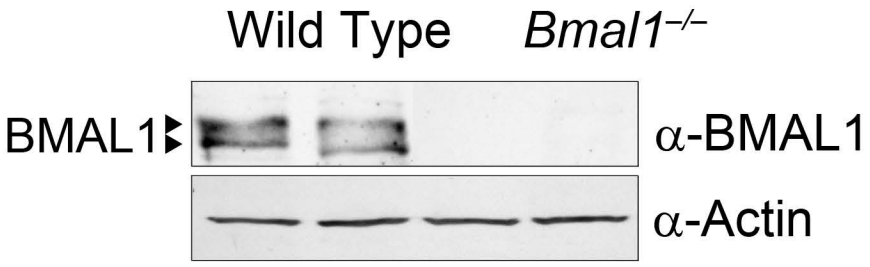
Supplementary Information

Cycles in spatial and temporal chromosomal organization driven by the circadian clock

Lorena Aguilar-Arnal, Ofir Hakim, Vishal R. Patel, Pierre Baldi, Gordon L. Hager and Paolo Sassone-Corsi

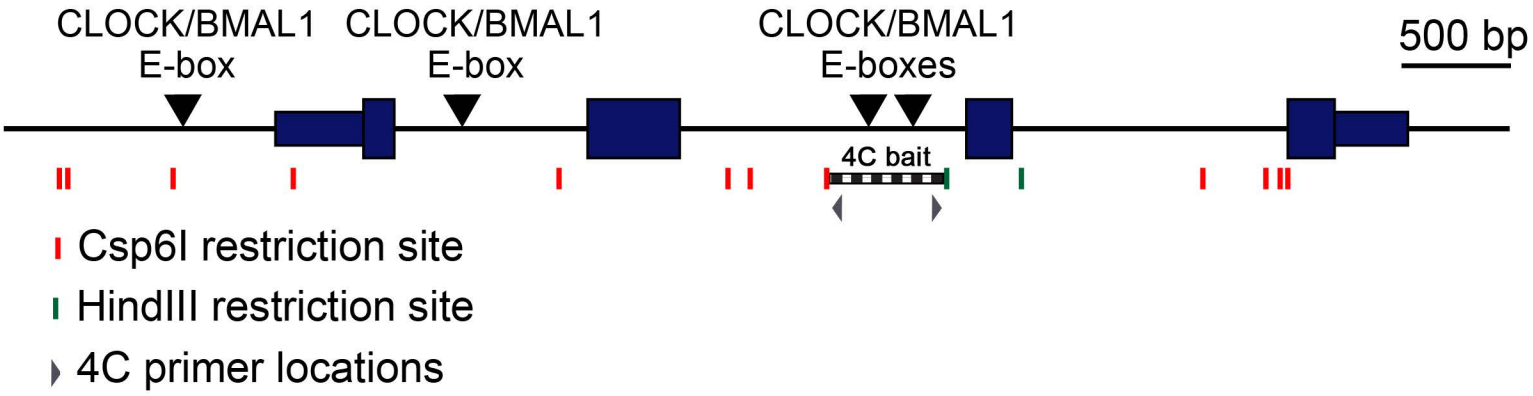
A

Total cell lysate:

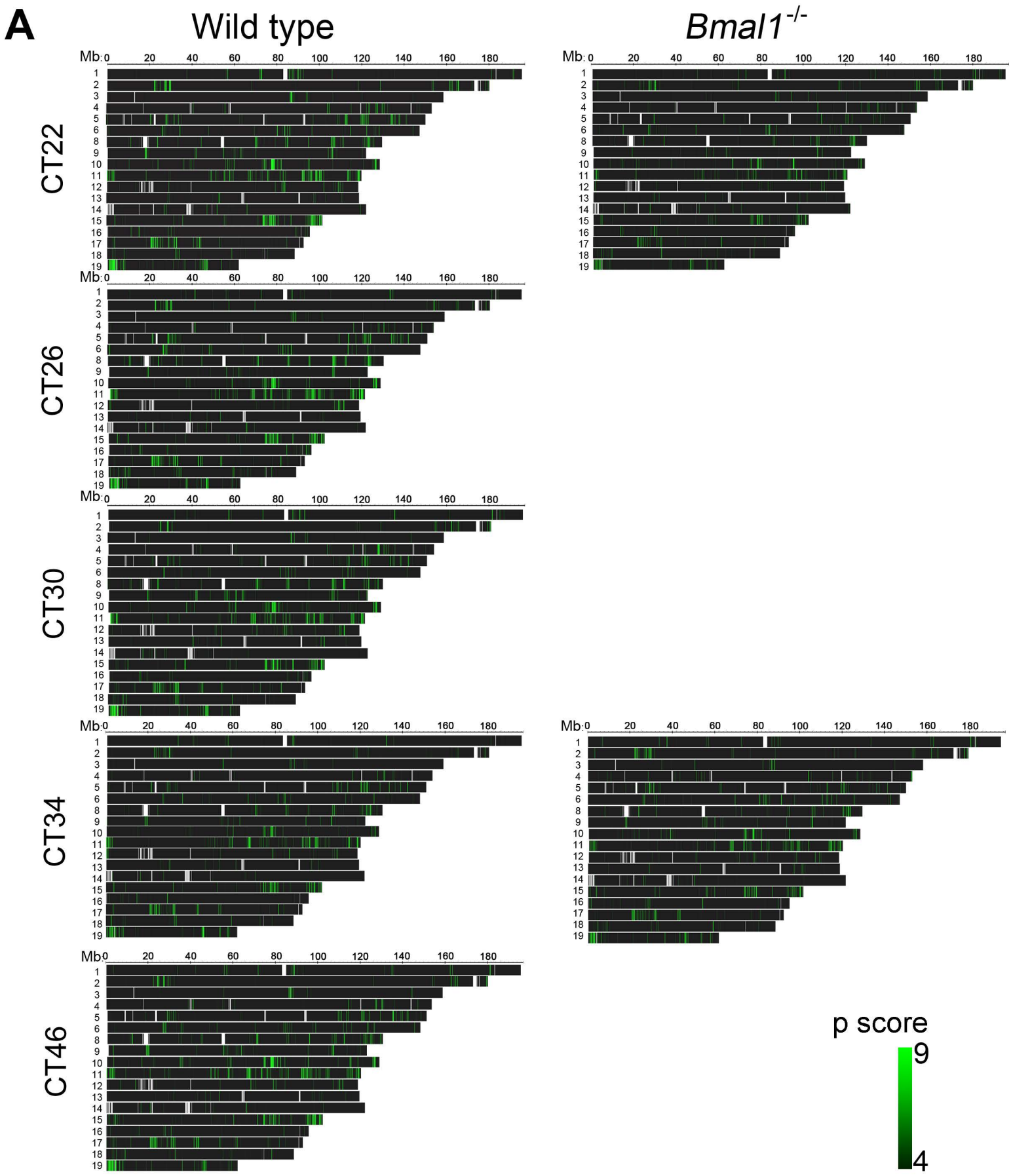


B

dbp locus on chromosome 7:

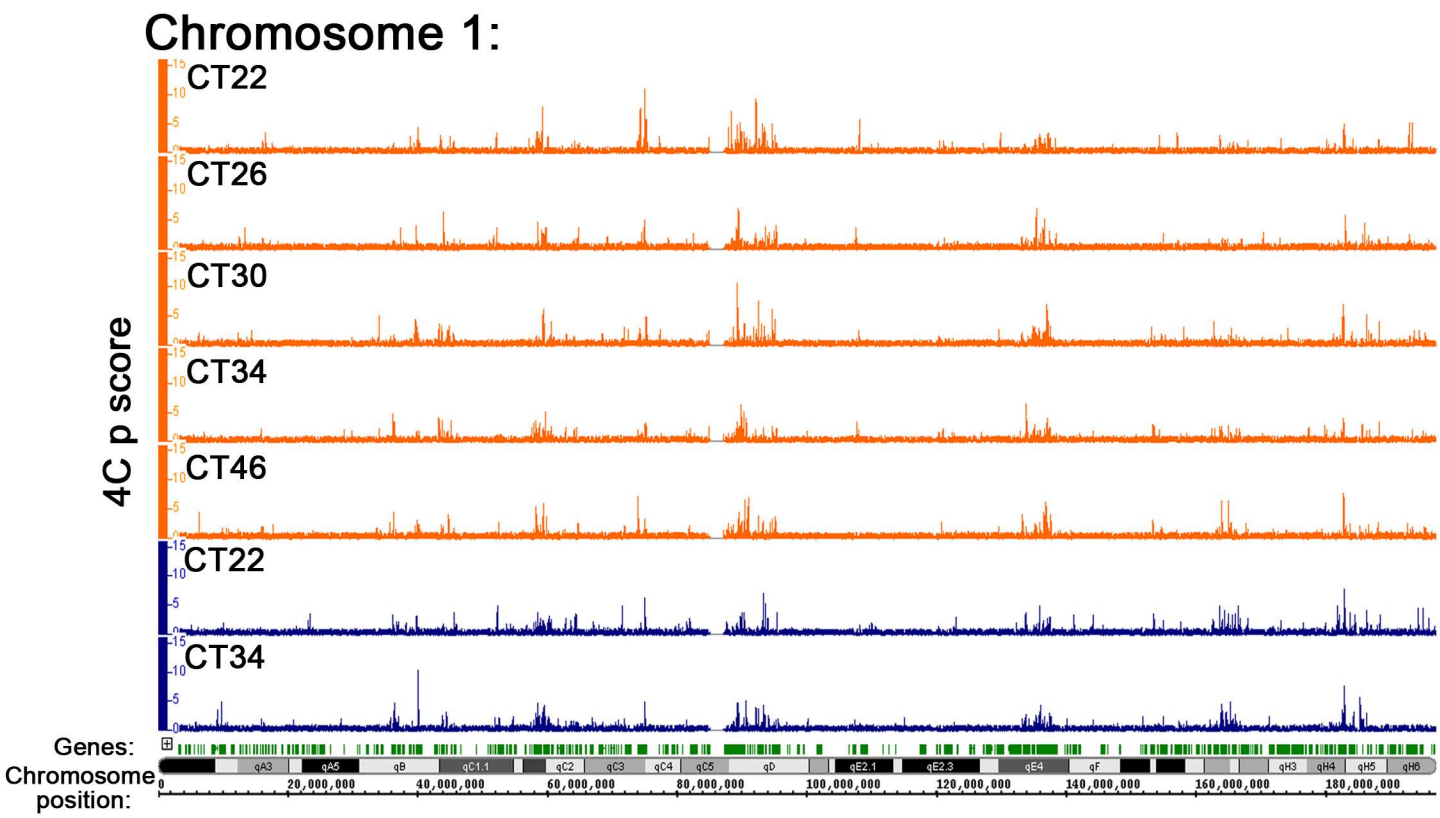


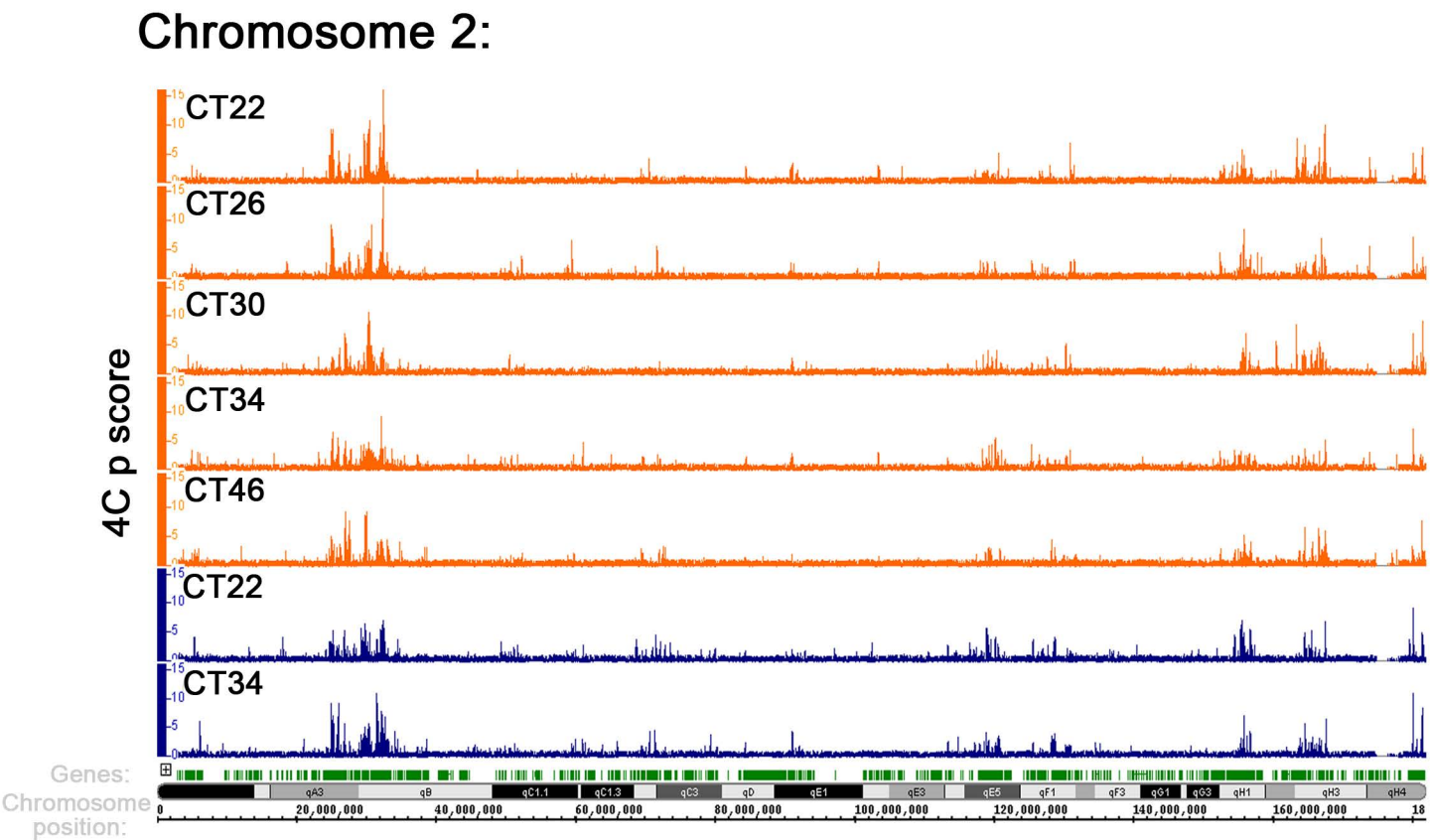
Supplementary Figure 1: 4C experimental design. **A**, Wild type and *Bmal1*$^{-/-}$ MEFs were used in this study. The BMAL1 protein levels were assayed by western blot on total cell extracts with anti-BMAL1 antibody (Abcam) in two independent samples. BMAL1 protein in the *Bmal1*$^{-/-}$ MEFs is absent, while actin protein levels are comparable to those shown in the wild type MEFs. **B**, Location of the 4C bait and HindIII/Csp6I restriction sites on *dbp* locus.



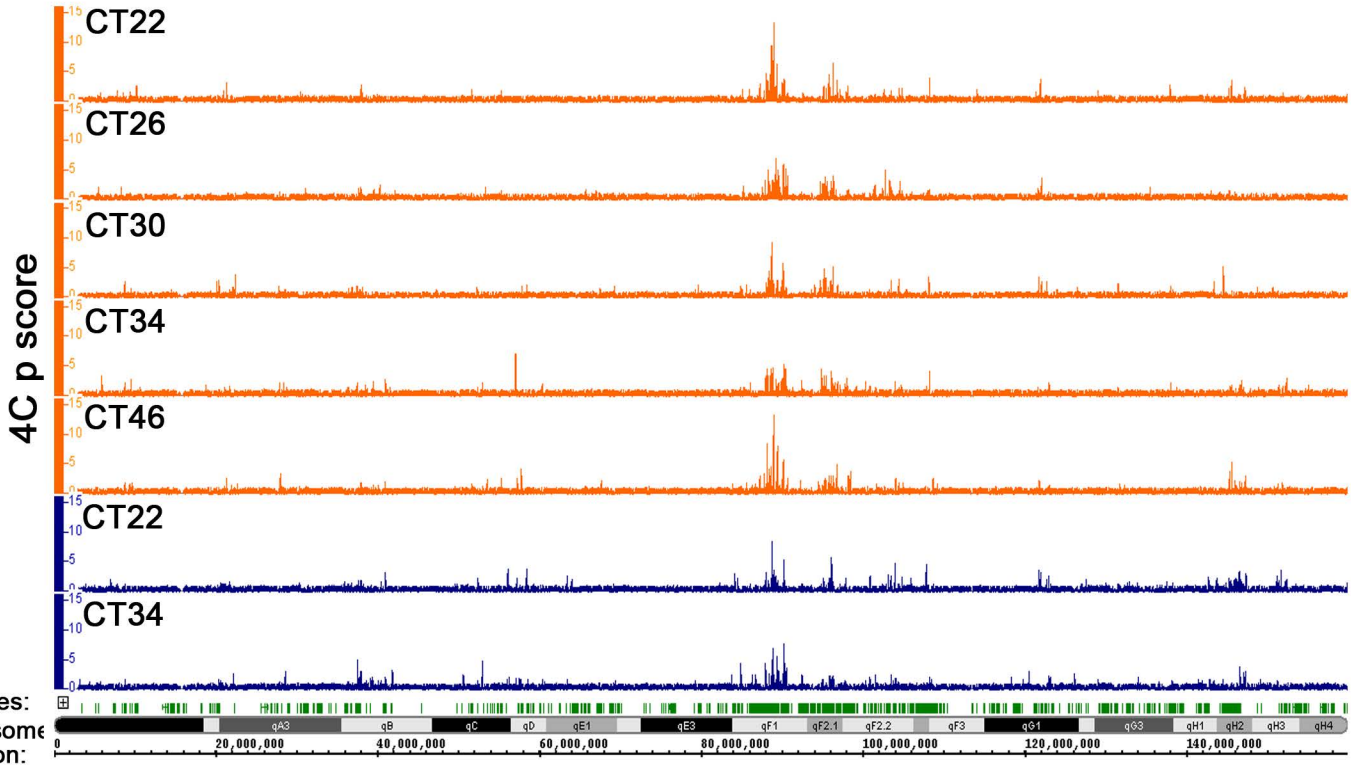
Supplementary Figure 2: Genomic heat maps of *dbp* contacts in *trans.* **A.** Frequencies of interaction for all the probes in the 4C microarrays are represented as different green intensities according to the indicated scale. Interaction frequencies are a representation of the intensity of the probe signal (log2 4C/genomic DNA). The two plots on the right column show the genomic interaction heat maps for *Bmal1*$^{-/-}$ MEFs. Hours after dexamethasone synchronization are indicated on the left (CT, circadian time). Black colored regions in the chromosomes do not interact with *dbp*. The genomic position in mm8 coordinates is indicated on the horizontal axis.

A

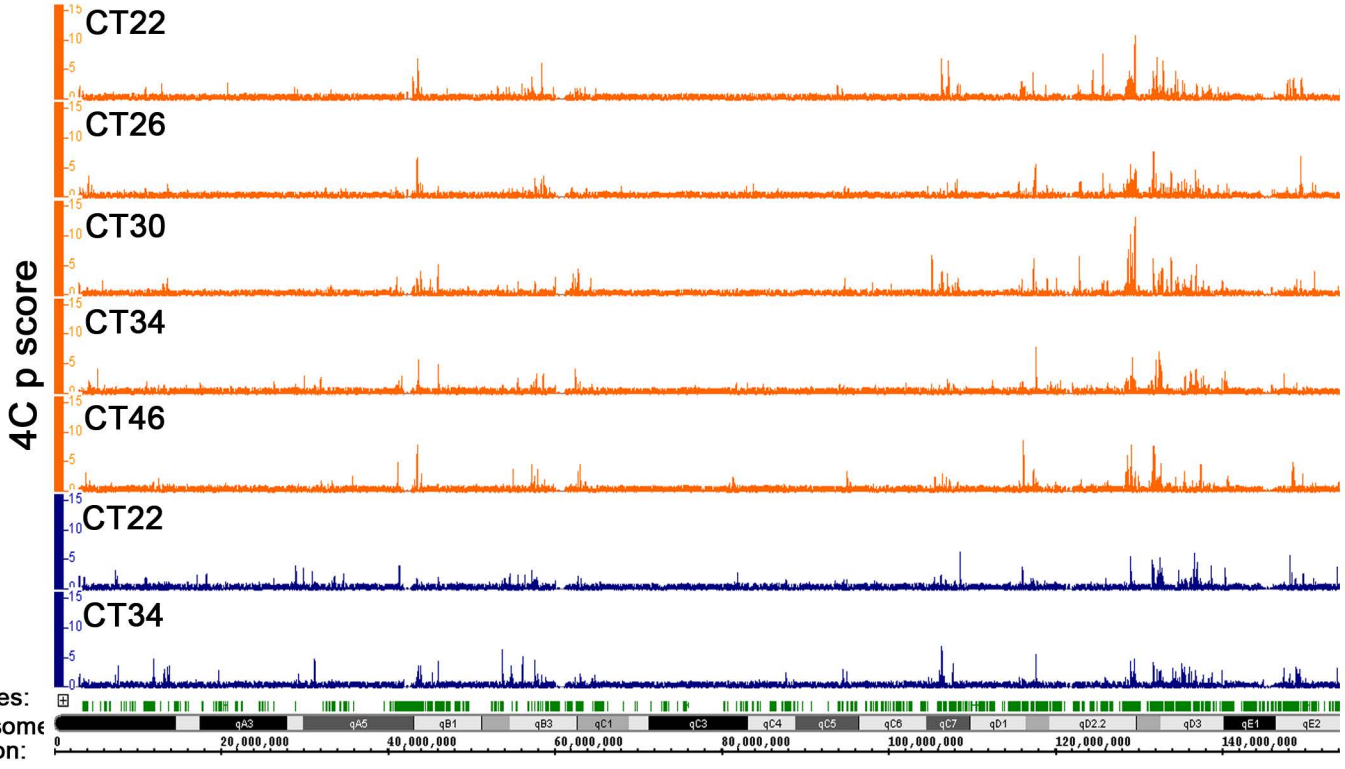

Chromosome 1:
CT22
CT26
CT30
CT34
CT46
CT22
CT34
4C p score
Genes:
Chromosome position:
0
20,000,000
40,000,000
60,000,000
80,000,000
100,000,000
120,000,000
140,000,000
160,000,000
180,000,000

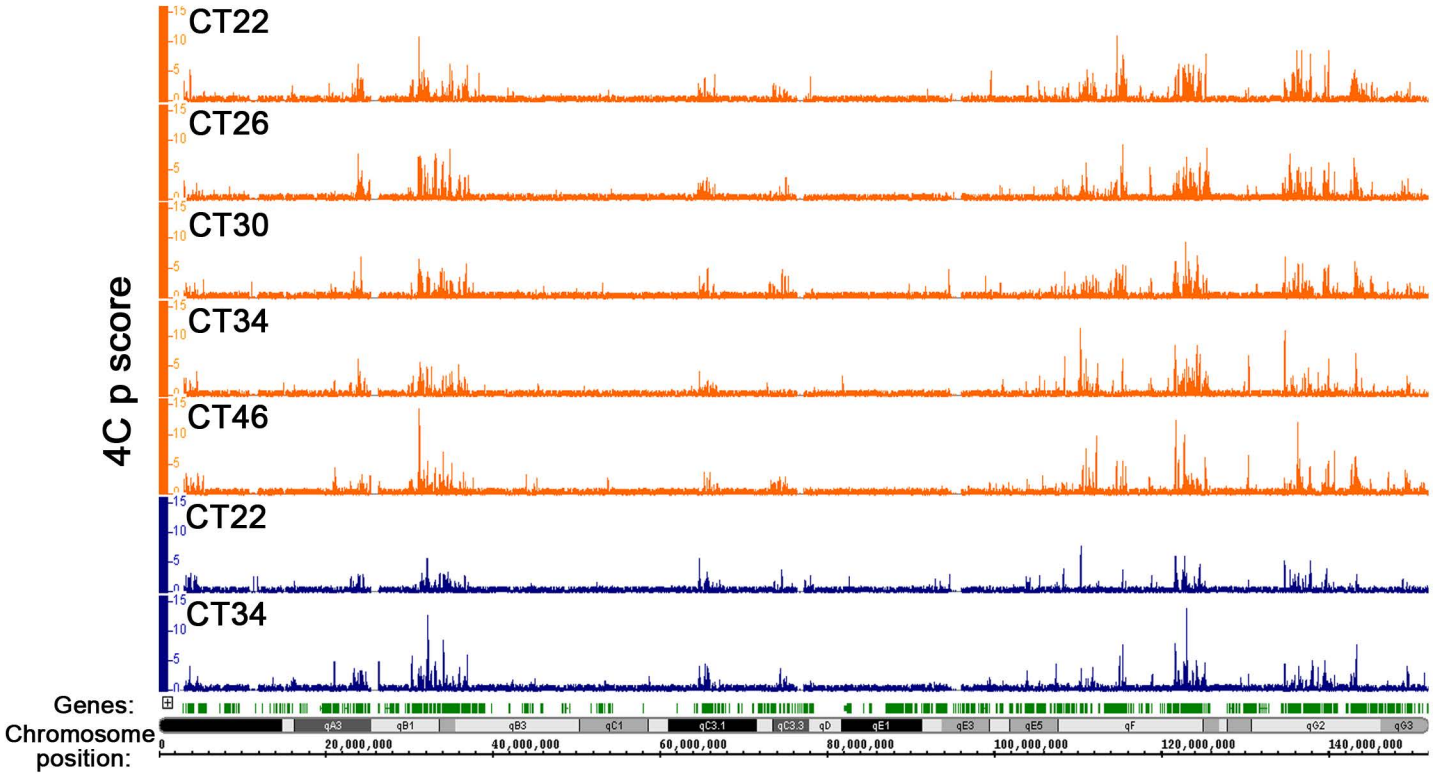

Chromosome 2:
CT22
CT26
CT30
CT34
CT46
CT22
CT34
4C p score
Genes:
Chromosome position:
0
20,000,000
40,000,000
60,000,000
80,000,000
100,000,000
120,000,000
140,000,000
160,000,000

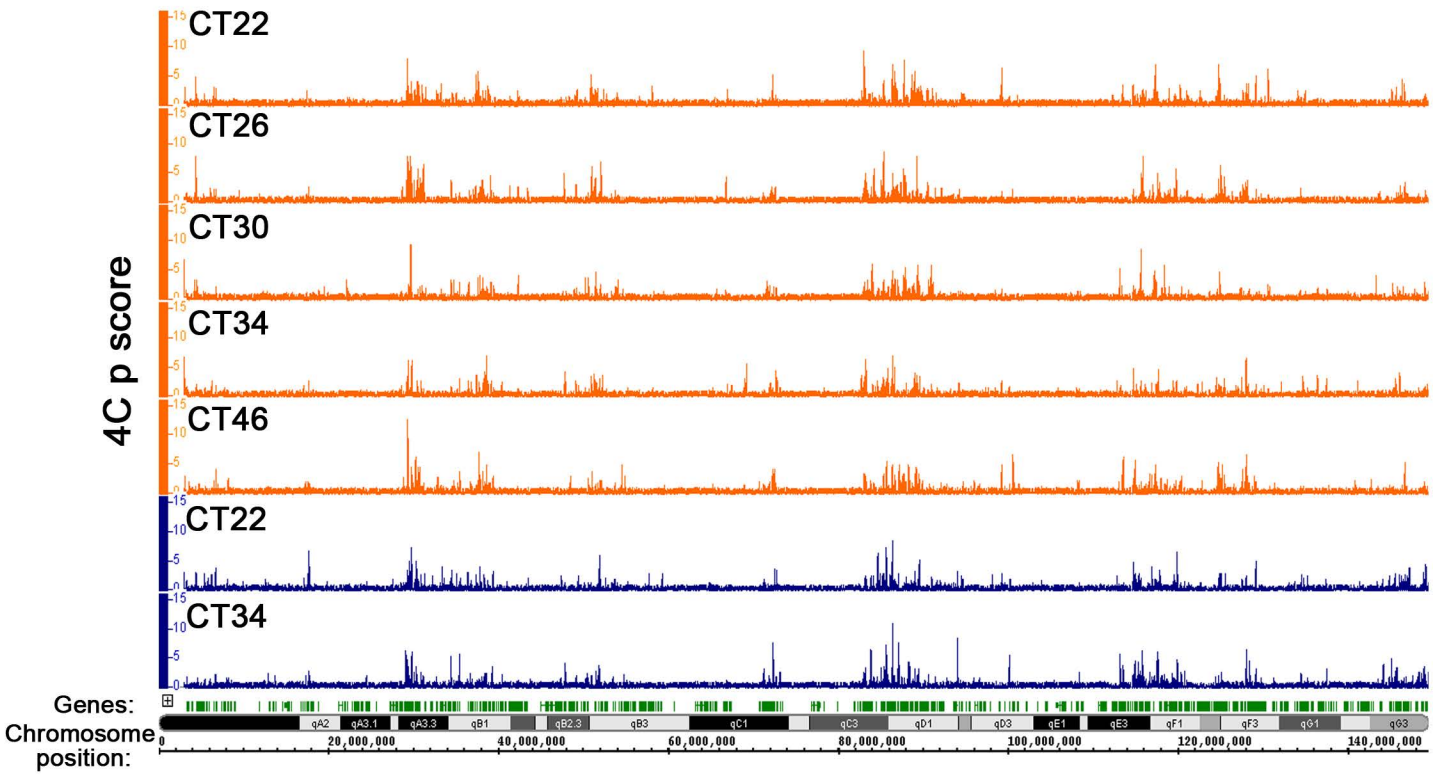
Chromosome 3:

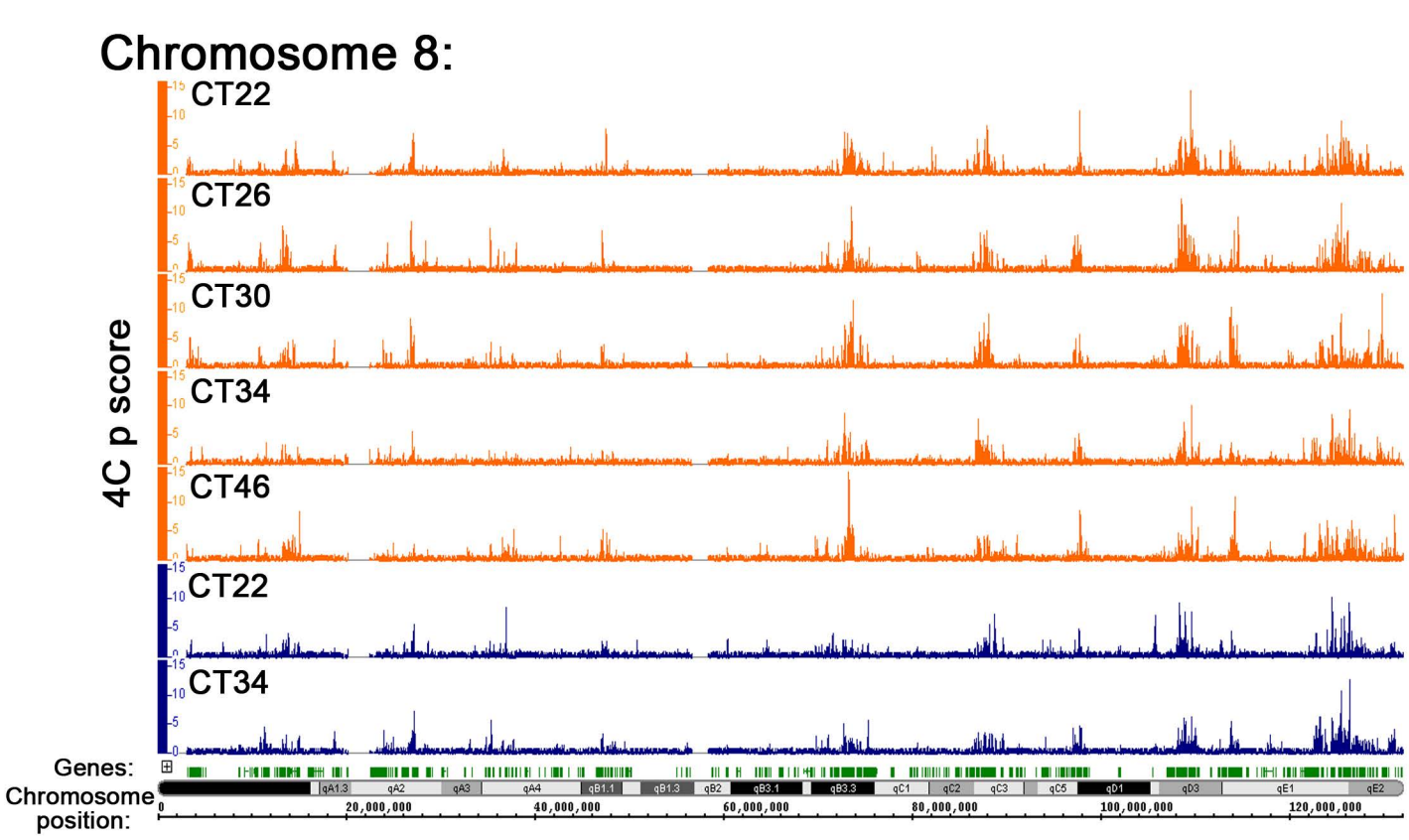


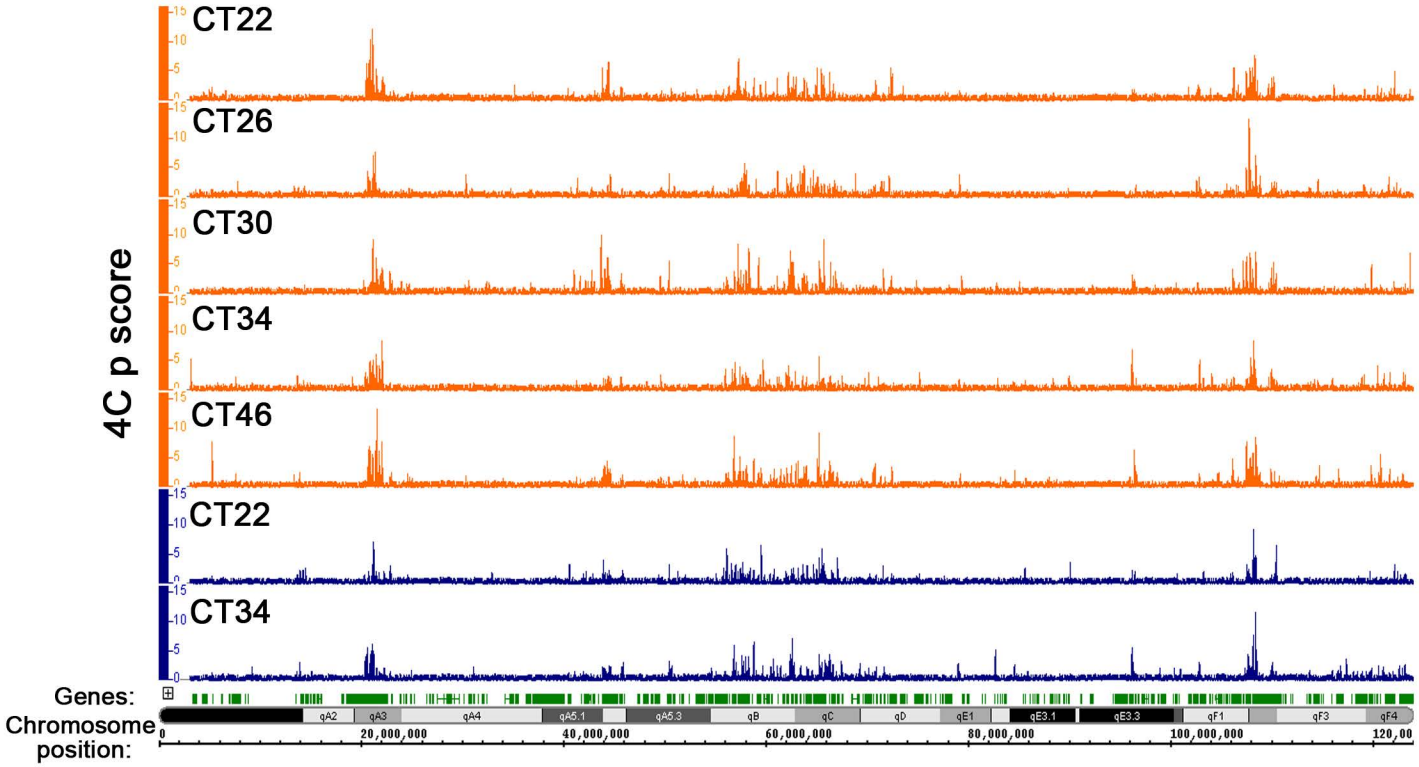
Chromosome 4:



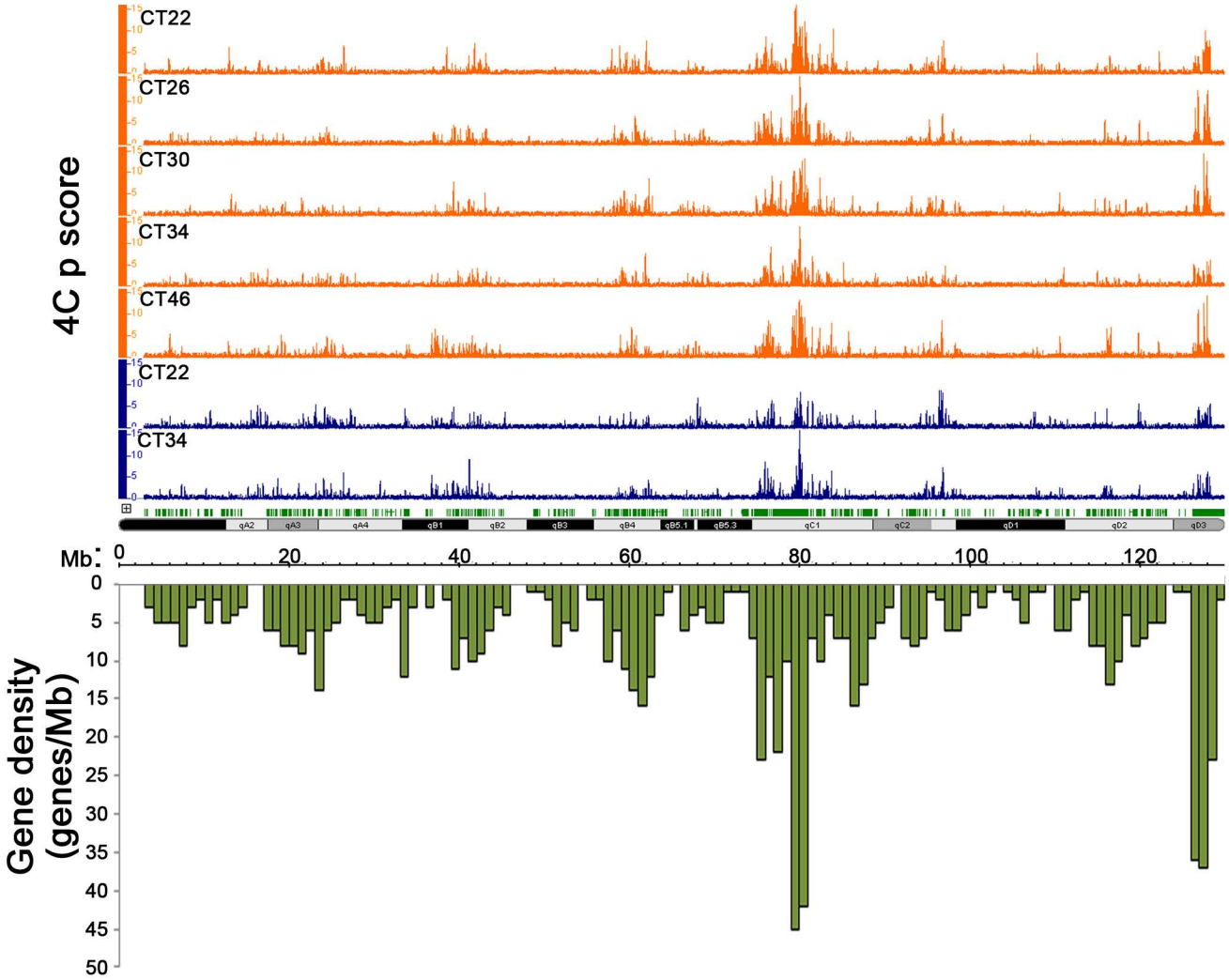

Chromosome 5:
4C p score
CT22
CT26
CT30
CT34
CT46
CT22
CT34
Genes:
Chromosome position:
0
20,000,000
40,000,000
60,000,000
80,000,000
100,000,000
120,000,000
140,000,000

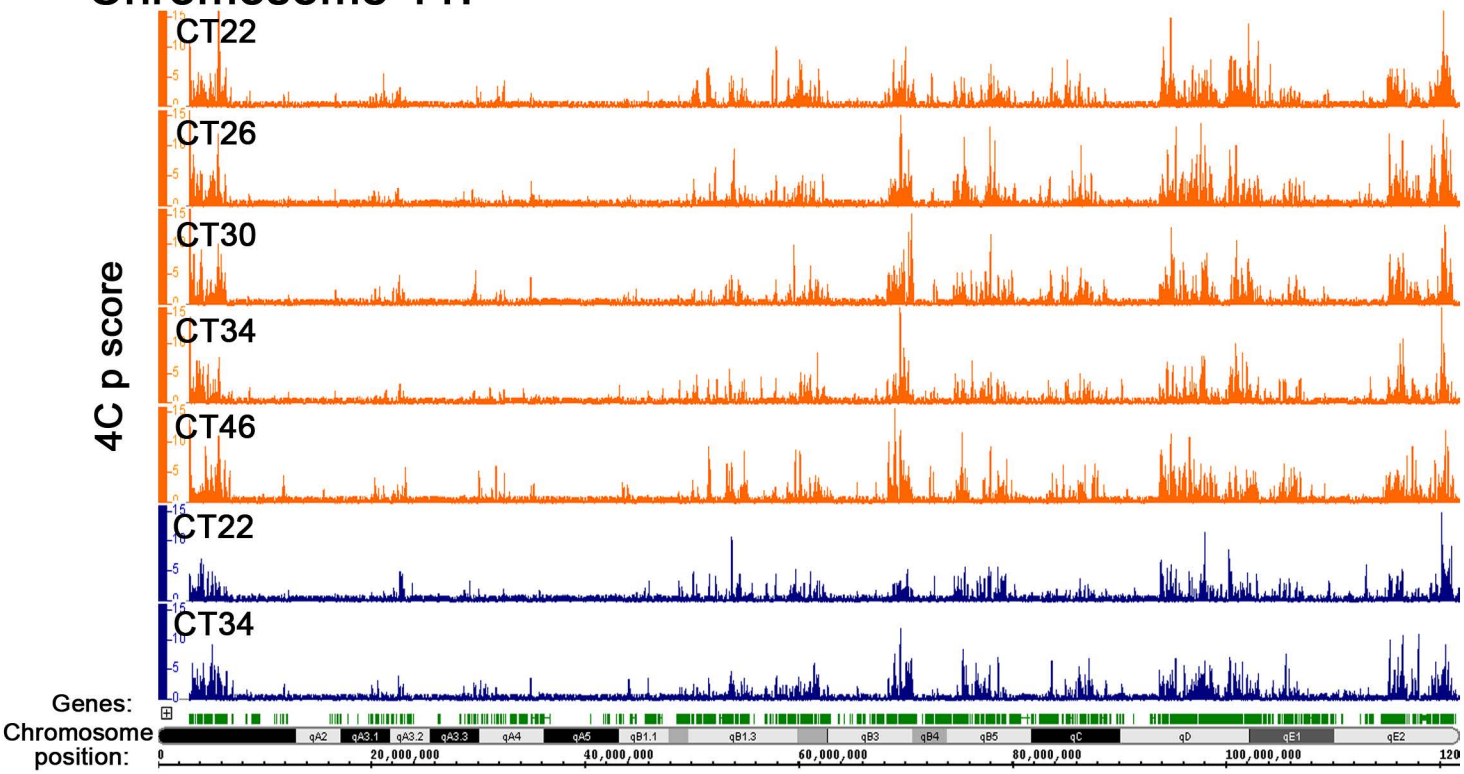

Chromosome 6:
4C p score
CT22
CT26
CT30
CT34
CT46
CT22
CT34
Genes:
Chromosome position:
0
20,000,000
40,000,000
60,000,000
80,000,000
100,000,000
120,000,000
140,000,000

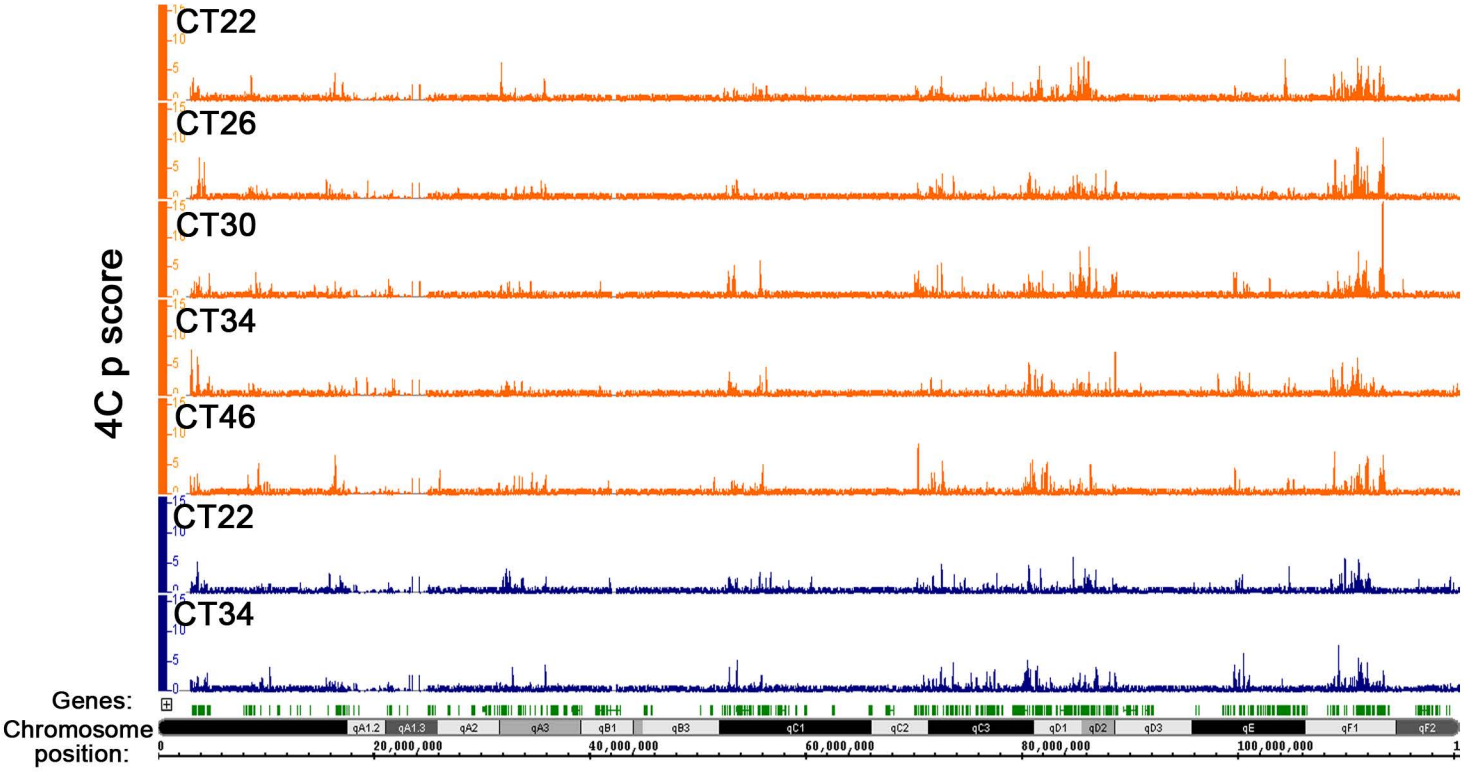

Chromosome 8:
4C p score
CT22
CT26
CT30
CT34
CT46
CT22
CT34
Genes:
Chromosome position:
qA1.3
qA2
qA3
qA4
qB1.1
qB1.3
qB2
qB3.1
qB3.3
qC1
qC2
qC3
qC5
qD1
qD3
qE1
qE2
0
20,000,000
40,000,000
60,000,000
80,000,000
100,000,000
120,000,000

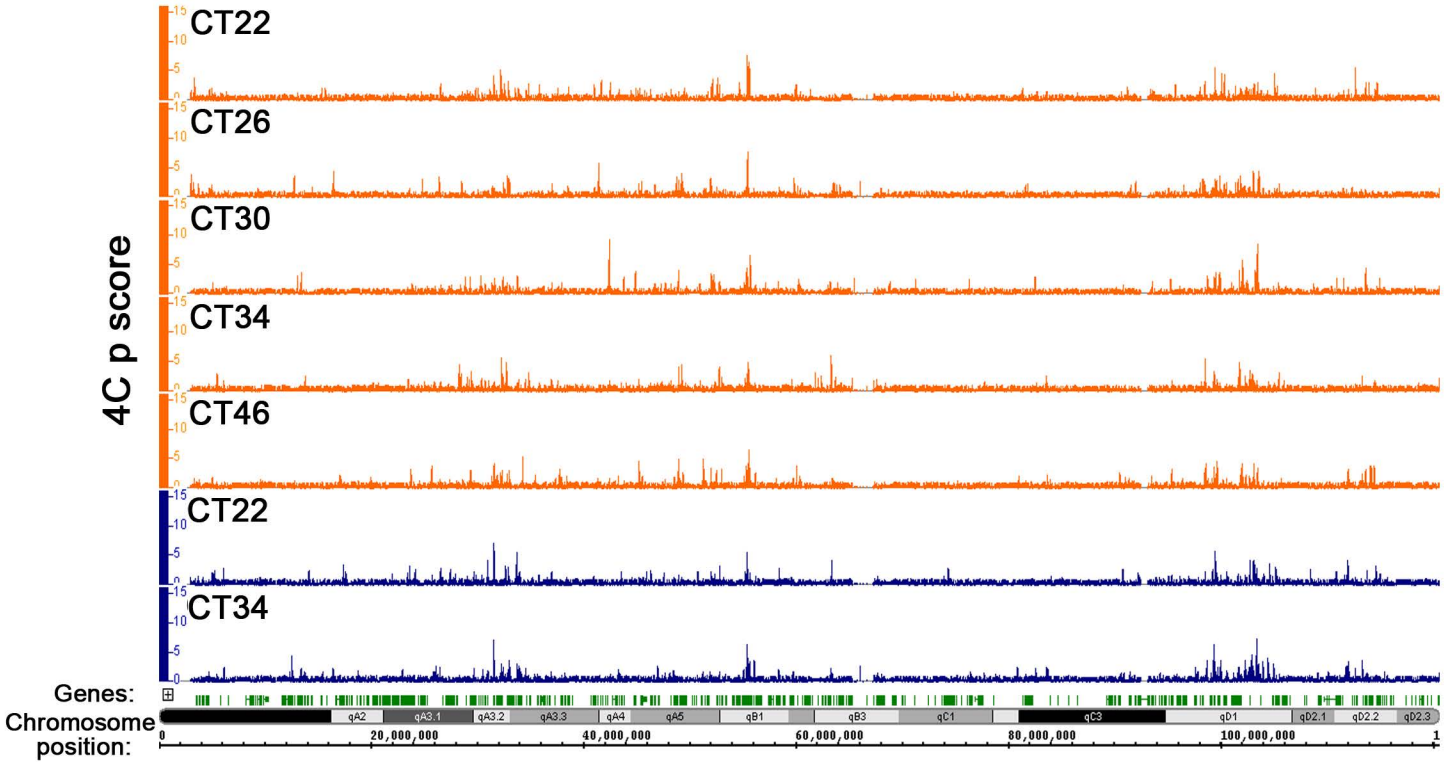

Chromosome 9:
4C p score
CT22
CT26
CT30
CT34
CT46
CT22
CT34
Genes:
Chromosome position:
qA2
qA3
qA4
qA5.1
qA5.3
qB
qC
qD
qE1
qE3.1
qE3.3
qF1
qF3
qF4
0
20,000,000
40,000,000
60,000,000
80,000,000
100,000,000
120,00

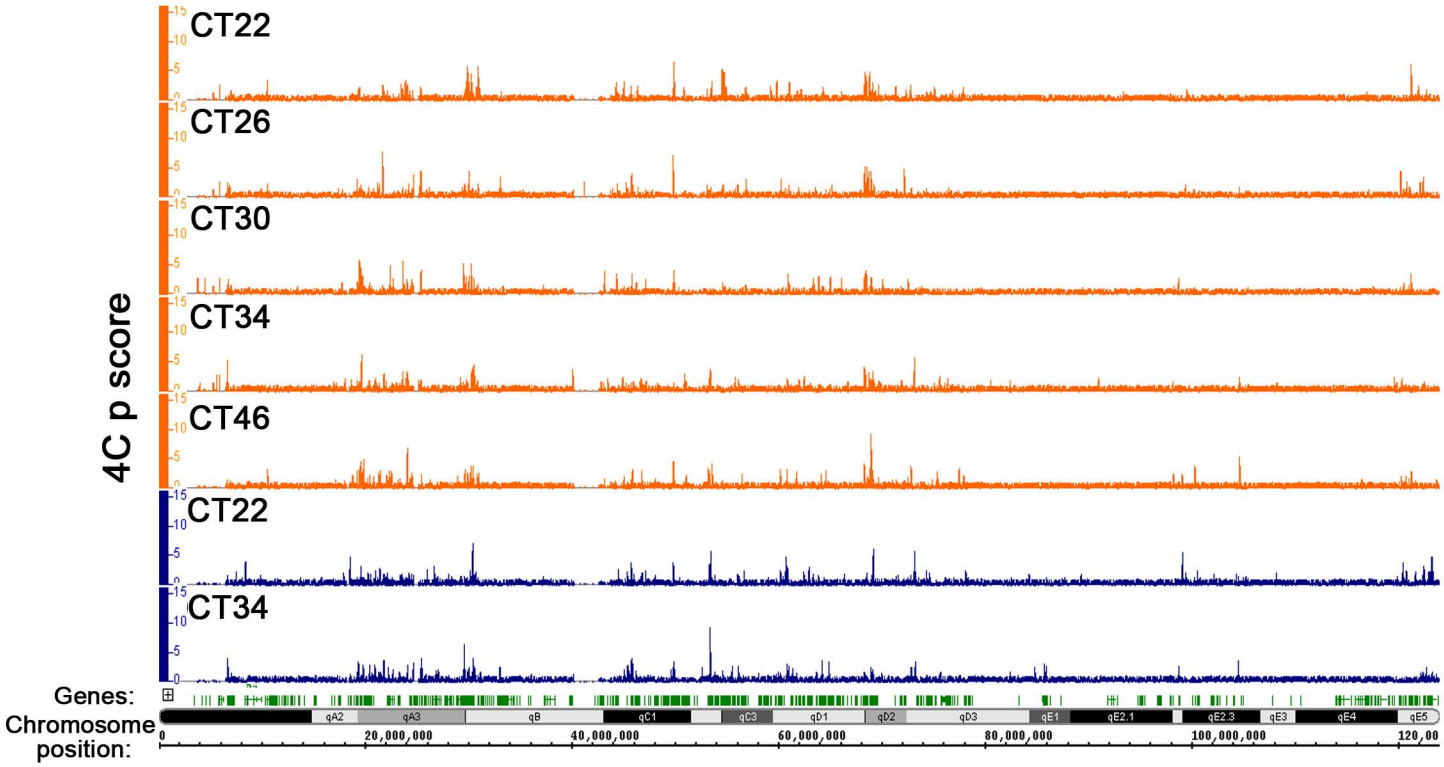
Chromosome 10:



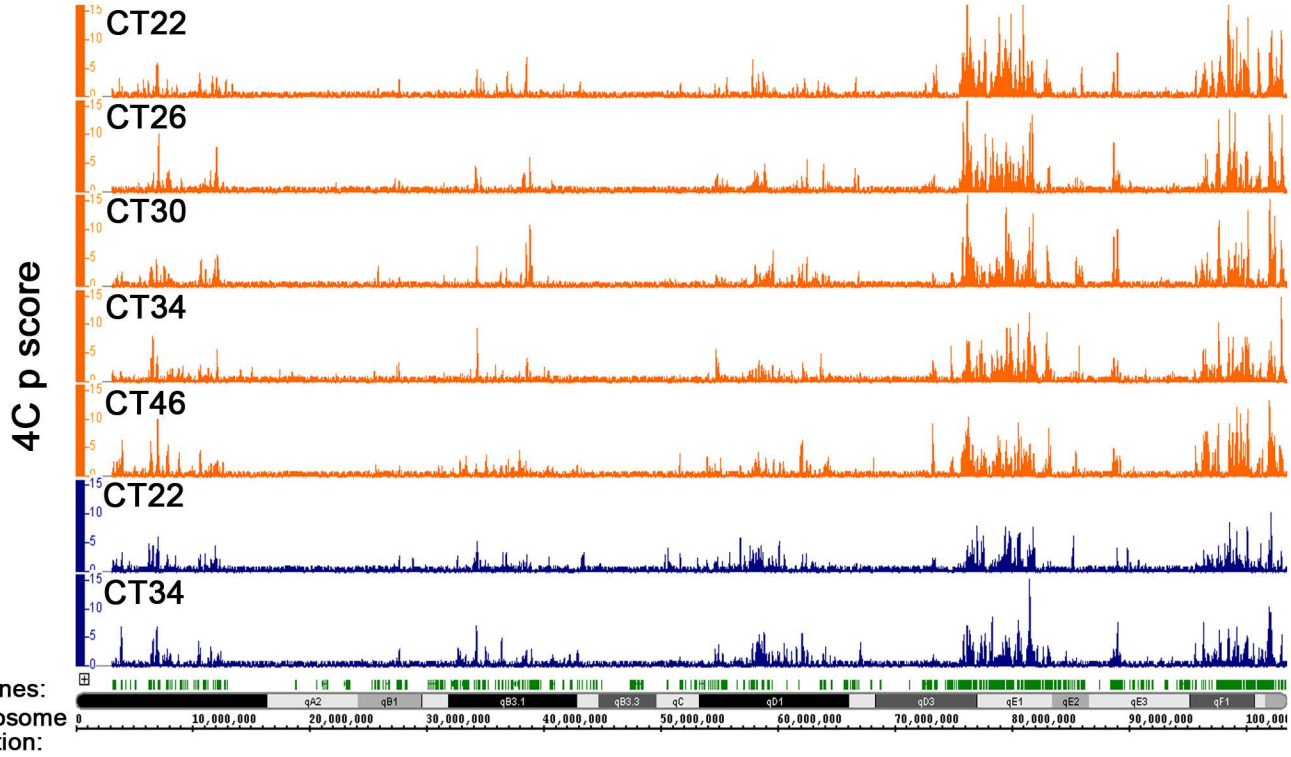

Chromosome 11:
4C p score
CT22
CT26
CT30
CT34
CT46
CT22
CT34
Genes:
Chromosome position:
qA2
qA3.1
qA3.2
qA3.3
qA4
qA5
qB1.1
qB1.3
qB3
qB4
qB5
qC
qD
qE1
qE2
0
20,000,000
40,000,000
60,000,000
80,000,000
100,000,000


Chromosome 12:
4C p score
CT22
CT26
CT30
CT34
CT46
CT22
CT34
Genes:
Chromosome position:
qA1.2
qA1.3
qA2
qA3
qB1
qB3
qC1
qC2
qC3
qD1
qD2
qD3
qE
qF1
qF2
0
20,000,000
40,000,000
60,000,000
80,000,000
100,000,000

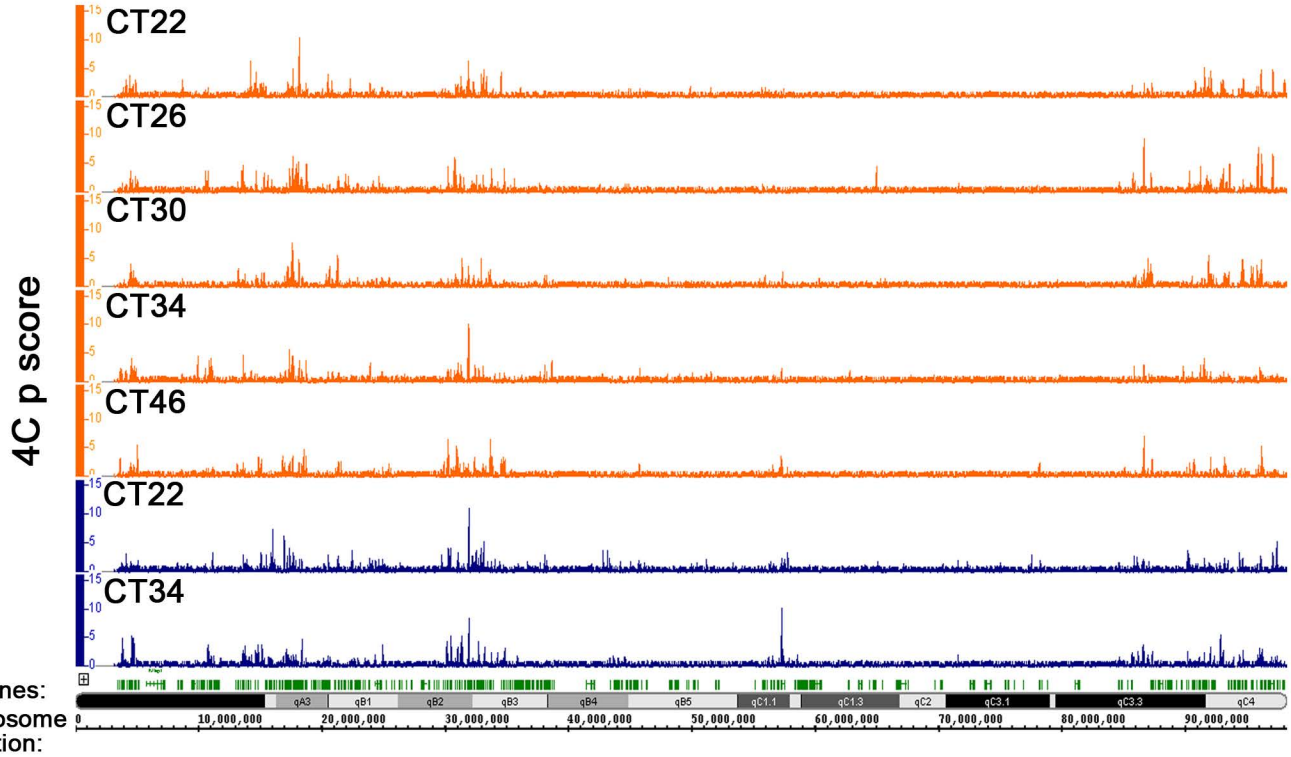

Chromosome 13:
4C p score
CT22
CT26
CT30
CT34
CT46
CT22
CT34
Genes:
Chromosome position:
qA2
qA3.1
qA3.2
qA3.3
qA4
qA5
qB1
qB3
qC1
qC3
qD1
qD2.1
qD2.2
qD2.3
0
20,000,000
40,000,000
60,000,000
80,000,000
100,000,000


Chromosome 14:
4C p score
CT22
CT26
CT30
CT34
CT46
CT22
CT34
Genes:
Chromosome position:
qA2
qA3
qB
qC1
qC3
qD1
qD2
qD3
qE1
qE2.1
qE2.3
qE3
qE4
qE5
0
20,000,000
40,000,000
60,000,000
80,000,000
100,000,000
120,00

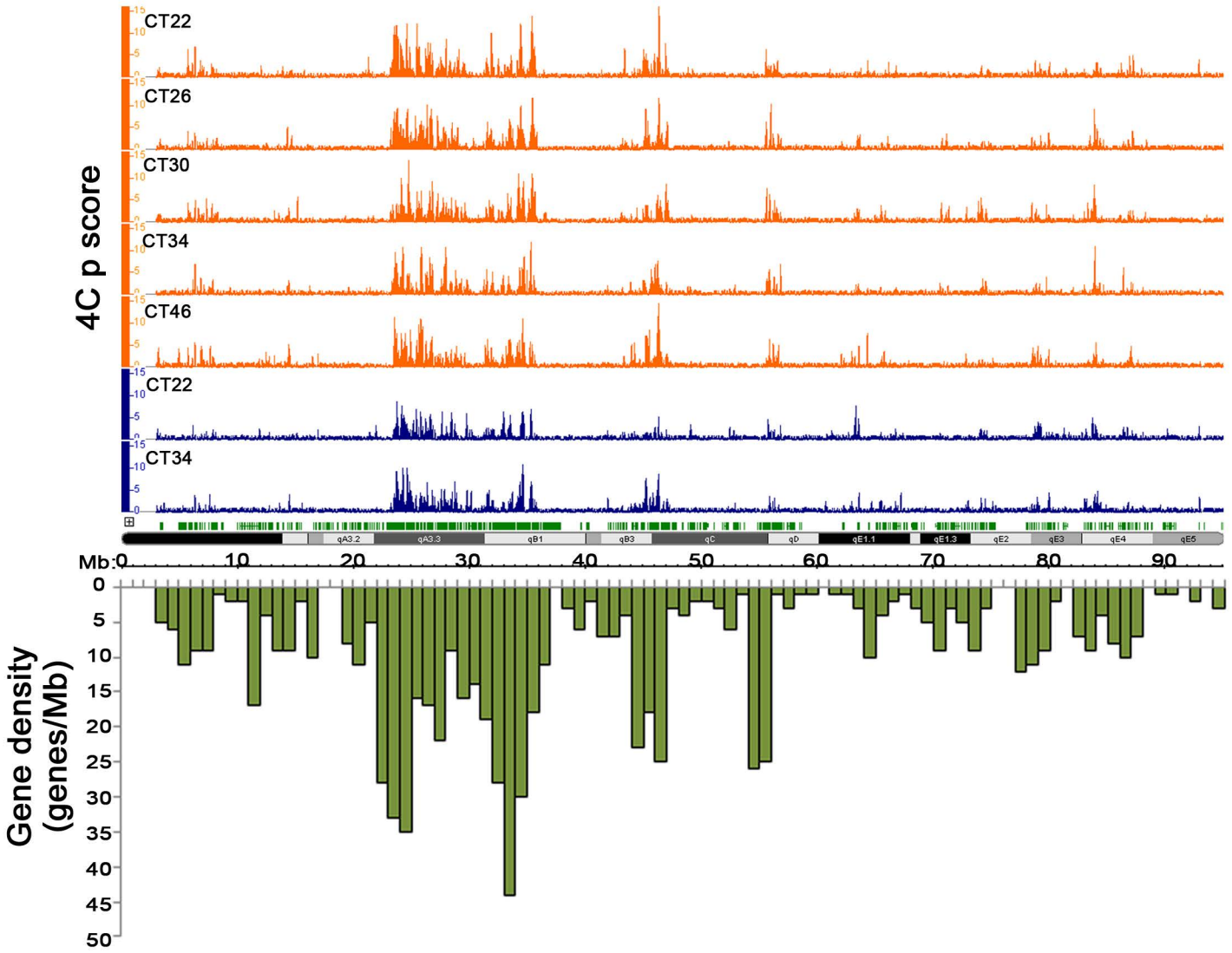
Chromosome 15:



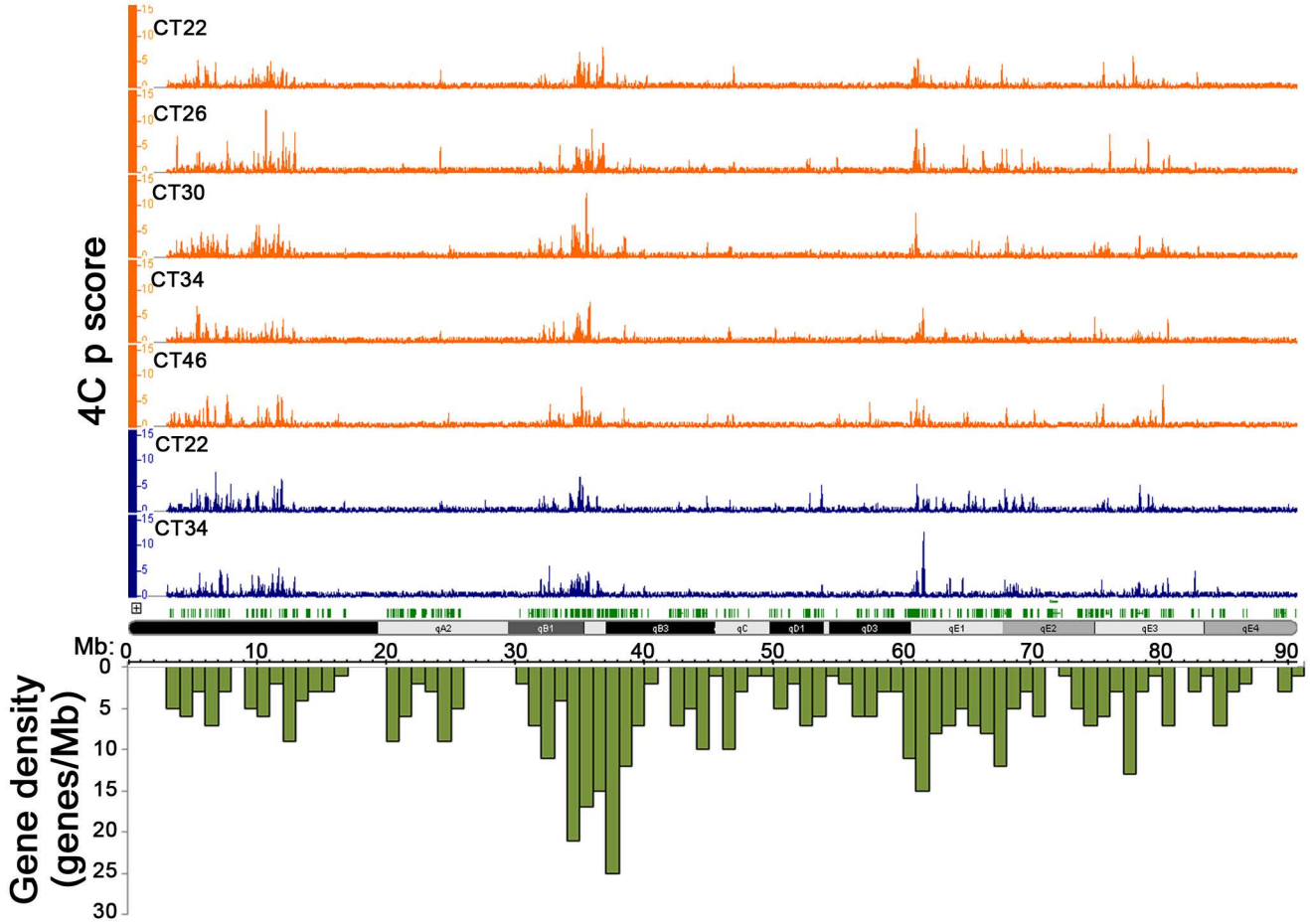
Chromosome 16:

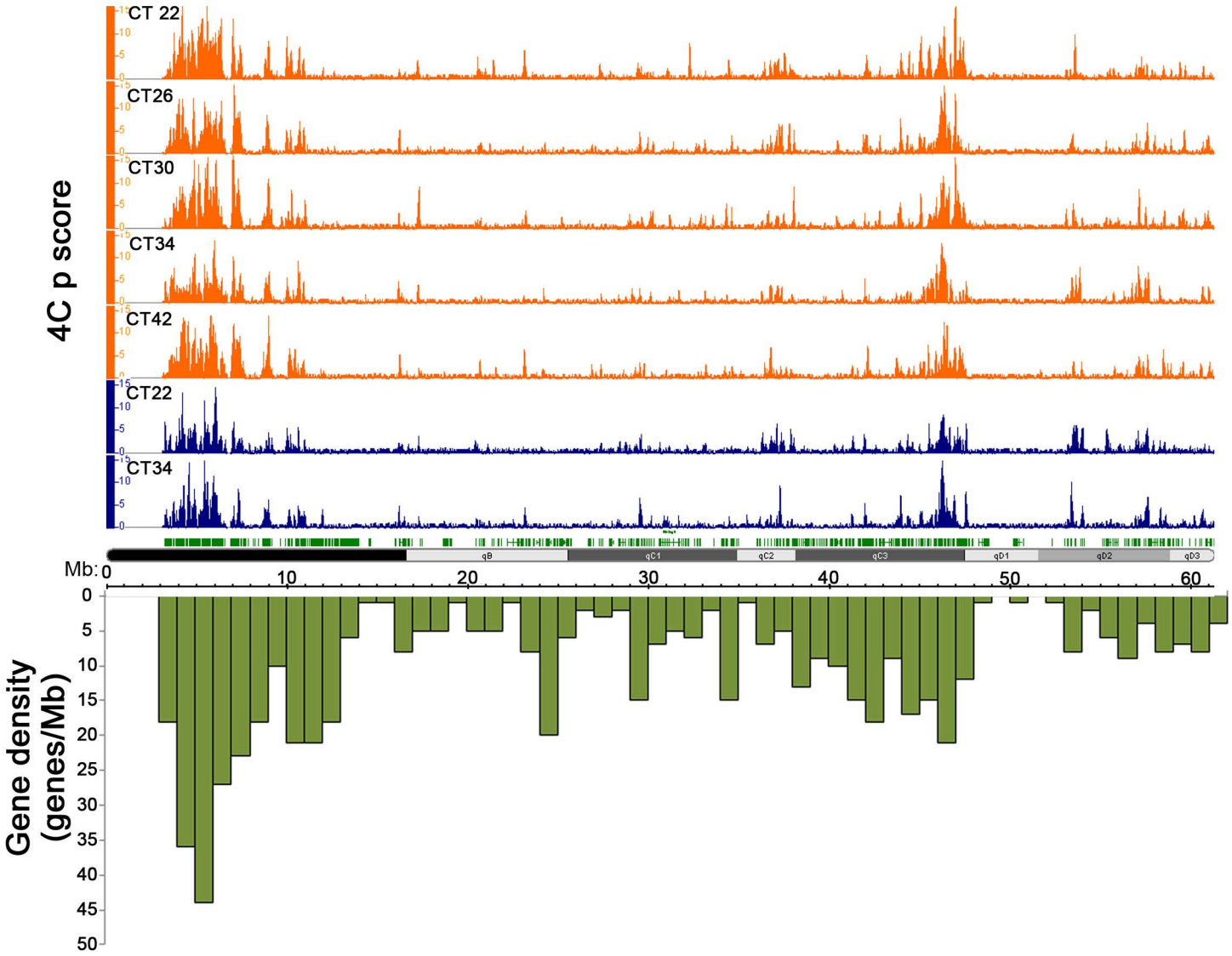


Chromosome 17:



Chromosome 18:





Supplementary Figure 3: Contact frequencies of *dbp* with trans chromosomes. **A**.Wild type (WT) (orange colored plots) and *Bmal1*$^{-/-}$ (blue colored plots) MEFs were analyzed by 4C at the indicated time points (CT, circadian time). The plots show a high correlation in the location of the contacts for all the chromosomes and conditions. The correlation of the contacts with gene density is shown for chromosomes 10, 17, 18 and 19 as an example. Highest running mean values of the 4C data (4C p score) are found at chromosomal regions with a high gene content (bottom histograms). The genomic position in mm8 coordinates is indicated on the horizontal axis.

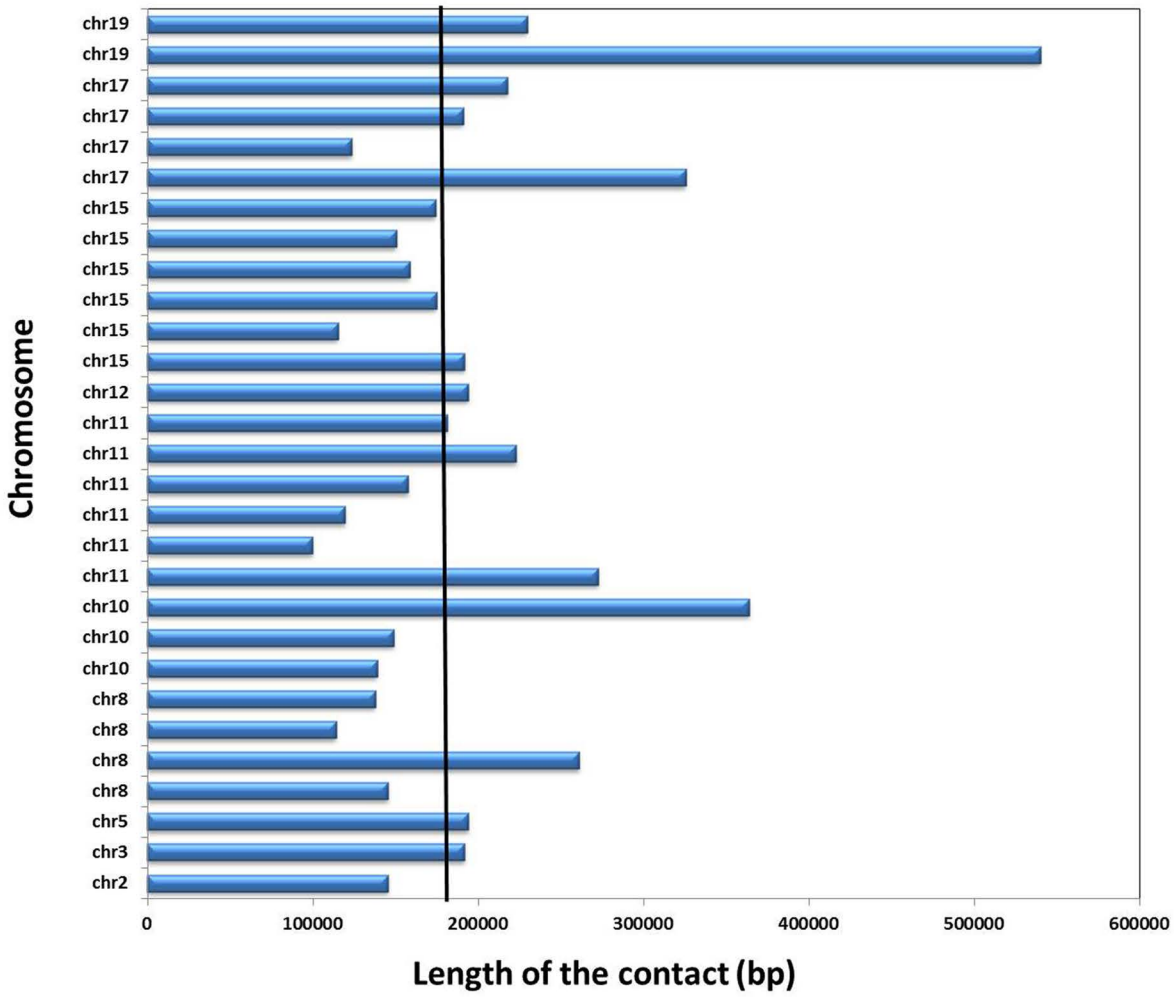
A

WT MEFs gene data		
	Total	Circadian expression
Genes	28,750	1,189
Genes on *dbp* 4C contacts	1,239 *	166 *
Genes on *dbp* circadian 4C contacts	255 *	18 *

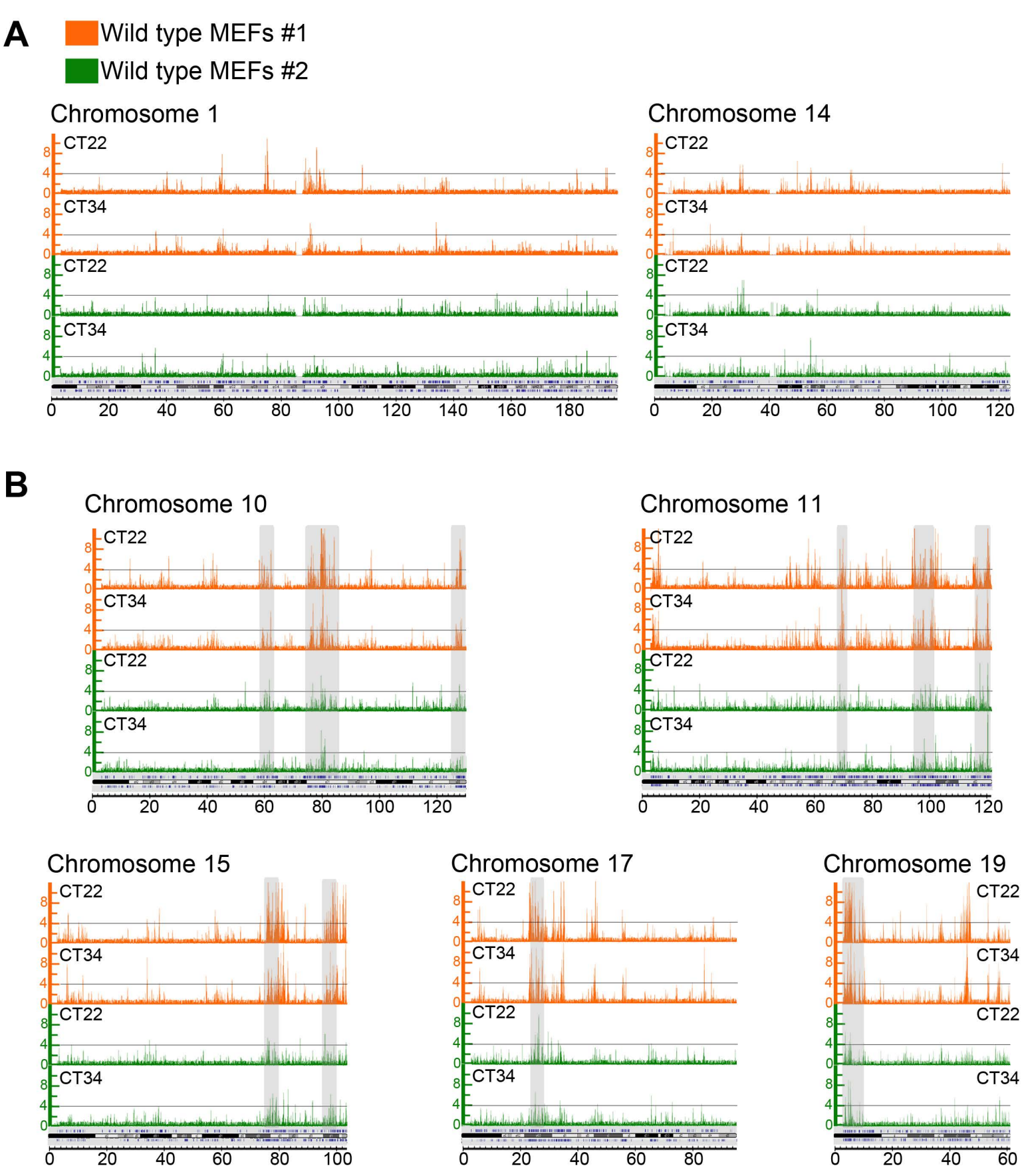
4C data						
	WT-only	*Bmal1*$^{-/-}$ - only	Both	TOTAL WT	TOTAL *Bmal1* $^{-/-}$	TOTAL contacts
dbp 4C contacts (all)	146	10	55	201	65	211
dbp circadian 4C contacts	29	0	0	29	0	29

* *dbp* itself is excluded from this count.

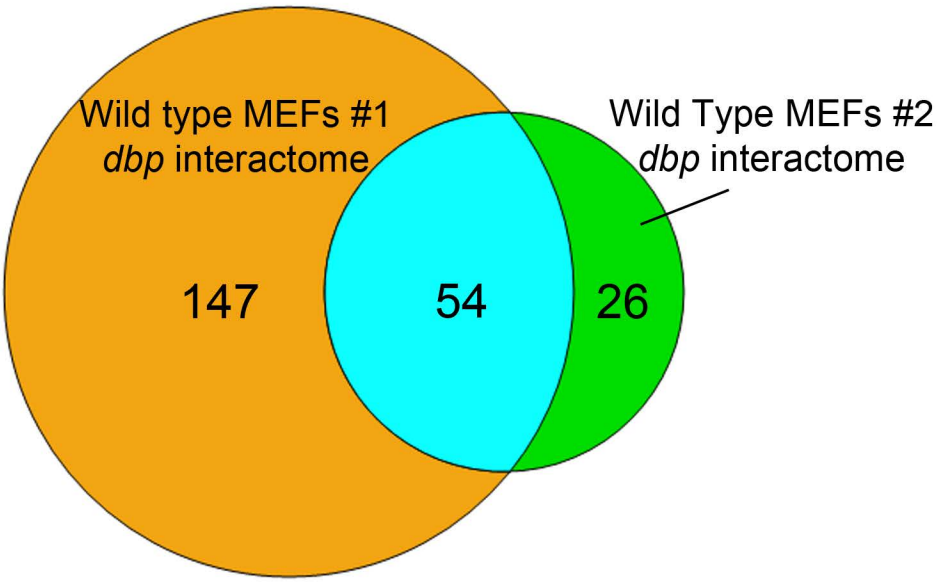
B



Supplementary Figure 4: Different features of the *dbp* interactome, **A**. Number of interacting regions, number of genes and how many of these oscillate. This Table illustrates the most relevant figures related to the *dbp* interactome (see the text for further discussion). **B.** Histograms representing the length of the contacts for the *dbp* circadian interactome. The vertical axis indicates the chromosome which contains the contact; these are organized according to their order of appearance in each chromosome. A black vertical line indicates the mean length for all the contacts at the *dbp* interactome.

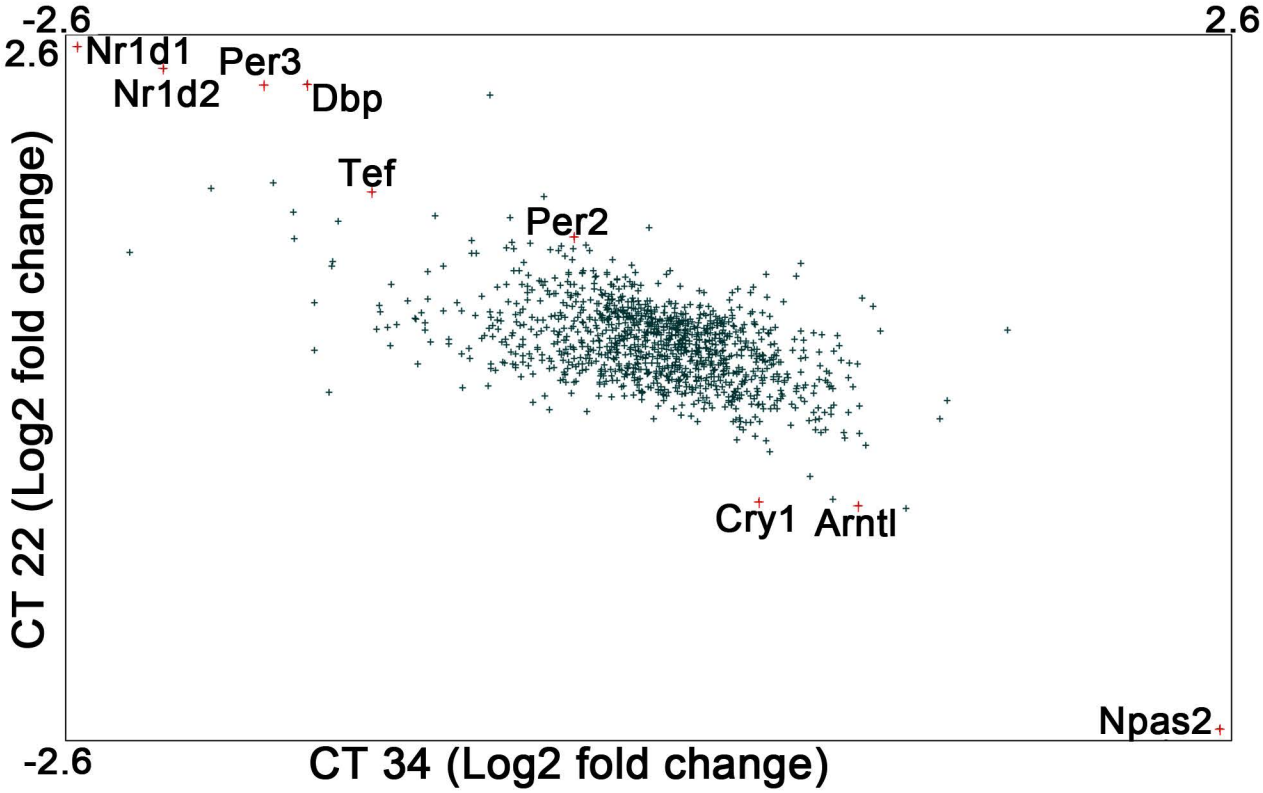

A
Wild type MEFs #1
Wild type MEFs #2
Chromosome 1
CT22
CT34
CT22
CT34
0 20 40 60 80 100 120 140 160 180
Chromosome 14
CT22
CT34
CT22
CT34
0 20 40 60 80 100 120
B
Chromosome 10
CT22
CT34
CT22
CT34
0 20 40 60 80 100 120
Chromosome 11
CT22
CT34
CT22
CT34
0 20 40 60 80 100 120
Chromosome 15
CT22
CT34
CT22
CT34
0 20 40 60 80 100
Chromosome 17
CT22
CT34
CT22
CT34
0 20 40 60 80
Chromosome 19
CT22
CT34
CT22
CT34
0 20 40 60

C

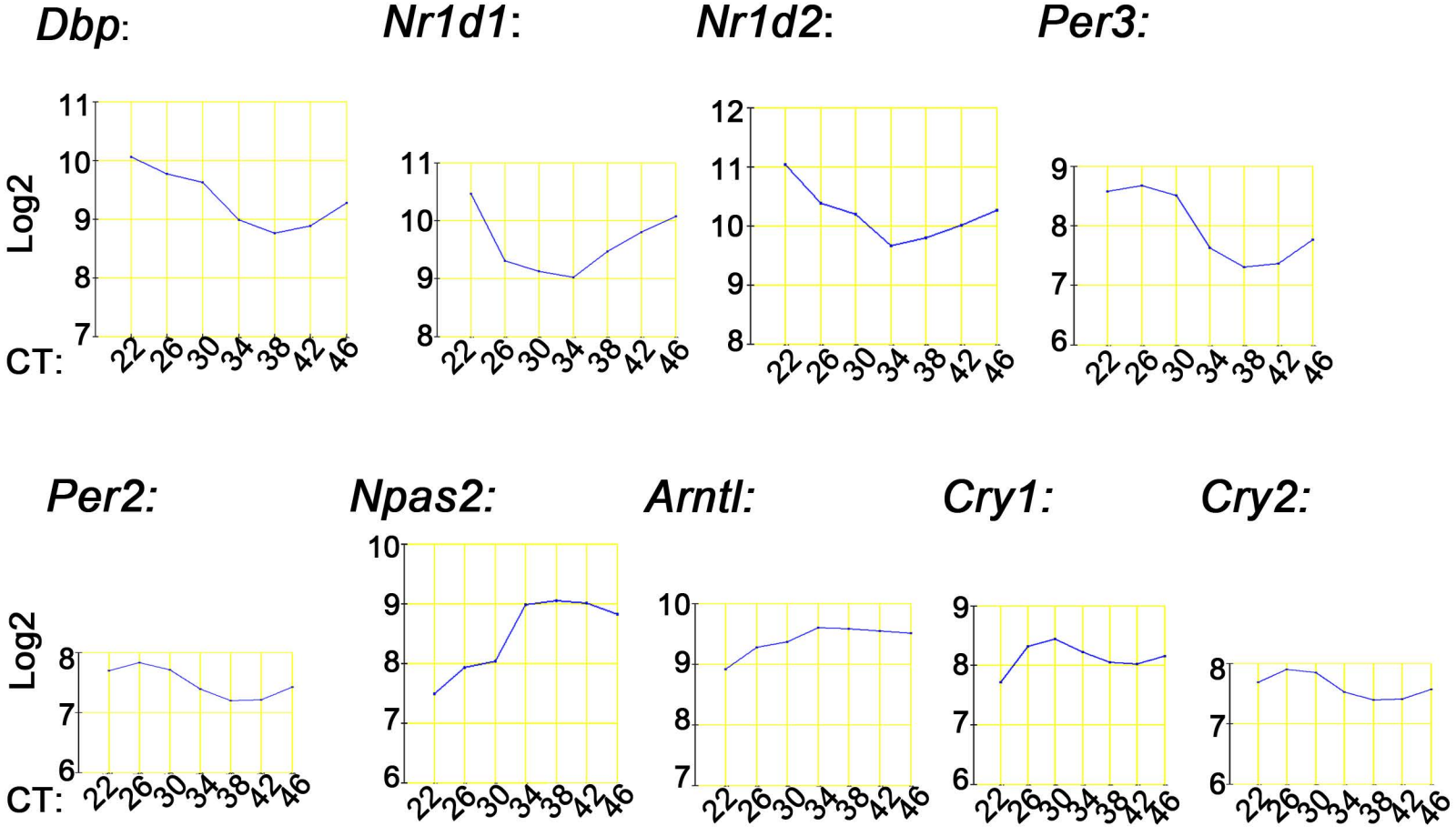


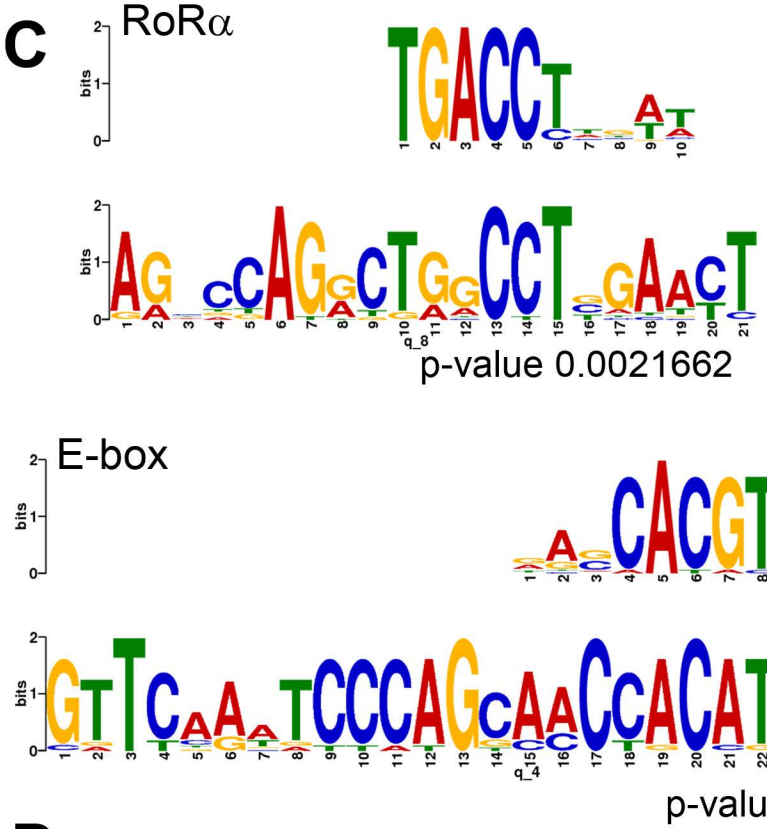
Supplementary Figure 5: Comparative analysis of two 4C biological replicates from two different wild type MEFs lines at CT 22 and CT 34. **A**, Comparative genome browser shots of the 4C p-scores from chromosomes 1 and 14. These chromosomes barely interact with the *dbp* bait. **B**, Comparative genome browsers shots from chromosomes 10, 11, 15, 17 and 19. These chromosomes interact with *dbp* at many locations (see Figure 1 and Supplementary Figure 3). Shaded regions indicate conserved interacting regions between both biological replicates. The genomic positions on mm8 coordinates are indicated on the bottom horizontal axis for each plot. **C**, Venn diagram showing the significant overlap within the *dbp*-interactome between both biological replicates. The numbers represent genomic regions which interact with *dbp*.

A


-2.6
2.6
2.6
Nr1d1
Per3
Nr1d2
Dbp
Tef
Per2
Cry1
Arntl
Npas2
-2.6
CT 22 (Log2 fold change)
CT 34 (Log2 fold change)

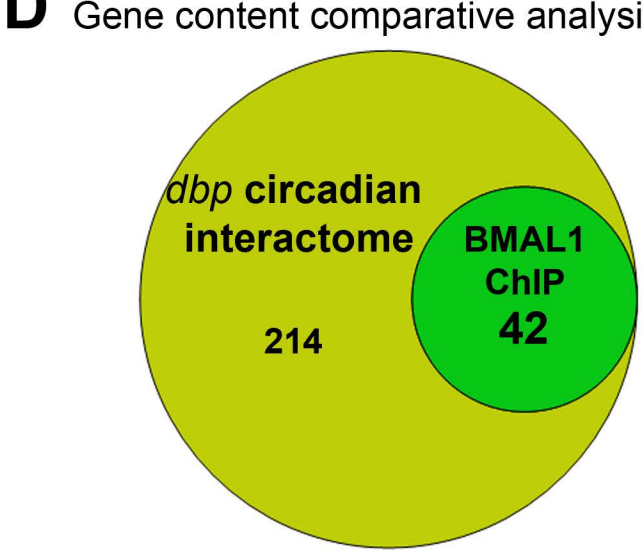
B


Dbp:
Nr1d1:
Nr1d2:
Per3:
Per2:
Npas2:
Arntl:
Cry1:
Cry2:
Log2
CT:
22 26 30 34 38 42 46



E

Gene	Circadian feature
gpd1	BMAL1 ChIP
il12rb1	BMAL1 ChIP
pde4c	BMAL1 ChIP
tlcd1	BMAL1 ChIP
r3hdm2	BMAL1 ChIP
acy3	BMAL1 ChIP
flot2	BMAL1 ChIP
lrrc48	CLOCK/PER2/CRY1 ChIP
ciapin1	CRY1/CRY2 ChIP
sbp	E-box -999
alg12	E-box -267
bak1	RoR element -247
itga3	D-box -431; E-box -714
ash1l	D-box -1617
kxd1	D-box -947
fam222b	D-box -563
gon4l	D-box -1606
gpr114	D-box -2157



Supplementary Figure 6: Circadian gene expression in wild type MEFs and analyses of genes at the *dbp* interactome. **A**, scatter plot of the gene expression data for circadian genes at CT 22 versus CT 34. Genes with higher fold change correspond to the circadian machinery (*Npas2*, *Arntl*, *Per2* and *3*, *Cry 1* and *2*, *Nr1d1, Nr1d2*), and well-known circadian transcription factors (*Dbp* and *Tef*). **B**, graphs representing the average Log2 values of the expression data from three microarrays for each analyzed time point. Circadian genes were selected and plotted as indicated. **C,** motif logos of the over-represented motifs identified by MEME analysis on the promoters of the circadian genes located at the *dbp* interactome. The height of the stack of letters is proportional to the information content as represented in bits, and the relative frequency of each base is given by its relative height. These motifs were queried against JASPAR database using Tomtom program to identify matches to known transcription factor binding motifs. E-box and RoRα elements were significantly identified. The p values below the matrices are related to the Tomtom algorithm. **D**, Venn diagram represents the number of genes that are located at the *dbp* circadian interactome and how many of them are enriched in BMAL1 at their promoters, as identified by BMAL1 ChIP-seq analyses[13,49] . **E**, the table illustrates specific features of the 18 circadian genes located at the *dbp* circadian interactome. ChIP-seq data has been previously reported [13,49]. The numbers indicate the positions of the selected motifs with respect to the TSS for each circadian gene.

SUPPLEMENTARY TABLE LEGENDS:

Supplementary Table 1: 4C genomic regions that interact in *trans* with *Dbp*. **Sheet 1**: the contacts identified in wild type MEFs. **Sheet 2**: The contacts detected in*Bmal1*$^{-/-}$ MEFs. Data corresponds to mm9 genome coordinates.

Supplementary Table 2: Gene content of *Dbp* contacts detected in wild type MEFs. The mm9 genome version is used.

Supplementary Table 3: *Dbp* circadian interactome. **Sheet 1** (*Dbp circadian contacts*): the genomic locations of the *Dbp* circadian interactome. **Sheet 2** (*genes on Dbp circadian contacts*): the gene content of the *Dbp* circadian interactome. The genomic coordinates are given based on the mm9 version.

Supplementary Table 4: Motif Map analyses on 4C contact regions. **Sheet 1** (*Motifs on contacts 4C regions)*: indicates the position of the motif in the *Dbp* contacts and the putative genes that are controlled by the promoter that contain each motif. **Sheet 2** (*Motif map enrichment table*): indicates significantly enriched motifs on the *Dbp* interactome over randomized data. **Sheet 3** (*Clock elements on circadian 4C*)*:* selection of e-boxes, RoRα elements and CREB motifs that are present on the *Dbp* circadian interactome, which is defined on Figure 2 of the main text. The headings of the columns are defined as follows:
MOTIF name: the transcription factor which binds the motif for each row.
MOTIF length: The length (in base pairs) of the position weight matrix used in for the motif identification process.
TF multiplicity: The number of different transcription factors that use this position weight matrix.
Uniprot_NAME: The name of the transcription factor as identified by the UniProt database.
Uniprot ID: The identifier of the transcription factor in the UniProt database.

MOTIF ID: The identifier of the matrix in the TRANSFAC/JASPAR/MotifMap databases.
NBBLS (conservation): Normalized Bayesian Branch Length Score. It is metric of conservation of the predicted transcription factor binding from 0 to 1. 1 means that is completely conserved. This data has been filtered to only keep sites with a NBBLS of at least 0.1, meaning that the site is conserved at least in two organisms in the phylogenetic alignment.
NLOD (Motif Matching): Normalized Log Odds. It is a metric of how well the predicted transcription factor binding site matches the position weight matrix. The data has been filtered to keep the NLOD between 0.8 and 1.
Chromosome: The chromosome in which the motif is located.
MOTIF-Start/ -Stop: The location in the genome (mm9 version) of the motif for each row.
GENES: List of putative genes regulated by the motif for each row. For each gene, several different identifiers are indicated.
BG. COUNT/ BG. TOTAL: The number of gene promoters in the genome which contain the motif for each row (BG count) and the total number of promoters in the MotifMap database (BG total).
EXP.COUNT / EXP. TOTAL: The number of gene promoters that lie in the 4C regions (EXP TOTAL) and how many of those contain the motif for each row (EXP. COUNT).
P VALUE: Two tailed Fisher exact test p-value.

Supplementary Table 5: p scores at the region analyzed by FISH. p scores represent contacts frequencies for the 4C probes located on the selected region for FISH analysis on chromosome 10 and presented in the browser shot (Figure 3A).

Supplementary Table 6: Circadian gene expression in wild type MEFs **Sheet 1:** JTK_cycle analysis or the circadian gene expression microarray experiment. **Sheet 2:** a list with the circadian genes on wild type MEFs.
Some of the headings for each column are described in the Affymetrix gene chip manual (http://www.affymetrix.com/support/technical/manual/taf_manual.affx#array_information).The rest corresponds to the parameters calculated by JTK-cycle, and are defined as follows:
JTK_ADJP: JTK p value (see Online Methods).

JTK_AMP: Amplitude of the oscillation.

JTK_LAG: Lag of the oscillation.

JTK_PER: Period of the oscillation.

JTK_BHQ: JTK q value (FDR metric).

Supplementary Table 7: Ontological analyses. **Sheet 1:** Gene Ontology (GO) analysis and significantly enriched amongst the circadian genes. **Sheet 2:** Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of circadian genes. **Sheet 3:** Gene Ontology analysis of the genes located in circadian 4C contacts. For all the analyses in this table, the p value is calculated using the Hypergeometric distribution and indicated in the "p-value" headed columns. A FDR correction is then applied (see Online Methods) and the results are indicated in the "Corrected p-value" headed columns.

Supplementary Table 8: Lists of circadian genes. **Sheet 1:** list of circadian genes that are located on *Dbp* 4C contacts. **Sheet 2:** list of circadian genes which can be found at *Dbp* circadian interactome.