**Major revise outline**

1. Method comparison. Will compare HIMs to clusters using only Hi-C interaction network to highlight the advantage of utilizing both Hi-C interaction network and GRN.
2. Validation. Will add motif and ChIP-seq enrichment for each HIM.
3. HIM dynamics. Will add HIM dynamic analysis centering on housekeeping genes/essential genes.
4. Writing. Will restructure results section to address highly correlated features and wordy and long discussion.

**Comments**

----------- Overall evaluation -----------  
MAJOR   
  
The authors present MOCHI, a method for identifying spatially close clusters of genes that are also regulated by shared subset of transcription factors. Using Hi-C data and regulatory networks curated by another study to characterize spatial and regulatory relationships, they identify clusters of genes that are regulated by a set of TFs, and call such modules heterogeneous interactome modules (HIM). Given the recent developments in the study of the 3D organization of the nucleus, the HIM concept is exciting. However, the authors characterize MOCHI in too many different ways and offer a numerous anecdotes about individual HIMs in different cell types; however, they omit offering a concise and thorough characterization of HIMs that would convince the reader that indeed such modules exist.

**Response:** These can be partly addressed by addressing the following comments.  
  
Authors characterize HIMs with respect to numerous features: A/B compartments, chromosomal subcompartments, super-enhancer contacts and essential genes. Overall conclusion seems to be the case that HIMs are localized largely at the active regions of the nucleus and close to nuclear speckles. However, these 'features' also correlate with each other and simply the transcriptional state of the gene, namely these genes in HIMS are likely actively transcribed. It seems the most compelling part of a HIM is not that it is a spatially close cluster of genes in an active region of the nucleus, but the sharing of regulatory TFs. I would recommend the authors to focus on this aspect more, presenting more evidence to back this up.

**Response:**

1. We will highlight that genes in HIMs are more actively transcribed as compared to genes not in HIMs, genes analyzed in the paper are all transcribed.
2. We will move some correlated features to supplementary file.
3. Regarding the sharing of regulatory TFs, we will add motif and ChIP-seq enrichment analysis as suggested by the reviewer in the last major comment.  
   Added following to the main text:  
   “To assess that the genes in a HIM are co-regulated by the same TF, we downloaded Chip-seq data from ENCODE for 26 TFs in GM12878 and K562 cell types. Overall, among the HIMs having these TFs, more than half (55.85\%) of them have majority of their genes with corresponding TF Chip-seq peaks (Fig. S1).”  
   Note that these HIMs do not show statistically significant enrichment in Chip-seq peaks because of small numbers (average 11) of genes in the HIMs (data not shown).

There is a rather wordy discussion of HIM anecdotes in most sections, which are reasonably backed up by some evidence but ultimately not enough to convince the reader that HIM is a widespread organization unit that has been missed so far. These anecdotes also make it hard to follow the arguments as they are rather long. I would recommend the authors to cut back on these and focus on the most compelling arguments that back up the HIM concept.

**Response:** I will think about how to restructure results section. One strategy is to move some highly correlated features such as replication timing into the supplementary file.  
  
In the section titled 'HIMs undergo changes across cell types', authors characterize HIMs across cell types, which are compelling analyses that can convince the reader of the HIM concept. However, In Fig3A, authors compare gene sets that have been assigned to HIMs across cell types. Since HIMs seem to largely pick up genes in active and transcribed regions, would this Venn diagram look similar if only transcription state of the each gene in each cell type was used to make this figure? Presumably, one can obtain similar GO enrichments by choosing genes that are specifically active in specific cell types. Furthermore, despite the fact that GM12878 and K562 are both blood cells, they seem to share roughly the same number of genes and with HeLA, though I can not tell entirely by eyeballing the Venn diagram. It would be nice to thoroughly characterize the degree of overlap between GM12878 and K562 vs other cell types. Presumably they should share more HIMs, being both blood cells.

**Response:**

1. GO is downgraded to the supplementary file. GO cannot be obtained after excluding cell type specific genes. Thus cell type specific GO terms are largely due to genes specifically expressed in the cell type.
2. I will compare % of genes assigned to both GM12878 and K562 with the % in other possible pairs of cell types.  
   Added: Pairwise comparison shows that the genes assigned to HIMs have the highest degree of overlap between GM12878 and K562 (both are blood cell lines) compared to the other cell types (Fig. S\ref{X}), suggesting that functionally similar cell types share similar HIMs.

The most compelling arguments for the existence of HIMs is presented on page 7. Intuitively, we expect some HIMs to be shared across cell types, and presumably HIMs that contain housekeeping and similar essential genes that should be expressed in most cell types should have similar regulatory elements. Authors attack this problem but they identify conserved HIMs from their own HIM sets and again cover anecdotes. A more convincing approach would be identify HIMs that contain housekeeping genes from orthogonal datasets and annotations; following that up with the characterization of such HIMs across cell types. Do they share TFs? Do they form the spatial clusters in each cell type? If HIM is indeed a nuclear genome organization unit, we would expect to see this. Years of chromatin organization studies have shown that compartment and topologically associating domains are largely conserved across cell types. HIMs likely should exhibit similar characteristics. If not, the contrary claim should be backed up.

**Response:** Iwill add the dynamic analysis centering on housekeeping/essential genes accordingly.

In the last paragraph of the section titled 'HIMs have strong preference in spatial locations in the nucleus', authors characterize HIMs at both gene and TF level, which is the most interesting analysis. Based on the heterogeneity, authors claim it is possible that the Hi-C interaction networks and GRNs are highly cell type-specific. This seems unlikely. I think a deeper elaboration of this analysis is necessary. Also, the text reads: "We found that the genes in the same HIM undergo moderate change from one cell type to another". It is unclear to me how there is a 'same HIM' in two cell types. HIMs are identified separately in each cell type. Are the most similar HIMs from two different cell types paired and this is what is referred to as "same HIM" above?

**Response:**

1. The sentence is revised by adding more details.   
   “the Hi-C interaction networks and GRNs are highly cell type-specific as 66% Hi-C interactions and 31.4% GRN interactions only exist in one cell type.”
2. “Genes in the same HIM” is clear to me. Google scholar shows that “nodes in the same community” + network appears ~1300 times, “nodes in the same cluster” + network appears ~3100 times. Do we need to revise this phase?

The authors tackle a very interesting potential concept of a HIM and offer a solid computational method to identify such units. There is a reasonable characterization of HIMs and their properties but the HIM as an organization unit is not validated. The study suffers from anecdotal discussions. I think the authors should focus on backing their HIM hypothesis by validating their central claims. For example, a motif or ChIP-seq enrichment for each HIM would ensure each HIM is indeed regulated by the identified TF nodes. Extending the analysis of cross cell type HIM analysis would be advisable. Do two HIMs from two cell types that share a significant number of genes? Such analyses will go farther to convince the reader.

**Response:**

1. To directly address this question, we will add ChIP-seq enrichment when applicable. Enhancers has multiple TFs.
2. It is challenging to test that the proportion of shared genes between two HIMs from two cell types is significant. I need more time to pick up a proper statistical test.

At this moment, this question can be indirectly addressed by the observation that 62.2%-64% of HIMs in each cell type share **at least half** of their genes with HIMs in another cell type. These HIMs are called conserved HIMs in our paper.

MINOR  
  
Page 2. Referring to motif M as 'motif' is slightly confusing, since TFs also bind conserved DNA sequences and the collection of such sequences are also called 'motif'.  I suggest calling it a graph or network motif.

**Response**: Done.  
  
Page 2-3. "for each pair of gene loci within 10Mb, we use the “observed over expected” (O/E) quantity (we use O/E>1 as the cutoff in this work but we found that our main results are largely consistent with different cutoffs; see Supplementary Results)" I believe this is the O/E of Hi-C contact matrix. Maybe clarify this as Hi-C has been barely introduced in the text at this point.

**Response**: (O/E) quantity 🡪 (O/E) quantity in the Hi-C data.  
  
Page 5. “HIMs tend to be near the interior of the nucleus where transcriptions are more active” - > “HIMs tend to be near the interior of the nucleus where there is more transcription”

**Response**: Done.

----------- Overall evaluation -----------  
The authors propose MOCHI, a graph clustering algorithm that aims to identify groups of loci and TFs where the loci are strongly interacting, and are regulated by the chosen TFs. The idea of jointly analyzing the chromosome conformation data and gene regulatory network is extremely important and this represents a step forward in that direction. The clustering approach is based on the definition of a 4-node subgraph made of 2 TFs and two loci. The algorithm is a recursive bi-partitioning of the graph that aims to minimize the cuts of those motifs.  
  
While I like the general idea of the paper, I find it suffers from several shortcomings:  
  
1) The algorithm does not provide any kind of statistical significance of the clusters (HIMs) found.

**Response:** I think the critic is unfair to spectral clustering methods including MOCHI because spectral clustering methods are not designed to quantify the statistical significance of individual clusters, directly or indirectly. We theoretically proved that the identified HIMs are near optimal. Do you think shall we address this comment?

Give a try.

2) The author should show that similar results could not be obtained by simply clustering the HiC network (e.g. not considering the regulatory network), or by simply using TADs as clusters.

**Response:** We will add this part as method comparison. Here are my ideas on how to do it.

1. Apply recursive spectral clustering (same as MOCHI) to cluster Hi-C network on each chromosome. The tuning parameter is selected such that the resulted clusters and HIMs should have similar distribution of gene number.
2. Compare identified clusters to HIMs regarding:
   1. There is no TF co-regulations; show case in one cell type
   2. Hi-C density (spatial proximity strength). Hi-C > Hi-C+GRN. Kindly expected.
   3. GRN density (regulation strength). We will control the number of TFs in each cluster to make a fair comparison.
   4. Motif density (co-regulation strength). We will control the number of TFs in each cluster to make a fair comparison.
   5. Hi-C+GRN > Hi-C in A percent, earlier replicated, gene distance.
   6. Hi-C+GRN~Hi-C in tads number, loop number

3) Much of the results study the set of genes/loci that get selected in HIMs, versus those that do not. I don’t find this type of global analysis particularly interesting, since HIMs are most likely to be found in active chromosome compartments (compartment A). The results about expression, replication timing stats, essentiality, and super-enhancers may all be due to the same enrichment for A compartments. Same thing for the cross-cell-type comparison.  
  
A useful negative control would preserve distance and HiC value. Much of the observed results may simply be due to the fact that HIMs presumably involve highly interacting regions.

**Response**: How about we discuss that this is challenging to create appropriate negative controls in the Discussion section to avoid similar comments?   
  
  
4) The analysis of PPIs within individual HIMs is more interesting, but should still be improved/clarified.  
  
"The density of the sub-PPI network is 0.389 and is 2.46 times higher than the average density (0.158) of the random cases, which is also the density of the whole PPI network.” —> The authors should only compare to the PPI network restricted to TFs expressed in K561, not the whole PPI network. The restricted PPI is presumably denser

**Response**: Highlighted that the whole PPI network is between TF proteins.

We also add the following sentence in the Collection and processing of data used in this study section: “Note that the GRNs have the same set of TF protein. Thus the whole PPI network is suitable for all the 5 different cell types.”  
  
  
5) The description of the algorithm is hard to follow.  
For example, Step 4 of the algorithm (Finding overlapping HIM) is not clear. It is stated as aiming to allow overlap between HIMs, but its description only talks about removing certain TFs. This section should be better explained.

**Response**: We add one sentence to explain how to get the ‘certain TFs’.

“Next, we pool together the TFs from the HIM and from its ancestor HIMs.”

Typos:  
  
"that (1) the gene loci have”

**Response:** Anyone clear where is the typo?  
  
"while the S that minimizes the function as the optimal solution"

**Response**: Done  
  
"Moreover, the conductance of the graph G, 'G, is defined as argminS'G(S)" => min, not argmin

**Response**: Done  
"ELF1" "ETV6" "IKZF2" "IRF5" "RUNX3" "if they are not spatially proximal to each other than expected"

**Response:** Anyone clear where is the typo?

**Additional**

**CTCF in the boundary between HIMs, and its relationship with super-enhancers**

**Super-enhancers in the discussion section at least**

**GM12878 related TFs**

**Genecard Clear evidence: ELF1, ETV6, IKZF1, IKZF2, NFATC1,**

**Not sure: IRF3, IRF5, RUNX3, STAT1, STAT5A, TCF12**

**No: ATF2, NR2F1, PAX8, PKNOX1,**

**Master TF in GM: "ELF1" "ETV6" "IKZF2" "IRF5" "RUNX3"**

**K562: "BHLHE40" "MNT" "STAT5A" "TAL1"**