Reviewer's report

Title: Uncovering the variation within epigenomic datasets using the Karhunen-Loeve transform

Version: 1 Date: 1 January 2015

Reviewer: Alejandra Cervera

Reviewer's report:

Major Compulsory Revisions:
None

Minor Essential Revisions:

1. In Conclusions, the authors state that “PCA to uncover latent principal components that were hidden in the data.” PCA or any of its variations, such as FPCA, is a good tool to reduce dimension, not necessarily to unveil latent information. To uncover latent variables, other methods are available, such as independent component analysis (ICA). Thus, I suggest authors rephrase this sentence to avoid the term “hidden”.

2. “We found that genomic features, such as a read coverage over a region, can be suitably modeled by a linear combination of B-spline functions defined by positioning equidistant knots along the peak regions” Why use B-splines? Could authors justify their criteria?

3. “We selected a region 1.5 kb around each TSS, applied B-spline smoothing using 150 B-spline basis functions, and calculated the functional standard deviation around the functional mean.” Why the threshold of 1.5kb and 150 B-splines?

4. Cluster analysis may be misleading if the wrong metric is applied. Did authors follow an Euclidean metric? It is not specified what metric (and other parameters such as centroids) was used for the hierarchical clustering.

Discretionary Revisions

First sentence in Background section I believe it is under dispute. The fact that encode found transcription factors all over the genome does not guarantee encoding for specific biochemical functions.

Cancer is not a fundamental biological process.

turnover -> turnover

several misplaced commas, in particular the oxford comma is not used evenly throughout the document.

Making the script available, perhaps as part of a vignette in NarrowPeaks, or as
supplementary information would provide additional value to the paper.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

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

**Declaration of competing interests:**

I declare that I have no competing interests.