Reviewer's report

Title: Tight Associations Between Transcription Promoter Type and Epigenetic Variation in Histone Positioning and Modification

Version: 2 **Date:** 27 January 2011

Reviewer number: 3

Reviewer's report:

- Major Compulsory Revisions

1) How did you define single TSS representing broad promoters? Dominant CTSS or something else? It is very important to know how the TSS was defined, because all plots have 1 bp resolution so the impact of definition is big.

2) What are the distributions of gene expression levels for genes with peak and broad promoters? Are they the same? Is it possible that H3K4me1, -2 and -3 and H3K9ac are associated with broad promoters more often just because they are generally more expressed? It would be nice to show plots for expression levels for two types of promoter.

3) It is not really clear how the percentages denoted on x-axis in Figures 1,2 and 4 were calculated. It should be better explained in the methods. Is it the following: for each genomic position the number of promoters having nucleosome on this position (or +/- 15 bp upstream or downstream) was divided by total number of promoters times 31 (since each nucleosome is represented in 31 position). If this is the case, than the maximal peaks, e.g. first downstream peak for broad promoters, are derived from less than 50% of the promoters, and the percentage of peak promoters contributing to the observed distribution is even smaller.

Is the annotation of x-axis maybe wrong and the numbers do not represent percentages (as denoted in the axis label) but fractions?

4) The method used to scan for TATA box motifs is not explained. What was the position-specific weight matrix used? Which cutoff was used? Figure S1 shows that TATA box is overrepresented in peak promoters, however, when calculated, the absolute number of broad promoters associated with TATA-box is larger than number of peak promoters associated with TATA-box.

5) What is the point of using ChIP-chip data for H3K9 acetylation? What do you want to show with the inserted figure in Figure 2D? Moreover, the ChIP-chip data comes from a different cell type. This part should be explained better and incorporated in the interpretation, or completely removed.

6) Why were the differences in the nucleosome proportions normalized only by the sum of proportions of broad promoters? This would be fine if the ratio of sums of proportions between broad and peak promoters was constant for all

modifications. Since this is not the case, such calculation is biased. The normalization of difference should include both sum of proportions of broad and peak promoters.

7) There are no results shown for the method explained in the last paragraph in the methods (comparison of nucleosome distributions (H3K4me3) located upstream of TSSs (positions from -150 to -100 bp) associated with broad promoters and those associated with peak promoters where downstream genes show similar expression level). Supplementary Figure 3 shows only distribution of expression levels in selected range for two types of promoters.

8) What do the numbers on x-axis in Figure 5 represent? How were absolute gene expression levels derived from microarray data?

9) How were you able to classify genes into two groups (with or without modificated histones) just by looking if they have any of 19 (or 18) different modifications? The division is too simplified. The genes having just one activating methylation were put together with the ones having only one repressive methylation and with all other genes having different kinds of methylation combinations. It is quite a peculiar idea to pool all modifications together and try to correlate with gene expression levels.

10) The suggestion in the discussion:

”Because Sp1 is ubiquitously expressed, it can bind to the nucleosome-free regions in any tissue; this may be one of the causes of the ubiquitous expression of genes downstream of broad promoters.”

implies a very simplified explanation of regulation of gene expression, and there is not enough evidence provided in this work to make such claims, especially since only minor fraction of broad promoters have Sp1 motif (as can shown in supplementary Figure 2A).

- Minor Essential Revisions

1) Abstract, last parahraph, last sentence

- not understandable, needs revising

2) Introduction, first paragraph

- ”intracellular” is unnecessary

- please use the plural forms of “transcript “and “histone”

3) Introduction, second paragraph, fifth sentence

- using both ”according to” and ”it is demonstrated” is redundant

- ’promotome’ is only one and it is a set of all promoters, so the construction of the sentence is quite confusing. It would be better to say ”human promoters can be classified into two types…” or ”human ’promotome’ can be divided into two

types of promoters…”

4) Introduction, third paragraph, first sentence

- better to use singular ”chromosomal DNA is packed…”

5) Results, first paragraph, fifth sentence

- use plural ”nucleosomes distribute around...”
- ”TSSs widely spread...” not ”spreads”

6) Results, second paragraph, third sentence

- ” harbors the histone variant H2A.Z” not ” harbors the variant histone H2A.Z”

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: No, the manuscript does not need to be seen by a statistician.