Supplementary Text T2

The strong correlation between Hi-C coverage and occupancy of restriction fragments, as shown in Supplementary Text T1, can be exploited to perform an assessment of expected coverage. Changes in Contained Restriction Fragments (CRF) distribution are likely to results in changes in coverage (see Supplementary Text T1), and change point analysis can be used to detect them. Given n change points $p_1, p_2, .. p_n$, we define *S* as the average length of the segments $l_1, l_2, .. l_{n+1}$, where l_i is defined as the segment starting at position p_{i-1} and ending at position p_i (the first and last segments are bounded by the first and last positions of the observation set, correspondingly). *S* is a measure for the spread of change points, where a more homogenous distribution will result in larger values of *S*. *S* will take the size of the observation set when there are no change points (fully homogenous) and will decrease as more change points are added (Figure T2.1). If the S value of an enzyme E1 is larger than the S value of a second Enzyme E2, we expect E1 to provide a more homogenous coverage. Following this assumption we have computed *S* values from change points detected for CRF observations (of 1Mb regions) and discriminated between HindIII and NcoI, based on the expected coverage homogeneity (for each chromosome). We then evaluated this discrimination by comparing to *S* values computed for *cis* 1D contact profiles (excluding self-interactions from the profile) generated from contact maps of GM06990 NcoI and HindIII replicates (contact maps generated as described in Supplementary Text T1). We found that for 18 out of the 23 chromosomes, the chosen enzyme (based on *S* values of CRF) also provided a better (or equal) coverage (equal or larger S value for *cis* contact profiles, Table T2.1). The final choice (HindIII), based on the number of chromosomes that achieved larger *S* values of *CRF* change points, was also the one that gave a more homogenous coverage overall. These results suggest that wavelet Poisson change point analysis can be used to estimate coverage, which is important both for improving and guiding experimental design as well as for analyzing coverage bias.

The different cutting preferences of HindIII and NcoI may be complementary in some cases and can be exploited for improving coverage homogeneity. To this end, we have further investigated the homogeneity of the separate and combined alternatives. We have computed *S* values for the combined (GM06990 NcoI and HindIII) *cis* 1D contact profiles and compared with the *S* values of the separate alternatives. For 3 chromosomes: 4, 16 and X, the combination resulted in larger *S* values (higher homogeneity). In 9 others, the combined S was equal to one of the separate alternatives, suggesting that overall the combined alternative is only partially advantageous is this case. Depending on the experiment goals, such analysis could be important for choosing an appropriate enzyme or a combination of enzymes.

Table T2.1. S values computed for each chromosome, from change point analysis for CRF and *cis* 1D contact profiles, with GM06990 NcoI and HindIII replicates. For 18 out of the 23 chromosomes, the chosen enzyme based on *S* CRF values also provided a better (or equal) coverage (equal or larger *S* values for contact profiles). CRF, Contained Restriction Fragments;

Chromosome	HindIII S (CRF)	HindIII S (contact profile)	NcoI S (CRF)	NcoI S (contact profile)
1	31.63	42.17	36.14	50.6
2	35.43	35.43	49.6	41.33
3	29.14	40.8	25.5	40.8
4	196	39.2	28	21.78
5	92.5	61.67	61.67	30.83
6	43.75	87.5	25	35
7	162	54	40.5	81
8	75	150	30	30
9	36	48	24	36
10	69.5	139	69.5	139
11	138	69	46	46
12	33.75	27	67.5	135
13	29.25	58.5	39	39
14	36.33	54.5	36.33	54.5
15	51.5	34.33	34.33	51.5
16	30.33	30.33	45.5	30.33
17	81	27	20.25	20.25
18	78	26	78	78
19	66	33	66	66
20	64	64	32	64
21	24	24	48	24
22	25.5	25.5	25.5	25.5
X	39.75	26.5	79.5	22.71

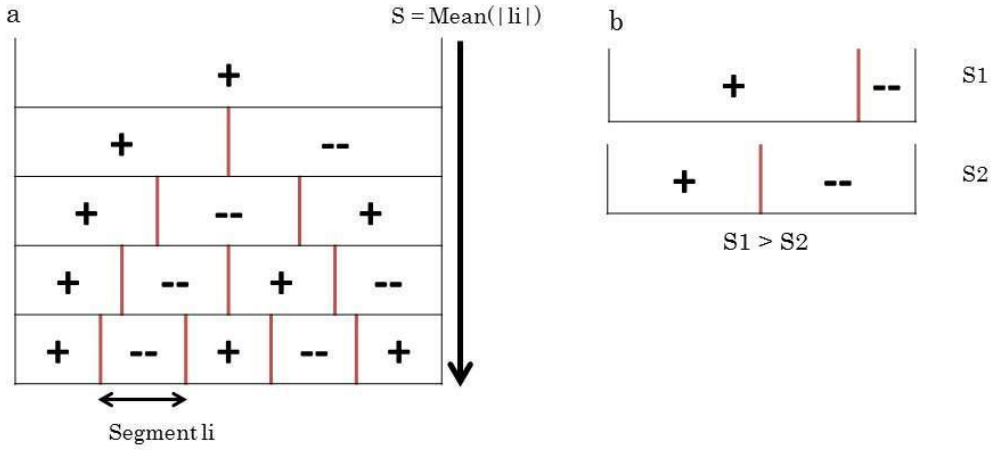

Fig. T2.1. *S* as a measure for change point spread and homogeneity. As change points are added (a), the S value (mean of segments' length) decreases, representing decreasing homogeneity. The spread of the change points also affect S where longer segments contribute to higher homogeneity and larger values of S (b).