Exploiting native forces to capture chromosome conformation in mammalian cell nuclei

Miss Lilija Brant, Theodore Georgomanolis, Mr. Milos Nikolic, Mr. Chris Brackley, Petros Kolovos, Wilfred van IJcken, Frank G. Grosveld, Davide Marenduzzo and Argyris Papantonis

Corresponding author: Argyris Papantonis, University of Cologne

---

Editor: Maria Polychronidou

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision   27 September 2016

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the reviewers who agreed to evaluate the study. As you will see below, the reviewers appreciate that the presented approach is a valuable contribution to the field. They raise however a series of concerns, which should be carefully addressed in a revision.

I think that the recommendations of the reviewers are quite clear so there is no need to repeat the points listed below. Please feel free to contact me in case you would like to discuss any of these points in further detail.

REFEREE REPORTS

Reviewer #1:

By comparing the 3C, 4C Hi-C results from native and crosslinked cells, they showed that most of the chromatin interactions are conserved in the absence of x-linking. In other words, most of the previously discovered chromatin interactions from formaldehyde crosslinking cells are reliable. This is an important contribution to the field as skepticism about 3C methods persist.
Another paper had previously reported Hi-C with un-crosslinked cells but seemed to have higher background. Here, the authors have done a lot of experiments with the un-crosslinked condition native 3C: i3C, i4C, iHi-C and iT2C. The results and methods will be useful to the field. It would be good to get more information about what interactions are different between x-linked and non x-linked methods. Are there any general features (e.g. pol2 occupancy, histone marks etc that would be predictive of differences?

Also, the TALE-iD technology is very interesting. This method could be developed further and published separately.

A few minor comments:

1. In Fig.EV1B, all the enzymes that didn't work are 6-bp cutters. Is this a coincidence or because of other reasons?

2. Fig.3C, why does the scrambled TALE also have enriched signal around the anchor?

3. Some of the names are not consistent. In Fig.EV8, the second track from top, does "i3C-seq" mean "i4C-seq" as used in other places? Also, "TALE-iD" in Fig.3A and main text, but "TALEN-iD" in the legend.

4. Fig.EV17D, switching site of the directionality index (DI) is not equal to TAD boundary. The DI also switches in the middle of TADs. To get the boundaries, the authors need to run HMM after calling DI.

Reviewer #2:

Brant et al. describe an improved technique to study DNA:DNA interactions at a global scale within individual nuclei without the need for chemical crosslinkers. They show that their technique provides data of similar quality to 3C or 4C and corroborates the results observed using 3C and 4C-based approaches. They show that their approach can be applied to all:all comparisons such as those in HiC, and provides results comparable to high-resolution modifications of the technique such as micro-C. They have two main conclusions: first, that the chemical biases inherent in the crosslinker do not significantly change the patterns of association discovered by these methods, and second, that crosslinkers may not be necessary for most studies of genome organization.
The authors provide a thorough analysis of the applicability and robustness of their technique. The description of a crosslink-free method is of significance and the observation of similar results to approaches that use crosslinking is reassuring and addresses a persistent question in the field. As such this study is a very valuable contribution. We have only few major points of criticism but we would like clarification of some.

Major points:
1) The authors observe that more than 40% of chromatin is lost after digestion and before sequencing. They claim that this represents nonspecific loss and use as evidence for this claim that there is relatively similar enrichment for various annotations in the lost and retained fractions. We would like some more information to clarify whether nuclear structure or DNA sequence contribute to whether a fragment is lost or retained. For example:
a. Does the size of an ApoI fragment contribute to its likelihood to be retained or lost?
b. Are known structural elements such as LADs or Nucleolar Organizing Regions enriched in one pool of fragments?
c. If a locus has mapped reads from the retained pool, is it more or less likely to also be present in the lost pool (are these non-overlapping sets)?
2) The authors observe a striking sparsity in the iT2C data, as compared to the conventional T2C data they present. We are curious what the sources, and consequences, of this are. In particular:
a. Do the two libraries have similar coverage, and similar QC scores (total reads, percent mapped without errors, percent duplicated etc)? If this is not the case, what does the conventional T2C data look like when you resample it down to levels comparable to the iT2C data (or vice versa)?
b. In the i3C data, the authors noted that a full 83% of sequenced reads came from loci with more than 100 reads per million. Is this also true of the iT2C data, and if so is it causing the observed sparseness? (If this is the case, could the graphs in figure 4A be renormalized to better reflect this?)
c. What is the signal:noise ratio in iT2C data as opposed to T2C data? Can this be quantified using the previously identified CTCF-CTCF loops?
3) In general, can the authors include in their supplemental material and in their materials and methods section a more thorough discussion of the number of reads sequenced, number of reads mapped, and quality control metrics performed on these reads?

Minor points:
1) The authors use the terminology "more focal" to describe their data without specifying precisely what they mean by this. Please provide a specific metric for this observation.
2) The authors observe that 40% of cis-contacted fragments are shared between traditional 4C and i4C. This is surprisingly low. Are most of the 60% non-overlapping cis-contacts found only in 4C data or only in i4C data?
3) Similarly, in light of this, how correlated are i4C and conventional 4C?
4) Could the authors provide some quantification of the extent to which i4C shows "significantly lower numbers of "uncut" and "self-ligation" reads... as well as more reads mapping within its TAD"? What is the overall percent change and what is the level of significance?
5) Could the authors specify how the i3C-based techniques specifically differ from in situ Hi-C without a crosslinker?
6) Using conventional 4C, can small changes in structure upon NF-kB binding be observed, or are these observations only present in i4C?

Reviewer #3:

The manuscript describes a modification to standard #C-related technology to carry out this procedure without formaldehyde fixation. The authors present solid evidence showing that the new procedure gives equivalent results than fixation-based methods, but with some notable advantages. Overall, the manuscript represents an important contribution to this rapidly moving field. There are a few concerns the authors should address:

1. Step 3 in the iC technique eliminates some DNA by centrifugation of nuclei. Figure 1C indicates that the material eliminated contains enhancers, promoters, insulators, etc. The authors should show more convincingly that this material does not represent important contacts.

2. In Figure 2A, I believe the authors show domains obtained from Rao et al. and refer to these domains as TADs. However, the domains identified using the Arrowhead algorithm are not TADs, rather, they are smaller domains referred to by Rao et al as "loops" or contact domains, which are smaller than TADs.

3. In supplementary Figure EV11 A, it appears that treatment with RNase I results in a general decrease in interaction frequency that the authors interpret as evidence that RNA plays an important role in maintaining interactions. An alternative explanation is that RNase treatment increases the amount of material release from nuclei in step 3 of the procedure, and that the effect is not specific.

4. In Figure EV19 B, it appears that all the compartmental interactions normally found in HiC disappear in the iHi-C procedure. The authors should comment on this. Authors should also comment on the multiple lines, some of them perpendicular to the diagonal, that appear in these figures. Do these lines correspond to rearrangements in the cells used for the experiments?

1st Revision - authors' response   08 November 2016

Text continued on next page.

Point-by-point response to reviewers' comments

Reviewer #1:

By comparing the 3C, 4C Hi-C results from native and crosslinked cells, they showed that most of the chromatin interactions are conserved in the absence of x-linking. In other words, most of the previously discovered chromatin interactions from formaldehyde crosslinking cells are reliable. This is an important contribution to the field as skepticism about 3C methods persist.
Another paper had previously reported Hi-C with un-crosslinked cells but seemed to have higher background. Here, the authors have done a lot of experiments with the un-crosslinked condition native 3C: i3C, i4C, iHi-C and iT2C. The results and methods will be useful to the field. It would be good to get more information about what interactions are different between x-linked and non x-linked methods. Are there any general features (e.g. pol2 occupancy, histone marks etc that would be predictive of differences? Also, the TALE-iD technology is very interesting. This method could be developed further and published separately.

We thank the reviewer for finding that our "*results and methods will be useful to the field*". As regards features that would be predictive of differences, there are no obvious ones, but it is safe to say that the absence of recognizable features at particular contacts is only applicable to the conventional method. These mostly represent "bystander" or "baseline" interactions and are simply absent from i3C data. Of course, we cannot rule out that some hitherto unknown mark/feature might support such differences. Finally, as regards the TALE-iD approach, we thank the reviewer for highlighting it. We are indeed trying to develop a more global variant of it, but this is very much still work in progress.

A few minor comments:
1. In Fig.EV1B, all the enzymes that didn't work are 6-bp cutters. Is this a coincidence or because of other reasons?

This is a coincidence. There are 4-bp cutters (e.g., *Mbo*I) that also do not work in our PB buffer and it is due to incompatibility with our "physiological" buffer's composition; for example, the pre-2013 batches of *Hind*III by New England Biolabs worked fine in PB, but the new batches that followed require different salts/pH/etc and simply do not cut efficiently.

2. Fig.3C, why does the scrambled TALE also have enriched signal around the anchor?

We thank the reviewer for prompting us to look into this. It seems that the primer pair targeting region "p1" was somewhat "overefficient"; we have now replaced this particular pair, and added some more targeting other *Dpn*I sites in the *ZFPM2* locus, showing that no such trend exists.

3. Some of the names are not consistent. In Fig.EV8, the second track from top, does "i3C-seq" mean "i4C-seq" as used in other places? Also, "TALE-iD" in Fig.3A and main text, but "TALEN-iD" in the legend.

We thank the reviewer for pointing out these oversights so we could correct them throughout the text; the correct terms are "i4C-seq" (not "i3C-seq") and "TALE-iD" (not "TALEN-iD").

4. Fig.EV17D, switching site of the directionality index (DI) is not equal to TAD boundary. The DI also switches in the middle of TADs. To get the boundaries, the authors need to run HMM after calling DI.

This is correct, and we did also apply such a Hidden Markov Model. However, at the resolution of T2C (10-kbp or less) the sub-domains called essentially coincide with the switching sites of the DI.

Reviewer #2:

Brant et al. describe an improved technique to study DNA:DNA interactions at a global scale within individual nuclei without the need for chemical crosslinkers. They show that their technique provides data of similar quality to 3C or 4C and corroborates the results observed using 3C-based approaches. They show that their approach can be applied to all:all comparisons such as those in HiC, and provides results comparable to high-resolution modifications of the technique such as micro-C. They have two main conclusions: first, that the chemical biases inherent in the crosslinker do not significantly change the patterns of association discovered by these methods, and second, that crosslinkers may not be necessary for most studies of genome organization. The authors provide a thorough analysis of the applicability and robustness of their technique. The description of a crosslink-free method is of significance and the observation of similar results to approaches that use crosslinking is reassuring and addresses a persistent question in the field. As such this study is a very valuable contribution. We have only few major points of criticism but we would like clarification of some.

We would like to thank the reviewer for finding that we "*provide a thorough analysis of the applicability and robustness*" of i3C, and that "*this study is a very valuable contribution*".

Major points:

1) The authors observe that more than 40% of chromatin is lost after digestion and before sequencing. They claim that this represents nonspecific loss and use as evidence for this claim that there is relatively similar enrichment for various annotations in the lost and retained fractions. We would like some more information to clarify whether nuclear structure or DNA sequence contribute to whether a fragment is lost or retained. For example:

a. Does the size of an ApoI fragment contribute to its likelihood to be retained or lost?

No, this does not seem to be the case. We used our sequencing data from retained and lost fragments after cutting with *Nla*III (we have not sequenced *Apo*I-cut chromatin), and identified the ~300,000 that are most enriched (more than 5-fold) in each fraction compared to the other. The size distributions were essentially identical between the two groups (with a median of ~275 bp), despite the broad range of fragment sizes (50-4000 bp). This information is now added to **Figure S1**.

b. Are known structural elements such as LADs or Nucleolar Organizing Regions enriched in one pool of fragments?

As above, this does not seem to be the case. We used the genomic coordinates of LADs and NADs from fibroblasts (unfortunately there exists no such dataset for HUVECs), and overlapped them with the most enriched ~300,000 fragments in the "lost" or "retained" fractions (as above). Looking at either the number of reads carried by the fragments in each fraction or their sizes, no significant different is observed. This is also now included in **Figure S1**, and points to the lost and retained pools being overlapping sets (see also below).

c. If a locus has mapped reads from the retained pool, is it more or less likely to also be present in the lost pool (are these non-overlapping sets)?

The reviewer raises a very good point. Indeed, the "lost" and "retained" fragment pools are sets that overlap by almost 70%, and very few fragments are consistently depleted from one fraction while always being present in the other. We exemplify this by comparing the two pools in two chromosomes that vary greatly in size and gene content – chromosomes 2 and 20. Our analysis (now included in **Figure S1**) shows that the distribution of fragments unique to each pool in respect to HUVEC ChromHMM segments is very similar. There is only a small de-enrichment of the

"retained" pool for repetitive elements, and a slight enrichment for "active promoters" and "strong enhancers". Overall, this points to a heterogeneity that should probably be expected of a cell population, and is in line with the conclusions drawn from single-cell Hi-C (Nagano *et al*, 2013).

2) The authors observe a striking sparsity in the iT2C data, as compared to the conventional T2C data they present. We are curious what the sources, and consequences, of this are. In particular:
a. Do the two libraries have similar coverage, and similar QC scores (total reads, percent mapped without errors, percent duplicated etc)? If this is not the case, what does the conventional T2C data look like when you resample it down to levels comparable to the iT2C data (or vice versa)?
The conventional and iT2C datasets display very similar coverage (~45 million read pairs each), mapping efficiencies (~65%), and content of duplicates (<10%; see **Table S4** for details). Thus, no resampling is needed – the sparsity is not at all due to technical discrepancies, but rather an inherent feature of the i3C method.

b. In the i3C data, the authors noted that a full 83% of sequenced reads came from loci with more than 100 reads per million. Is this also true of the iT2C data, and if so is it causing the observed sparseness? (If this is the case, could the graphs in figure 4A be renormalized to better reflect this?)
We have added the relevant graph for the iT2C data (in comparison to conventional T2C) in **Figure S21**. Here, the conventional approach has ~90% interactions carrying <10 reads per million, whereas iT2C only ~60%. This is in agreement with a loss of "bystander/baseline" interactions from the iT2C data and explains the sparse motifs in the interaction matrices. Nonetheless, the data in **Figure 4A** were already presented normalized and plotted at the same scale. Hence, the matrices shown are not skewed as a result of these differences (e.g., **Figures S21** and **S22**).

c. What is the signal:noise ratio in iT2C data as opposed to T2C data? Can this be quantified using the previously identified CTCF-CTCF loops?
We have approached this calculation in different ways, but, because of the virtually empty matrix surrounding CTCF-CTCF contact bins (giving a denominator value close to 0), the ratio calculated is probably overstated. Thus, unfortunately, we cannot offer a numerical metric, but we believe that the matrices compared, for example, in **Figure S21F** exemplify this signal-to-noise improvement.

3) In general, can the authors include in their supplemental material and in their materials and methods section a more thorough discussion of the number of reads sequenced, number of reads mapped, and quality control metrics performed on these reads?
We apologize for the oversight. Although these metrics were included for all i4C experiments, the iT2C data were left out – we now include these as **Table S4**, and have also added extra details on quality control metrics for NGS reads and their mapping in the **Appendix Methods** section.

Minor points:
1) The authors use the terminology "more focal" to describe their data without specifying precisely what they mean by this. Please provide a specific metric for this observation.
The reviewer is right in pointing out the lack of a specific metric. We now provide this in **Figure S6** by comparing the breadth of interactions called by the *foursig* algorithm in all i4C or conventional 4C data generated using the *SAMD4A* TSS as a viewpoint. Both the distribution and the median size of contacted fragments are significantly smaller in i4C.

2) The authors observe that 40% of cis-contacted fragments are shared between traditional 4C and i4C. This is surprisingly low. Are most of the 60% non-overlapping cis-contacts found only in 4C data

or only in i4C data?

The 40% overlap in *cis*-contacted fragments between conventional 4C and i4C refers to the analysis of fragments carrying >100 rpm (in an effort to focus on "stronger" interactions). If we look at the overlap of all *cis*-contacted fragments the number will increase to >65% — still, we must note that the vast majority of differences is found at fragments outside the viewpoint's TAD.

3) Similarly, in light of this, how correlated are i4C and conventional 4C?

Pairwise correlations of i4C and conventional 4C datasets (and also between i4C replicates) are shown in **Figure S5B** and **S7D**; for these all Spearman's correlation coefficients are typically >0.7.

4) Could the authors provide some quantification of the extent to which i4C shows "significantly lower numbers of "uncut" and "self-ligation" reads... as well as more reads mapping within its TAD"? What is the overall percent change and what is the level of significance?

For uncut and self-ligation reads, the levels drop from 4-5% in conventional 4C to <2% in i4C (often to <1%). For reads mapping within TADs, the levels rise from ~15% in conventional 4C to >20% in i4C, while we observe ~10% fewer *trans*-contacted fragments. All these changes are statistically significant (*P*-value <0.01; Student's unpaired t-test), and the data is included in **Figure S6**.

5) Could the authors specify how the i3C-based techniques specifically differ from in situ Hi-C without a crosslinker?

We have tried to address this in two ways. First, we tried to reproduce the uncrosslinked *in situ* Hi-C protocol to generate 4C data from (the description in the original Rao *et al* paper only states that the procedure is identical to the crosslinked *in situ* Hi-C, but "with more gentle handling"; this means that both heating is used and SDS is added to uncrosslinked nuclei). In our hands, this did not work efficiently, even with short exposures to heating and lower SDS titers. Nonetheless, we did produce some 4C-seq data for the *SAMD4A* viewpoint (**Figure S13**), which when compared to the i4C approach seem markedly depleted of major interactions. We attribute this to the harsher handling of the *in situ* protocol, and also to the lack of a physiological buffer. Second, we looked at the published Hi-C data generated without crosslinking (Rao *et al*, 2014). In their paper, Rao *et al* generated 5 such libraries from uncrosslinked lymphoblasts; 4 of these were on cells embedded in agar plugs, and 1 in agar-free cells. Although this last library is the one that mostly resembles our approach, the data therein is sparse. Thus, wanting to do this nice work justice, we only compare our iHi-C to data from Hi-C on agar-embedded cells (where cut chromatin cannot escape nuclei). This led to three main observations: (*1*) the large-scale, low-resolution, structure of A/B compartments is sufficiently retained in agar-embedded cells, but is very faint in iHi-C; (*2*) at high-resolution the agar-embedded maps are more noisy and only few strong structural features stand out; (*3*) CTCF-CTCF loops are more difficult to detect in the agar-embedded Hi-C data, in contrast to the high local enrichments seen in iHi-C (**Figure S23**). These observations are now added to the main text (pg. 8-9), and we believe they highlight how the two approaches differ.

6) Using conventional 4C, can small changes in structure upon NF-kB binding be observed, or are these observations only present in i4C?

The vast majority of spatial interactions seen by conventional 4C in three loci on chr14 (*BMP4*, *CDKN3*, and *SAMD4A*; see **Figure S17**) remain unchanged upon TNF stimulation. This appears to be less so when using i4C for the same viewpoints, where some distinct changes (that also match the regulation of these three loci) were detected. We exemplify this by directly comparing raw profiles (**Figures S16** and **S17**) and by performing differential analyses of interactions (**Figure S19** and **S20**).

Reviewer #3:

The manuscript describes a modification to standard #C-related technology to carry out this procedure without formaldehyde fixation. The authors present solid evidence showing that the new procedure gives equivalent results than fixation-based methods, but with some notable advantages. Overall, the manuscript represents an important contribution to this rapidly moving field.

We would like to thank this reviewer for acknowledging that our work "*represents an important contribution to this rapidly moving field*".

There are a few concerns the authors should address:
1. Step 3 in the i3C technique eliminates some DNA by centrifugation of nuclei. Figure 1C indicates that the material eliminated contains enhancers, promoters, insulators, etc. The authors should show more convincingly that this material does not represent important contacts.

As this is a concern raised by both reviewers, we tried to address it in a number of ways, on top of the analysis that was already included in the manuscript. We have now added new analyses (**Figure S1**), i3C-qPCR validations (**Figure S3**), plus an iHi-C experiment comparing contacts in the "lost" and "retained" fractions" (**Figure S24**); they all support the notion that few (if any) "meaningful" contacts are specific to the material lost upon chromatin cutting of uncrosslinked nuclei.

2. In Figure 2A, I believe the authors show domains obtained from Rao et al. and refer to these domains as TADs. However, the domains identified using the Arrowhead algorithm are not TADs, rather, they are smaller domains referred to by Rao et al as "loops" or contact domains, which are smaller than TADs.

The domains showed in **Figure 2A** (and elsewhere, where domains are outlined on top of Hi-C data from Rao *et al*) were identified using the "directionality index/HMM" approach at 40-kbp resolution data (Dixon *et al*, 2012). This is now clarified in the corresponding figure legends.

3. In supplementary Figure EV11A, it appears that treatment with RNase I results in a general decrease in interaction frequency that the authors interpret as evidence that RNA plays an important role in maintaining interactions. An alternative explanation is that RNase treatment increases the amount of material release from nuclei in step 3 of the procedure, and that the effect is not specific.

The reviewer is right in pointing this possibility out. We have now performed an experiment to test this: we apply i3C to HUVECs by cutting chromatin with *Apo*I and post-treating (or not) with RNase A. Then, using a Qubit device, we find a ~1.5-fold increase in the amount of released chromatin; this is now added to the manuscript (pg. 6 and **Figure S13E**).

4. In Figure EV19B, it appears that all the compartmental interactions normally found in HiC disappear in the iHi-C procedure. The authors should comment on this. Authors should also comment on the multiple lines, some of them perpendicular to the diagonal, that appear in these figures. Do these lines correspond to rearrangements in the cells used for the experiments?

Indeed the large-scale compartments typically seen in Hi-C are not observed in iHi-C (and this is not at all connected to sequencing depth or the resolution at which the data are portrayed). We now point this out in the main text. As regards the lines perpendicular to the diagonal, these are also a feature of the conventional Hi-C, not an aberration of iHi-C. They are seen in multiple loci across the genome and typically represent interactions between large H3K27me3-marked regions. Thus, they do not represent genomic rearrangements in HUVECs (in fact, we also see then in IMR90s, in the agar-embedded lymphoblasts assayed by Rao *et al*, and in conventional Hi-C from three single HUVEC donors; *own unpublished data*).

Thank you again for sending us your revised manuscript. We now satisfied with the modifications made and I am pleased to inform you that your paper has been accepted for publication.

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ⬇

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

USEFUL LINKS FOR COMPLETING THIS FORM

Corresponding Author Name: Argyris Papantonis
Journal Submitted to: Molecular Systems Biology
Manuscript Number: MSB-16-7311

Reporting Checklist For Life Sciences Articles (Rev. July 2015)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures
1. Data
The data shown in figures should satisfy the following conditions:
➔ the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
➔ figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
➔ graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
➔ if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be justified
➔ Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

➔ a specification of the experimental system investigated (eg cell line, species name).
➔ the assay(s) and method(s) used to carry out the reported observations and measurements
➔ an explicit mention of the biological and chemical entity(ies) that are being measured.
➔ an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.

➔ the exact sample size (n) for each experimental group/condition, given as a number, not a range;
➔ a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
➔ a statement of how many times the experiment shown was independently replicated in the laboratory.
➔ definitions of statistical methods and measures:
• common tests, such as t-test (please specify whether paired vs. unpaired), simple χ2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
• are tests one-sided or two-sided?
• are there adjustments for multiple comparisons?
• exact statistical test results, e.g., P values = x but not P values < x;
• definition of 'center values' as median or average;
• definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

Please ensure that the answers to the following questions are reported in the manuscript itself. We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

In the pink boxes below, provide the page number(s) of the manuscript draft or figure legend(s) where the information can be located. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable).

B- Statistics and general methods

Please fill out these boxes ⬇ (Do not worry if you cannot see all your text once you press return)

Question	Answer
1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	Typically, in this kind of work, biological duplicates or triplicates suffice. In some instances we have exceeded this number.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	NA
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	NA
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	NA
For animal studies, include a statement about randomization even if no randomization was used.	NA
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	In the three different conditions tested for TALE-iD (Figure 3) the investigator performing the qPCR analyses was blind as regards the identity of the test/ctrl samples.
4.b. For animal studies, include a statement about blinding even if no blinding was done	NA
5. For every figure, are statistical tests justified as appropriate?	Yes, for all main and Appendix figures.
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	Yes they do. Distributions were plotted to ensure this.
Is there an estimate of variation within each group of data?	Yes.
Is the variance similar between the groups that are being statistically compared?	Yes.

C- Reagents

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	Citations and/or catalogue numbers are provided for all antibodies used (in the Methods section).
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	Here we mostly used HUVEC and IMR-90 human primary cells, which are provided by outlets ensuring authentication and no-contamination (Lonza Inc. and the Coriell repository, respectively). The K562 leukemia cells and E14 mouse ES cells we also used were routinely tested for mycoplasmam contamination.

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	NA
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	NA
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	NA

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	NA
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	NA
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	NA
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	NA
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	NA
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	NA
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	NA

F- Data Accessibility

18. Provide accession codes for deposited data. See author guidelines, under 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	All i3C data generated here is available at the EBI Array Express archive under accession number E-MTAB-4719. Data generated using the conventional 3C approach is available at the Sequence Read Archive under the accession number SRP066044.
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	See answer above.
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	NA
21. As far as possible, primary and referenced data should be formally cited in a Data Availability section. Please state whether you have included this section. Examples: **Primary Data** Wetmore KM, Deutschbauer AM, Price MN, Arkin AP (2012). Comparison of gene expression and mutant fitness in Shewanella oneidensis MR-1. Gene Expression Omnibus GSE39462 **Referenced Data** Huang J, Brown AF, Lei M (2012). Crystal structure of the TRBD domain of TERT and the CR4/5 of TR. Protein Data Bank 4O26 AP-MS analysis of human histone deacetylase interactions in CEM-T cells (2013). PRIDE PXD000208	NA
22. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	The computational algorithms used for i3C data analysis, as well as that used for in silico modeling, have already been published and are properly referenced throughout the manuscript.

G- Dual use research of concern

23. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	NA