Electronic Supplementary Material (ESI) for Molecular BioSystems.

Supplementary Text T5

We provide below 2 examples for the impact of our correction on spatial modelling. We have used here 1 megabase (1Mb) contact maps of GM06690 (summing the NcoI and HindIII replicates), generated as described in the paper and in Supplementary Text T4 and Supplementary Text T1, with and without fragment filtration, and corrected contact maps, generated with our variance stabilization correction (VS) and with YT [1] , HiCNorm [2] and ICE [3]. For the purpose of spatial modelling, each contact map was further converted to a distance matrix as described in the paper (taking its inverse). Finally, a 3-dimensional model was produced for each contact map with multi-dimensional scaling, using the smacof R package [4].

Example 1

Using FISH data of markers for chromosome 14 in GM06690, from [5], we show that our method provides a better fit with the intra-chromosomal distances. Here, 3 FISH markers were considered: L2, L3 and L4, where the approximated ratio between L2-L3 and L2-L4 distances (based on FISH distance curves [5]) is 1.3-1.6. In the observed model the ratio between L2-L3 and L2-L4 distances was 2.86 (3.02 with fragment filtration). YT, HiCNorm and ICE provided a slightly lower ratio but still above 2 (Table T5.1). Instead, our correction (VS) provided a ratio of 1.34 (1.24 with fragment filtration) with a much better fit to the real intra-chromosomal distance.

Example 2

Similarly to example 1, we use FISH data available from [5] of markers in chromosome 14: L1, L2, L3. Here the ratio between the L2-L3 and L1-L3 distances is approximately 1-1.3 (based on FISH distance curves [5]). VS provided the best fit with this ratio (1.15 and 1.12 with and without fragment filtration) while YT, HiCNorm and ICE provided an over-estimation similarly to the observed data (Table T5.1).

Table T5.1. Distances between markers, calculated for spatial models based on the observed and corrected distance matrices. The ratio with the best fit with the data (in both cases achieved with VS) is highlighted in each example.

	Observed	Observed [with fragment filtration]	VS	VS [with fragment filtration]	YT	HiCNorm	ICE
Example 1							
Distance L2-L3	1.26	0.88	1.21	1.09	0.94	1.26	1.04
Distance L2-L4	0.44	0.29	0.90	0.88	0.44	0.54	0.45
Ratio	2.86	3.02	**1.34**	**1.24**	2.13	2.33	2.31
Example 2							
Distance L3-L2	1.26	0.88	1.21	1.09	0.94	1.26	1.04
Distance L3-L1	0.43	0.29	1.05	0.97	0.45	0.47	0.56
Ratio	2.93	3.02	**1.15**	**1.12**	2.08	2.68	1.86

References

1. Yaffe E and Tanay A: **Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture**. *Nat. Genet*. 2011, **43**:1059–65.
2. Hu M, Deng K, Selvaraj K, Qin Z, Ren B, and Liu SJ: **HiCNorm: removing biases in Hi-C data via Poisson regression**. *Bioinformatics* 2012, **28**:3–5.
3. Imakaev M, Fudenberg G, Mccord RP, Naumova N, Goloborodko A, Lajoie, BR, Dekker J and Mirny LA: **Iterative correction of Hi-C data reveals hallmarks of chromosome organization**. *Nat Methods* 2012, **9**: 999–1003.
4. Leeuw JD and Mair P: **Multidimensional Scaling Using Majorization: SMACOF in R**. *Journal of Statistical Software* 2009, **31**:1-30. URL http://www.jstatsoft.org/v31/i03/.
5. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny

LA, Lander ES and Dekker J: **Comprehensive mapping of long-range interactions reveals folding principles of the human genome**. *Science* 2009, **326**:289–93.