STAG2 loss-of-function affects short-range genomic contacts and modulates the basal-luminal transcriptional program of bladder cancer cells

Laia Richart, Eleonora Lapi, Vera Pancaldi, Mirabai Cuenca-Ardura, Enrique Carrillo-De-Santa Pau, Miguel Madrid-Mencía, Hélène Neyret-Kahn, François Radvanyi, Juan Antonio Rodríguez, Yasmina Cuartero, et al.

https://doi.org/10.1093/nar/gkab864

STAG2 loss-of-function affects short-range genomic contacts and modulates the basal-luminal transcriptional program of bladder cancer cells

Laia Richart[1,*,†], Eleonora Lapi[1,2,†], Vera Pancaldi[3,4,5], Mirabai Cuenca-Ardura[1], Enrique Carrillo-de-Santa Pau[1], Miguel Madrid-Mencía[3,4,5], Hélène Neyret-Kahn[6,7], François Radvanyi[6,7], Juan Antonio Rodríguez[8], Yasmina Cuartero[8], François Serra[3], François Le Dily[8], Alfonso Valencia[3,9], Marc A. Marti-Renom[8,9,10,11,*] and Francisco X. Real[1,2,11,*]

[1]Epithelial Carcinogenesis Group, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain, [2]Center for Biomedical Research Network (CIBERONC), 28029 Madrid, Spain, [3]Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain, [4]Centre de Recherches en Cancérologie de Toulouse (CRCT), UMR1037 Inserm, ERL5294 CNRS, 31037 Toulouse, France, [5]University Paul Sabatier III, Toulouse, France, [6]Institut Curie, PSL Research University, CNRS, UMR144, Equipe Labellisée Ligue Contre le Cancer, 75005 Paris, France, [7]Sorbonne Université, UPMC Université Paris 06, CNRS, UMR144, 75005 Paris, France, [8]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), 08028 Barcelona, Spain, [9]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain, [10]Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), 08003 Barcelona, Spain and [11]Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain

Received February 05, 2021; Revised September 08, 2021; Editorial Decision September 10, 2021; Accepted September 29, 2021

ABSTRACT

Cohesin exists in two variants containing STAG1 or STAG2. *STAG2* is one of the most mutated genes in cancer and a major bladder tumor suppressor. Little is known about how its inactivation contributes to tumorigenesis. Here, we analyze the genomic distribution of STAG1 and STAG2 and perform STAG2 loss-of-function experiments using RT112 bladder cancer cells; we then analyze the genomic effects by integrating gene expression and chromatin interaction data. Functional compartmentalization exists between the cohesin complexes: cohesin-STAG2 displays a distinctive genomic distribution and mediates short and mid-ranged interactions that engage genes at higher frequency than those established by cohesin-STAG1. STAG2 knockdown results in down-regulation of the luminal urothelial signature and up-regulation of the basal transcriptional program, mirroring differences between STAG2-high and STAG2-low human bladder tumors. This is accompanied by rewiring of DNA contacts within topological domains, while compartments and domain boundaries remain refractive. Contacts lost upon depletion of STAG2 are assortative, preferentially occur within silent chromatin domains, and are associated with de-repression of lineage-specifying genes. Our findings indicate that STAG2 participates in the DNA looping that keeps the basal transcriptional program silent and thus sustains the luminal program. This mechanism may contribute to the tumor suppressor function of STAG2 in the urothelium.

INTRODUCTION

STAG2 encodes a subunit of the cohesin complex and is one of the most mutated genes in human cancer (1). Among

*To whom correspondence should be addressed. Tel: +34 917328000; Email: freal@cnio.es
Correspondence may also be addressed to Laia Richart. Email: laia.richart-gines@curie.fr
Correspondence may also be addressed to Marc A. Marti-Renom. Email: martirenom@cnag.crg.eu
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Present addresses:
Laia Richart, Institut Curie, Paris Sciences et Lettres Research University, Sorbonne University, 75005 Paris, France.
Laia Richart, INSERM U934/CNRS UMR3215, 75005 Paris, France.
Enrique Carrillo-de-Santa Pau, Computational Biology Group, Precision Nutrition and Cancer Program, IMDEA Food Institute, 28049 Madrid, Spain.

cohesin-associated genes, it harbors the highest frequency of predicted pathogenic mutations (2). Focal deletions on the X chromosome involving *STAG2* were first detected in glioblastomas (3). Subsequently, massive parallel sequencing allowed the detection of point mutations in a variety of human tumors including urothelial bladder cancer (UBC) (3–7), Ewing sarcoma (8,9), myelodysplastic syndrome, and acute myeloid leukemia (AML) (10,11). The majority of *STAG2* mutations reported in cancer lead to a premature stop codon and the absence of protein (12) but loss of expression can also result from gene deletion and/or methylation (6,13,14). Mutations in *STAG2* were first reported almost a decade ago and there is increasing evidence for a role of *STAG2* as a tumor suppressor; yet the mechanisms whereby its inactivation contributes to cancer remain elusive. In UBC, *STAG2* mutations occur mainly in indolent tumors (4–7). By contrast, in Ewing sarcoma they are associated with aggressive neoplasms (8,9), emphasizing the need to perform functional analyses in appropriate model systems to identify potential tissue-specific effects.

The cohesin complex is composed of SMC1, SMC3, RAD21 and either STAG1 or STAG2. As a result, two versions of the complex exist in somatic cells with potentially distinct biological functions (15). In this regard, knockout mouse models have revealed that STAG1 plays a predominant role in telomeric cohesion, while STAG2 plays a more important role in cohesion at chromosome arms or in centromeric regions (16–18). The well-established role of cohesin in chromosome segregation initially suggested that *STAG2* inactivation in cancers might be associated with aneuploidy (3). However, genetic analyses of UBC and AML strikingly showed that *STAG2*-mutant tumors were genomically stable (4,19), supporting the importance of additional molecular mechanisms. These findings are in contrast with a higher rate of somatic copy number changes - but not ploidy - in Ewing sarcoma (9).

There is increasing evidence that cohesin participates in a variety of processes beyond chromosome segregation, including DNA repair and replication, chromatin organization, and gene regulation (12). An increased understanding of these processes has emerged from the use of Chromosome Conformation Capture (3C) technologies (20) including Hi-C (21), which revealed that genomes are folded into complex, hierarchically organized, 3D structures playing a key role in essential processes (e.g. transcription). These structures span a wide range of length scales: from large chromosomal domains and compartments, ∼1Mb self-interacting domains (Topologically Associating Domains or TADs), to DNA loops connecting promoters and gene regulatory elements (22). Cohesin, together with the chromatin insulator CTCF, contributes to TAD border definition by means of loop extrusion (23–27), as well as to intra-TAD promoter-enhancer interactions (28,29). A large fraction of genomic sites targeted by cohesin are simultaneously bound by STAG1, STAG2 and CTCF, yet a few STAG1-only and STAG2-only sites occur in the genome (17,30,31). The latter are depleted of CTCF and are enriched in enhancers and transcription factor binding sites. Knockdown experiments showed that, upon STAG2 depletion, cohesin-STAG1 does not bind to the STAG2-only sites (30–33), suggesting that cohesin-STAG2 has distinct distribution and discrete functions whose role in tumor suppression is yet to be determined.

Despite UBC having the highest frequency of *STAG2* mutations, there are no studies on the role of STAG2 in urothelial cells at the genomic level. UBC is a heterogeneous cancer with two broad histological subtypes (34,35): low-grade/papillary (75–80% of cases) which tend to be non-muscle-invasive—and solid/muscle-invasive (20–25% of cases). The latter can present with variable phenotypes: some tumors preserve urothelial/luminal identity while others show basal/squamous characteristics. Up to 40% of papillary tumors harbor *STAG2* inactivation, which is significantly associated with activating *FGFR3* mutations and low levels of genomic instability (4,7). Conversely, *STAG2* mutations occur only in 12–15% of muscle-invasive tumors. Among them, they are enriched in tumors with urothelial/luminal differentiation and *FGFR3* mutations (36), suggesting that they represent the invasive counterpart of a subset of papillary tumors. Several questions arise from these clinical-molecular observations, including the mechanistic basis of the association with specific transcriptomic signatures and urothelial differentiation.

Here, we set out to explore the contribution of STAG2 to genome organization and the urothelial transcriptional program in RT112, a well-characterized bladder cancer cell line displaying luminal features, mutant *FGFR3,* and wild-type *STAG2* (37). Using Hi-C in combination with chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) for cohesin subunits and RNA-Seq, we show that the cohesin-STAG2 complex is important for the formation and/or maintenance of DNA contacts within TADs but not for the integrity of their boundaries. Chromosomal interactions mediated by cohesin-STAG2 are short and mid-ranged and engage promoters and gene bodies with higher frequency than those mediated by cohesin-STAG1, in agreement with the concept of compartmentalization in the molecular processes undertaken by the two types of cohesin complexes (16–18,30,32,33). Depletion of STAG2 leads to rewiring of short and mid-range contacts and concomitant changes in the expression of selected luminal/basal signature genes.

MATERIALS AND METHODS

Cell lines

RT112 bladder cancer cells used at CNIO and Institut Curie were from the same original stock; HEK293T cells (transformed human embryonic kidney) were from the ATCC. Cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% heat-inactivated FBS (Fetal Bovine Serum) and 1 mM sodium pyruvate.

Plasmids and lentiviral infections

Mission shRNAs (Sigma) were used for RNA interference. Two STAG2-targeting shRNAs were selected based on silencing efficacy and compared to a control non-targeting shRNA. Infectious lentiviruses were produced in HEK293T cells by FuGene-mediated transfection of the lentiviral construct together with the packaging plasmids psPAX2 and pCMV-VSVG. After transfection (48h), the medium was

collected twice for an additional 48 h. Viral supernatants were filtered and either frozen down in aliquots or applied to target cells in the presence of 5 mg·ml^{-1} of polybrene. Cells were harvested after 48 h of puromycin selection (2 mg·ml^{-1}) in serum-free medium. Gene silencing experiments were performed at high cell density and in the absence of serum to avoid cell cycle-dependent effects and to obtain homogeneous cell populations.

Western blotting

Cell pellets were lysed in RIPA buffer supplemented with protease and phosphatase inhibitors. Following sonication, clearing by centrifugation, and quantification, proteins were subjected to SDS-PAGE. Samples were run under reducing conditions and then transferred to nitrocellulose membranes, which were blocked with TBS-Tween, 5% skim milk. Membranes were subsequently incubated with primary antibodies against STAG2 (Santa Cruz, ref. sc-81852, 1:500) or Vinculin (Sigma-Aldrich, ref. V9131-2ML, 1:2000). After washing with TBS-Tween, membranes were incubated with horseradish peroxidase-conjugated secondary antibodies (Dako, 1:10 000) and washed. Reactions were detected using enhanced chemiluminescence.

ChIP sequencing for cohesin subunits and downstream analysis

ChIP-Seq was performed on RT112 in duplicates. Briefly, cells (4×10^7) were washed with PBS, trypsinized, and resuspended in 20 ml of growing media supplemented with 1% formaldehyde for 15 min at room temperature (RT). After quenching with glycine (0.125 M final concentration), fixed cells were washed twice with PBS containing protease inhibitors, pelleted, and resuspended in lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris–HCl pH = 8.1) at 2×10^7 cells/ml. Chromatin was sonicated in a Covaris instrument for 30 min (20% duty cycle; 6% intensity; 200 cycle), yielding DNA fragments of 300–500 bp. Sonicated samples were centrifuged to pellet debris. Chromatin was quantified on a Nanodrop and a 30 μL aliquot of this material was used as input. Chromatin was diluted with buffer (1% Triton X-100, 2 mM EDTA pH 8, 150 mM NaCl, 20 mM Tris–HCl pH 8.1) supplemented with a protease inhibitor cocktail and PMSF. Samples were pre-cleared with a mix of protein A/G agarose beads (previously washed and blocked with 5% BSA) for 1 h at 4°C. After centrifugation, supernatant was divided into aliquots of 500 μg and incubated with 25 μg of antibody [anti-STAG1 (source: 18), anti-STAG2 (source: 18), anti-SMC1 (source: 18), non-related IgG]. After overnight incubation at 4°C, 100μL of pre-blocked protein A/G agarose beads were added for 2 h at 4°C on a rotating platform to collect the immune complexes. Then, beads were sequentially washed with 1 mL of the following buffers: low-salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris–HCl pH 8.1, 150 mM NaCl), high-salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris–HCl pH = 8.1, 500 mM NaCl), LiCl wash buffer (0.25 M LiCl, 1% NP-40, 1% Na deoxycholate, 1 mM EDTA, 10 mM Tris–HCl pH = 8.1), and TE 1× (10 mM Tris–HCl pH 7.5, 1 mM EDTA). DNA was recovered in elution buffer (1% SDS, 0.1 M $NaHCO_3$) and cross-linking was reversed by overnight incubation at 65°C. RNA and proteins were sequentially digested with 20 μg of RNAse and 40 μg of proteinase K. DNA was purified by phenol–chloroform extraction and resuspended in TE 0.5×.

For library preparation, 5 ng of DNA per condition were used. Samples were processed through sequential enzymatic treatments of end-repair, dA-tailing, and ligation to adapters with 'NEBNext Ultra II DNA Library Prep Kit for Illumina' (New England BioLabs, ref. E7645). Adapter-ligated libraries were completed by limited-cycle PCR and extracted with a single double-sided SPRI size selection. Resulting average fragment size was 370 bp, from which 120 bp corresponded to adaptor sequences. Libraries were applied to an Illumina flow cell for cluster generation and sequenced on an Illumina NextSeq 500.

Conversion of Illumina BCL files to bam format was performed with the Illumina2bam tool (Wellcome Trust Sanger Institute - NPG). RUbioSeq (v3.8.1; 38) was used with default parameters to check sequencing quality, align reads to the human reference genome (hg19), normalize library sizes, and calculate ChIP-Seq peaks. Differential peaks between STAG1 and STAG2 were calculated with the DiffBind R package (39). We used the dba.count function to include peaks in the analysis that appear at least in one sample from STAG1, STAG2 or SMC1 ChIP-Seq experiments. Then, peaks were length normalized to 500 bp, extending 250 bp up- and down- stream of the peak summit to keep the peaks at a consistent width. Read counting in peaks was done with ChIP-Seq alignments normalized by library size. Differential enrichment peaks analyses were carried out with dba.contrast and dba.analyze functions. Three categories of cohesin-bound genomic positions were identified with dba.report function, based on statistical differences in read densities for STAG1 and STAG2: common peaks with no statistical differences between STAG1 and STAG2 read densities; STAG1-enriched (STAG1 > STAG2) peaks with FDR < 0.05 and higher STAG1 read density; STAG2-enriched (STAG2 > STAG1) peaks with FDR < 0.05 and higher STAG2 read density. Peak annotation over genomic elements was done with HOMER (v4.8.3; 40). RPKM-normalized bigwig files were generated with DeepTools (v3.0.2) bamCoverage. Heatmaps and density plots were carried out with DeepTools computeMatrix and plotHeatmap around the center of peaks. Motif enrichment was done with MEME-ChIP (v4.12.0; 41) using default parameters.

To integrate the ChIP-Seq and Hi-C results, coordinates of cohesin binding sites were moved to the hg38 assembly using the LiftOver tool by UCSC (https://genome.ucsc.edu/cgi-bin/hgLiftOver).

ChIP-Seq for histone modifications and chromatin state assignment

ChIP-Seq for histone marks and CTCF were performed using RT112 cells. Cells were crosslinked directly in culture medium with formaldehyde (1% final concentration) for 10 min at RT. The reaction was stopped by adding Glycine (0.125 M final concentration) for 10 minutes at RT. Fixed cells were rinsed three times with PBS containing protease

inhibitors, pelleted, and resuspended in lysis buffer (10mM EDTA, pH = 8, 50mM Tris-HCl pH = 8, SDS 1%). After centrifugation, ChIPs were performed using the ChIP-IT High Sensitivity kit (Active Motif, ref. 53040), following manufacturer's instructions. Chromatin was sonicated in a Diagenode Picoruptor sonicator for 10 min (30s ON/30s OFF). Sheared chromatin was immunoprecipitated using the following antibodies: H3K4me1 (Abcam, ref. ab8895), H3K4me3 (Abcam, ref. ab8580), H3K27me3 (Active Motif, ref. 39155), H3K27Ac (Abcam, ref. ab4729), H3K9Ac (Millipore, ref. 07-352), H3K9me3 (Active Motif, ref. 39161) and CTCF (Millipore, ref. 07-729).

ChIP-Seq libraries were prepared using the NEXTflex ChIP-Seq Kit (Bioo Scientific, ref. 5143-02) following the manufacturer's protocol with some modifications. Briefly, 10 ng of ChIP enriched DNA were end repaired using T4 DNA polymerase, Klenow DNA polymerase and T4 PNK, then size selected and cleaned-up using Agencourt AMPure XP beads (Beckman, ref. A63881). A single 'A' nucleotide was added to the 3′ ends of the blunt DNA fragments with a Klenow fragment (3′ to 5′exo minus). The ends of the DNA fragments were ligated to double stranded barcoded DNA adapters (NEXTflex ChIP-Seq Barcodes, Bioo Scientific, ref. 514120) using T4 DNA Ligase. The ligated products were enriched by PCR (2 min at 98°C; [30 s at 98°C, 30 s at 65°C, 60 s at 72°C] x 14 cycles; 4 min at 72°C) and cleaned-up using Agencourt AMPure XP beads. Prior to sequencing, DNA libraries were checked for quality and quantified using a 2100 Bioanalyzer (Agilent). Libraries were loaded in the flow cell at 8 pM concentration and clusters were generated using the Cbot and sequenced on the Illumina Hiseq2500 as single-end 50 base reads following Illumina's instructions.

Sequence reads were mapped to reference genome hg19 using Bowtie 1.0.0 with the following parameters -m 1 –strata –best -y -S -l 40 -p 2. Peak detection was performed using MACS2 (model based analysis for ChIP-Seq v2.1.0.20140616) software under settings where an input sample was used as a negative control. We used a default cut-off and -B option for broad peaks.

ChromHMM was used to identify chromatin states. The genome was analyzed at 200 bp intervals and the tool was used to learn models from the six histone marks, CTCF ChIP-Seq reads files and corresponding Input controls. A model of 10 states was selected and applied to all samples. Nine of the 10 states identified were then given functional annotation based on histone marks enrichment.

Cohesin binding sites enrichment at repeat families

The three sets of cohesin binding sites -common, STAG1-enriched, and STAG2-enriched - were intersected with the Repeatmasker annotation and the number of overlapped binding sites per family calculated. For comparison, 100 random controls were generated per set of cohesin binding sites by randomizing them over each chromosome and avoiding unmappable regions of the genome. Enrichment values were calculated by dividing the number of real cohesin target sites overlaps with the mean number of overlaps in the random controls.

RNA sequencing and analysis

RNA-Seq of control and STAG2-silenced RT112 cells was performed in triplicates (1×10^6 cells per sample). Total RNA was extracted using TRIzol (ThermoFisher, ref. 15596026) and purified with the RNeasy Mini Kit (Qiagen, ref. 74104), according to manufacturer's instructions.

For library preparation, 1 μg of total RNA per condition, each containing an equal amount of ERCC ExFold RNA Spike-In Mix 2 (Ambion, ref. 4456739), was used. Average sample RNA Integrity Number was 9.4 (range 9.0–9.8) when assayed on an Agilent 2100 Bioanalyzer. The PolyA + fraction was purified and randomly fragmented, converted to double stranded cDNA and processed through subsequent enzymatic treatments of end-repair, dA-tailing, and ligation to adapters as in Illumina's 'TruSeq Stranded mRNA Sample Preparation Part # 15031047 Rev. D' kit (this kit incorporates dUTP during 2nd strand cDNA synthesis, which implies that only the cDNA strand generated during 1st strand synthesis is eventually sequenced). The adapter-ligated library was completed by PCR with Illumina PE primers. The resulting purified cDNA library was applied to an Illumina flow cell for cluster generation and sequenced on an Illumina HiSeq 2500.

Conversion of Illumina BCL files to bam format was performed with the Illumina2bam tool (Wellcome Trust Sanger Institute – NPG). Nextpresso 1.9 was used for downstream RNAseq analysis (42). Raw reads were aligned to the human reference genome (hg19).

Motif enrichment analysis of differentially expressed genes was done using HOMER findMotifs (40) on promoters, defined as TSS ±2 kb.

Hi-C library preparation and analysis

RT112 cells were arrested in G1 by culturing at high density and low serum (1%). Hi-C was performed as previously described (43) with some modifications. Purified DNA was fragmented to obtain fragments of an average size of 300–400 bp using a Bioruptor Pico (Diagenode; 8 cycles; 20 s ON/60 sec OFF). 3 μg of DNA per condition were used for library preparation. Biotinylated DNA was pulled down with Dynabeads MyOne T1 streptavidin beads. End-repair, A-tailing and the Illumina adaptors ligation were performed on beads. Libraries were amplified by 10 cycles of PCR and purified using AMPure XP beads. The concentration and size distribution of the Hi-C library after PCR amplification were determined using a Qbit fluorometer and visual exploration in an agarose gel. Hi-C libraries were then paired-end sequenced on an Illumina NextSeq500 (200M reads per library).

Data were processed for read quality control, mapping, interaction detection, filtering, and matrix normalization using TADbit (44) (Supplementary Figure S3A). First, reads were quality-controlled using the TADbit implementation of FastQC for Hi-C datasets. Average PHRED scores were > 25 throughout paired-end reads, indicative of good quality (Supplementary Figure S3B). Then, we used a fragment-based strategy for mapping the remaining reads to the reference human genome (GRCh38). Non-informative contacts including self-circles, dangling-ends, errors,

random breaks or duplicates were filtered out, resulting in 158–197 M valid interactions per condition (Supplementary Table S3). These were then used to generate genome-wide interaction maps at 100 and 20 kb resolution to segment the genome into A/B compartments, demarcate TADs, and identify changes in chromatin looping.

The reproducibility of Hi-C datasets was calculated on pairs of raw Hi-C contact matrices at 40 kb resolution using a stratum-adjusted correlation coefficient (SCC) (45) and a neighborhood size parameter of 5. This calculation was applied to all chromosomes and averaged to obtain a final reproducibility score.

A/B compartments were identified with HOMER (40) by calculating the first two eigenvectors of vanilla-normalized 100 kb contact matrices for every chromosome. Chromosome bins with positive PC1 values and high gene density were considered to be part of compartment A, while bins with negative values and low gene density were assigned to compartment B. For chromosome 4, there was no clear separation between the first and second eigenvector profiles, so the values of the second eigenvector were considered for further analysis. The Y chromosome was excluded from the analysis of genomic bins switching compartments.

Compartment strength was calculated as previously described (46) with minor modifications. We first generated O/E Hi-C matrices at a resolution of 100kb and then sorted the bins according to their associated PC1 values. Compartment strength is the tally of contacts between bins with the same type (A with A or B with B) over the tally of contacts between bins of a different type (A with B). The higher the value, the better the segregation between compartments A and B. To generate saddle plots, we discretized the eigenvalues into 50 categories based on quantiles for each sample and chromosome. Saddle plots represent the aggregated observed vs. expected contacts (log10 transformed) for each of the 50×50 possible combinations. The diagonal pattern of contact enrichment is the consequence of same-to-same compartment interactions. The strength of compartmentalization is included in the resulting saddle plots and was defined as AA (lower right corner)/AB (upper right corner) and BB (upper left corner)/BA (lower left corner). The values used for each corner were calculated as a mean value of 10 bins.

TADs were identified with the TAD detection algorithm implemented in TADbit using vanilla-normalized 100 kb contact matrices. TAD border localization and strength were calculated to evaluate their conservation upon depletion of STAG2. The insulation score of TAD borders was computed using HOMER findTADsAndLoops (40) while excluding regions with known segmental duplications (extracted from the genomicSuperDups.txt and gap.txt files available at UCSC).

To study changes in chromatin looping, we first identified significant interactions (FDR < 0.1) in matrices of 20kb resolution of control (shNT) and STAG2-depleted cells with HOMER analyzeHiC (40). We then employed HiCcompare (47) to investigate the statistical significance of changes in interaction frequency between control and STAG2-silenced cells. Interactions classified as gained or lost complied with the following criteria: adjusted $P < 0.05$ and $\log_2$ FC in interaction frequency $< -1 \mid > 1$. Motif enrichment analysis at the peaks of interactions and pairwise enrichment analysis was done using HOMER findMotifsGenome (40).

Assortativity of regions altered upon STAG2 silencing

To assess whether lost/gained contacts upon STAG2 silencing are proximal in the 3D space, chromatin assortativity (ChAs) analysis was used (48), whereby the genome is represented as a network of nodes (chromatin fragments, here corresponding to Hi-C bins) which are connected if a Hi-C contact between them is observed. Networks were displayed using Cytoscape organic layout.

To select significant Hi-C contacts we identified significant Hi-C interactions at specific binning resolution (20 kb, 100 kb, 1 Mb) by comparing each dataset to the background of the same datasets. Each sample is considered independent. We can thus define a control network and compare it to the STAG2 silenced network and identify connections that are lost and gained from one condition to the next. Briefly, ChAs is a measure of correlation of feature values across all edge pairs in a network and allows us to see whether nodes with a specific property tend to interact more with each other than expected at random. We asked whether assortativity of chromatin regions affected by STAG2 removal is particularly strong for contacts spanning specific ranges of genomic distance by filtering the contacts by distance spanned, eliminating contacts spanning progressively longer distances, thus generating networks in which the minimum distance spanned by any contact is 50kb, 100 kb, 250 kb, 500 kb, 750 kb, 1 Mb. Finally, we mapped network nodes to genes by finding which Hi-C bins were overlapping promoters and thus filtered 'in-silico' promoter-promoter networks from our Hi-C networks at the different resolutions. We then assigned to all the nodes in the P-P network the value of $\log_2$FC of the gene between STAG2 KO and WT and ChAs of fold change. We repeated the calculation of assortativity in networks on which we permuted the expression values on the network nodes to generate a null distribution of ChAs values and show the distribution of these random ChAs values in the plots.

RESULTS

STAG2-enriched cohesin localizes to active enhancers and promoters independently of CTCF

We profiled the genome-wide distribution of STAG2 and STAG1 in RT112 cells by ChIP-Seq. SMC1, common to both cohesin complexes, was used as control. Three categories of cohesin-bound genomic positions were identified based on differential binding of STAG2 and STAG1: common ($n = 35\,321$), STAG1-enriched ($n = 5007$), and STAG2-enriched ($n = 2330$) (Figure 1A and Supplementary Figure S1A). Common positions were occupied by either complex variant and showed comparably high read density for both STAG1 and STAG2. STAG1-enriched positions showed higher read density for STAG1 than for STAG2. In contrast, STAG2-enriched positions had higher STAG2 read density than STAG1 but showed the lowest read den-

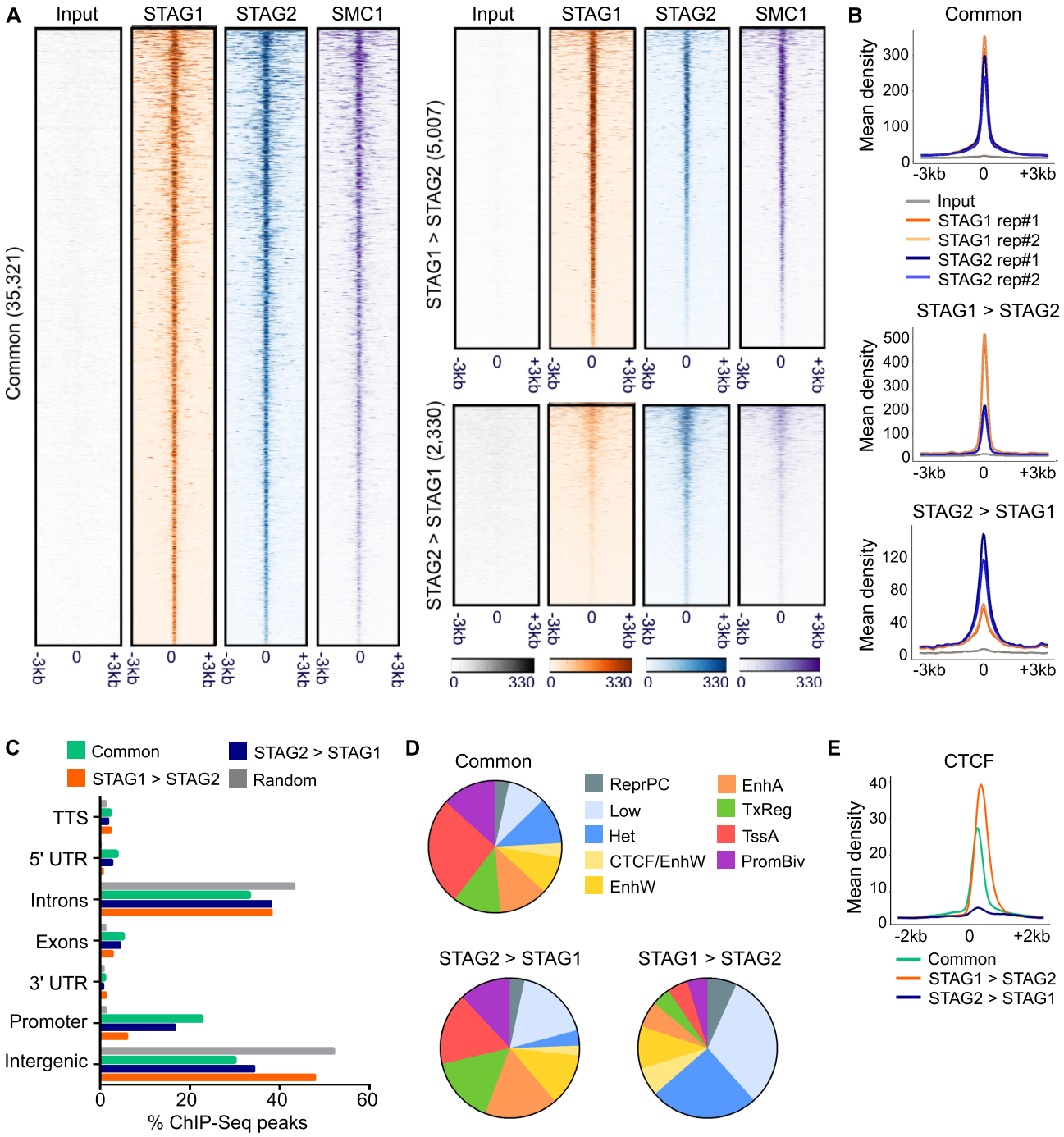

Figure 1. STAG1 and STAG2 show both overlapping and unique distributions over genomic elements and chromatin states in RT112 cells. (**A**) ChIP-Seq read density heatmaps for STAG1, STAG2, and SMC1 at common, STAG1-enriched (STAG1 > STAG2), and STAG2-enriched (STAG2 > STAG1) cohesin positions within a peak-centered 6kb window. (**B**) Read density distribution for STAG1 and STAG2 at common, STAG1-enriched, and STAG2-enriched positions within a peak-centered 6kb window. (**C**) Bar-plot diagram showing the distribution of common, STAG1-enriched, and STAG2-enriched cohesin positions over genomic elements. (**D**) Distribution of cohesin-bound genomic sites throughout chromatin states identified in RT112 cells by ChromHMM and based on combinations of histone modifications and CTCF (see Supplementary Figure S1C for definition of chromatin states). (**E**) Peak-centered enrichment plot for CTCF over the three categories of cohesin-bound positions showing relative depletion in STAG2-enriched sites.

sity of all categories. SMC1 was present in the three categories of cohesin-bound positions (Figure 1A and Supplementary Figure S1A). Peak-centered ChIP-Seq read density plots revealed a pattern of sharp and narrow peaks for STAG1 and STAG2 around common and STAG1-enriched cohesin positions. Peaks at STAG2-enriched sites were broader (Figure 1B), suggesting higher cell-to-cell variability or greater dynamics of this complex variant (30,49).

Analysis of enrichment over genomic elements revealed no differences in the distribution of STAG1 and STAG2 when considered independently (Supplementary Figure S1B). STAG1-enriched positions were comparatively more abundant in intergenic regions whereas common and STAG2-enriched peaks showed higher overlap with promoters, exons, and 5′ untranslated r[+egions (5′ UTR) (Figure 1C and Supplementary Figure S1B). We then investigated cohesin enrichment over 9 chromatin states defined by combinations of histone modifications and CTCF in RT112 cells (Supplementary Figure S1C). Similar to the overlap with genomic features, distribution of STAG1 and STAG2 within alternative chromatin states was highly comparable, unless their relative enrichment was taken into consideration (Figure 1D and Supplementary Figure S1D). STAG1-enriched positions characteristically overlapped with transcriptionally inactive chromatin marked by H3K27me3 ('ReprPC'), H3K9me3 ('Het'), or low levels of the assessed histone modifications ('Low'), in addition to chromatin domains bound by CTCF ('CTCF/EnhW'). Conversely, both common and STAG2-enriched sites were mostly distributed over transcriptionally active genes, promoters, and enhancers (Figure 1D). Peak-centered density plots highlight the differential dependency of the three categories of cohesin-bound genomic positions on CTCF (Figure 1E). Consistently, motif analysis showed that STAG1-enriched positions were significantly enriched for the CTCF binding motif, whereas STAG2-enriched positions displayed binding motifs of transcription factors participating in cancer, including ASCL1 and KLF5 (Supplementary Figure S1E) (50–52). We expanded the characterization of the three types of cohesin-binding sites by computing the overlap with repeat elements and found that, overall, cohesin sites are depleted of such sequences (Supplementary Figure S1F). Furthermore, no remarkable differences were detected between the sets of cohesin target sites for most repeat elements apart from telomeres and SINE-VNTR-Alus (SVA). The depletion of STAG2-enriched sites from telomeric sequences is consistent with prior observations on the differential requirements between STAG1 and STAG2 for the maintenance of telomere and centromere cohesion, respectively (16). The positive enrichment of SVAs among STAG2-enriched sites is of interest since these are evolutionary young repetitive elements that occasionally insert into genes and cause disease (53). Overall, our findings on the differential distribution of cohesin-binding sites over (epi)genomic features are largely in agreement with previous results in other cell types (30,31). However, unlike in mES cells (31), we did not detect higher overlap with H3K27me3 domains among STAG2-enriched sites (Figure 1D).

Chromosomal compartments and TAD boundaries are resilient to STAG2 depletion

To assess the contribution of STAG2 to chromatin architecture and transcriptional regulation in RT112 cells, we efficiently silenced STAG2 with two shRNAs (sh1 and sh2)—using a non-targeting shRNA as control (shNT) (Figure 2A)—and performed Hi-C and RNA-Seq experiments. Reproducibility among Hi-C datasets, measured with a stratum-adjusted correlation coefficient (SCC) (45), was higher when comparing sh1 versus sh2 than when comparing each of the shRNAs with the control condition, suggesting that the changes in chromatin 3D configuration introduced by the two independent shRNAs are comparable (Figure 2B).

Given the role of cohesin in sister chromatid segregation (54), we first assessed whether STAG2 depletion affected ploidy. Visual exploration of the number of reads per bin in genome-wide 100 kb contact matrices did not unveil gross genomic differences in ploidy between control and STAG2-depleted cells (Supplementary Figure S2).

To determine whether STAG2 loss resulted in major changes in chromatin organization, we first explored the effect on genomic compartments. The genome is segregated into two major compartments, named A and B, that differ in their gene density, epigenetic modifications, and transcriptional output. Overall, the A compartment contains transcribed genes and active histone modifications while the B compartment encompasses lesser genes in a transcriptionally repressed state (21,22). A and B compartments were defined using the first principal component (PC) obtained by eigenvector decomposition of normalized Hi-C matrices at 100 kb resolution (Figure 2C and D) in combination with information on gene density and transcriptional activity (Figure 2E). In control cells, genomic regions assigned to compartment A were comparatively gene-rich (7760 genes in A versus 2491 genes in B) and genes therein were expressed at significantly higher levels (Figure 2E). Compartment strength was comparable between shNT and shSTAG2 RT112 cells (Figure 2F) and segregation into A/B compartments was highly correlated, with 45.3% of the genome consisting of constitutive A-type domains and 54.7% classified as B-type (Figure 2G). In agreement with the preferential association of cohesin with genes and their regulatory elements, most of its binding sites were found in compartment A, and no major differences were observed between the three categories of cohesin-bound genomic sites (common: 67%, STAG1 enriched: 66%, STAG2 enriched: 67%). Upon STAG2 knock-down, only 0.6–0.8% and 0.75–1.2% of the genome underwent A-to-B and B-to-A compartment changes, respectively (Figure 2G). A significant degree of overlap was detected between STAG2-silenced cells in terms of genomic regions 'flipping' compartments, suggesting small but consistent effects (Figure 2G). Of those genes spanning regions with differential compartment assignation between control and STAG2-silenced cells (Figure 2G and Supplementary Table S1), only 26–31% were expressed and, in general, their expression level was comparatively low in control RT112 (median expression in FPKM of genes in A: 14.3, B: 11, A to B: 8.6, B to A: 8.8). Gene ontology (GO) analysis of genes transitioning from A

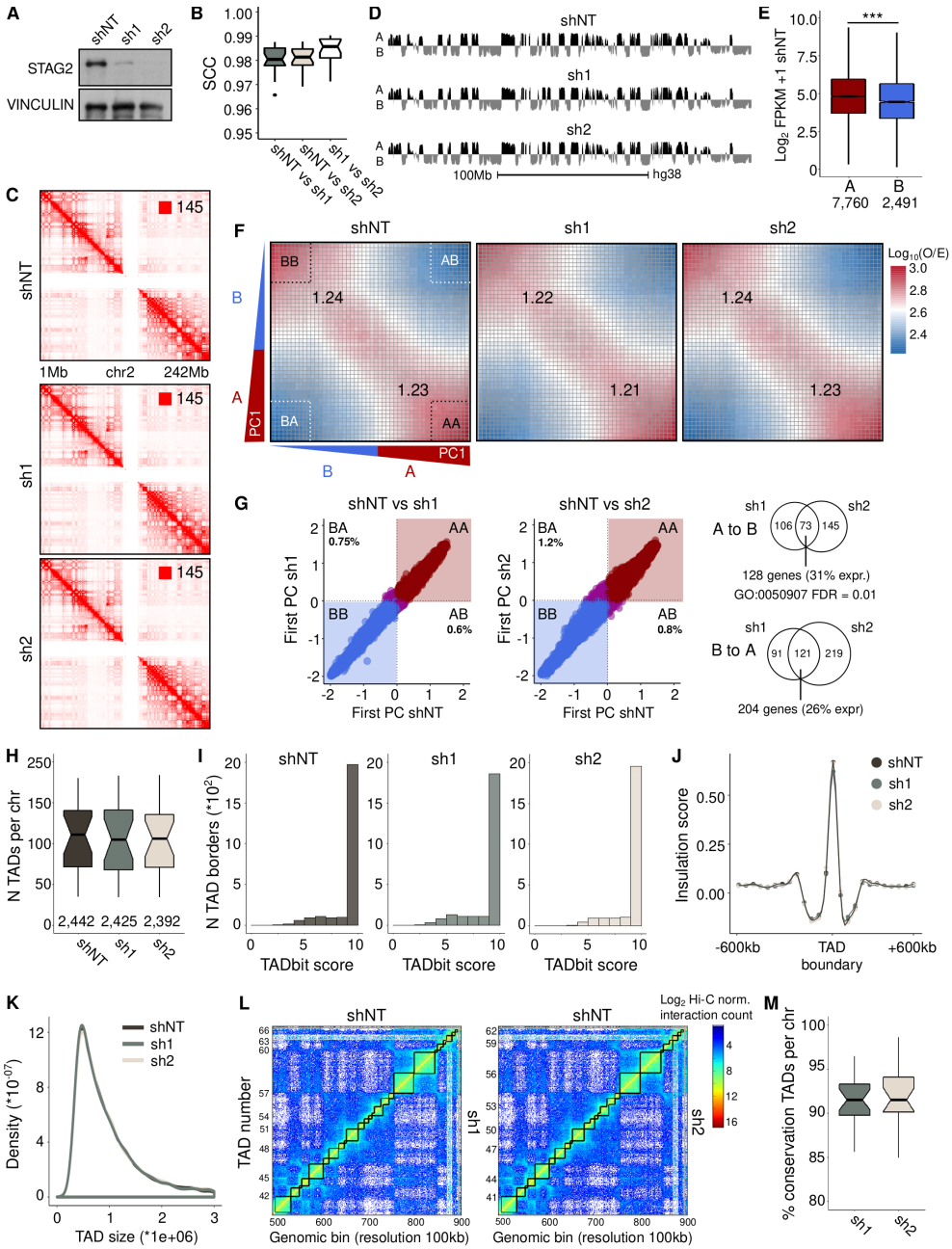

Figure 2. STAG2 loss in RT112 cells does not interfere with A/B compartments or TAD boundaries. (**A**) Western blot analysis of control (shNT) and STAG2-silenced RT112 cells showing efficient depletion of STAG2 at the protein level. (**B**) Reproducibility, measured by a stratum-adjusted correlation coefficient (SCC), between pair-wise comparisons of Hi-C datasets. (**C**) Hi-C matrices for chr2 at 500 kb resolution in cells transduced with control or STAG2-targeting shRNAs. The darker red reflects a greater frequency of interaction. (**D**) Compartment tracks for chr2 at 100 kb resolution as determined by the values of the first principal component (PC1) in control and STAG2-silenced cells. (**E**) Expression, as defined by RNA-Seq ($\log_2$ FPKM), of genes within compartments A and B. As expected, genes assigned to compartment A are more transcriptionally active than genes in compartment B. *t*-test: $^{***}P < 0.001$. (**F**) Compartmentalization saddle plots: average intra-chromosomal interaction frequencies between 200kb bins, normalized by expected interaction frequency based on genomic distance. Bins are sorted by their PC1 value derived from control cells Hi-C data. Preferential B-B interactions are in the upper left corner, and preferential A-A interactions are in the lower right corner. Numbers in corners represent the strength of AA interactions as compared to AB interactions and BB interactions over BA interactions. (**G**) Scatterplot of PC1 values of the eigenvectors of intrachromosomal interaction matrices for control and STAG2-silenced cells. The Venn diagrams show the overlap in terms of compartment-switching bins between sh1 and sh2. The number of genes mapped to genomic bins switching compartments is also indicated. Only one GO term is significantly enriched (FDR < 0.01) among genes switching from A to B: GO:0050907 (detection of chemical stimulus involved in sensory perception). (**H**) Effect of STAG2-depletion on the number of TADs per chromosome. Boxplot notches represent the confidence interval around the median. The number of total TADs is indicated below the boxplots. (**I**) Histograms depicting the strength of the TAD borders detected in control and STAG2-silenced cells, according to the TADbit score. (**J**) Average insulation profile around TAD boundaries (±600 kb) in control and STAG2-silenced cells. (**K**) Density plot depicting the distribution of TAD sizes identified in control and STAG2-silenced cells. (**L**) Hi-C normalized interaction matrices for chr2 at 100kb resolution comparing TAD organization in control and STAG2-silenced cells. (**M**) Effect of STAG2-depletion on conservation of TAD borders. Boxplot notches represent the confidence interval around the median.

to B revealed only one significantly enriched pathway with little connection to bladder cancer biology: GO:0050907 (detection of chemical stimulus involved in sensory perception). The significant enrichment for this GO term results from its constituent genes clustering linearly in the genome and therefore being embedded in one chromatin fragment with differential compartment identity.

We then explored the possibility that STAG2 depletion might interfere with the organization of the genome into TADs. Using normalized 100kb contact maps, we identified a total of 2442, 2425 and 2392 TAD borders in shNT, sh1, and sh2 cells, respectively (Figure 2H). The strength of the TAD borders was comparable among conditions as shown by the TADbit and insulation scores (Figure 2I and J). The size of TADs was also similar among conditions (Figure 2K) and, as revealed by alignment of their boundaries, they were highly conserved (average conservation with respect to shNT: sh1 90.8%, sh2 91.5%) (Figure 2L and M). The small decrease in TAD number in STAG2-depleted cells might result from merging of adjacent TADs (Figure 2L). In agreement with previous work (30,31,33,49), our results indicate that megabase-scale architectural compartments and TAD borders are resilient to reduced STAG2 protein levels and can be sustained by cohesin-STAG1 alone.

Loss of STAG2 leads to rewiring of DNA loops

Cohesin contributes to the 3D conformation of chromatin at the submegabase scale in a cell-type specific fashion through both long-range constitutive interactions and short-range promoter-enhancer contacts that regulate transcription (55). STAG2 knock-down resulted in modest, statistically significant (FDR < 0.05), changes in expression levels of a subset of genes (sh1, $n = 510$; sh2, $n = 438$), with a similar number of up- and down-regulated transcripts (Figure 3A). Gene expression changes induced by either shRNA were positively correlated and 20–32% of significantly up- and down-regulated genes were common to both shRNAs (Figure 3B). Conditional *Stag2* deletion in the mouse blood compartment is associated with transcriptional dysregulation and impaired differentiation of hematopoietic stem cells (32). We thus explored the effects of depleting STAG2 on RT112 differentiation by assessing the differential expression of gene signatures characteristic of muscle-invasive UBC molecular subtypes (36). GSEA (Gene Set Enrichment Analysis) revealed consistent and significant up-regulation of genes linked to the basal/squamous class, and a significant down-regulation of genes linked to the luminal papillary class, in cells transduced with either STAG2-targeting shRNA (Figure 3C). These findings suggest that STAG2 loss leads to defective maintenance of the luminal differentiation transcriptional program (Figure 3C). GSEA further revealed a significant overlap between genes down-regulated in STAG2-silenced RT112 cells and genes differentially expressed in '*STAG2* low' (bottom quartile) versus '*STAG2* high' (top quartile) samples from the UROMOL study, involving 476 cases of non-muscle invasive bladder cancer (56) (Figure 3D–G), thus validating the relevance of our *in vitro* system.

We hypothesized that STAG2 knockdown might affect transcription by interfering with formation of chromatin loops engaging specific promoters and their regulatory regions. To address this question, we first identified significant DNA contacts at 20 kb resolution with HOMER (40) in control and STAG2-depleted cells and observed an overall increase in the genomic distance spanned by DNA loops in STAG2-silenced cells resulting from a loss of short-range (<250 kb) and a concomitant increase in long-range interactions (>1 Mb) (Figure 4A and B).

To dissect the effects on loop formation, we intersected the interactions in control cells with the three subsets of cohesin-binding sites identified in our ChIP-Seq experiments -common, STAG1-enriched, and STAG2-enriched-. Interactions overlapping STAG2-enriched binding sites displayed some unique characteristics, indicating functional compartmentalization. Pairwise enrichment analysis showed that DNA sequences bound by STAG2-cohesin interact more frequently with regions bound by STAG2-cohesin than by other cohesin complex combinations, suggesting a certain degree of spatial segregation between the DNA loops established by the different cohesin complexes (Figure 4C). Cohesin-overlapping interactions were, overall, of high interaction frequency and spanned genomic distances within 1Mb, or the average TAD size (Figure 4D and E). Noticeably, STAG2-enriched overlapping DNA loops displayed the highest interaction frequencies and were distinctively short-ranged (Figure 4D and E). In addition, while all cohesin-overlapping interactions showed a preference for engaging genes and preferentially accumulated in compartment A (Figure 4F and G), those associated with STAG2-enriched cohesin showed the highest enrichment in promoters and overlapped with genes expressed at significantly higher levels than those involved in interactions overlapping other cohesin binding sites (Figure 4H).

We next defined a collection of interactions significantly 'lost' and 'gained' upon knockdown with both shRNAs (adjusted $P < 0.05$ and $\log_2 FC < -1 \mid \log_2 FC > 1$) (Figure 5A). Importantly, changes in interaction frequency of DNA contacts in cells transduced with sh1 were positively and significantly correlated with changes in cells transduced with sh2, further supporting the reproducibility of the Hi-C data (Figure 5B). A total of 1686 lost and 1881 gained interactions were identified that showed consistent changes in both STAG2-targeting shRNAs (Figure 5C). Of note, lost and gained contacts showed differential properties, with lost interactions spanning shorter genomic distances than gained interactions (Figure 5D). Motif analysis of genomic regions engaged by control, lost, and gained interactions further revealed that motifs for CTCF and its paralog CTCFL were among the top four motifs enriched in control and gained interactions but not among those lost (Figure 5E and Supplementary Table S2). Lastly, both lost and gained interactions were distributed throughout A/B compartments and chromatin states (Figure 5F and G), but lost interactions showed comparatively higher overlap with the B compartment (Figure 5F) and constitutive and facultative heterochromatin ('ReprPC' and 'Het') (Figure 5G).

To assess whether the loci that lose or gain interactions upon STAG2 silencing are located close to each other in 3D, we performed a chromatin assortativity analysis whereby the genome is represented as a network of interacting nodes and each node is a chromatin fragment (48). Loci involved

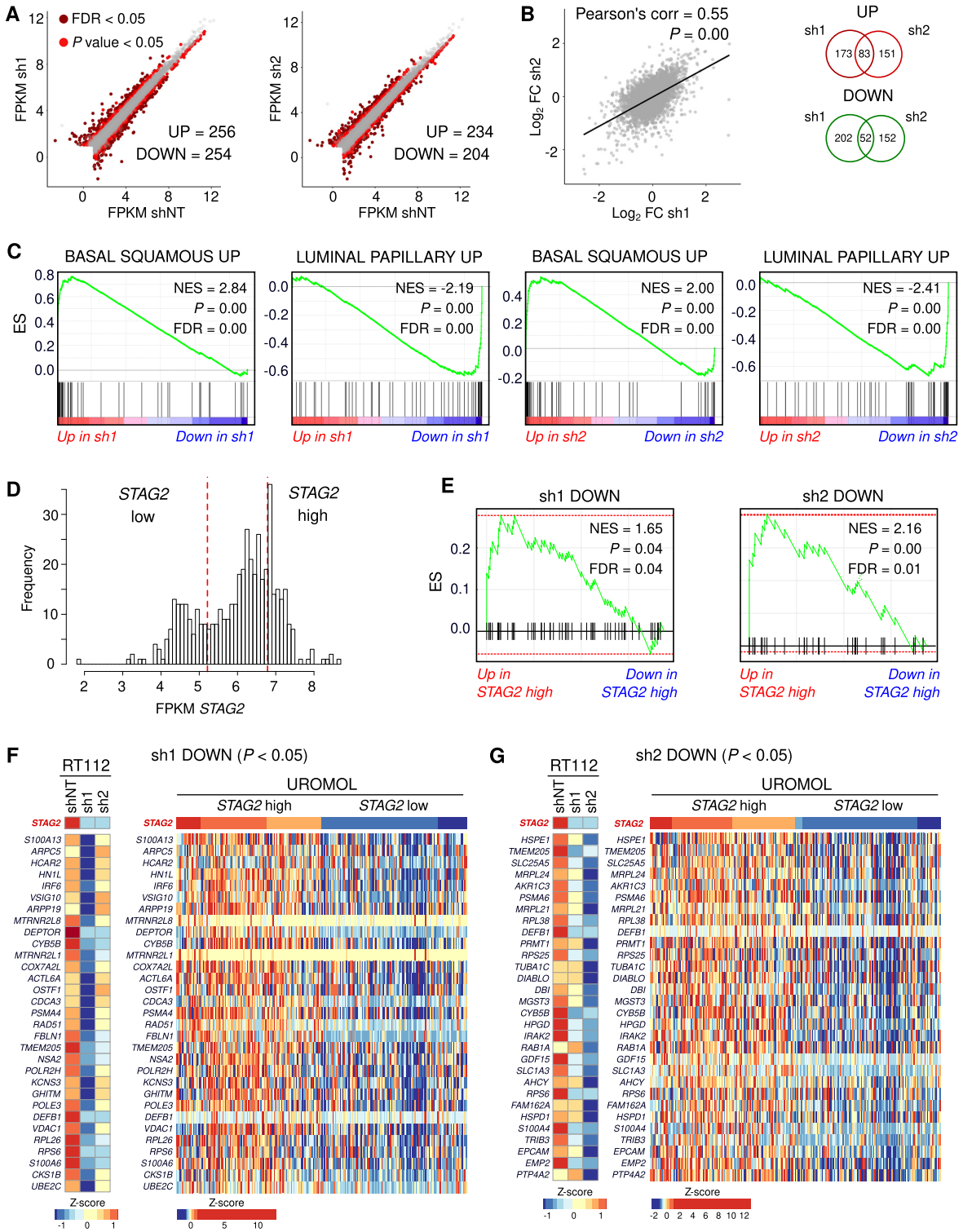

Figure 3. STAG2 depletion leads to deregulation of the basal/luminal transcriptional programs in RT112. (**A**) Scatter plots of expression values (FPKM) of genes in control versus STAG2-silenced cells. Statistically significant differentially expressed genes are highlighted in dark (FDR < 0.05) or light red ($P < 0.05$). (**B**) Scatter plot showing a positive and significant correlation between gene expression changes in sh1 and sh2 (left). Venn diagrams displaying the overlap between sh1 and sh2 in terms of significant up- and down-regulated genes. (**C**) GSEA enrichment plots of gene sets associated with the luminal and basal subtypes of muscle-invasive UBC showing significant deregulation in STAG2-silenced RT112 cells. (**D**) Distribution of STAG2 expression (FPKM) in the UROMOL cohort of 476 UBC samples (56), highlighting the thresholds of the first and fourth quartiles (119 samples per group). We defined 'STAG2 high' cases as those with expression values in the fourth quartile, and 'STAG2 low' cases as those with STAG2 levels in the first quartile. (**E**) GSEA enrichment plots for genes down-regulated in STAG2-silenced cells in 'STAG2 high' versus 'STAG2 low' tumor samples. (**F**, **G**) Heatmaps displaying relative expression values (Z-score of FPKM) of genes significantly down-regulated in RT112 cells with sh1 (F) or sh2 (G) and in 'STAG2 low' versus 'STAG2 high' tumor samples.

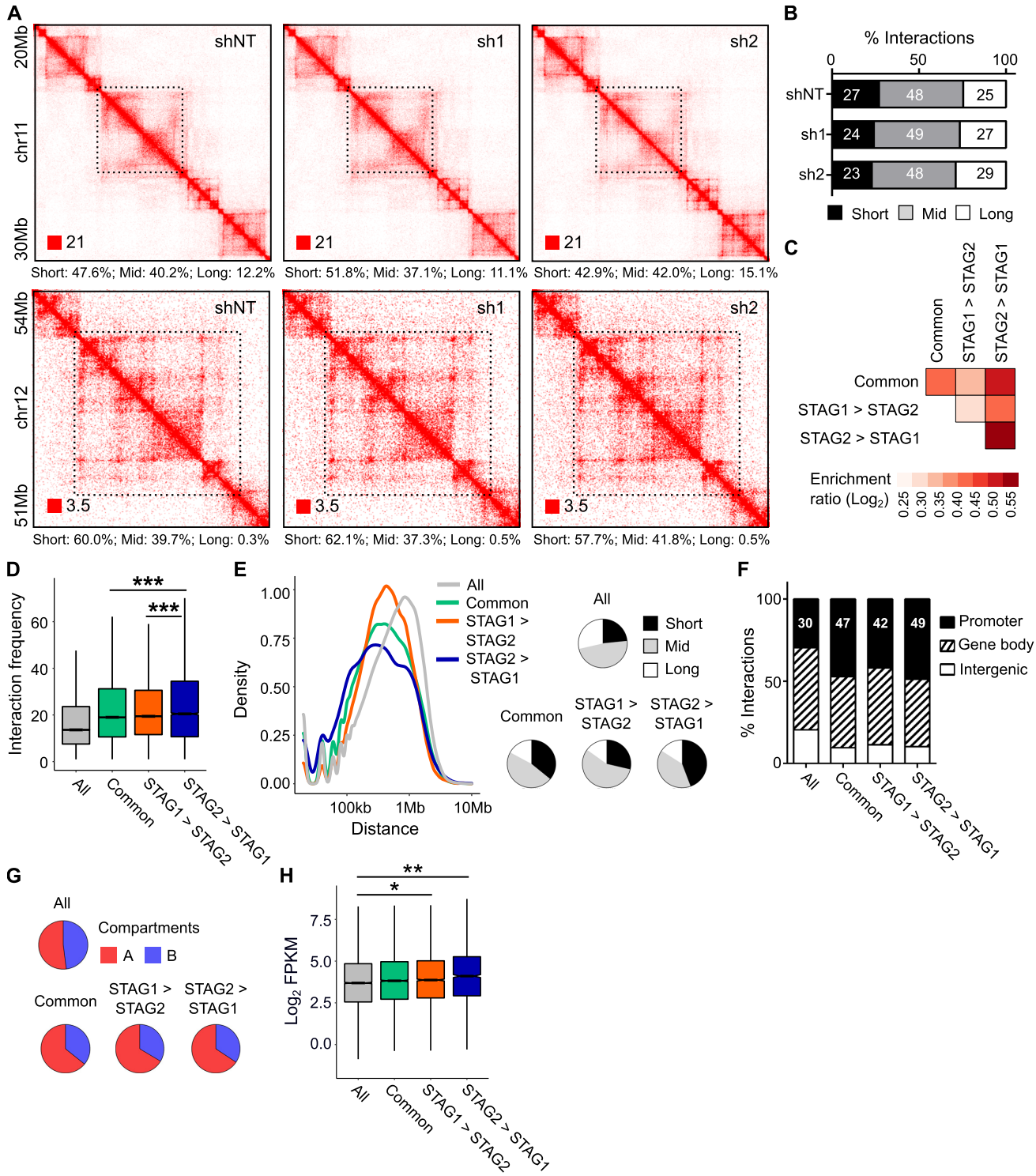

Figure 4. STAG2-enriched cohesin mediates short and mid-range contacts that engage actively transcribed genes. (**A**) Snapshots of coverage-normalized contact matrices at 25 kb (upper) and 50 kb (lower) resolution in control and STAG2-depleted cells illustrating the loss of short-ranged contacts and the formation of long-ranged contacts upon STAG2 silencing. The darker red reflects a greater frequency of interaction. Contact matrices were visualized in Juicebox (59). Below each contact matrix, the proportion of short (<250 kb), mid (250 kb–1 Mb) and long (>1 Mb) interactions for that genomic window is shown. (**B**) Proportion of short, mid and long interactions in control and STAG2-silenced cells. (**C**) Pairwise feature enrichment at the anchors of interactions overlapping cohesin-binding sites. (**D**) Interaction frequency of DNA contacts overlapping cohesin binding sites in control cells. (**E**) Left: density plot showing the distribution of genomic distances spanned by interactions overlapping cohesin binding sites in control cells. The distribution of genomic distances spanned by all interactions is included as control. Right: pie charts plotting the proportion of short (<250 kb), mid (250 kb–1 Mb), and long-ranged (>1 Mb) contacts among interactions overlapping cohesin binding sites. (**F**) Distribution of all and cohesin-overlapping interactions, in control cells, over genomic elements and (**G**) A/B compartments. (**H**) RNA-Seq expression values ($\log_2$ FPKM) of genes engaged by all or cohesin-overlapping interactions, in control cells. *t*-test: $^*P < 0.05$; $^{**}P < 0.01$.

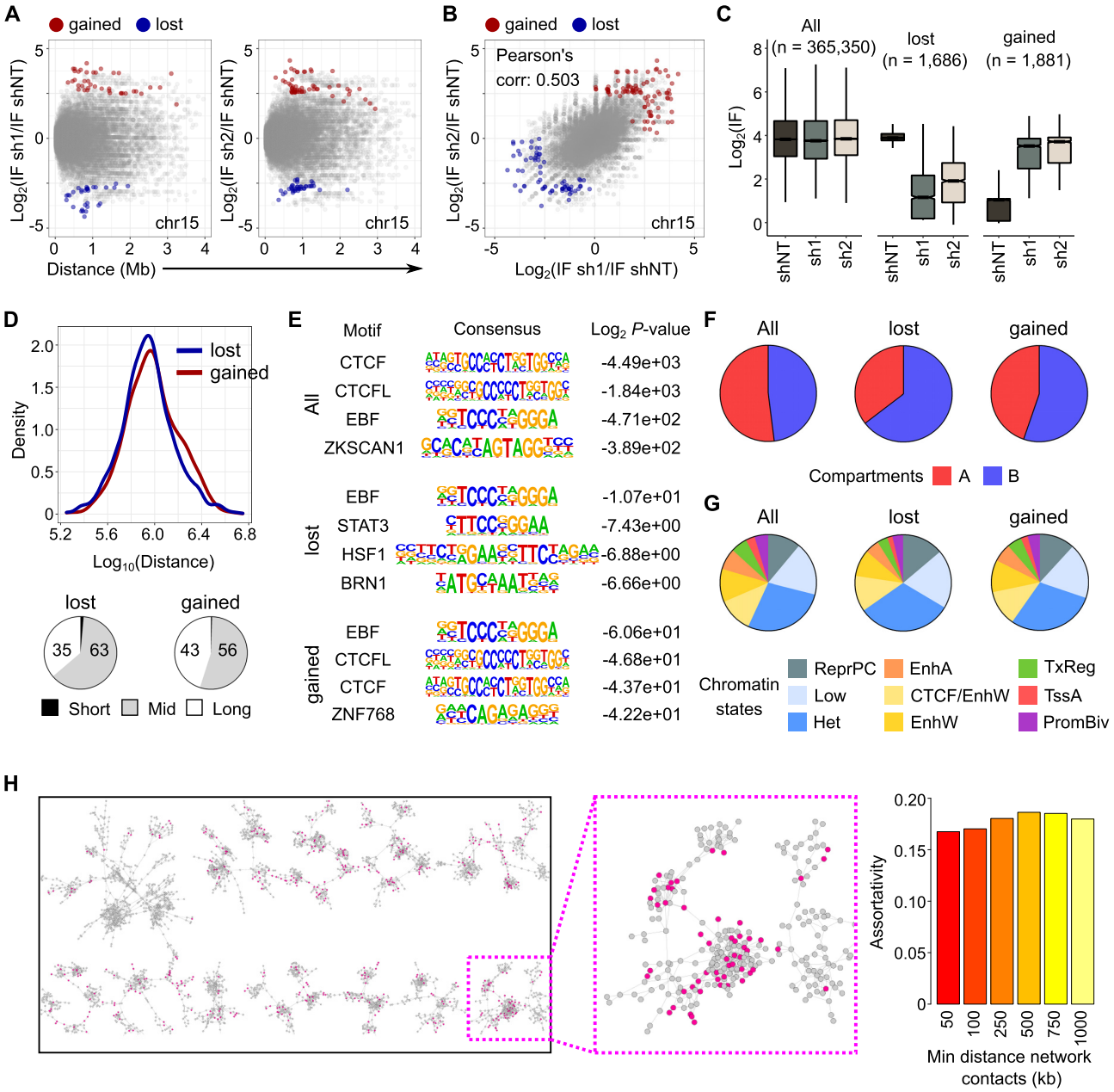

Figure 5. STAG2 silencing is accompanied by rewiring of DNA contacts. (**A**) MD plots depicting the fold change in interaction frequency of DNA contacts in control (shNT) versus STAG2-silenced cells in relation to the linear distance between interacting regions for chromosome 15. Interactions showing statistically significant differences (adjusted $P < 0.05$) are highlighted in red (gained) or blue (lost). (**B**) Scatter plot comparing changes in interaction frequency between cells transduced with sh1 versus cells transduced with sh2. Gained and lost interactions are highlighted in red and blue, respectively. (**C**) Interaction frequency of lost and gained contacts in shNT and STAG2-silenced cells. (**D**) Distance between peaks of lost and gained DNA contacts. Gained interactions span longer distances than lost interactions. (**E**) Top four scoring motifs enriched in the subsets of interactions defined in C. For a more extensive list of significantly enriched motifs, see Supplementary Table S2. (**F**) Distribution of control, lost, and gained interactions over compartments and (**G**) chromatin states in RT112 cells (see Supplementary Figure S1C). (**H**) Left: chromatin contact network generated from the 20kb resolution Hi-C interaction map of control cells, showing in pink the nodes involved in contacts that are lost upon STAG2 silencing. Right: chromatin assortativity of nodes that lose contacts as the network is filtered eliminating contacts spanning short distances.

in lost interactions were more assortative than expected by chance, suggesting that they tend to be close to each other in the 3D space. Assortativity of these regions was stable on networks in which contacts spanning shorter distances were eliminated, reaching a maximum for interactions that span less than 1Mb (intra-TAD) (Figure 5H). On the contrary, gained interactions did not show any patterns of assortativity, suggesting a lack of a functional consistency between loci affected by gained interactions.

The differences in genomic distance, motif enrichment, and assortativity between lost and gained interactions suggest that loss of interactions would result from direct loss of STAG2-cohesin, which mediates formation of short and mid-range DNA loops without participation of CTCF. By contrast, gained interactions would likely arise from redistribution or increased accumulation at pre-existing cohesin-binding sites of STAG1-cohesin, which is involved in formation of longer DNA loops together with CTCF.

Differences in chromatin looping arising from loss of STAG2 result in transcriptional changes

We next explored whether differential chromatin looping translated into transcriptional changes. First, we conducted a motif analysis of genomic regions spanned by lost and gained interactions and compared the results to the transcription factor target sequences enriched among significantly de-regulated genes in STAG2-silenced cells (Figure 6A). Some motifs were only enriched in regions involved in differential looping but not in de-regulated genes (collection #1) and, vice versa, other motifs were only significant among differentially expressed genes (collection #2). Interestingly, we could also detect a high proportion of motifs enriched in both differential DNA loops and up-regulated genes (collection #3). Overall, this analysis suggests that rewiring of chromatin looping is concomitant with, at least partially, transcriptional changes. In addition, it indicates that not all alterations in gene expression found in STAG2-depleted cells are due to differential DNA looping but rather to secondary epigenetic events.

Intersection of the contact and RNA-Seq data showed that lost interactions overlapped promoters and bodies of genes consistently expressed at low levels (Figure 6B and C). Interestingly, loss of DNA contacts engaging promoters, but not gene bodies, led to statistically significant up-regulation of gene expression (Figure 6B and C). Gained interactions overlapping promoters also occurred on low-expressed genes but had no measurable impact on gene expression (Figure 6B). In contrast, genes displaying increased interactions throughout their gene bodies were of average expression levels and became significantly down-regulated in STAG2-silenced cells (Figure 6C). These effects are exemplified by *TNC and COL17A1*, two genes associated with the basal/squamous UBC molecular subtype. Both genes are expressed at low levels, are targeted by PRC2, and, upon STAG2 silencing, display loss of interactions around their promoter regions that is concomitant to transcriptional up-regulation (Figure 6D and E).

DISCUSSION

The mechanisms whereby STAG2 acts as a tumor suppressor gene and contributes to cancer are not well established and are likely to be diverse. The lack of association between *STAG2* inactivation and aneuploidy/genomic instability in AML and UBC strongly suggests its participation through effects other than chromosome segregation (4,19). Recent evidence on the role of cohesin in higher-order chromatin structure and on the distinct functions of STAG1 and STAG2 in several cell types, mainly in the haematopoietic lineage, has provided support to the hypothesis that changes in gene expression may play a crucial role in the tumor suppressive role of STAG2. Yet, this notion is challenged by the fact that, in numerous cellular systems, suppression of STAG2 activity results in only modest changes at the global transcriptome level (30,31,33).

Importantly, these questions have not been addressed in UBC, the tumor with the highest prevalence of *STAG2* mutations. A major limitation has been the lack of adequate models. Most UBC lines are derived from muscle-invasive UBC and, therefore, few of them harbour *STAG2* mutations. Furthermore, until recently it has not been possible to permanently maintain normal urothelial cells in culture (57). Therefore, we aimed at assessing the effects of STAG2 knockdown on one of the most commonly used luminal-type UBC lines: RT112. An important finding of our study is that the small fraction of the transcriptome undergoing changes upon STAG2 silencing overlaps significantly with genes that are differentially expressed in UBC with low vs. high STAG2 levels. These observations strongly support the suitability of RT112 cells to explore the mechanisms through which STAG2 contributes to UBC.

We find that among RT112 cells—as in other cell types—a small subset of cohesin-bound sites are STAG2-enriched. These genomic sites are over-represented in promoters and enhancers of transcriptionally active genes and their sequences are depleted of CTCF while being enriched in tissue-specific transcription factors (i.e. TFAP2A, KLF5). We also find that STAG2 depletion does not result in major changes in the genome compartmentalization and has limited effects on megabase organization except for the merging of a limited number of adjacent TADs. However, DNA loops spanning varying distances are rewired in the absence of STAG2, in agreement with observations made in other cellular systems (30–33,49). An equivalent number of interactions is lost or gained and, interestingly, the two sets of contacts display differential features: lost interactions span shorter genomic distances, are depleted of CTCF motifs, and are more assortative than gained interactions. This suggests that, while lost interactions result from direct loss of STAG2-cohesin, gained interactions arise from redistribution of the remaining STAG1-cohesin complexes. Supporting this notion, increased STAG1 levels were detected at common cohesin binding sites upon silencing of STAG2 (30). These structural changes appear to be functionally relevant, with genes whose promoters are associated with lost interactions being up-regulated in STAG2-depleted cells. Recently, it was reported that STAG2-cohesin promotes PRC1 recruitment and thereby contributes to Polycomb domain compaction and formation of long-range contacts

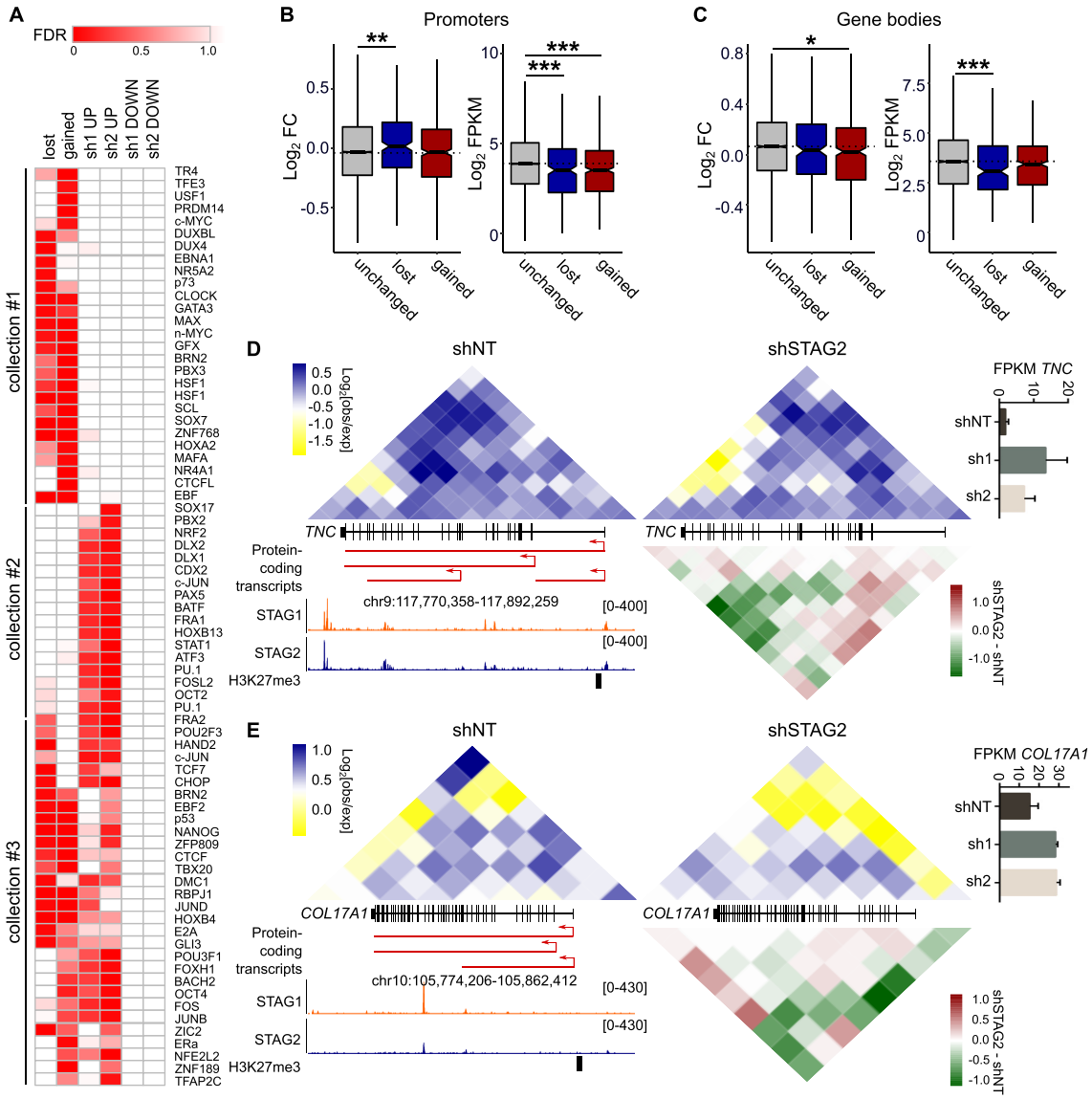

Figure 6. Rewiring of chromatin looping arising from loss of STAG2 results in transcriptional changes. (**A**) Heatmap representing the statistical significance (FDR) of motif analyses of promoters of differentially expressed genes and genomic positions involved in differential DNA looping in STAG2-silenced cells. We identify three collections of motifs: #1, motifs enriched in differential interactions; #2, motifs enriched in differentially expressed genes; and #3, motifs enriched in both interactions and transcriptionally deregulated genes. (**B**) Average fold change in gene expression values (FPKM) of genes engaged by control and differential interactions overlapping promoters or (**C**) gene bodies. Boxplot notches represent the confidence interval around the median. *t*-test: $*P < 0.05$; $**P < 0.01$; $***P < 0.001$. (**D**, **E**) Hi-C contact matrices at the *TNC* and COL17A1 loci in control and STAG2-silenced cells. The contact matrices for the *STAG2*-silenced condition are the result of averaging the matrices for sh1 and sh2. Snapshots of the ChIP-Seq tracks for STAG1 and STAG2, differential contact matrices, H3K27me3 peaks, and gene expression values (FPKM) are included. Loss of interactions overlapping the promoter of *TNC* and *COL17A1* upon STAG2 silencing correlates with a consistent increase in gene expression. Error bars represent mean ± SEM.

between those domains in mouse ES cells (31). Consistently, contacts lost upon *STAG2* down-regulation in our system preferentially overlap the transcriptionally silent B compartment and heterochromatin domains, in addition to engaging low expressed genes. Overall, our data suggests that loss of STAG2 in bladder cancer cells leads to loss of short and mid-ranged DNA contacts, mostly in transcriptionally silent chromatin domains, leading to decompaction and transcriptional derepression of lineage-specifying Polycomb target genes such as those of the basal transcriptional program.

Interestingly, two of the transcription factors whose binding motifs are enriched in STAG2-only positions - TFAP2A and KLF5 - show a tissue-restricted expression pattern, with high levels in squamous epithelia such as the skin and the esophagus (https://www.gtexportal.org). Activation of basal/squamous differentiation programs is a feature of a specific subset of UBC that displays down-regulation of luminal genes, reflecting loss-of-identity, and designated as basal/squamous-like (BASQ). There is increasing evidence that, in several epithelial tumors, the canonical vs. basal programs are regulated in a complex manner through epige-

netic mechanisms and appear as a continuum rather than as discrete phenotypes. In bladder cancer cells, TFAP2A is repressed by PPARg, a major regulator of luminal-type tumors, and is up-regulated in BASQ-type tumors (58). In addition, the KLF4-driven regulon is selectively activated in BASQ tumors (36). Despite the association of *STAG2* mutations with papillary tumors, STAG2 knockdown resulted in the down-regulation of the luminal signature. These apparently paradoxical findings are, nevertheless, in agreement with the observation that, among MIBC, the UBC subtype displaying the highest prevalence of *STAG2* mutations also shows a higher activation score of the basal differentiation signature than other luminal tumors (36). Furthermore, among NMIBC, *STAG2* mutations are enriched in class 3 tumors which - while displaying luminal urothelial features - are enriched in basal markers (56). We therefore hypothesize that STAG2 plays a tumor suppressor role by establishing and/or maintaining the DNA looping required for the Polycomb-mediated repression of the basal differentiation program in urothelial cells and thus stabilize their luminal identity.

DATA AVAILABILITY

ChIP-Seq data on cohesin subunits has been deposited in the GEO database under accession number GSE111913. ChIP-Seq data on histone modifications and CTCF for RT112 are deposited in the GEO database under accession number GSE104804. RNA-Seq data for RT112 cells next-generation sequencing data are available in NCBI GEO, under accession number GSE154878. Hi-C data are available in NCBI GEO, under accession number GSE155380. The authors declare that all other data supporting the findings of this study are within the manuscript and its supplementary files.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank all the staff of the GenomEast high throughput sequencing facility of the IGBMC in Strasbourg (CNRS UMR7104, Inserm U1258, Université de Strasbourg) for the sequencing of ChIPs of histone marks and CTCF; the UROMOL investigators and Lars Dyrskjot for providing RNA-Seq data; Ana Losada for providing reagents; and Mark Kalisz, Ana Losada, and Ana Cuadrado for critical review of the manuscript.

Author contributions: L.R., E.L., M.A.M.-R. and F.X.R. initiated and designed the studies. L.R. and Y.C. performed the Hi-C experiment and L.R. analysed and integrated it with the RNA-Seq and ChIP-Seq data. E.L. and M.C. performed RNA-Seq and cohesin ChIP-Seq experiments and E.C.P., E.L. and L.R. analysed the data. V.P., M.M.M. and J.A.R. performed the assortativity analyses and helped with the comparison of the RNA-Seq data on RT112 cells and the human tumor data. HNK and FR conducted the ChIP-Seq experiments for histone modifications and CTCF and the analysis of chromatin states in RT112 cells. F.L.D. supervised and provided guidance on the preparation of the Hi-C libraries. F.S. supervised and provided guidance on the analysis of the Hi-C data. L.R., E.L., V.P., M.A.M.-R. and F.X.R. interpreted the results and wrote the manuscript. M.A.M.-R. and F.X.R. supervised the whole research and provided guidance. All authors had access to the final manuscript and approved the submission of the final version of the paper.

FUNDING

Fundación Científica de la Asociación Española Contra el Cáncer (to F.X.R., E.L., in part); V.P. is supported by INSERM, the Fondation Toulouse Cancer Santé and Pierre Fabre Research Institute as part of the Chair of Bioinformatics in Oncology of the CRCT; Bioinfo4women programme at the Barcelona Supercomputing Center; European Union's H2020 Framework Programme through the ERC [609989 to M.A.M.-R., in part]; Spanish Ministerio de Ciencia, Innovación y Universidades [BFU2017-85926-P to M.A.M.-R.]; C.N.I.O. is supported by Ministerio de Ciencia, Innovación y Universidades as a Centro de Excelencia Severo Ochoa [SEV-2015-0510]; C.R.G. acknowledges support from 'Centro de Excelencia Severo Ochoa 2013–2017' [SEV-2012-0208]; Spanish ministry of Science and Innovation to the EMBL partnership and the CERCA Programme/Generalitat de Catalunya (to C.R.G.); C.R.G. also acknowledges support of the Spanish Ministry of Science and Innovation through the Instituto de Salud Carlos III, the Generalitat de Catalunya through Departament de Salut and Departament d'Empresa i Coneixement; Spanish Ministry of Science and Innovation with funds from the European Regional Development Fund (ERDF) corresponding to the 2014–2020 Smart Growth Operating Program (to C.N.A.G.). Funding for open access charge: Own funds.

Conflict of interest statement. None declared.

REFERENCES

1. Lawrence,M.S., Stojanov,P., Mermel,C.H., Robinson,J.T., Garraway,L.A., Golub,T.R., Meyerson,M., Gabriel,S.B., Lander,E.S. and Getz,G. (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*, **505**, 495–501.
2. Romero-Pérez,L., Surdez,D., Brunet,E., Delattre,O. and Grünewald,T.G.P. (2019) STAG mutations in cancer. *Trends Cancer*, **5**, 506–520.
3. Solomon,D.A., Kim,T., Diaz-Martinez,L.A., Fair,J., Elkahloun,A.G., Harris,B.T., Toretsky,J.A., Rosenberg,S.A., Shukla,N., Ladanyi,M. *et al.* (2011) Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science*, **333**, 1039–1043.
4. Balbás-Martínez,C., Sagrera,A., Carrillo-de-Santa-Pau,E., Earl,J., Márquez,M., Vazquez,M., Lapi,E., Castro-Giner,F., Beltran,S., Bayés,M. *et al.* (2013) Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. *Nat. Genet.*, **45**, 1464–1469.
5. Solomon,D.A., Kim,J.-S., Bondaruk,J., Shariat,S.F., Wang,Z.-F., Elkahloun,A.G., Ozawa,T., Gerard,J., Zhuang,D., Zhang,S. *et al.* (2013) Frequent truncating mutations of STAG2 in bladder cancer. *Nat. Genet.*, **45**, 1428–1430.
6. Guo,G., Sun,X., Chen,C., Wu,S., Huang,P., Li,Z., Dean,M., Huang,Y., Jia,W., Zhou,Q. *et al.* (2013) 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.
7. Taylor,C.F., Platt,F.M., Hurst,C.D., Thygesen,H.H. and Knowles,M.A. (2014) 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.

8. Brohl,A.S., Solomon,D.A., Chang,W., Wang,J., Song,Y., Sindiri,S., Patidar,R., Hurd,L., Chen,L., Shern,J.F. *et al.* (2014) The genomic landscape of the Ewing Sarcoma family of tumors reveals recurrent STAG2 mutation. *PLoS Genet.*, **10**, e1004475.
9. Crompton,B.D., Stewart,C., Taylor-Weiner,A., Alexe,G., Kurek,K.C., Calicchio,M.L., Kiezun,A., Carter,S.L., Shukla,S.A., Mehta,S.S. *et al.* (2014) The genomic landscape of pediatric Ewing sarcoma. *Cancer Discov.*, **4**, 1326–1341.
10. Kon,A., Shih,L.-Y., Minamino,M., Sanada,M., Shiraishi,Y., Nagata,Y., Yoshida,K., Okuno,Y., Bando,M., Nakato,R. *et al.* (2013) Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. *Nat. Genet.*, **45**, 1232–1237.
11. Thota,S., Viny,A.D., Makishima,H., Spitzer,B., Radivoyevitch,T., Przychodzen,B., Sekeres,M.A., Levine,R.L. and Maciejewski,J.P. (2014) Genetic alterations of the cohesin complex genes in myeloid malignancies. *Blood*, **124**, 1790–1798.
12. De Koninck,M. and Losada,A. (2016) Cohesin mutations in cancer. *Cold Spring Harb. Perspect. Med.*, **6**, a026476.
13. Shen,C.-H., Kim,S.H., Trousil,S., Frederick,D.T., Piris,A., Yuan,P., Cai,L., Gu,L., Li,M., Lee,J.H. *et al.* (2016) Loss of cohesin complex components STAG2 or STAG3 confers resistance to BRAF inhibition in melanoma. *Nat. Med.*, **22**, 1056–1061.
14. Kim,Y., Shi,Z., Zhang,H., Finkelstein,I.J. and Yu,H. (2019) Human cohesin compacts DNA by loop extrusion. *Science*, **366**, 1345–1349.
15. Remeseiro,S. and Losada,A. (2013) Cohesin, a chromatin engagement ring. *Curr. Opin. Cell Biol.*, **25**, 63–71.
16. Canudas,S. and Smith,S. (2009) Differential regulation of telomere and centromere cohesion by the Scc3 homologues SA1 and SA2, respectively, in human cells. *J. Cell Biol.*, **187**, 165–173.
17. Remeseiro,S., Cuadrado,A., Gómez-López,G., Pisano,D.G. and Losada,A. (2012) A unique role of cohesin-SA1 in gene regulation and development. *EMBO J.*, **31**, 2090–2102.
18. Remeseiro,S., Cuadrado,A., Carretero,M., Martínez,P., Drosopoulos,W.C., Cañamero,M., Schildkraut,C.L., Blasco,M.A. and Losada,A. (2012) Cohesin-SA1 deficiency drives aneuploidy and tumourigenesis in mice due to impaired replication of telomeres. *EMBO J.*, **31**, 2076–2089.
19. Welch,J.S., Ley,T.J., Link,D.C., Miller,C.A., Larson,D.E., Koboldt,D.C., Wartman,L.D., Lamprecht,T.L., Liu,F., Xia,J. *et al.* (2012) The origin and evolution of mutations in acute myeloid leukemia. *Cell*, **150**, 264–278.
20. Dekker,J., Rippe,K., Dekker,M. and Kleckner,N. (2002) Capturing chromosome conformation. *Science*, **295**, 1306–1311.
21. Lieberman-Aiden,E., van Berkum,N.L., Williams,L., Imakaev,M., Ragoczy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
22. Rowley,M.J. and Corces,V.G. (2018) Organizational principles of 3D genome architecture. *Nat. Rev. Genet.*, **19**, 789–800.
23. Sanborn,A.L., Rao,S.S.P., Huang,S., Durand,N.C., Huntley,M.H., Jewett,A.I., Bochkov,I.D., Chinnappan,D., Cutkosky,A., Li,J. *et al.* (2015) Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E6456–E6465.
24. Guo,Y., Xu,Q., Canzio,D., Shou,J., Li,J., Gorkin,D.U., Jung,I., Wu,H., Zhai,Y., Tang,Y. *et al.* (2015) CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell*, **162**, 900–910.
25. Fudenberg,G., Imakaev,M., Lu,C., Goloborodko,A., Abdennur,N., Mirny,L.A., Fudenberg,G., Imakaev,M., Lu,C., Goloborodko,A. *et al.* (2016) Formation of chromosomal domains by loop extrusion. *CellReports*, **15**, 2038–2049.
26. Rao,S.S.P., Huang,S.-C., Glenn St Hilaire,B., Engreitz,J.M., Perez,E.M., Kieffer-Kwon,K.-R., Sanborn,A.L., Johnstone,S.E., Bascom,G.D., Bochkov,I.D. *et al.* (2017) Cohesin loss eliminates all loop domains. *Cell*, **171**, 305–320.
27. Kim,Y., Shi,Z., Zhang,H., Finkelstein,I.J. and Yu,H. (2019) Human cohesin compacts DNA by loop extrusion. *Science*, **366**, 1345–1349.
28. Kagey,M.H., Newman,J.J., Bilodeau,S., Zhan,Y., Orlando,D.A., van Berkum,N.L., Ebmeier,C.C., Goossens,J., Rahl,P.B., Levine,S.S. *et al.* (2010) Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, **467**, 430–435.
29. Schaaf,C.A., Kwak,H., Koenig,A., Misulovin,Z., Gohara,D.W., Watson,A., Zhou,Y., Lis,J.T. and Dorsett,D. (2013) Genome-wide control of RNA polymerase II activity by cohesin. *PLos Genet.*, **9**, e1003382.
30. Kojic,A., Cuadrado,A., De Koninck,M., Giménez-Llorente,D., Rodríguez-Corsino,M., Gómez-López,G., Le Dily,F., Marti-Renom,M.A. and Losada,A. (2018) Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. *Nat. Struct. Mol. Biol.*, **25**, 496–504.
31. Cuadrado,A., Giménez-Llorente,D., Kojic,A., Rodríguez-Corsino,M., Cuartero,Y., Martín-Serrano,G., Gómez-López,G., Marti-Renom,M.A. and Losada,A. (2019) Specific contributions of Cohesin-SA1 and Cohesin-SA2 to TADs and polycomb domains in embryonic stem cells. *Cell Rep.*, **27**, 3500–3510.
32. Viny,A.D., Bowman,R.L., Liu,Y., Lavallée,V.-P., Eisman,S.E., Xiao,W., Durham,B.H., Navitski,A., Park,J., Braunstein,S. *et al.* (2019) 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.
33. Casa,V., Moronta Gines,M., Gade Gusmao,E., Slotman,J.A., Zirkel,A., Josipovic,N., Oole,E., van IJcken,W.F.J., Houtsmuller,A.B., Papantonis,A. *et al.* (2020) Redundant and specific roles of cohesin STAG subunits in chromatin looping and transcriptional control. *Genome Res.*, **30**, 515–527.
34. Gui,Y., Guo,G., Huang,Y., Hu,X., Tang,A., Gao,S., Wu,R., Chen,C., Li,X., Zhou,L. *et al.* (2011) Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat. Genet.*, **43**, 875–878.
35. McConkey,D.J., Lee,S., Choi,W., Tran,M., Majewski,T., Lee,S., Siefker-Radtke,A., Dinney,C. and Czerniak,B. (2010) Molecular genetics of bladder cancer: Emerging mechanisms of tumor initiation and progression. *Urol. Oncol.*, **28**, 429–440.
36. Kamoun,A., de Reyniès,A., Allory,Y., Sjödahl,G., Robertson,A.G., Seiler,R., Hoadley,K.A., Groeneveld,C.S., Al-Ahmadie,H., Choi,W. *et al.* (2020) A consensus molecular classification of muscle-invasive bladder cancer. *Eur. Urol.*, **77**, 420–433.
37. Earl,J., Rico,D., Carrillo-de-Santa-Pau,E., Rodríguez-Santiago,B., Méndez-Pertuz,M., Auer,H., Gómez,G., Grossman,H.B., Pisano,D.G., Schulz,W.A. *et al.* (2015) The UBC-40 Urothelial Bladder Cancer cell line index: a genomic resource for functional studies. *BMC Genomics*, **16**, 403.
38. Rubio-Camarillo,M., López-Fernández,H., Gómez-López,G., Carro,Á., Fernández,J.M., Torre,C.F., Fdez-Riverola,F. and Glez-Peña,D. (2017) RUbioSeq+: a multiplatform application that executes parallelized pipelines to analyse next-generation sequencing data. *Comput. Methods Programs Biomed.*, **138**, 73–81.
39. Ross-Innes,C.S., Stark,R., Teschendorff,A.E., Holmes,K.A., Ali,H.R., Dunning,M.J., Brown,G.D., Gojis,O., Ellis,I.O., Green,A.R. *et al.* (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*, **481**, 389–393.
40. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
41. Machanick,P. and Bailey,T.L. (2011) MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*, **27**, 1696–1697.
42. Graña,O., Rubio-Camarillo,M., Fernandez-Riverola,F., Pisano,D.G. and Glez-Peña,D. (2017) Nextpresso: next generation sequencing expression analysis pipeline. *Curr. Bioinform.*, **13**, 583–591.
43. Rao,S.S.P., Huntley,M.H., Durand,N.C. and Stamenova,E.K. (2014) Article A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665–1680.
44. Serra,F., Baù,D., Goodstadt,M., Castillo,D., Filion,G.J. and Marti-Renom,M.A. (2017) Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS Comput. Biol.*, **13**, e1005665.
45. Yang,T., Zhang,F., Yardımcı,G.G., Song,F., Hardison,R.C., Noble,W.S., Yue,F. and Li,Q. (2017) HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.*, **27**, 1939–1949.
46. Amat,R., Böttcher,R., Le Dily,F., Vidal,E., Quilez,J., Cuartero,Y., Beato,M., de Nadal,E. and Posas,F. (2019) Rapid reversible changes

in compartments and local chromatin organization revealed by hyperosmotic shock. *Genome Res.*, **29**, 18–28.
47. Stansfield,J.C., Cresswell,K.G., Vladimirov,V.I. and Dozmorov,M.G. (2018) HiCcompare: an R-package for joint normalization and comparison of HI-C datasets. *BMC Bioinformatics*, **19**, 279.
48. Pancaldi,V., Carrillo-de-Santa-Pau,E., Javierre,B.M., Juan,D., Fraser,P., Spivakov,M., Valencia,A. and Rico,D. (2016) Integrating epigenomic data and 3D genomic structure with a new measure of chromatin assortativity. *Genome Biol.*, **17**, 152.
49. Wutz,G., Ladurner,R., St Hilaire,B.G., Stocsits,R.R., Nagasaka,K., Pignard,B., Sanborn,A., Tang,W., Várnai,C., Ivanov,M.P. *et al.* (2020) ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesin-STAG1 from WAPL. *Elife*, **9**, e52091.
50. Du,C., Gao,Y., Xu,S., Jia,J., Huang,Z., Fan,J., Wang,X., He,D. and Guo,P. (2016) KLF5 promotes cell migration by up-regulating FYN in bladder cancer cells. *FEBS Lett.*, **590**, 408–418.
51. Qin,J., Zhou,Z., Chen,W., Wang,C., Zhang,H., Ge,G., Shao,M., You,D., Fan,Z., Xia,H. *et al.* (2015) BAP1 promotes breast cancer cell proliferation and metastasis by deubiquitinating KLF5. *Nat. Commun.*, **6**, 8471.
52. Sabari,J.K., Lok,B.H., Laird,J.H., Poirier,J.T. and Rudin,C.M. (2017) Unravelling the biology of SCLC: implications for therapy. *Nat. Rev. Clin. Oncol.*, **14**, 549–561.
53. Ostertag,E.M., Goodier,J.L., Zhang,Y. and Kazazian,H.H.J. (2003) SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am. J. Hum. Genet.*, **73**, 1444–1451.
54. Morales,C. and Losada,A. (2018) Establishing and dissolving cohesion during the vertebrate cell cycle. *Curr. Opin. Cell Biol.*, **52**, 51–57.
55. Phillips-Cremins,J.E., Sauria,M.E.G., Sanyal,A., Gerasimova,T.I., Lajoie,B.R., Bell,J.S.K., Ong,C.-T., Hookway,T.A., Guo,C., Sun,Y. *et al.* (2013) Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, **153**, 1281–1295.
56. Hedegaard,J., Lamy,P., Nordentoft,I., Algaba,F., Høyer,S., Ulhøi,B.P., Vang,S., Reinert,T., Hermann,G.G., Mogensen,K. *et al.* (2016) Comprehensive transcriptional analysis of early-stage urothelial carcinoma. *Cancer Cell,* **30**, 27–42.
57. Santos,C.P., Lapi,E., Martínez de Villarreal,J., Álvaro-Espinosa,L., Fernández-Barral,A., Barbáchano,A., Domínguez,O., Laughney,A.M., Megías,D., Muñoz,A. *et al.* (2019) Urothelial organoids originating from Cd49f(high) mouse stem cells display Notch-dependent differentiation capacity. *Nat. Commun.*, **10**, 4407.
58. Yamashita,H., Kawasawa,Y.I., Shuman,L., Zheng,Z., Tran,T., Walter,V., Warrick,J.I., Chen,G., Al-Ahmadie,H., Kaag,M. *et al.* (2019) Repression of transcription factor AP-2 alpha by PPARγ reveals a novel transcriptional circuit in basal-squamous bladder cancer. *Oncogenesis*, **8**, 69.
59. Durand,N.C., Robinson,J.T., Shamim,M.S., Machol,I., Mesirov,J.P., Lander,E.S. and Aiden,E.L. (2016) Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.*, **3**, 99–101.