Cell Reports, Volume *28*

Supplemental Information

Modeling Single-Molecule Conformations of the *HoxD* Region in Mouse Embryonic Stem and Cortical Neuronal Cells

Simona Bianco, Carlo Annunziatella, Guillaume Andrey, Andrea M. Chiariello, Andrea Esposito, Luca Fiorillo, Antonella Prisco, Mattia Conte, Raffaele Campanile, and Mario Nicodemi

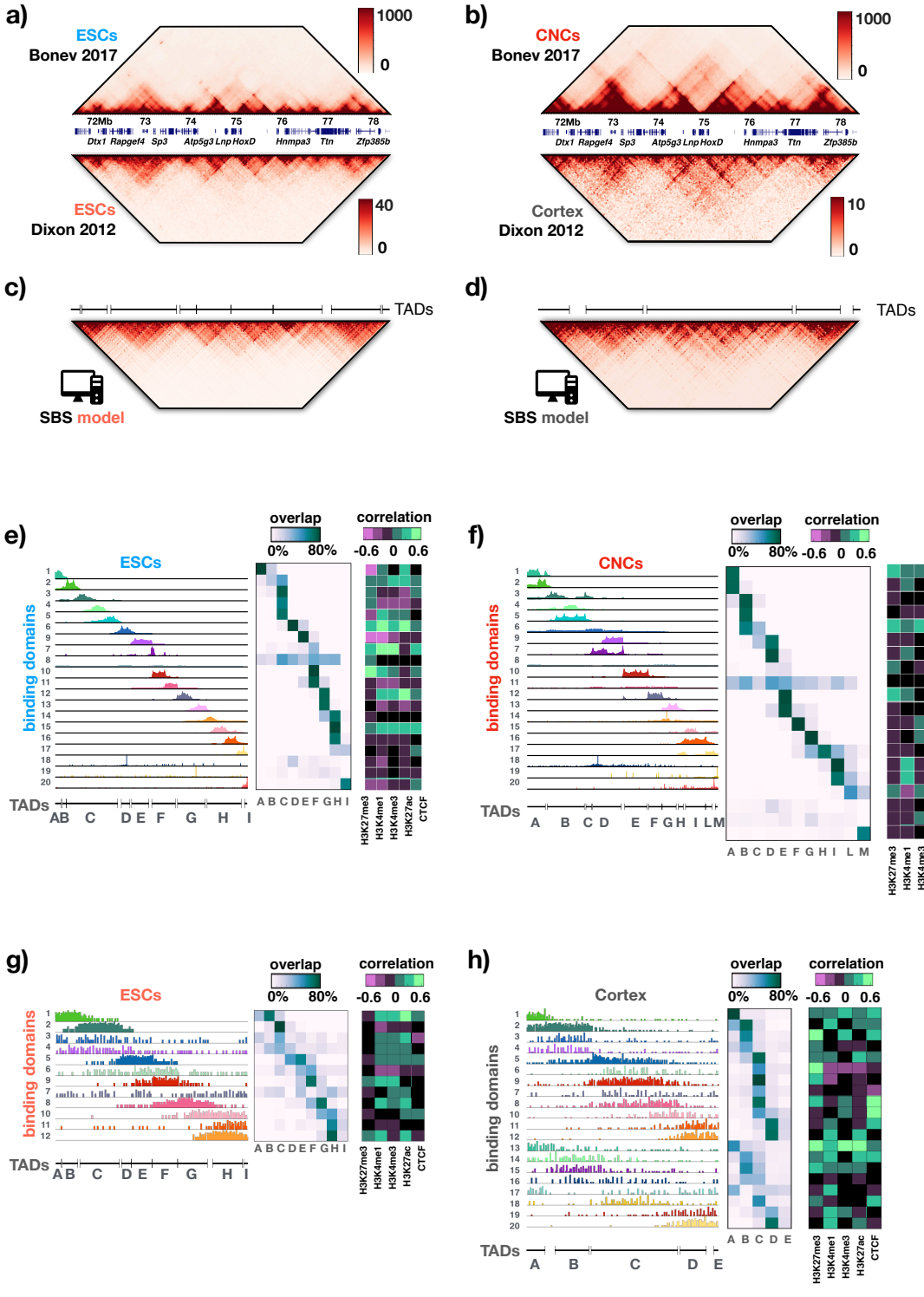
a)
ESCs
Bonev 2017
1000
0
72Mb 73 74 75 76 77 78
Dtx1 Rapgef4 Sp3 Atp5g3 Lnp HoxD Hnmpa3 Ttn Zfp385b
ESCs
Dixon 2012
40
0
b)
CNCs
Bonev 2017
1000
0
72Mb 73 74 75 76 77 78
Dtx1 Rapgef4 Sp3 Atp5g3 Lnp HoxD Hnmpa3 Ttn Zfp385b
Cortex
Dixon 2012
10
0
c)
TADs
SBS model
d)
TADs
SBS model
e)
ESCs
overlap
0% 80%
correlation
-0.6 0 0.6
binding domains
TADs
AB C DE F G H I
A B C D E F G H I
H3K27me3
H3K4me1
H3K4me3
H3K27ac
CTCF
f)
CNCs
overlap
0% 80%
correlation
-0.6 0 0.6
binding domains
TADs
A B C D E F G H I L M
A B C D E F G H I L M
H3K27me3
H3K4me1
H3K4me3
H3K27ac
CTCF
g)
ESCs
overlap
0% 80%
correlation
-0.6 0 0.6
binding domains
TADs
AB C DE F G H I
A B C D E F G H I
H3K27me3
H3K4me1
H3K4me3
H3K27ac
CTCF
h)
Cortex
overlap
0% 80%
correlation
-0.6 0 0.6
binding domains
TADs
A B C D E
A B C D E
H3K27me3
H3K4me1
H3K4me3
H3K27ac
CTCF

Figure S1. Comparison of Hi-C datasets at the murine *HoxD* locus and SBS models in ESCs and Cortex cells at 40kb resolution. (related to Fig. 1)

Hi-C dataset at the *HoxD* region from Bonev et al. 2017 binned at 40kb resolution (**a,b top**) compared to the 40kb resolution dataset from Dixon et al. 2012 (**a,b bottom**). The Pearson correlation coefficient between the two datasets at 40kb in ESCs (**a**) is r=0.84, and correlation between in-vivo Cortex data (Dixon et al. 2012) and in-vitro differentiated Cortical Neuronal cells (Bonev et al. 2017) at 40kb (**b**) is r=0.82.

SBS model average contact matrix of the *HoxD* region in ESCs (**c**) derived from 40kb Hi-C data from Dixon et al. 2012 (**a, bottom**) have a Pearson correlation r=0.96 and a distance corrected correlation r'=0.70. 40kb. SBS model average contact matrix in Cortex (**d**) derived from 40kb Hi-C data from Dixon et al. 2012 (**b, bottom**) have a Pearson and a distance corrected correlation r=0.92 and r'=0.71. Black segments show the TADs of the region (Dixon et al. 2012).

Genomic location of the model different binding domains (colors) in ESCs (**g, left**) and CNCs (**h, left**), inferred from 5kb resolution Bonev et al. 2017 Hi-C data, their overlaps with TADs as defined in Bonev et al. 2017 (**g,h, middle**) and their significant correlations with available chromatin marks (ENCODE; Bonev et al. 2017) (**e,f, right**). The color scheme of the domains in CNCs (h, left) is chosen by the highest overlap with the domains in ESCs (g, left), just for visualization purposes. The domains reported in Figure 1a and 1d are those providing the main contribution to contact formation (as highlighted in Figure 1b and 1e).

Genomic location of the model different binding domains (colors) in ESCs (**e, left**) and Cortex cells (**f, left**), their overlaps with TADs (**e,f, middle**) and their significant correlations with available ENCODE marks (**e,f, right**). The color scheme of the domains in Cortex (f, left) is chosen by the highest overlap with the domains in ESCs (e, left), just for visualization purposes.

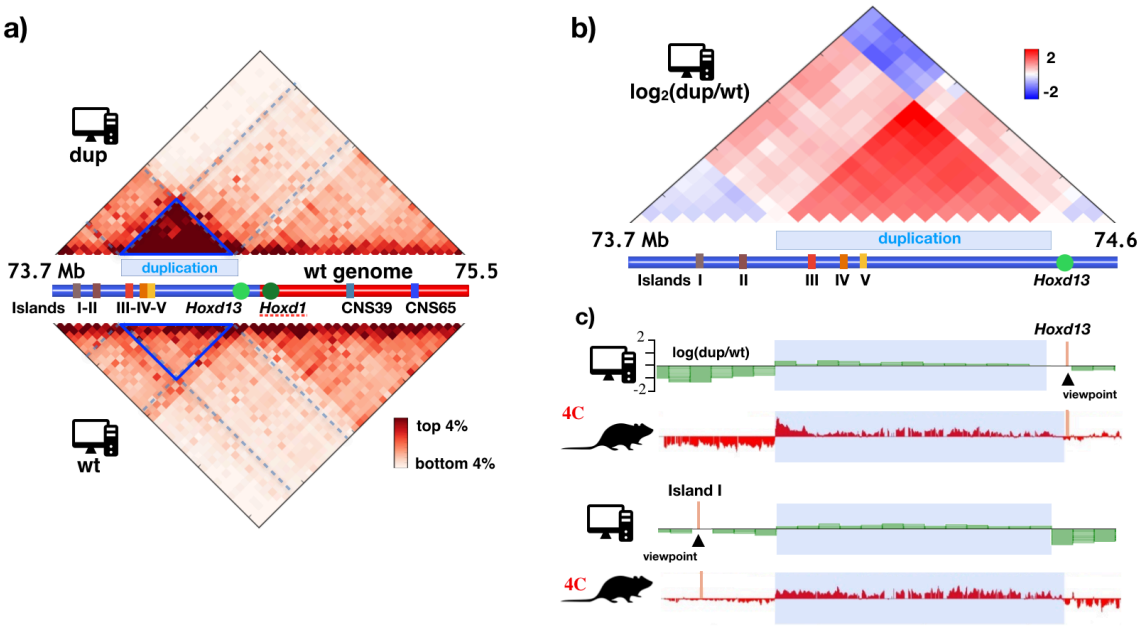

Figure S2. Prediction of the effect of the Nsi-SB duplication from the 40kb polymer model in ESCs. (related to Fig. 2)

Implementation of the Nsi-SB duplication in the SBS polymer model of the ESCs WT locus at 40kb resolution (Figure S9a). A pattern of increased contacts is predicted in the contact map (**a**) of the mutated system (blue triangles). The log2 ratio of the mutated and wild type signals (**b**) better highlights the interaction changes. Model predictions compare well with published, higher resolution 4C data (**c**, Montavon et al. 2012) from the *Hoxd*13 and Island I viewpoints in murine limb cells (adapted from Montavon et al. 2012).

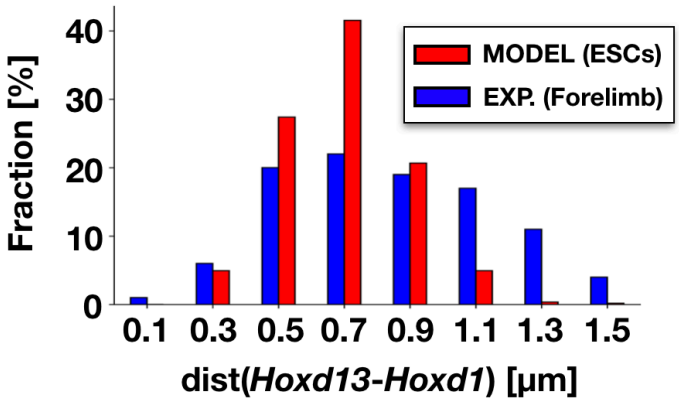

Figure S3. Model v.s. FISH distance distribution. (related to Fig. 3)

The figure shows FISH data (Fabre et al. 2015) about the *Hoxd13-Hoxd1* distance distribution in Forelimb tissue (blue) and in ESCs 5kb resolution model distances (red). The shape of the distribution is statistically similar (p-value=0.1, Kolmogorov-Smirnov test) after a scaling factor (in Forelimb the mean distance is approx. 700nm and in ESCs it is 350nm).

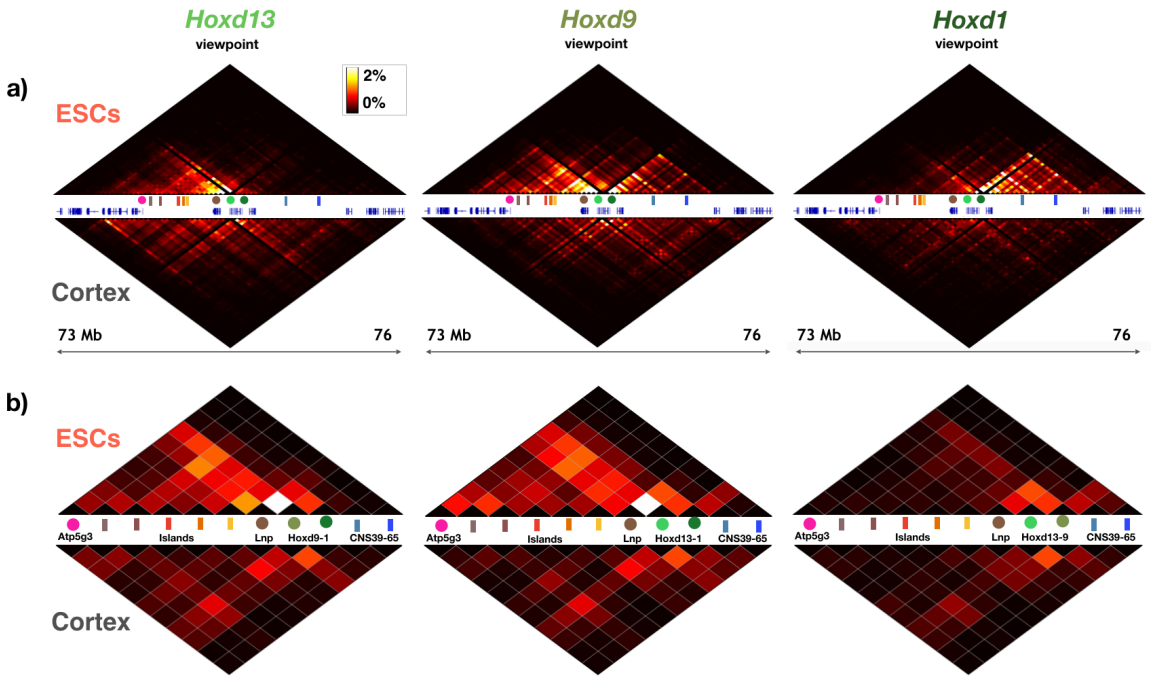

Figure S4. Triple contact probabilities of genes and regulators at the 40kb resolution *HoxD* locus in ESCs and Cortex tissue. (related to Fig. 4)

(**a**) Single-cell triple contact probability from the viewpoint of *Hoxd1*, *Hoxd9* and *Hoxd13* derived from the SBS model based on Dixon et al. 2012 40kb Hi-C data. In ESCs, *Hoxd13* and *Hoxd1* form triplets especially within respectively centromeric and telomeric TADs, while *Hoxd9* form triplets with both. Conversely, in Cortex they all share broader interactions within a larger meta-TAD. (**b**) Promoter specific subset of triplets are formed by genes and regulators within the *HoxD* region.

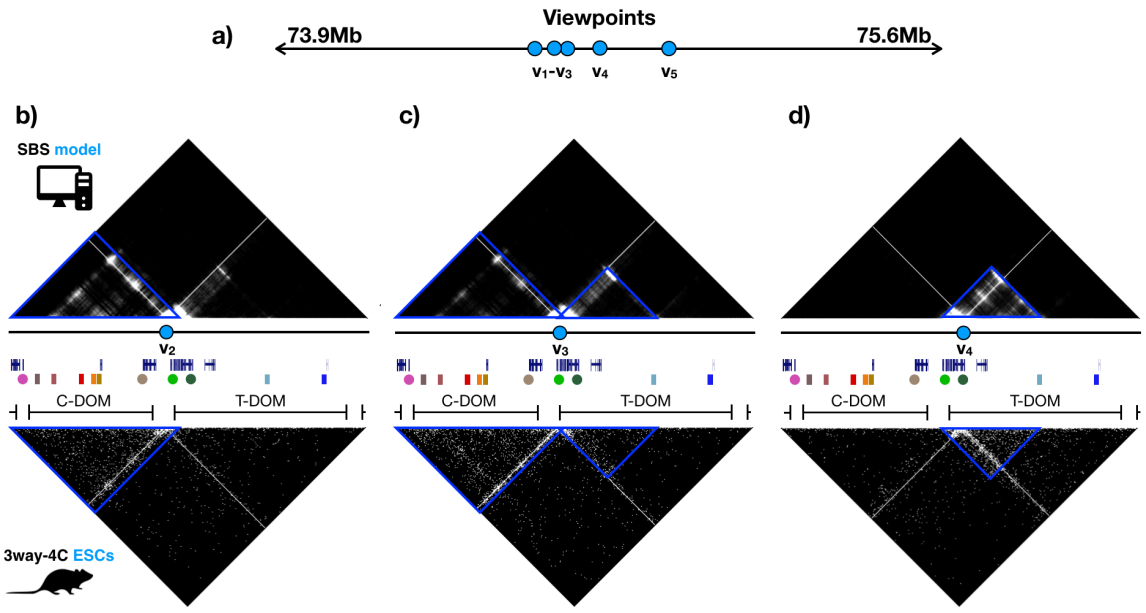

Figure S5. Comparison of model predicted and experimental triplets at the *HoxD* locus in ESCs. (related to Fig. 5)

(**a**) 3-way 4C data viewpoints at the *HoxD* locus from (Olivares-Chauvet et al. 2016). (**b**) Comparison of the triplet contact probabilities predicted from the SBS model based on Bonev et al. 2017 high resolution data (top) and experimental 3-way 4C data (Olivares-Chauvet et al. 2016, bottom) in ESCs, from viewpoint, v2, (**c**) viewpoint v3 (*Hoxd13*) and (**d**) viewpoint v4 (*Hoxd1*).