our  $\Delta$ STAG2 cells, suggesting that this may indeed be the case. Additionally, the fact that we observed significantly weaker loop signals in STAG2 depleted cells suggests that although not defective, loop formation is significantly reduced, supporting this hypothesis. However, one would predict that this would not lead to the formation of larger domains, but rather simply the loss of small loop domain structures. Nevertheless, it is also possible that STAG1 containing cohesin (STAG1/cohesin), forms more stable DNA bound complexes and/or is more processive (i.e. extrudes DNA through its pore/ring structure faster) than STAG2 containing cohesin (STAG2/cohesin) [20]. In this instance, STAG1/cohesin would remain bound at more distant CTCF sites at the boundaries of large primary TADs, thereby leading to loss of smaller domains and the formation of larger domains in the absence of STAG2/cohesin.

In any event, it is clear that the loss of smaller domains and the generation of larger domains have a wider functional impact in our model system. Indeed, we observed altered genomic compartmentalisation in our  $\Delta$ STAG2 cells compared to their STAG2-WT counterparts; compartmental switching occurred at approximately 18% of defined compartments, with slightly more B-A (10%) than A-B (8%) switches occurring. Much of the switching appeared to be the result of “slippage” of compartmental edges, consistent with the formation of larger domains, where adjacent TADs represented the boundary between different compartment types. A similar pattern of slippage at compartmental edges and subsequent compartmental switching was observed in WAPL depleted cells, further linking the formation of larger loop domain

structures [27]. The loss of WAPL led to lack of removal of both STAG1/cohesin and STAG2/cohesin and therefore may have an exacerbated effect on genome compartmentalisation. However, this data was generated from HAP1 cells, with a near haploid DNA content, and it is unclear what effect this may have on the gene expression and compartmentalisation requirements of these cells.

A significantly altered gene expression profile was observed in  $\Delta$ STAG2 cells as a result of changes in compartmentalisation associated with chronic STAG2 depletion. Although it has been reported that a cancer-associated STAG2 mutation can support sister chromatid cohesion but was unable to repress transcription at DSBs [7]; this may be due to stalled replication forks and collapse of the interaction between the cohesin and the replication machinery [29].

In our study we observed deregulation of the expression of early HOXA genes; aberrant HOXA gene expression is a feature of different sub-types of AML [22, 30]. In the STAG2-WT cells, the HOXA1-7 genes were expressed at very low levels and this region is marked by a low level/absence of H3K27ac and a significant level of H3K27me3, which indicates a repressed compartment with the boundary maintained by a TAD lying upstream of the transcriptionally active late HOXA genes (HOXA9-13). However, in the absence of STAG2, the interactions that form and maintain the HOXA1-7 containing TAD are reduced, with formation of a larger TAD containing all of the HOXA1-A10. This new, larger domain also coincided with increased H3K27ac and decreased H3K27me3 across the early HOXA gene region (HOXA1-7) and was coupled with increased expression of these early HOXA genes. During limb development and haematopoiesis, HOX genes are expressed in lineage and stage-specific combinations with gene expression within each HOX cluster generally occurring in a linear fashion, from early to late HOX genes, which allows these genes to “share” enhancers. Most developmental processes are accompanied by two successive waves of HOX gene transcription, with genes in the early and late waves of transcription residing within separate but adjacent TADS [31, 32]. The progression of expression from the early wave of HOX gene expression to late wave of HOX gene expression is controlled by shifting the TAD boundary to allow the “switching” of genes near the TAD boundaries from the TAD containing the early expressed genes, to the TAD containing the late expressed genes [31, 32]. In cells chronically depleted of STAG2 this shift is, at least in part, reversed resulting in increased expression of the early HOXA genes, allowing these genes to interact with enhancers controlling the expression of the late HOXA genes, leading to upregulated expression of this region. The expression of the early HOXA genes may contribute



to a block in differentiation, contributing to the development of STAG2 mutated AML. Indeed, the acute siRNA mediated depletion of the core cohesin complex protein RAD21 led to enhanced self-renewal of hematopoietic stem and progenitor cells (HSPCs) driven by de-repression of HOXA7 and HOXA9 in vitro [33].

The altered gene expression across the HOXA cluster observed in our model system was only one example of deregulated gene expression arising from the loss of STAG2. A number of downregulated genes within the MAPK signaling pathway were associated with loss of smaller loop domains in  $\Delta$ STAG2 cells. Intriguingly, our  $\Delta$ STAG2 cells were highly sensitive to MEK1/2 inhibition, suggesting that the chronic loss of STAG2 and the associated deregulation of MAPK signaling genes had a functional impact on these cells that may be therapeutically targeted [28, 34].

Taken together, our data demonstrates that the chronic loss of STAG2, in a clinically relevant genetic AML background, led to altered 3D genomic structure, causing altered transcriptional programming, which may contribute to disease development, through the upregulation of early HOXA genes. Our data also supports the suggestion [35] that vulnerabilities in cohesion mutations may present novel therapeutic targeting strategies, in this case through targeting the deregulated MAPK signalling pathway with MEK inhibitors.

## Conclusions

Mutations in STAG2, a member of the Cohesin complex, occur in myelodysplastic syndromes (MDS) and acute myeloid leukaemia (AML). We have used functional genomics to investigate the chronic loss of STAG2 on genome structure and transcriptional programming. The chronic loss of STAG2 led to a shift in domain loop size and altered genome compartmentalisation. Consequently, the genomic changes resulted in altered gene expression, including deregulation of the HOXA locus, which may contribute to disease development. The altered genomic architecture also resulted in altered MAPK signalling and sensitivity to MEK inhibition which may be a therapeutic strategy for the treatment of STAG2 mutant leukaemia.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12967-020-02500-y>.

**Additional file 1: Figure S1.** Genomic distribution of STAG1 and STAG2. Heatmap profiles of binding patterns for STAG1 and STAG2 across the genome in both STAG2-WT and  $\Delta$ STAG2 cells. Binding profiles are plotted from -2.5 – 2.5Kb either side of CTCF binding sites (CTCF), transcription start sites (TSS), promoters and enhancers. Colour scale depicts fold binding enrichment at these sites. Locations of all called binding peaks are

provides in Table S1. **Figure S2.** Further examples of altered 3D structure within the  $\Delta$ STAG2 genome. a HiChIP Contact matrixes displaying interactions over a 1 Mb region 29.2 Mb into chromosome 8 (this region encompasses the DUSP4 gene). b Virtual 4C plot displaying interactions over the 1 Mb region depicted above. c HiChIP Contact matrixes displaying interactions over a 2 Mb region 60 Mb into chromosome 18 (this region encompasses the BCL2 gene). d Virtual 4C plot displaying interactions over the 2 Mb region depicted above. **Figure S3.** Altered gene expression in an AML patient cohort. Box and whisker plots of gene expression levels (log2) of STAG2 and genes in the HOXA locus between STAG2 mutant patients (n=6) relative to STAG2 wild-type (n=177) AML patients (GSE68833). **Figure S4.** Altered gene expression in an MDS patient cohort. Box and whisker plots of gene expression levels (log2) of STAG2 and genes in the HOXA locus between STAG2 mutant patients (n=6) relative to STAG2 wild-type (n=83) MDS patients (GSE58831). **Figure S5.** Altered chromatin structure surrounding MAPK signaling genes DUSP4 and MMP9 and DUSP4 expression in STAG2 wild-type and mutant patient samples. a Virtual 4C plot displaying interactions over a 1.2 Mb region of chromosome encompassing the DUSP4 gene. The V4c plot is anchored upstream of the DUSP4 gene. b Virtual 4C plot displaying interactions over a 1Mb region of chromosome encompassing the MMP9. c Box and whisker plots of gene expression levels (log2) of DUSP4 between STAG2 mutant patients (n=6) relative to STAG2 wild-type (n=177) AML patients (GSE68833). d Box and whisker plots of gene expression levels (log2) of DUSP4 between STAG2 mutant patients (n=6) relative to STAG2 wild-type (n=83) MDS patients (GSE58831). **Figure S6.** Quantification of MEK signaling and apoptosis following MEK inhibition in SATG2-WT and  $\Delta$ STAG2 cells. (A-C) Densitometry based quantification of pERK (C), Cleaved PARP (D) and Cleaved Caspase 3 (E), from the Representative Western Blot shown in figure 7G.

**Additional file 2: Table S1.** ChIP-seq derived binding peak coordinates for STAG1, STAG2 and CTCF in STAG2-WT and  $\Delta$ STAG2.

**Additional file 3: Table S2.** Deregulated gene expression in  $\Delta$ STAG2 versus STAG2-WT cells.

**Additional file 4: Table S3.** Significantly deregulated Disease Functions/pathways identified through Ingenuity Pathway Analysis (IPA) of significantly altered gene expression (+/- Log2 fod change) in  $\Delta$ STAG2 cells.

## Abbreviations

$\Delta$ STAG2: OCI-AML3 cells with the STAG2 mutation; AML: Acute Myeloid Leukaemia; CTCF: CCCTC binding factor; DMSO: Dimethyl sulfoxide; DSMZ: Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Germany; MDS: Myelodysplastic Syndromes; WT-STAG2: Wild-type OCI-AML cells.

## Acknowledgements

We are grateful to the Genomics Core Technology Unit, PJGCC, in particular Mr Marc-Aurel Fuchs, for access to, and advice using, the Illumina Miseq and Nextseq 500 sequencers and the High Performance Computing Cluster. We would also like to thank all the members of the Blood Cancer Research Laboratory, PGJCCR, for insightful comments on experimental progress and the manuscript direction.

## Authors' contributions

KIM, AT and KIS developed the project, JSS, KIS and KIM designed experiments. JSS performed model generation, ChIP-Seq, RNA-Seq and HiChIP. KML performed additional experiments. JSS, KML, SGC, CMC, KIS and KIM analysed and interpreted data. KML, FGL, SSM, AT, KIS and KIM provided intellectual input. KIM & AT acquired funding. JSS, KIS and KIM wrote the manuscript with input from all authors. All authors read and approved the final manuscript.

## Funding

This work was supported by a Leukaemia and Lymphoma NI (LLNI) Golden Anniversary Victoria Montgomery Studentship (R2053CNR) and LLNI development grant (R2536CNR).

## Availability of data and materials

All genomic data, HiChIP, ChIP-Seq and RNA-seq data files, are accessible at GEO Series record GSE111537.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

All authors consent to the publication of this manuscript.

**Competing interests**

The authors declare no competing interests.

**Author details**

<sup>1</sup> Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast BT9 7AE, Northern Ireland, UK. <sup>2</sup> Wellcome Sanger Institute, Cambridge, UK. <sup>3</sup> Division of Cancer and Stem Cells, School of Medicine, Wolfson Centre for Stem Cells, Tissue Engineering & Modelling (STEM), University of Nottingham, Nottingham NG7 2RD, UK.

Received: 26 May 2020 Accepted: 24 August 2020

Published online: 03 September 2020

**References**

- Estey EH. Acute myeloid leukemia: 2013 update on risk-stratification and management. *Am J Hematol*. 2013;88(4):318–27.
- Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med*. 2016;374(23):2209–21.
- Thota S, Viny AD, Makishima H, Spitzer B, Radivoyevitch T, Przychodzen B, et al. Genetic alterations of the cohesin complex genes in myeloid malignancies. *Blood*. 2014;124(11):1790–8.
- Hill VK, Kim JS, Waldman T. Cohesin mutations in human cancer. *Biochim Biophys Acta*. 2016;1866(1):1–11.
- Fisher JB, McNulty M, Burke MJ, Crispino JD, Rao S. Cohesin mutations in myeloid malignancies. *Trends Cancer*. 2017;3(4):282–93.
- Steensma DP, Bejar R, Jaiswal S, Lindsley RC, Sekeres MA, Hasserjian RP, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*. 2015;126(1):9–16.
- Meisenberg C, Pinder SI, Hopkins SR, Wooller SK, Benstead-Hume G, Pearl FMG, et al. Repression of transcription at DNA breaks requires cohesin throughout interphase and prevents genome instability. *Mol Cell*. 2019;73(2):212–22.
- Merkenschlager M, Odom DT. CTCF and cohesin: linking gene regulatory elements with their targets. *Cell*. 2013;152(6):1285–97.
- Phillips JE, Corces VG. CTCF: master weaver of the genome. *Cell*. 2009;137(7):1194–211.
- Peric-Hupkes D, van Steensel B. Linking cohesin to gene regulation. *Cell*. 2008;132(6):925–8.
- Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods*. 2016;13(11):919–22.
- Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. 2016;3(1):95–8.
- Tiacci E, Spanhol-Rosseto A, Martelli MP, Pasqualucci L, Quentmeier H, Grossmann V, et al. The NPM1 wild-type OCI-AML2 and the NPM1-mutated OCI-AML3 cell lines carry DNMT3A mutations. *Leukemia*. 2012;26(3):554–7.
- Wang M, Yang C, Zhang L, Schaar DG. Molecular mutations and their cooccurrences in cytogenetically normal acute myeloid leukemia. *Stem Cells Int*. 2017;2017:6962379.
- Kojic A, Ana C, De Koninck M, Gomez-Lopez G, Rodriguez-Corsino M, Le Dily F, et al. Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. *bioRxiv*. 2017:166264.
- Barrington C, Finn R, Hadjur S. Cohesin biology meets the loop extrusion model. *Chromosome Res*. 2017;25(1):51–60.
- Gheldorf N, Leleu M, Noordermeer D, Rougemont J, Reymond A. Detecting long-range chromatin interactions using the chromosome conformation capture sequencing (4C-seq) method. *Methods Mol Biol*. 2012;786:211–25.
- Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159(7):1665–800.
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326(5950):289–93.
- Kojic A, Cuadrado A, De Koninck M, Gimenez-Llorente D, Rodriguez-Corsino M, Gomez-Lopez G, et al. Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. *Nat Struct Mol Biol*. 2018;25(6):496–504.
- Remeseiro S, Cuadrado A, Gomez-Lopez G, Pisano DG, Losada A. A unique role of cohesin-SA1 in gene regulation and development. *EMBO J*. 2012;31(9):2090–102.
- Grubach L, Juhl-Christensen C, Rethmeier A, Olesen LH, Aggerholm A, Hokland P, et al. Gene expression profiling of Polycomb, Hox and Meis genes in patients with acute myeloid leukaemia. *Eur J Haematol*. 2008;81(2):112–22.
- Gerstung M, Pellagatti A, Malcovati L, Giagounidis A, Porta MG, Jadersten M, et al. Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. *Nat Commun*. 2015;6:5901.
- Balko JM, Schwarz LJ, Bhola NE, Kurupi R, Owens P, Miller TW, et al. Activation of MAPK pathways due to DUSP4 loss promotes cancer stem cell-like phenotypes in basal-like breast cancer. *Cancer Res*. 2013;73(20):6346–58.
- Rao SSP, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, et al. Cohesin loss eliminates all loop domains. *Cell*. 2017;171(2):305–20.
- Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature*. 2017;551(7678):51–6.
- Haarhuis JHI, van der Weide RH, Blomen VA, Yanez-Cuna JO, Amendola M, van Ruiten MS, et al. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell*. 2017;169(4):693–707.
- Hansen AS, Cattoglio C, Darzacq X, Tjian R. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus*. 2018;9(1):20–322.
- Mondal G, Stevers M, Goode B, Ashworth A, Solomon DA. A requirement for STAG2 in replication fork progression creates a targetable synthetic lethality in cohesin-mutant cancers. *Nat Commun*. 2019;10(1):1686.
- Dickson GJ, Liberante FG, Kettyle LM, O'Hagan KA, Finnegan DP, Bullinger L, et al. HOXA/PBX3 knockdown impairs growth and sensitizes cytogenetically normal acute myeloid leukemia cells to chemotherapy. *Haematologica*. 2013;98(8):1216–25.
- Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, et al. A switch between topological domains underlies HoxD genes colinearity in mouse limbs. *Science*. 2013;340(6137):1234167.
- Lonfat N, Duboule D. Structure, function and evolution of topologically associating domains (TADs) at HOX loci. *FEBS Lett*. 2015;589(20):2869–76.
- Fisher JB, Peterson J, Reimer M, Stelloh C, Pulakanti K, Gerbec ZJ, et al. The cohesin subunit Rad21 is a negative regulator of hematopoietic self-renewal through epigenetic repression of Hoxa7 and Hoxa9. *Leukemia*. 2017;31(3):712–9.
- Rottenberg S, Jonkers J. MEK inhibition as a strategy for targeting residual breast cancer cells with low DUSP4 expression. *Breast Cancer Res*. 2012;14(6):324.
- Cuartero S, Innes AJ, Merkenschlager M. Towards a Better Understanding of Cohesin Mutations in AML. *Front Oncol*. 2019;9:867.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.