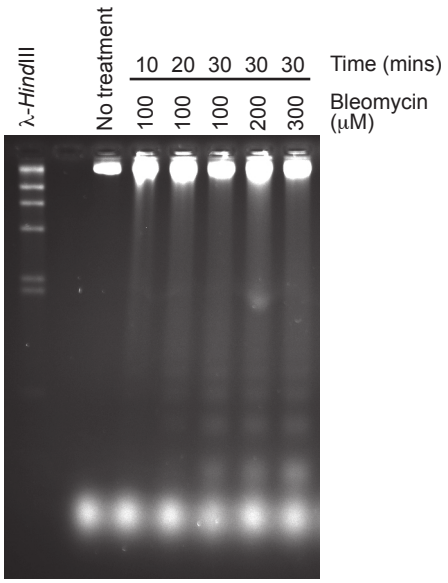
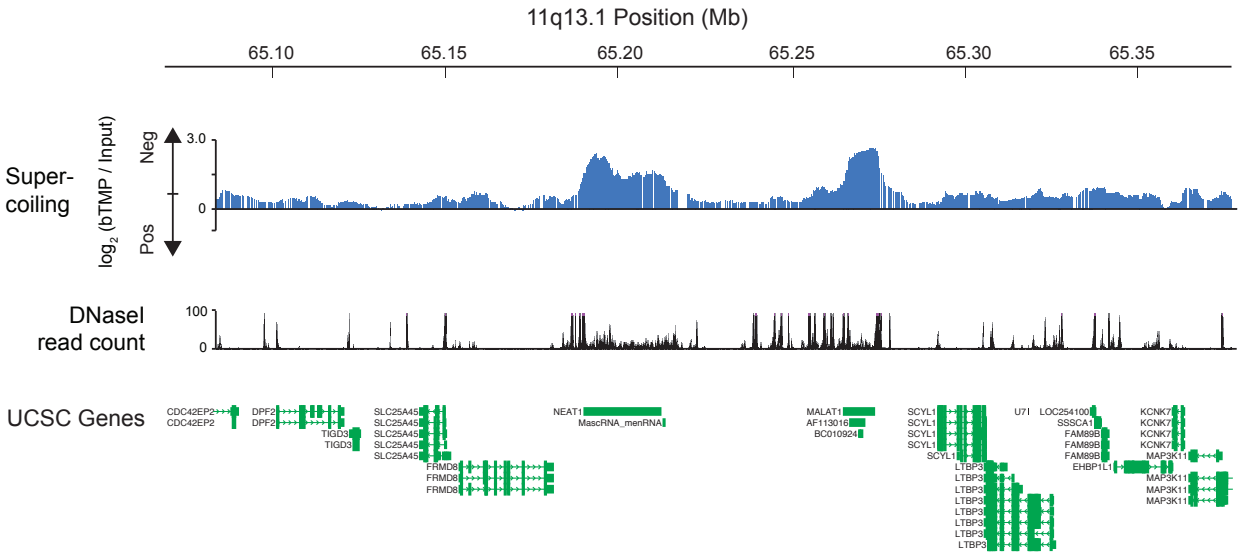
Transcription forms and remodels supercoiling domains unfolding large scale chromatin structures

Catherine Naughton, Nicolaos Avlonitis, Samuel Corless, James G. Prendergast, Ioulia K. Mati, Paul P. Eijk, Scott L. Cockroft, Mark Bradley, Bauke Ylstra, Nick Gilbert

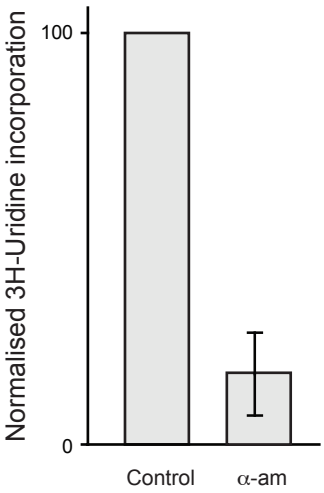
Supplementary online material



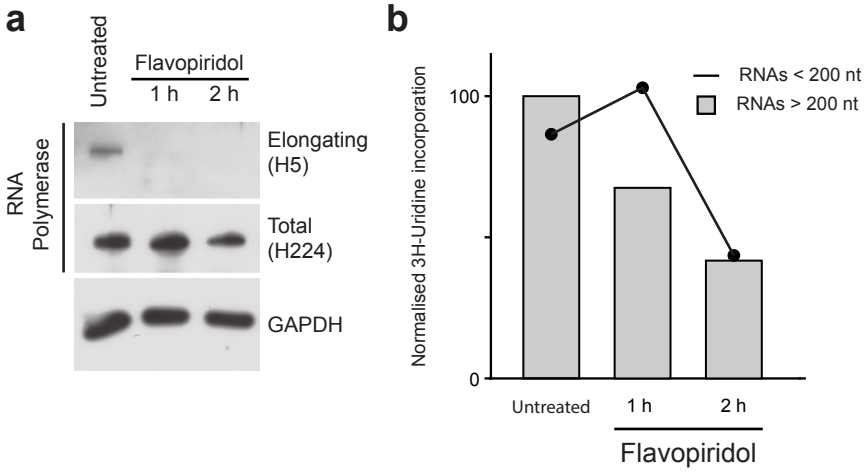
Supplementary Figure 1 Bleomycin nicks genomic DNA. Alkali agarose gel showing bleomycin nicking of DNA. Cells were treated with different concentrations of bleomycin, a gycopeptide antibiotic, for times shown and DNA was purified from the cells and analyzed on an alkaline agarose gel. Analysis of the DNA profile shows that bleomycin introduces increasing numbers of nicks over time. Markers are λ-*Hind*III.



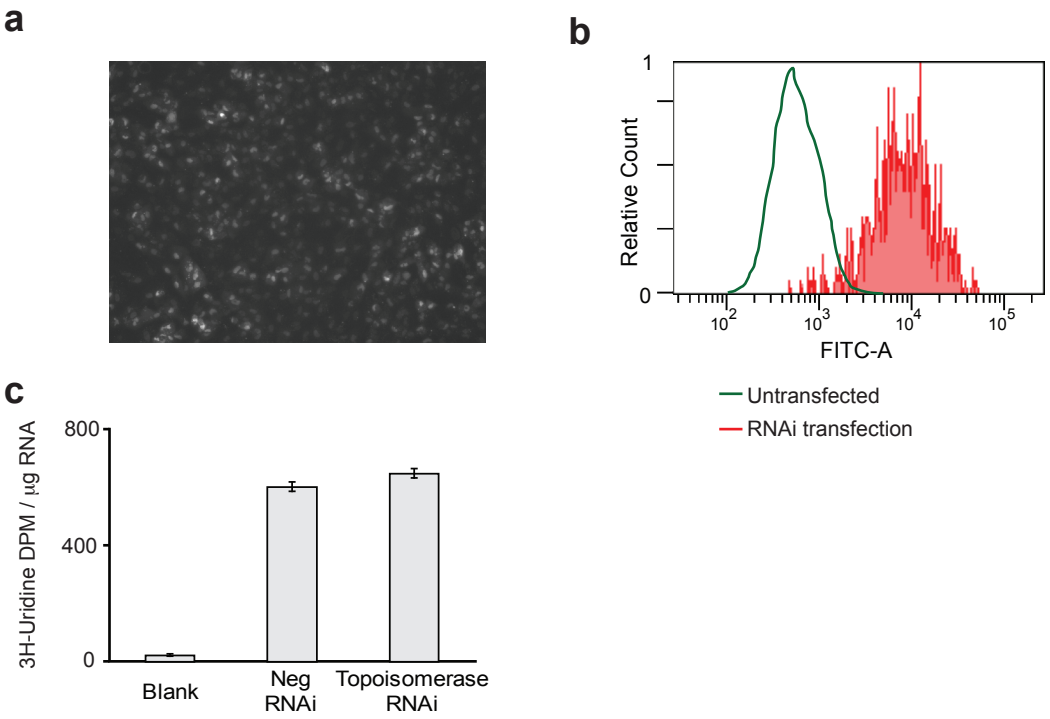
Supplementary Figure 2 DNA supercoiling around the transcriptionally active *NEAT1* and *MALAT1* genes at 11q13.1. Microarray data showing bTMP binding as log2(bTMP/Input) indicative of DNA supercoiling at 11q13.1. To analyze DNA supercoiling cells were treated with biotinylated trimethylpsoralen (bTMP) and UV cross-linked. DNA was purified, enriched for biotin using streptavidin beads, amplified and hybridized to genomic microarrays vs. input control. Increased binding of psoralen corresponds to an enrichment of negative supercoiling and under-wound DNA. The *NEAT1* and *MALAT1* non-coding RNA genes are highly transcribed and are enriched in under-wound DNA. The DNaseI read count from RPE1 cells was obtained from the UCSC genome browser and is shown with the positions of UCSC genes. The genes are enriched in DNaseI sensitive sites but there are pronounced DNaseI peaks that are not associated with peaks of psoralen binding. The expressed *SCYL1* and *LTBP3* genes are distal to the *MALAT1* gene and are convergently transcribed. Although the genes are associated with an increase in under-wound DNA there is not a pronounced peak of positive DNA supercoiling in between the genes.



Supplementary Figure 3 Global transcription inhibition by α-amanitin. Bar graph showing 3H-Uridine incorporation into RNA before and after α-amanitin (α-am) treatment. Cells were treated with the transcription inhibitor α-amanitin for 5 hrs. 30 minutes before harvesting cells were pulse labeled with 3H-Uridine. RNA was isolated and labeled RNA was measured by scintillation counting. The graph shown is normalized for the total amount of RNA and the control sample is scaled to 100.



Supplementary Figure 4 Short RNAs are synthesized by the initiating form of RNA polymerase. (**a**). Western blot showing the global levels of elongating and total RNA polymerase after treatment with flavopiridol. GAPDH was used as a loading control. (**b**). Graph showing incorporation of 3H-Uridine into short (<200 nt) and long (>200 nt) RNA transcripts after 30 min pulse labeling to measure RNA synthesis after transcription inhibition by flavopiridol. The graph shown is normalized for the total amount of RNA and the control sample is scaled to 100. After inhibition of transcription elongation there is a decrease in long RNA transcription but a lag before a reduction in short RNA transcription suggesting that short RNAs are produced by the initiating form of RNA polymerase.



Supplementary Figure 5 Effect of topoisomerase RNAi on transcription. (**a**). Micrograph showing cells transfected with siRNA against topoisomerase I, IIα and IIβ and a fluorescent (FITC) reporter oligo. (**b**). FACS profile of RNAi transfected cells. Fluorescent cells were analyzed and sorted. (**c**). Bar graph showing 30 min pulse 3H-Uridine incorporation into RNA 48 hrs after topoisomerase knockdown by RNAi. This data demonstrates that under these conditions topoisomerase inhibition did not have a pronounced effect on transcription.

Supplementary Table 1

Chromosome regions analyzed

Nimblegen and Agilent arrays were custom designed to be representative of the entire genome. Probe spacing and gene density (Genes per Mb) are averaged across the regions.

Region	Make	Region Start (bp)	Region End (bp)	Size (Mb)	Probes	Probe spacing	Genes per Mb
11p	Nimblegen	129,000	51,600,000	51.5	258409	199	22
11q	Nimblegen	54,700,000	134,950,000	80.3	408679	197	25
17q24.2	Nimblegen	63,700,000	68,700,000	5	25122	199	18
21q22.2	Nimblegen	38,500,000	43,500,000	5	26672	187	19
11p15.5	Agilent	1	2,800,000	2.8	26126	107	50
11p15.1	Agilent	17,417,960	19,417,960	2	15176	132	35
11p14.1	Agilent	27,100,001	32,600,000	5.5	43382	127	13
Enr312 (Chr11)	Agilent	131,031,152	131,732,236	0.7	6965	101	14
Enr332 (Chr11)	Agilent	64,120,923	64,720,922	0.6	5501	109	50
Xq13.1	Agilent	68,369,744	70,369,744	2	11047	181	26
Xq25	Agilent	119,145,001	123,045,000	3.9	21830	179	17

Supplementary Table 2

Cosmid and Fosmid probes

Details of the cosmid and fosmid probes used for FISH to investigate large scale chromatin compaction

Position	Whitehead name	Sequence Name	Start position (bp)	End position (bp)	Midpoint (bp)	Refs
11p15.5	CI 11p15-49				736369	26, 27
11p15.5	H19-IGF2				2070000	26,27
11p14.1	W12-906E12	G248P8190C6	29706206	29752652	29729429	
11p14.1	W12-1673H3	G248P83109D2	31217238	31257378	31237308	
11P15.1	W12-2224K21	G248P87004F11	17510383	17549607	17529995	
11P15.1	W12-3168N23	G248P8017G12	19055816	19093921	19074869	
Xq13.1	W12-2977E1	G248P89772C1	69082819	69124819	69103819	27
Xq13.1	W12-1660A18	G248P87866A9	71070920	71107485	71089203	27

All positions are NCBI37 hg19 assembly of the human genome

Supplementary Note

Microarray hybridization, data processing and analysis

We used additional ChIP-Seq and DNaseI-Seq datasets from the ENCODE project in the UCSC genome browser. For this study we used DNaseI sites mapped in RPE1 cells (Stamatoyannopoulous Lab, UW), CTCF binding sites mapped in RPE1 cells (Stamatoyannopoulous Lab, UW) and CBP-p300 binding sites mapped in A549 epithelial cells (Snyder Lab, Stanford).

The boxplots shown are all a standard format. The midline is the median, the bottom and the top of the box are the lower and upper quartiles respectively. The whiskers correspond to 1.5x the IQR. The asterisks correspond to standard p-values

Ns P > 0.05

* P < 0.05

** P < 0.01

*** P < 0.001

**** P < 0.0001

Immunoprecipitated DNA was amplified using whole genome amplification (Sigma). 500 ng DNA was random prime labeled with Cy3 or Cy5 (ENZO) and purified on a Mine lute PCR purification column (Qiagen). Essentially input DNA (no pull down) was labeled in one color and sample DNA (pull down with streptavidin-coated beads) was labeled in another color. Labeled DNA was diluted in hybridization buffer (Agilent) and hybridized to either Agilent 180 K custom made arrays or Nimblegen 720 K customer arrays for 24 hrs. Slides were washed according to the manufacturer's instructions and scanned on an Agilent Microarray scanner at 2 micron resolution generating a TIFF file.

All co-ordinates are NCBI37 hg19. Spot signal intensity was extracted from the TIFF files using Feature Extraction software (Agilent arrays) or Nimblegen software (Nimblegen arrays) and was pre-processed in R using the RINGO bioconductor package to give the raw Cy5 and Cy3 signal intensities for each spot. Individual Cy5 and Cy3 channels were normalized to each other and between arrays using a variance stabilizing algorithm (for bTMP arrays) and loess normalized and scaled (for ChIP arrays) using the standard Bio conductor LIMMA package. All arrays were quality controlled by checking array hybridization patterns, analyzing signal profiles and using MA plots. bTMP binding, as a measure of DNA supercoiling, was analyzed and presented as log2(bTMP signal/input signal) for each data point which is a standard approach for all two color microarray experiments[1]. bTMP binding to naked DNA was analyzed and normalized, using the same experimental approach as for bTMP binding in cells, giving $\log2(bTMP_{genomic\ DNA}/Input)$. To correct for topology independent effects of the bTMP drug, bTMP binding to genomic DNA was subtracted from bTMP binding in cells i.e. $\log2(bTMP_{cells}/Input) - \log2(bTMP_{genomic\ DNA}/Input)$ giving the normalized log2(bTMP/Input binding). This ratio calculates the fold enrichment of binding for the sample with

respect to the input signal along the arrays and therefore gives a quantitative measure of where the bTMP or protein is bound. Positive numbers indicate relative enrichment of bTMP whilst negative numbers show a relative depletion of bTMP binding. All experiments were repeated in at least duplicates and all independent data sets gave comparable results. Data points between biological replicates were combined using means or medians as appropriate. For data analysis log2(sample/input) data was loaded in to the ZOO package in R and for display the data was smoothed using a rolling median.

To identify supercoiling boundaries a similar approach was taken as described previously[2], but using a 300 probe edge filter. The supercoiling boundaries were defined using the same threshold for all genomic loci and were defined as being “under-wound”, “over-wound” or “stable” based on the difference in normalized log2(bTMP/input) signal between the control and α-amanitin treated samples in the region. Differences in domain properties were compared using t-tests and significant differences between adjacent domains were confirmed by using a series of t-tests.

To investigate the overlap between supercoiling boundaries and topological boundaries (as described in Dixon et al.[3]) each supercoiling boundary was analyzed to determine if it was located within +/-20 kb of a topological boundary. Likewise each CTCF binding site was tested for whether it was located +/-20 kb of a supercoiling boundary. To map CTCF binding sites around supercoiling boundaries the number of CTCF binding sites was calculated in 10 kb windows upstream and downstream of every supercoiling boundary. The median number of CTCF binding sites was then calculated for each window around every boundary. P-values were calculated by using random permutation analysis.

To meta-analyze bTMP and protein binding around transcription start sites, CTCF binding sites, DNaseI sites or other specific features we calculated the distance between each probe and the nearest feature being analyzed using a custom script in Perl. The median signal was then calculated in windows (usually 1 kb or 2 kb) around every boundary. Simulations of randomly generated data gave a flat line. To ensure peaks were significant we used t-tests between the peaks and the simulated data (in all cases $P < 2.2 \times 10^{-16}$). To analyze GC changes around supercoiling boundaries a 50 kb windows was taken around each boundary, similarly orientated, quintile normalized and the data across each window was averaged.

1. Quackenbush,J. Microarray data normalization and transformation. *Nat. Genet.* **32 Suppl:496-501.**, 496-501 (2002).

2. Guelen,L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature.* **453**, 948-951 (2008).

3. Dixon,J.R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 485, 376-380 (2012).