

HaploC-tools reveal haplotype-specific chromosome conformation and chromatin states

Corresponding Author: Professor Giovanni Ciriello

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Liu et al. have created a computational pipeline that integrates SNP calling and phasing from Hi-C data before generating pairwise interaction maps. Then, it uses these maps to identify structural variants and haplotype-specific differences in 3D chromatin architecture. They then apply this pipeline to Hi-C datasets from 64 cancer cell lines and 47 control cell lines, highlighting differences in haplotype-specific Hi-C maps among some samples.

Their software package, HaploC-tools, provides a more tailored and user friendly wrapper of other software tools such as FreeBayes and HapCUT2. While it would be possible to perform these same analyses with custom pipelines and in-house code, HaploC-tools makes these types of analyses more accessible.

Their biological observations about heterochromatin decompaction and B-to-A compartment switching in cancer cell lines are interesting, though not entirely novel or surprising, but they do demonstrate the utility of applying HaploC-tools to human samples.

One major comment is that it would be helpful for an end user to know what Hi-C coverage is required from human samples to detect haplotype-specific differences in 3D chromatin organization at various levels of resolution. What fraction of reads contain informative heterozygous sites and can be assigned reliably to one haplotype or the other by HaploC-tools? If performing a new Hi-C experiment or searching for published datasets, what minimum coverage should be targeted? Downsampling existing datasets can help to answer this.

Specific comments:

-Line 14-15: haplotype-specific Hi-C maps have been explored in highly heterozygous mouse crosses, and even in humans in Rao et al. 2014. Longzhi Tan's work also looks at this in humans with single-cell Hi-C. The authors should cite and refer to these works in the text.

-In the results section, it would help to explicitly highlight the technical advances implemented in HaploC-tools beyond just wrapping other software tools.

-Why FreeBayes? Why not GATK?

-For the initial assessment of accuracy, it would be nice to see how the additive maps are concordant with un-phased maps, given that they should be showcasing the same underlying information.

-Most Hi-C maps in the figures (e.g. Fig 2) are missing scale bars, and need coordinates and labeled ticks. It is currently difficult to interpret them.

-The authors should state more clearly in the main text how CTCF binding affinities were predicted. It is confusing as written.

-Figure 3 would benefit from alignment of panels d and e, to visualize how the ChIP-seq track aligns with the Hi-C map.

-4e-j It would be helpful to show the 1D ChIP-seq tracks aligned to the Hi-C maps

-Identifying aberrant 3D organization relative to "consensus" 3D architecture seems problematic, as healthy cell types arising from different tissues can have highly variable 3D structures. As a negative control, healthy cells should be held out from generating the consensus, to see how many "aberrant" structures are identified in these healthy cells.

-The rationale for the SET1DB KO is unclear. The authors saw that cancers seem to deviate from the norm more in B compartments relative to A compartments. This affects both haplotypes, but the claim is that this is happening in cancer cells because there is differential H3K9me3 turnover dependent on haplotype?

-Is H3K9me3 really reduced in both haplotypes? 5E seems shows a mode with Fold-Change above 0. Showing haplotypes

plotted against each other here is confusing. It would help to show the effects of the KO on H3K9me3 at large before doing hap vs hap comparisons.

(Remarks on code availability)

We reviewed the Github and README but did not attempt to install or run the code. It is helpful that the authors provide a link to demo data.

Reviewer #2

(Remarks to the Author)

COMMENTS FOR THE AUTHORS

The manuscript by Liu et al presents a novel suite of bioinformatic tools (HaploC-tools) to phase and analyze haplotype-specific Hi-C maps. The tool is very carefully designed to dissect several levels of information on chromatin 3D organization, and to facilitate comparison with other types of high-throughput sequencing genomics data. Overall the tool seems very powerful, allowing to achieve novel insights into allele-specific regulation of genome functionality. Indeed also the sample results shown on cancer cell lines relating heterochromatin perturbations to allele-specific events are very intriguing, although they would deserve a dedicated study by themselves. I have just a few comments that probably can be easily addressed by the authors. But I think they would improve the clarity and fairness of the conclusions.

SPECIFIC COMMENTS

- In the main text, the authors state that HaploC implements "a custom-optimized version of HapCUT2" (line 67). My understanding is that the specific optimizations are those described in the methods section (lines 320-338). I think it may be useful to at least briefly summarize in the main text about the nature of such optimizations (i.e., assessing consistency between inferred haplotypes across multiple bin sizes, then assembling blocks by minimizing H-trans). In this "optimized" procedure, there's also a step where chromosomes with low consistency are flagged as "failed" and removed from subsequent analyses. I wonder if this may thus yield more favourable results for HaploC in its comparison against HapCUT2. This may deserve at least an explicit comment in the text for fairness.

- At line 79, I read "average consistency of 52%, as expected by chance". I am not sure if this means that at the population level, a 52% concordance of haplotypes is expected, or if this is based on some randomization or other statistical considerations. Maybe a revised phrasing could clarify and avoid ambiguity in this sentence.

- At line 82: "HaploC successfully phased 94% of SNPs". This "successfully" refers to the fact that the assignment of SNPs to a specific haplotype satisfied the thresholds used in the HaploC algorithm itself. My understanding from other points in the manuscript is that the authors also examined some biological samples for which there is a "ground truth" reference haplotype assembly. If this is the case, it may be fairer to also report the percentage of "success" against the ground truth. If instead all available Hi-C datasets are done on samples where the haplotype is "inferred" by other algorithms, an additional comparison may not be conclusive and a more unambiguous phrasing should help clarity and fairness.

Also related to this concept, the data on U266 and WSU-DLCL2 cell lines very evidently and certainly support the presence of the structural variants commented on in the text. However, I was wondering if any other independent study has confirmed (possibly with other experimental techniques) if these cell lines indeed harbor such structural variants. I.e. I was wondering if there is a consensus "ground truth" on such reported translocations.

- Some earlier publications, e.g. I recall in particular a work by Bing Ren's group (Gorkin et al Genome Biol. 2019 - <https://doi.org/10.1186/s13059-019-1855-4>), did some attempts at linking specific population-level sequence variants with chromatin 3D architecture variants. If Liu et al. found any overlap of their allele-specific variants with such population-level variants linked to 3D chromatin architecture it may be an additional independent validation that gives higher strength to the method.

- About the results reported in Figure 5h, maybe I am not correctly reading the color map, but my understanding is that this example locus is not showing a switch from B2.2. to A1.1. Other panels from Figure 5 and their related comments in the text seemed to suggest that a loss of H3K9me3 could push such types of loci directly to A1.1. Did I understand this wrong? Maybe a more explicit phrasing and discussion could avoid a misinterpretation of this potentially relevant biological point.

- For what concerns HaploCI and HaploCalder, the authors proposed a clever approach to estimate the random expectation of delta-insulation or delta-rank in compartment scores, based on a random assignment of Hi-C reads to two haplotypes, then computing pairwise differences in insulation and compartment between the randomized haplotype Hi-C maps. This is basically building an empirical "null" distribution for the delta-insulation or delta-rank in compartment scores. Then the authors propose to use the 80th percentile to rescale such scores. Finally, although only for the HaploCalder tool, a specific threshold is applied on the rescaled delta-rank to select "significant" differences. What I don't understand completely is why the authors didn't use directly a more stringent threshold on the null distribution (e.g. 95th percentile) to select significant differences. I guess that would also give a direct estimation of the false positive error rate from a statistical point of view. Also related to this, is there any theoretical or empirical way to actually estimate the false discovery rate achieved with the

additional thresholds currently applied and described in the methods section? Anyway, the control of false positives should be good, considering the comparison with dcHIC where the results were filtered with adjusted p-value < 0.0001.

- To be fair, it should be mentioned that dcHIC is a tool that can also handle replicates in the analysis of differential compartmentalization. Thus, the analysis of compartment differences in 1 vs 1 haplotype phased matrices may not be the best scenario for dcHIC to perform well.

- some typos or formatting problems:

--> ref 36 the journal name should be "Bioinformatics"

--> line 663 I think there's an inversion in the order of sentences as it is referring to "the above 8 cell lines", whereas the cell lines are listed in the following sentence.

--> line 665 there's a reference to "[sup table]" that is probably a placeholder not updated to the final supplementary items numbering

(Remarks on code availability)

I reviewed the github repository content but I didn't try to actually execute the code.

The repository comes with detailed step by step instructions about how to install the required packages including dependencies.

Overall the repository and documentation are well curated.

However, some specific parameters or analysis steps could benefit from additional details in the documentation. For example, what are all the options available for the "enzyme" parameter in the working directory configuration file. Also, what should be indicated in case of mixed (dual enzymes) digestion, e.g. with Arima genomics kind of kits. To this concern, the "HaploC-tools/blob/main/HaploC/main/config_tmplate.txt" file seems to have slightly different options compared to what described in "<https://github.com/CSOgroup/HaploC-tools/wiki/Prepare-working-directory>". e.g. "n_bwa_thread" vs "thread4bwa", not clear which is the right one.

I could not find a point where some parameters usually relevant for Hi-C should be defined. e.g. I could not find a point where the Hi-C matrix "bin size" should be defined.

Based on the main manuscript I understand that HaploCI should return a set of regions with significant changes in chromatin insulation score, where significance is tested against randomized haplotypes. However, in the description of HaploCI output (<https://github.com/CSOgroup/HaploC-tools/wiki/HaploCI>), the "significance" of results is not really mentioned, thus it is not clear if these results should be further filtered.

There's a similar ambiguity in the wiki page describing the output of HaploCalder (<https://github.com/CSOgroup/HaploC-tools/wiki/HaploCalder>). However, in this case there's a mention that "significant" changes are reported. But it's not clear how they were filtered (what statistical thresholds) and it is also not clear how the output of files *diff_rank* should be related to the statistical significance.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have adequately addressed all of my concerns.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

COMMENT FOR THE AUTHORS

The authors did a great job in improving clarity and several details of the manuscript, which was already very good and thorough in the initial version.

They addressed in detail all major and minor concerns raised by the reviewers.

If possible, I would only suggest to incorporate in the methods section the explanation of hicCalder thresholding method, currently included in the reply to reviewers only:

"Regarding the two-step thresholding adopted in HaploCalder, [...] we should note that the first thresholding (80th percentile of the null distribution) is applied to the delta rank computed at the bin level. After such correction, the residual signal is segmented across the genome, average delta ranks are computed for each segment, and a second threshold is applied to select those segments (not bins) that exhibit a delta rank above the second threshold. A more stringent threshold at the first step (e.g. 95th percentile of the null) would further reduce the variance of the delta rank distribution and potentially negatively affect the segmentation procedure. The two-step approach that we chose guarantees to retain a relatively large number of segments with non-zero delta ranks, allowing the user to select the significant ones by modulating the second threshold."

(Remarks on code availability)

The reviewer did a thorough job also in addressing the previously noted details in the code. Now it is improved and clear.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Liu et al. have created a computational pipeline that integrates SNP calling and phasing from Hi-C data before generating pairwise interaction maps. Then, it uses these maps to identify structural variants and haplotype-specific differences in 3D chromatin architecture. They then apply this pipeline to Hi-C datasets from 64 cancer cell lines and 47 control cell lines, highlighting differences in haplotype-specific Hi-C maps among some samples.

Their software package, HaploC-tools, provides a more tailored and user-friendly wrapper of other software tools such as FreeBayes and HapCUT2. While it would be possible to perform these same analyses with custom pipelines and in-house code, HaploC-tools makes these types of analyses more accessible.

We thank the reviewer for their positive feedback, acknowledging the relevance of our work, and for providing useful comments to improve it. We believe we have now addressed all their concerns as described in more detail below.

Their biological observations about heterochromatin decompaction and B-to-A compartment switching in cancer cell lines are interesting, though not entirely novel or surprising, but they do demonstrate the utility of applying HaploC-tools to human samples.

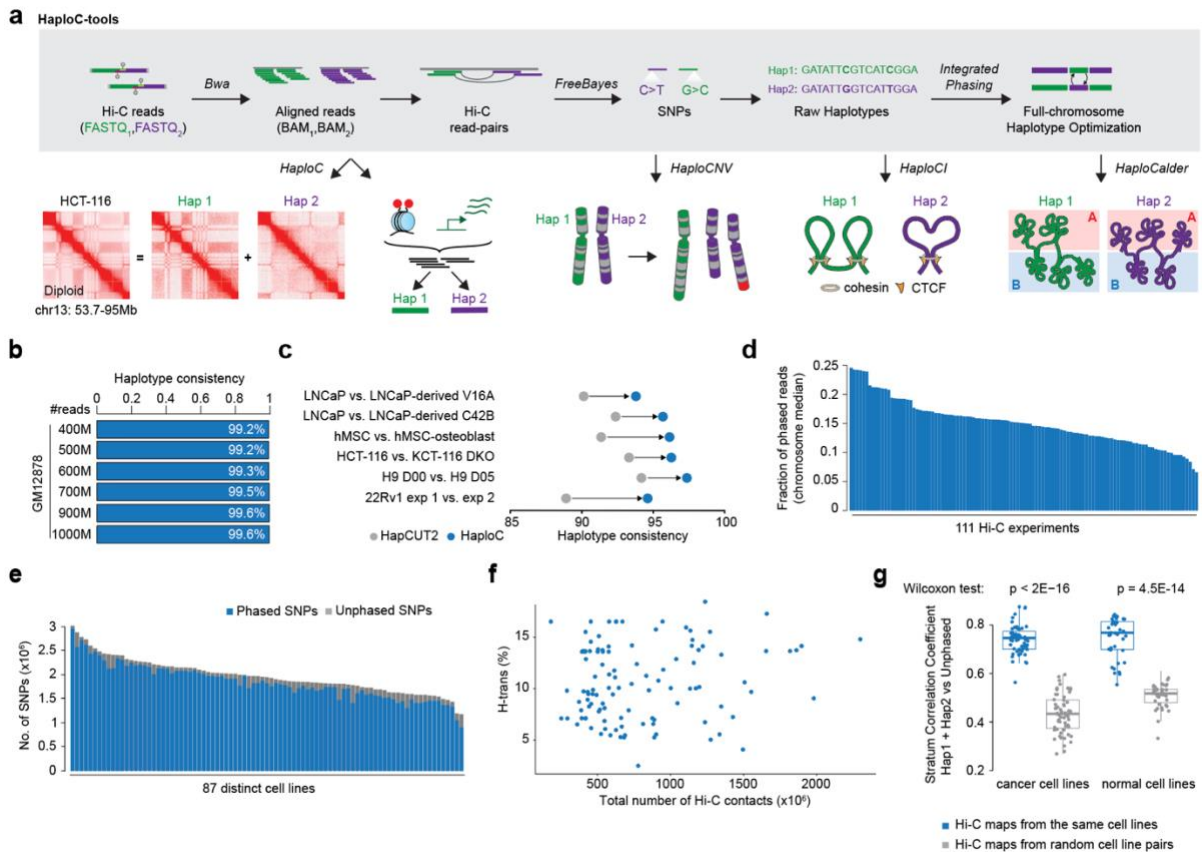
One major comment is that it would be helpful for an end user to know what Hi-C coverage is required from human samples to detect haplotype-specific differences in 3D chromatin organization at various levels of resolution. What fraction of reads contain informative heterozygous sites and can be assigned reliably to one haplotype or the other by HaploC-tools? If performing a new Hi-C experiment or searching for published datasets, what minimum coverage should be targeted? Downsampling existing datasets can help to answer this.

This is indeed an important point and something that we hadn't fully explored before. We have now expanded our analyses to assess to which data resolution affects our results.

First, we computed the fraction of phased contacts (maternal + paternal Hi-C matrices) over the total number of contacts across all the Hi-C experiments that we analyzed and observed that, on average, 15% of the contacts are successfully phased (see new **Figure 1d**). Even though the phasing procedure itself is efficient even in cases where the number of contacts is as low as 400 million (as demonstrated by the GM12878 downsampling analysis in **Figure 1b** and the percentage trans-haplotype contacts in **Figure 1f**), we

generally don't recommend going below this number when planning a new Hi-C experiment to be phased, especially for downstream analyses, as further discussed below. We copy here the updated **Figure 1** for the reviewer's perusal (also comprising a new panel 'g' in response to another comment from the reviewer – see below).

Figure 1



Next, we assessed the robustness of HaploC downstream analyses (HaploCI, HaploCalder, HaploCNV) to data resolution. To do so, we downsampled the Hi-C dataset for the HCT116 cell line to compare results obtained at its original resolution (~1500M contacts), at intermediate resolution (~1000M resolution), and at low resolution (~500M contacts).

For HaploCI, we evaluated the consistency of the differential insulation scores for bin sizes of 50kb or 25kb by correlating values obtained at original resolution with those obtained at intermediate and low resolutions. Average correlation values across chromosomes were consistently high, although for low resolution maps binned at 25kb the correlation decreased from 0.8 to 0.56 (**Table R1**).

For HaploCalder, we computed the fraction of haplotype-specific compartments (HSC) identified with the original resolution that could be recovered with intermediate or low resolutions. HSCs were retained if their difference compartment rank (Δ rank) was greater than 0.15 (as in the manuscript) or 0.25, to assess concordance of “stronger” HSCs (0.25 indicatively corresponds to a difference of 2 out of 8 sub-compartments). Using a 0.15 minimum Δ rank HaploCalder recovered 70.5% of HSC at intermediate resolution, and 57% at low resolution, while for $\min(\Delta\text{rank}) > 0.25$, more than 70% of HSC were recovered at both resolutions (**Table R1**).

Lastly, we applied HaploCNV to original, intermediate, and low-resolution Hi-C maps and assessed concordancy of both total and haplotype-specific copy number status. Strikingly, the concordance was nearly perfect for all comparisons (**Table R1**).

We now discuss these results and guidelines in the text, and have added **Table R1** as a new **Supplementary Table 4**.

Comparisons between results obtained at original resolution (1500M contacts) and:	Intermediate resolution (1000M contacts)	Low resolution (500M contacts)
Correlation of differential contact insulation (HaploCI) 50kb bin-size	0.84	0.8
Correlation of differential contact insulation (HaploCI) 25kb bin-size	0.73	0.56
Percentage of concordant HSC (min size = 100 kb, min Δ rank = 0.15)	71%	57%
Percentage of concordant HSC (min size = 100 kb, min Δ rank = 0.25)	77%	71%
Percentage of concordant total copy number status	99.9%	99.9%
Percentage of concordant haplotype-specific copy number status	99.9%	98.7-99.5%

Table R1: Comparison of results obtained by HaploCI, HaploCalder, and HaploCNV using HCT116 Hi-C maps at original resolution (1500M contacts), intermediate resolution (1000M contacts), and low resolution (500M contacts)

Specific comments:2

-Line 14-15: haplotype-specific Hi-C maps have been explored in highly heterozygous mouse crosses, and even in humans in Rao et al. 2014. Longzhi Tan's work also looks at this in humans with single-cell Hi-C. The authors should cite and refer to these works in the text.

We thank the reviewer for bringing those studies to our attention. We have now added the indicated references and a few more to highlight more recent studies focusing on haplotype-specific chromatin conformation.

-In the results section, it would help to explicitly highlight the technical advances implemented in HaploC-tools beyond just wrapping other software tools.

We thank the reviewer for suggesting to stress the advances introduced by HaploC. We would like to highlight here that beyond wrapping and optimizing existing tools to perform whole-chromosome phasing of Hi-C maps, HaploC-tools effectively introduced 3 new methods that leverage haplotype-resolved Hi-C maps to infer copy number variants (HaploCNV) determine differential contact insulation between haplotypes (HaploCI) and differential compartment organizations between haplotypes (HaploCalder).

-Why FreeBayes? Why not GATK?

FreeBayes is known for its portability across systems and speed. We agree that GATK is more suitable in a clinical setting, when one would expect challenging variants with low frequency (sub-clonal events). However, variant calling in HaploC-tools is aimed towards germline variants, which should, in general, be easily determined by all variant callers. FreeBayes performs this task efficiently while being a self-contained binary that does not require Java, thus smoothing the installation of HaploC-tools for the users. Overall, FreeBayes is an efficient and functional option to improve portability.

- For the initial assessment of accuracy, it would be nice to see how the additive maps are concordant with unphased maps, given that they should be showcasing the same underlying information.

We agree with the reviewer that the sum of haplotype-resolved maps should be highly concordant with the original unphased map, even though comprising only a fraction of the initial set of reads (recall that haplotype-resolved maps only contain phased reads, which are on average 15% of the total).

To test this hypothesis, we assess the concordance between the sum of phased contact maps and the original unphased Hi-C map using the Stratum Correlation Coefficient (SCC), which measures the correlation between two matrices. Despite the great difference in data resolution, SCC values computed for unphased and phased maps from the same cell line were significantly higher than values obtained for random pairs of phased and unphased maps, for both normal and cancer cell lines, reaching a median correlation value of ~0.76 (**Figure R1**). These comparisons indicate that phased maps retain chromatin conformation features of the original unphased Hi-C maps. The same data is plotted in a similar format in the new **Figure 1g**.

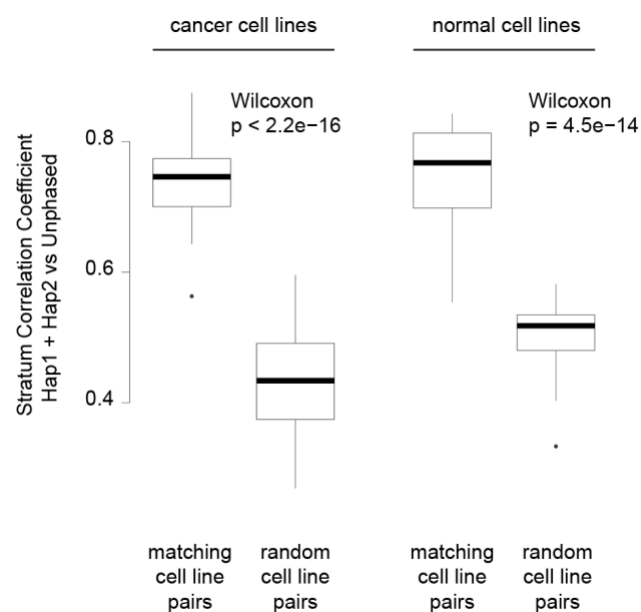


Figure R1: Stratum Correlation Coefficients between contact maps obtained by summing haplotype-specific Hi-C maps and the original unphased contact maps

-Most Hi-C maps in the figures (e.g. Fig 2) are missing scale bars and need coordinates and labeled ticks. It is currently difficult to interpret them.

The reviewer is correct, several information was indeed missing. We have now edited Figure 2 to report the adopted bin size and color scale legend. The coordinates are not reported because full chromosome maps are shown, as it is now explicitly indicated.

-The authors should state more clearly in the main text how CTCF binding affinities were predicted. It is confusing as written.

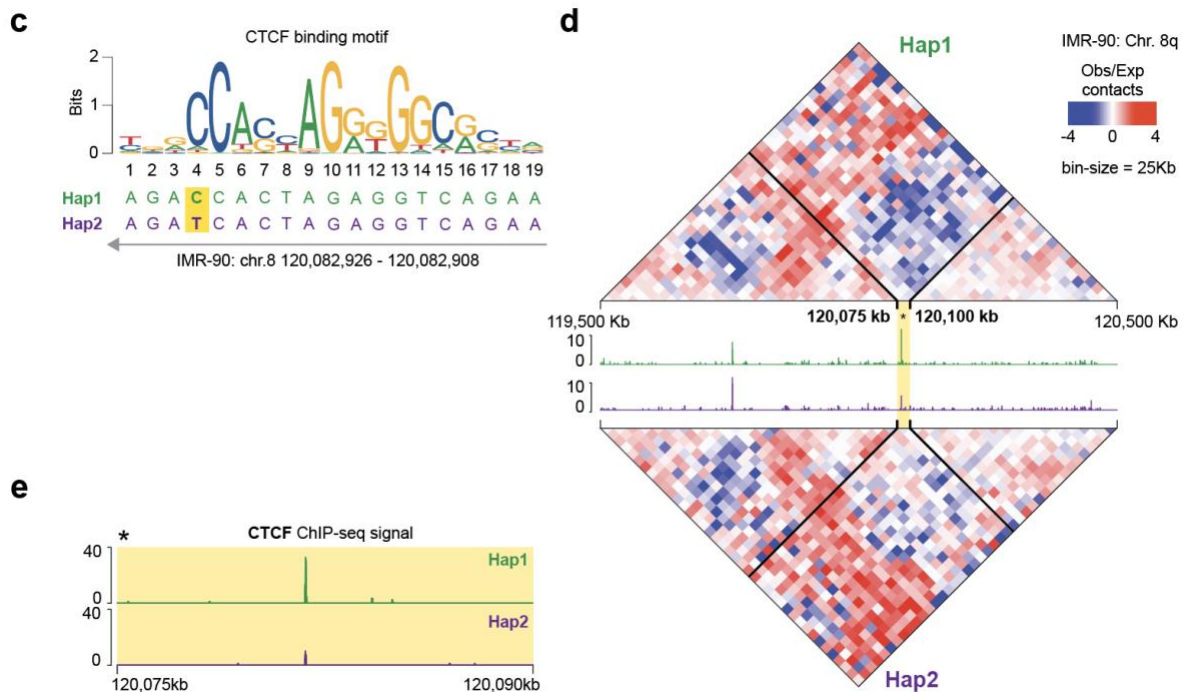
We edited the method section of our study by adding the following to the methods section:

“CTCF binding affinity estimation

The binding affinity score for each SNP-containing DNA sequence (19 bp in length) was calculated using the TFBSTools R package⁴⁷, based on the position weight matrix (PWM) for the CTCF transcription factor obtained from the JASPAR database. For each haplotype, sequences corresponding to both the reference and alternative SNP alleles were scanned with the PWM, and the resulting match scores—normalized to the maximum possible score for the matrix—were used as the CTCF binding affinity scores.”

- Figure 3 would benefit from alignment of panels d and e, to visualize how the ChIP-seq track aligns with the Hi-C map.

We have now updated Figure 3 to show CTCF ChIP-seq tracks for both maternal and paternal haplotypes aligned with the Hi-C map, highlighting and zooming in on the bin where differential CTCF binding was observed. The updated figure panels are also shown below for the reviewer.



-4e-j It would be helpful to show the 1D ChIP-seq tracks aligned to the Hi-C maps

We agree with the reviewer that haplotype-specific ChIP-seq tracks will further highlight the correspondence between changes of compartments between haplotypes and concordant changes between histone modifications. As these analyses require matching datasets for the selected cell models, we updated Figure 4 to show 4 examples of haplotype-specific compartments, 2 for HCT-116 and 2 for RPE cell lines, each example accompanied by haplotype-specific tracks for H3K27ac and H3K9me3. These examples confirm concordant changes between chromatin compartment and histone modifications between haplotypes.

-Identifying aberrant 3D organization relative to “consensus” 3D architecture seems problematic, as healthy cell types arising from different tissues can have highly variable 3D structures. As a negative control, healthy cells should be held out from generating the consensus, to see how many “aberrant” structures are identified in these healthy cells.

We agree with the reviewer that calculating a consensus compartment profile inevitably will average heterogeneous compartment calls across different models, possibly resulting in poorly representative compartment profiles. Moreover, this analysis is superfluous since the enrichment for haplotype-specific compartments in B subcompartments is clearly evident when examining compartments in matching tissues. For this reason, we decided to remove the figure panel related to the consensus compartments, and show only the enrichments resulting when comparing cancer cell lines with matched normal tissues.

-The rationale for the SET1DB KO is unclear. The authors saw that cancers seem to deviate from the norm more in B compartments relative to A compartments. This affects both haplotypes, but the claim is that this is happening in cancer cells because there is differential H3K9me3 turnover dependent on haplotype?

We apologize for the confusion and will try here to clarify the rationale behind this analysis. As the reviewer pointed out, we observed that compartment differences between haplotypes more frequently occur in regions that, in matching tissues, are found in the B subcompartments. This was shown specifically for 4 different cancer cell lines compared with matching normal tissues. This observation prompted the hypothesis that haplotype-specific compartments emerge from haplotype-specific loss of H3K9me3.

To test this hypothesis, we resorted to Hi-C experiment in syngeneic models where *SETDB1*, the enzyme that is responsible for depositing H3K9me3, was knock-out (KO). Loss of *SETDB1* indeed results in loss of H3K9me3. In many cases, this loss occurs in

both haplotypes, but this is not always true. Possibly because of incomplete KO, or less efficient compensatory enzymes, we found that upon *SETDB1* KO, many loci lose H3K9me3 in only one homologous chromosome but not in the other, here also referred as “asymmetric loss of H3K9me3”. Upon comparing haplotype-specific Hi-C maps *SETDB1* KO cells, we found that asymmetric losses of H3K9me3 result in compartment differences consistent with our hypothesis that monoallelic loss of H3K9me3 is sufficient to drive haplotype differences in compartment organization. We hope this clarifies the rationale of our analysis.

- Is H3K9me3 really reduced in both haplotypes? 5E seems to show a mode with Fold-Change above 0. Showing haplotypes plotted against each other here is confusing. It would help to show the effects of the KO on H3K9me3 at large before doing hap vs hap comparisons.

The reviewer correctly spotted a modest shift above zero, suggestive of a counterintuitive gain of H3K9me3 upon *SETDB1* KO in A549 cells. To confirm that this effect wasn't due to our phasing of the ChIP-seq signal, we went back to the original study and retrieved the original data from the corresponding GEO repository (GSE168233). By comparing the RPKM normalized tracks before and after *SETDB1* KO (without phasing), we confirmed broad loss of H3K9me3 across multiple genomic loci, but also a modest positive gain in another set of loci, as shown by the resulting bimodal distribution of log2 H3K9me3 fold-changes (**Fig. R2**).

However, we believe this is due to the choice of normalization strategy, rather than a true gain of H3K9me3. Indeed, lacking an exogenous spike-in, the authors resorted to RPKM normalization, which assumes the same total number of reads in both control and *SETDB1* KO cells. Given numerous and genome-wide losses of H3K9me3 following *SETDB1* KO, RPKM normalization likely inflated the read counts at those regions that did not lose the histone mark, causing this modest positive shift. Despite this potential artifact, all our conclusions remain valid as they exclusively concern regions exhibiting H3K9me3 loss (the magnitude of which may even be stronger if a normalization based on exogenous spike-in were used).

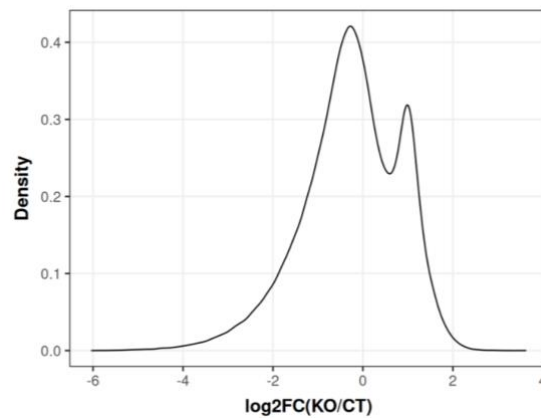


Figure R2: H3K9me3 fold-change distribution (log2) between SETDB1 KO and control A549 cells.

Reviewer #1 (Remarks on code availability):

We reviewed the Github and README but did not attempt to install or run the code. It is helpful that the authors provide a link to demo data.

We thank the reviewers for the positive feedback about our code repository. In response to specific requests from the Reviewer #2, we further updated the GitHub page providing additional information in the corresponding wiki.

Reviewer #2 (Remarks to the Author):

The manuscript by Liu et al presents a novel suite of bioinformatic tools (HaploC-tools) to phase and analyze haplotype-specific Hi-C maps. The tool is very carefully designed to dissect several levels of information on chromatin 3D organization, and to facilitate comparison with other types of high-throughput sequencing genomics data. Overall the tool seems very powerful, allowing to achieve novel insights into allele-specific regulation of genome functionality. Indeed also the sample results shown on cancer cell lines relating heterochromatin perturbations to allele-specific events are very intriguing, although they would deserve a dedicated study by themselves. I have just a few comments that probably can be easily addressed by the authors. But I think they would improve the clarity and fairness of the conclusions.

We thank the reviewer for the positive feedback and for acknowledging both the utility of our suite of tools and the biological relevance of the results we showed. We also thank them for the useful comments to improve our study, which we have now fully addressed as discussed below.

SPECIFIC COMMENTS

- In the main text, the authors state that HaploC implements "a custom-optimized version of HapCUT2" (line 67). My understanding is that the specific optimizations are those described in the methods section (lines 320-338). I think it may be useful to at least briefly summarize in the main text about the nature of such optimizations (i.e., assessing consistency between inferred haplotypes across multiple bin sizes, then assembling blocks by minimizing H-trans).

The reviewer is correct; those are the main optimization steps that we introduced in the HapCUT2 pipeline. We now explicitly mention them in the results section.

In this "optimized" procedure, there's also a step where chromosomes with low consistency are flagged as "failed" and removed from subsequent analyses. I wonder if this may thus yield more favourable results for HaploC in its comparison against HapCUT2. This may deserve at least an explicit comment in the text for fairness.

This is correct, and comparing HapCUT2 and HaploC on a different set of chromosomes may in principle favor HaploC. However, we made sure that in the comparisons shown in Figure 1c, none of the chromosomes failed, hence those comparisons were made on the same set of chromosomes for both pipelines.

- At line 79, I read "average consistency of 52%, as expected by chance". I am not sure if this means that at the population level, a 52% concordance of haplotypes is expected, or if this is based on some randomization or other statistical considerations. Maybe a revised phrasing could clarify and avoid ambiguity in this sentence.

We apologize for the confusion. Here, we did not rely on any statistical test or empirical randomization; we simply meant to say that given a set of SNPs, two random haplotype assemblies will, on average, be consistent for 50% of the loci. However, as this may indeed generate confusion, we amended the sentence to report only the observed consistency values.

- At line 82: "HaploC successfully phased 94% of SNPs". This "successfully" refers to the fact that the assignment of SNPs to a specific haplotype satisfied the thresholds used in the HaploC algorithm itself. My understanding from other points in the manuscript is that the authors also examined some biological samples for which there is a "ground truth" reference haplotype assembly. If this is the case, it may be fairer to also report the percentage of "success" against the ground truth. If instead all available Hi-C datasets are done on samples where the haplotype is "inferred" by other algorithms, an additional comparison may not be conclusive and a more unambiguous phrasing should help clarity and fairness.

To clarify our description of the results and avoid confusion, we removed the adverb "successfully", explicitly mentioning when inferred haplotypes were compared to a ground truth estimated from orthogonal data, which was done specifically for the GM12878 cell line (see **Figure 1b**).

Also related to this concept, the data on U266 and WSU-DLCL2 cell lines very evidently and certainly support the presence of the structural variants commented on in the text. However, I was wondering if any other independent study has confirmed (possibly with other experimental techniques) if these cell lines indeed harbor such structural variants. I.e. I was wondering if there is a consensus "ground truth" on such reported translocations.

This is a good point and something we looked at in more detail. First, for what concerns the WSU-DLCL2 model, both translocations t(4:14) and t(14:18) are reported by the vendor (see <https://www.dsmz.de/collection/catalogue/details/culture/ACC-575>). Notably, as we previously showed in a separate study (<https://pubmed.ncbi.nlm.nih.gov/33972799/>), the breakpoint of the t(14:18) translocation occurs in the BCL2 3'-UTR and is associated with enhancer hijacking and upregulation

of BCL2, which is a hallmark of B-cell lymphoma (<https://pubmed.ncbi.nlm.nih.gov/3874430/>).

For what concerns the U266 cell line, we noticed that the chromosomal rearrangements observed in this dataset (deposited at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA666272>) are different from those observed in another publicly available Hi-C dataset (<https://www.nature.com/articles/s41467-017-01793-w>, Supplementary Information Figure 3a). Further comparisons of the two datasets revealed additional inconsistencies, suggesting a possible mislabeling of the cell line in one of the two studies. For this reason, even though the translocation that we initially reported is clearly evident from the Hi-C map, as the reviewer also acknowledged, we prefer to remove this example from the Figure 2, and we decided to show instead a map of Chr. 11 of the HeLa cell line where chromothripsis is observed in one haplotype. Evidence of chromothripsis in Chr. 11 was previously reported by deep RNA and DNA sequencing, providing orthogonal evidence to our results (<https://genomebiology.biomedcentral.com/articles/10.1186/gb-2013-14-4-111>).

Some earlier publications, e.g. I recall in particular a work by Bing Ren's group (Gorkin et al Genome Biol. 2019 - <https://doi.org/10.1186/s13059-019-1855-4>), did some attempts at linking specific population-level sequence variants with chromatin 3D architecture variants. If Liu et al. found any overlap of their allele-specific variants with such population-level variants linked to 3D chromatin architecture, it may be an additional independent validation that gives higher strength to the method.

We thank the reviewer for bringing this study to our attention. While Gorkin and colleagues explored a link between genotype and chromatin 3D conformation, they did so by comparing chromatin 3D data of the same cell type (lymphoblastoid cells) among 20 distinct individuals. In contrast, we compared chromatin 3D *phased* data between homologous chromosomes of the same individual (for many cell lines). This distinction was important for 2 main reasons: **1)** differences between individuals may be bigger because associated with homozygous genetic variants (and other sources of variability), and **2)** by comparing unphased Hi-C data, Gorkin and colleagues compared Hi-C maps at higher resolution than what can be done for phased data (recall that on average only 15% of the reads can be phased, even when almost all SNPs are covered). Nevertheless, we compared our results to theirs for both chromatin insulation and compartments.

First, Gorkin and colleagues classified “strong” vs “weak” CTCF binding sites (also referred to as “motif alleles”) and reported significant differences in loop strength between individuals with homozygous strong alleles (SS), heterozygous alleles (SW), or homozygous weak alleles (WW), with decreasing loop strength from SS to SW and WW

(see **Figure 4a-d** of the Gorkin manuscript). This analysis is similar to the one we did in **Figure 3** and, indeed, what they classified as “strong” motif allele corresponds to the canonical CTCF binding motif that we used in our differential insulation analysis (see our **Figure 3c**) and we were able to show that SNPs converting a strong/canonical motif into a weak motif were associated with significant differences in contact insulation between homologous chromosomes (**Figure 3b**), concordant with the observation across individuals reported in the Gorkin study. A detailed example of this phenomenon is shown in **Figure 3c-e** for a SNP in the IMR90 cell line. We added a reference to the Gorkin study to highlight the consistency between what they reported among individuals and what we show between haplotypes.

Next, to assess a link between genotype differences in the Gorkin dataset and haplotype-specific compartments, we examined in more detail 10 out of 20 samples from their cohort, which were also included in our phased dataset. Within these samples, we examined regions that were classified as “hypervariable” in the Gorkin study, based on the first principal component of the observed/expected Hi-C matrix, which the authors used to call A and B compartments. Across all samples, we consistently observed that regions exhibiting compartment differences between haplotypes (detected by HaploCalder) had significantly higher overlap with hypervariable regions (detected in the Gorkin study) than regions exhibiting no compartment differences between haplotypes (**Fig. R3**). This result suggests that heterozygous genotypes may contribute to different chromatin configurations between homologous chromosomes in these samples. While strengthening the consistency between the two studies, this result should be taken with caution as, unlike for CTCF binding motifs, we lack a clear rationale or mechanism to impute compartment differences to specific SNPs. Moreover, given that the Gorkin study focuses on chromatin differences among individuals rather than among homologous chromosomes of the same individual, the scopes of the two studies are not exactly the same. For these reasons, we would prefer not to include this result in our manuscript unless the reviewer strongly feels otherwise.

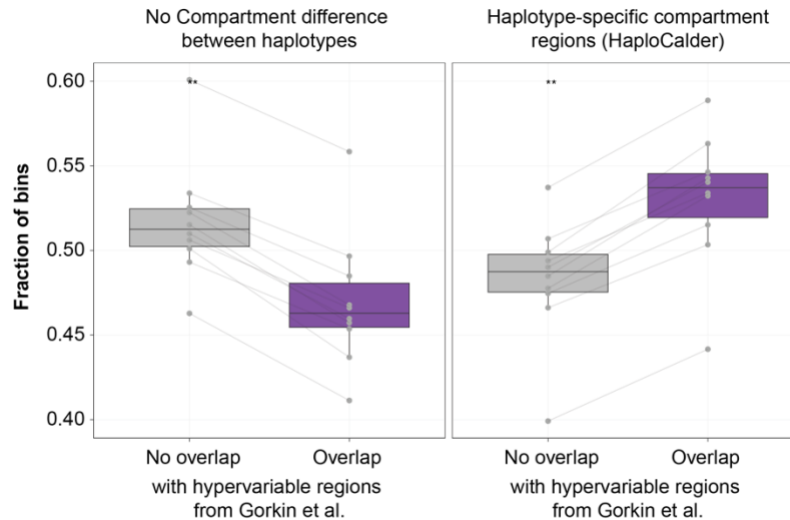


Fig. R3: For 10 LCL samples (points), fraction of bins (10kb) completely overlapping (purple boxplot) or not (gray boxplot) PC1 variable regions reported in Gorkin et al., when no compartment differences were observed between homologous chromosomes (left) and when instead significant differences were detected by HaploCalder (right). ** Wilcoxon p-value < 0.01

- About the results reported in Figure 5h, maybe I am not correctly reading the color map, but my understanding is that this example locus is not showing a switch from B2.2. to A1.1. Other panels from Figure 5 and their related comments in the text seemed to suggest that a loss of H3k9me3 could push such types of loci directly to A1.1. Did I understand this wrong? Maybe a more explicit phrasing and discussion could avoid a misinterpretation of this potentially relevant biological point.

The reviewer is correct, the compartment changes in **Figure 5h** are not necessarily between B.2.2 and A.1.1, which is the strongest possible compartment change. However, we never stated that loss of H3k9me3 always drives B.2.2 to A.1.1 compartment shifts, but rather that it drives a switch from B sub-compartments to A sub-compartments, or an “inactive-to-active” switch, which is to be intended as a change towards a “more active” compartment. We also added labels to the compartment annotation bars in **Figure 5h** to make the extent of the compartment switch clearer.

- For what concerns HaploCI and HaploCalder, the authors proposed a clever approach to estimate the random expectation of delta-insulation or delta-rank in compartment scores, based on a random assignment of Hi-C reads to two haplotypes, then computing pairwise differences in insulation and compartment between the randomized haplotype Hi-C maps. This is basically building an empirical “null” distribution for the delta-insulation or delta-rank in compartment scores. Then the authors propose to use the 80th percentile to rescale such scores. Finally, although only for the HaploCalder tool, a specific threshold

is applied on the rescaled delta-rank to select "significant" differences. What I don't understand completely is why the authors didn't use directly a more stringent threshold on the null distribution (e.g. 95th percentile) to select significant differences. I guess that would also give a direct estimation of the false positive error rate from a statistical point of view.

We thank the reviewer for the positive feedback. Regarding the two-step thresholding adopted in HaploCalder, the reviewer is in principle correct, one could increase the value of the first threshold and decrease or completely remove the second. However, we should note that the first thresholding (80th percentile of the null distribution) is applied to the delta rank computed at the bin level. After such correction, the residual signal is segmented across the genome, average delta ranks are computed for each segment, and a second threshold is applied to select those segments (not bins) that exhibit a delta rank above the second threshold. A more stringent threshold at the first step (e.g. 95th percentile of the null) would further reduce the variance of the delta rank distribution and potentially negatively affect the segmentation procedure. The two-step approach that we chose guarantees to retain a relatively large number of segments with non-zero delta ranks, allowing the user to select the significant ones by modulating the second threshold.

From a practical perspective, reported HS-C regions can also be further evaluated using orthogonal evidence to confirm their validity, for example, by integrating orthogonal data sources such as epigenetic differences, as we demonstrated in this study. Thus, a less stringent percentile provides higher sensitivity at the discovery stage while still allowing downstream validation to establish specificity.

Also related to this, is there any theoretical or empirical way to actually estimate the false discovery rate achieved with the additional thresholds currently applied and described in the methods section? Anyway, the control of false positives should be good, considering the comparison with dcHIC where the results were filtered with adjusted p-value < 0.0001.

This is indeed a good point and something that can be estimated directly from our model. As a proof of principle, we estimated the false discovery rate (FDR) associated with the thresholds used in this study for HaploCalder in two cell lines: HCT116, a cancer cell line with a high abundance of regions exhibiting haplotype-specific compartments (HS-C), and IMR-90, a normal cell line with few HS-C regions. The FDR was computed as follows:

1. Identify HS-C regions with HaploCalder between two phased Hi-C maps generated with real haplotypes,

2. Identify HS-C regions with HaploCalder between two phased Hi-C maps generated with random haplotypes (n = 45 random haplotype pairs obtained by pairing 10 randomly assembled haplotypes)

The FDR is then defined by the ratio between the average number of HS-C detected when comparing random haplotypes (false positives) and the number of HS-C detected by comparing the real haplotypes (false + true positives). (This procedure was initially introduced [here](#).)

Interestingly, in both cell lines, the threshold that we selected, 0.15, was associated with an FDR < 0.1 (**Fig. R4**). Note that this threshold approximately corresponds to the minimal difference of compartment rank between two loci that are in a different sub-compartment, according to a 8 sub-compartment classification, which is the default number of sub-compartments identified by Calder.

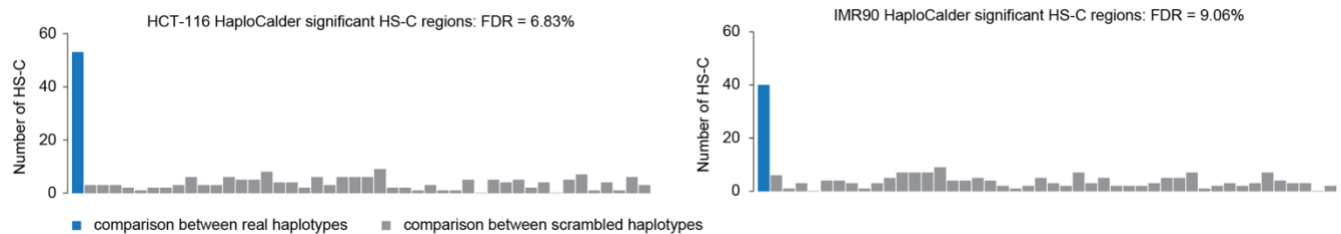


Fig. R4 (now also **Supplementary Fig. 4a**): Number of significant HS-C detected by HaploCalder in HCT116 (left) and IMR90 (right) when comparing real (blue) or scrambled (gray) haplotypes.

This analysis has now been included as **Supplementary Figure 4a**.

- To be fair, it should be mentioned that dChIC is a tool that can also handle replicates in the analysis of differential compartmentalization. Thus, the analysis of compartment differences in 1 vs 1 haplotype phased matrices may not be the best scenario for dChIC to perform well.

The reviewer is correct. We added the following sentence when describing the comparison between the two tools: “It should be noted that even though dChIC allows to compare multiple replicates for each condition, here it was used to compare two samples, one per condition.”

- some typos or formatting problems:

--> ref 36 the journal name should be "Bioinformatics"

--> line 663 I think there's an inversion in the order of sentences as it is referring to "the above 8 cell lines", whereas the cell lines are listed in the following sentence.

--> line 665 there's a reference to "[sup table]" that is probably a placeholder not updated to the final supplementary items numbering

We thank the reviewer for spotting those typos, which have now been fixed.

Reviewer #2 (Remarks on code availability):

I reviewed the github repository content but I didn't try to actually execute the code.

The repository comes with detailed step-by-step instructions about how to install the required packages, including dependencies.

Overall, the repository and documentation are well curated.

However, some specific parameters or analysis steps could benefit from additional details in the documentation.

For example, what are all the options available for the "enzyme" parameter in the working directory configuration file? Also, what should be indicated in case of mixed (dual enzymes) digestion, e.g. with Arima genomics kind of kits. To this concern, the "HaploC-tools/blob/main/HaploC/main/config_tmplate.txt" file seems to have slightly different options compared to what is described in "<https://github.com/CSOgroup/HaploC-tools/wiki/Prepare-working-directory>". e.g. "n_bwa_thread" vs "thread4bwa", not clear which is the right one.

We thank the reviewer for the care taken in checking our code. We now list the enzymes supported in the documentation page of HaploC (<https://github.com/CSOgroup/HaploC-tools/wiki/Prepare-working-directory>).

The enzyme support is provided internally by the juicer pipeline, which indeed includes Arima enzymes, as now explicitly stated. The discrepancy between the parameter names correctly reported by the reviewer was caused by an obsolete config template that was not in use. We have now updated the file to avoid confusion and ensure consistency with our documentation.

I could not find a point where some parameters usually relevant for Hi-C should be defined. e.g. I could not find a point where the Hi-C matrix "bin size" should be defined.

The bin size is indeed a critical parameter, even though we should clarify that in HaploC-tools, this parameter is not a one-size-fits-all. Indeed, HaploC-tools generates contact maps for both unphased and phased Hi-C datasets at multiple bin resolutions ranging between 10kb and 1Mb. Different maps can then be used in different downstream analyses. For example, HaploCalder relies on compartment calls generated with Calder (version 2 - <https://github.com/CSOgroup/CALDER2>), which implements an adaptive strategy to define the bin-size. Based on this strategy, Calder2 automatically increases the bin-size if the matrix resolution of a given chromosome is insufficient to adopt the smallest option (10kb). For HaploCI instead we fixed the bin-size at 25kb, which, for most Hi-C experiments, was a good compromise between exploring genomic loci with high granularity and retaining a sufficient number of contacts to robustly estimate insulation differences in phased contact maps.

Nevertheless, to increase transparency and flexibility for the users, we now included the option to select the bin resolutions to be used, allowing the users to tune this parameter based on their needs and data quality. These updates are documented here: <https://github.com/CSOgroup/HaploC-tools/wiki/Prepare-working-directory>.

Based on the main manuscript, I understand that HaploCI should return a set of regions with significant changes in chromatin insulation score, where significance is tested against randomized haplotypes. However, in the description of HaploCI output (<https://github.com/CSOgroup/HaploC-tools/wiki/HaploCI>), the "significance" of results is not really mentioned; thus, it is not clear if these results should be further filtered. There's a similar ambiguity in the wiki page describing the output of HaploCalder (<https://github.com/CSOgroup/HaploC-tools/wiki/HaploCalder>). However, in this case, there's a mention that "significant" changes are reported. But it's not clear how they were filtered (what statistical thresholds), and it is also not clear how the output of files **diff_rank** should be related to the statistical significance.

The reviewer is correct, and we apologize for not having properly described the output files of HaploC-tools. Indeed, we did not introduce specific statistical tests to derive a p-value in neither HaploCI nor HaploCalder, but rather we designed empirical null models based on scrambled haplotypes to derive the *expected* effect size and used this value to normalize the *observed* effect size.

For HaploCI, we decided to return the entire differential contact insulation profile, which users can visualize (e.g. using IGV) and analyze as they prefer. We prefer to avoid introducing a hard cut-off in this case for three main reasons: first, there is not an immediate interpretation of the magnitude of these values (unlike in the case of differential compartments), second, contact insulation scores may not be exactly

comparable between loci with different coverage, and, third, it allows the users to perform exploratory analyses, integrate results from multiple replicates and/or orthogonal data to corroborate insulation differences and select the most relevant to study.

For HaploCalder, instead, we introduced a cut-off to define regions exhibiting haplotype-specific compartments (HS-C). This threshold was set at 0.15, which approximately corresponds to the minimal compartment rank difference that exist between two loci that are in different sub-compartments, based on an 8 sub-compartment classification (which is the default number of sub-compartments used by Calder2). As shown in response to a previous question from the reviewer, we can leverage haplotype permutations to estimate the FDR of haplotype-specific compartment changes, and we showed that this cut-off was associated with an $FDR < 0.1$. Nevertheless, to avoid confusion, we rephrased the description of HaploCalder output in the wiki page, removing the term “significant” and adding a reference to our method section to clarify how regions were selected. In any case, the complete set of compartment differences are provided in output to allow the users to perform further exploratory analyses.

We hope we were able to clarify these points, and that the reviewer will appreciate the updates that were made to our code repository.