Published in final edited form as:

Nat Protoc. 2020 March ; 15(3): 1209–1236. doi:10.1038/s41596-019-0285-9.

Measurement of differential chromatin interactions with absolute quantification of architecture (AQuA-HiChIP)

Berkley E. Gryder[1,3,*], **Javed Khan**[1,*], **Benjamin Z. Stanton**[2,3,*]

[1]Genetics Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

[2]Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, OH, USA.

[3]These authors contributed equally: Berkley E. Gryder, Benjamin Z. Stanton.

Abstract

Methods developed to capture protein-anchored chromatin interactions (chromatin interaction analysis by paired-end tag sequencing and HiChIP) have yielded tremendous insights into the 3D folding principles of the genome, but are normalized by sequencing depth and therefore unable to accurately measure global changes in chromatin interactions and contact domain organization. We herein describe the protocol for absolute quantification of chromatin architecture (AQuA)–HiChIP, an advance that allows the absolute differences in protein-anchored chromatin interactions between samples to be determined. With our method, defined ratios of mouse and human fixed nuclei are mixed and subjected to endonuclease digestion. Chromatin contacts are captured by biotin-dATP incorporation and proximity ligation, followed by gentle shearing, ChIP, biotin capture and paired-end sequencing. 3D contacts are counted from paired-end tags (PETs) from the human genome and are normalized to the total PETs from the mouse genome. As orthogonal normalization allows observation of global changes, the approach will enable more quantitative insights into the topological determinants of transcriptional control and tissue-specific epigenetic memory. With our approach, we have discovered that rapid histone deacetylase inhibition disrupts super enhancer function by creating many new aberrant contacts. The code for data analysis is available at https://github.com/GryderArt/AQuA-HiChIP. This protocol reports both experimental

[*]**Correspondence and requests for materials** should be addressed to B.E.G. berkley.gryder@nih.gov or J.K. khanjav@nih.gov or B.Z.S. benjamin.stanton@nationwidechildrens.org.

Author contributions

B.Z.S, B.E.G. and J.K. conceived of the project. B.Z.S. and B.E.G. performed AQuA-HiChIP experiments. B.E.G. built the AQuA analysis pipeline, B.Z.S. and B.E.G. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41596-019-0285-9.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability

All code used herein is either provided by other research laboratories (see links throughout the protocol for any given step) or is custom scripted in R (available here: https://github.com/GryderArt/AQuA-HiChIP). The code in this protocol has been peer reviewed.

and bioinformatic details to perform AQuA-HiChIP, going from cell culture to ranking chromatin interactions within 6 d.

Introduction

We have developed a method called ‘AQuA (absolute quantification of chromatin architecture)-HiChIP’ to enable direct instead of relative comparisons of chromatin interactions. The utility is such that for chromatin alterations that affect global modification states (i.e., drugs and knockouts of epigenetic regulators), the method can detect absolute changes in chromatin contacts by spike-in of nuclei to derive normalized paired end tags (PETs). One of the central confounding problems in epigenetics is the global nature of chromatin regulation: how can the regulatory effects of epigenetic perturbations be defined, if the resulting changes are global, and dividing contact counts per million mapped reads hides the total change? Our approach addresses this unmet need and allows absolute quantification of 3D chromatin contacts. In our work, we focus on histone modification states, which in various combinations are present at the terminus of every signaling pathway.

From the initial sequencing of the human genome[1], it took several years to gain insight into the organization of the epigenome[2–5]. Following the Human Genome Project (e.g., ENCODE[6–8]), new methods (Hi-C and chromatin interaction analysis by paired-end tag sequencing[9,10]) enabled definition of chromatin interactions and of the epigenetic landscape on which histone marks were placed. Recently, in situ Hi-C and HiChIP have enabled the generation of higher-resolution chromatin interaction maps[11] and protein-mediated chromatin interactions[12], respectively. Our method builds upon approaches based on sequencing of protein-mediated interactomes[13] and uses chromatin from an orthologous species to identify changes in structural features (e.g., enhancer-enhancer and enhancer-promoter interactions), loops and compartments[14]. Our method will be directly applicable to central questions, including characterization of the fundamental regulatory events for (i) transcriptional control, (ii) replication timing, or (iii) tissue-specific memory. Furthermore, our method will enable quantitative comparisons of architectural features across tissues or alleles, rather than merely comparing the rank-ordered interactivity within them. These are fundamental questions at the frontier of epigenomics and lead to critical questions about the etiological roles of architecture in human malignancy: what makes the architecture of a cancer genome different from a healthy cell? Currently, it is difficult to directly and absolutely compare the epigenome folding of different cell types if relative normalization is carried out based on total interactive PETs. This concept also applies to methods for -omics approaches studying protein localization across the genome (i.e., epigenomics), where traditional ChIP-seq methods have limitations in revealing global changes to the 2D landscape[15]. ChIP with reference exogenous genome normalization (ChIP-Rx[15]), where the researchers spike in orthologous species chromatin during the ChIP reaction, has facilitated identification of global changes in chromatin state. It follows that the same logic applies to the 3D chromatin interactions occurring across loci, especially if global changes occur[11,16,17]. Thus, we developed and herein report AQuA-HiChIP, which couples protein-centric 3D epigenome sequencing to internal normalization with orthologous chromatin (Fig. 1).

Overview of the AQuA-HiChIP procedure

Here, we present a protocol for AQuA-HiChIP, a modified HiChIP[12] that uses exogenous chromatin from an orthogonal species to accurately detect global changes in chromatin interaction frequencies. First-generation HiChIP is a method that combines in situ Hi-C[11] and ChIP-seq[5] such that cells are fixed to covalently link DNA/chromatin/proteins, permeabilized with gentle heating and SDS, followed by restriction digestion, Klenow extension with biotin-dATP, proximity ligation, sonication and immunoprecipitation-sequencing. The traditional first-generation HiChIP method uses streptavidin beads to immobilize and enrich for DNA-biotin after immunoprecipitation, followed by elution from the beads with dimethylformamide solution, indexing and paired-end sequencing. Our method builds on this protocol with simultaneous preparation of mouse and human nuclei in the permeabilization step, and using predefined ratios of mouse:human cells, the enzymatic steps are controlled for with 3:1 human/mouse cell equivalents. Finally, in the biotin-enrichment steps, we have modified the first-generation HiChIP by performing end repair, A-tailing, adaptor ligation and PCR indexing on the streptavidin beads to avoid DNA loss in the elution step, and to also avoid the use of dimethylformamide in sequencing preparation. This second-generation technology, AQuA-HiChIP, can generate high-quality interaction maps with as few as 100 million reads total (from both species, human and mouse herein) per sample. The procedure can be divided into five main components: (i) mouse and human cell preparation; (ii) permeabilization, restriction digest, biotin incorporation and proximity ligation; (iii) sonication and immunoprecipitation; (iv) biotin capture and on-bead library generation; and (v) cross-species bioinformatics for AQuA-HiChIP normalization.

Our method integrates advances from ChIP-Rx[15], which uses digested *Drosophila* chromatin for normalization via an H2Av ChIP (*Drosphila*) concurrently with a ChIP for a protein of interest in the mammalian chromatin. We achieve absolute normalization of chromatin interactions with concurrent in-situ Hi-C on mouse and human cells simultaneously, with enrichments from protein epitopes. The advance of incorporating mixtures of *Drosophila* chromatin with human chromatin (ChIP-Rx) enabled definitive mapping of 2D changes in chromatin modification states. While the use of smaller genomes (e.g., *Drosophila*) is efficacious for ChIP-seq, because small genomes afford more relative sequencing reads per locus (per library), normalization of architecture and looping with orthologous chromatin necessitated a more similar genome (e.g., mouse) where chromatin interactions are more highly conserved (~60–70%; ref. [11]). With protein-centric comparisons of chromatin interactions, many histone marks and chromatin regulatory proteins are highly conserved as well between mouse and human, enabling strategic use of antibody cross-species reactivity to normalize orthologous chromatin immunoprecipitations. For species-specific normalization, we define mouse PETs and human PETs within each set of experimental conditions. Because the ratio of the cell equivalents for input mouse/human chromatin is held constant, while the treatment for the human cells differs (e.g., dimethylsulfoxide (DMSO) control versus histone deacetylase (HDAC) inhibition), our normalization across conditions is based on comparisons to mouse PETs within each condition. The normalization method is robust to differing ratios of mouse/human cells and is robust to differences in duplicate read quantity introduced during library amplification. We show

the utility for understanding chromatin interactions by directly comparing interacting PETs (‘valid contacts’) across species from pooled nuclear extracts.

Applications and outlook

AQuA-HiChIP is most suitable in cases where global gains or losses of chromatin interactions may occur through perturbation, such as (i) rapid degradation approaches (e.g., auxin-inducible degradation[18,19] and dTag systems[20]) that reduce the absolute levels of the immunoprecipitation target protein, or (ii) a drug that acutely perturbs the binding interface of chromatin-associated factors across the genome (e.g., chromatin reader domain inhibitors for proteins such as BRD4 and PBRM1), or (iii) a perturbation that globally alters the post-translational modification status of histones (e.g., lymphocyte activation[21] or HDAC inhibition; Fig. 2). The pioneering of this technique was pursued to answer a conundrum of why HDAC inhibition and increased acetylation caused selective downregulation of core regulatory transcription factors (cf. Fig. 3c in ref. [22]). These technologies of rapid and global perturbation are important for facilitating enhanced understanding of direct causal mechanisms in chromatin biology, but results of 3D genome sequencing requiring normalization by read depth or total interacting PETs, in the absence of exogenous spike-in, can lead to masking global changes in the 3D landscape.

As mentioned previously, analogous concepts in the context of ChIP-seq have been addressed with global normalization methods and are especially important for detecting rapid small molecule–induced perturbation of histone marks[15]. As protein binding to the genome in three dimensions can reveal architectural patterns not available from 2D chromatin sequencing, we suggest that this technique will facilitate analysis of protein-anchored architecture of the epigenome. We also feel that further applications to non–protein-associated 3D genome sequencing with Hi-C (ref. [9,11]) will be interesting for extensions and generalizations of this approach to studying the global chromatin landscape.

Limitations

A primary limitation is that this technique requires the availability of a cross-reactive antibody that binds to both mouse and human targets. For instance, cancer type–specific translocation proteins (i.e., PAX3-FOXO1 or EWS-FLI1) or cell type–specific proteins cannot be assayed for their 3D connectome with spiked-in chromatin, because mouse cells lack the antibody target. In principle, this could be overcome by overexpressing the protein of interest exogenously in mouse cells.

Experimental design

Spike-in cells

We typically add 2 million mouse NIH3T3 cells to 6 million human cells. The proportion of mouse to human cells does not need to be any particular ratio, as long as the proportion is identical between the multiple conditions being compared. We suggest adding identical volumes of the same mouse cell resuspension to each human cell sample and that human cell samples have identical cell counts. It is critical that the human:mouse ratio is kept constant

across experiments to directly and absolutely compare differential chromatin interactions. We encourage >2 million cell equivalents of mouse chromatin, to ensure robust preparation of chromatin contacts (non-redundant interactive mouse PETs). However, as long as this ratio is held constant across experiments, AQuA normalization will be robust. In our method, species-specific normalization is achieved through defining the AQuA factor (ratio of human:mouse PETs), and multiplying by this ratio allows quantification of interaction frequencies across the genome. Consider, for instance, the situation in which a 6 million:2 million human:mouse ratio is used in one experiment, where AQuA-HiChIP gives an AQuA factor (human/mouse valid contacts) of 3 for the control and 12 for the treatment. The same experiment, with a human:mouse cell ratio of 10:2 changes the AQuA factors to 5 (control) and 20 (treatment). Yet, the AQuA normalization in both cases yields an AQuA impact ratio (AQuA factor of treated sample divided by AQuA factor of untreated sample) of 4, revealing that there is an increase of 400% in the valid contacts mediated by the protein of interest in the human cells.

Shearing

For planning any chromatin-sequencing experiments in which chromatin is sheared, sonication is a critical step that needs to be monitored closely. Before initiating AQuA-HiChIP, it is essential to test the sonication platform, to ensure that fixed chromatin can be sheared such that the DNA fragment length is between 100 and 700 bp (see example in Box 1). Optimization of the sonication conditions can be performed by fixing cells (10 min, 1% (wt/vol) formaldehyde) and proceeding directly to sonication in 700 µl of TE buffer with protease inhibitors for 2, 4, 6, 8, 10 and 12 min total shearing time (30 s ‘on’ and 30 s ‘rest’). After each cycle, remove 12.5 µl for reverse cross-linking (add to 12.5 µl of nuclear lysis buffer), add 1 µl of proteinase K and incubate overnight at 65 °C for each time point. MinElute-purifying these samples and comparing them on E-Gel EX will give an indication of shearing efficiency and generalizability of the Active Motif Epi-shear sonicator to related platforms.

Chromatin immunoprecipitation

Another critical variable to optimize before initiation of the AQuA experiment is performing traditional ChIP with an antibody of interest. In our studies, we have performed ChIP-quantitative polymerase chain reaction (qPCR) to ensure enrichments, and also validated this antibody in mouse cells[23]. In our experience, some commercial antibodies initially seem successful for ChIP-qPCR experiments but may produce low signal upon indexing of ChIP DNA and sequencing. For experimental design, we recommend addressing this issue in two ways. First, it is important to design and test several target loci for ChIP-qPCR with a new antibody, to measure whether the enrichment is greater than 10-fold for a target region over a control region. Second, if the 10-fold enrichment threshold is met, index the ChIP DNA for sequencing to and mapping to the genome browser for inspection. In our experience, comparisons with published data of the same or similar ChIP target in the UCSC (University of California, Santa Cruz) or Washington University genome browsers (https://genome.ucsc.edu) or the Broad Institute’s IGV (Integrative Genomics Viewer) desktop software (https://software.broadinstitute.org/software/igv/) are useful indicators for whether a ChIP-seq experiment indeed was successful. Additional validation can be performed with

knockout experiments or inhibitor experiments that disrupt the placement of the ChIP target and testing the anticipated loss of ChIP efficiency in these conditions. Performing replicates on samples that have been fixed separately or originated from at least two distinct passage numbers is also essential for achieving robustness.

Bioinformatic considerations

The paired-end sequencing results in two fastq files (R1 and R2) can be processed using HiC-Pro[24], a valuable and well-supported pipeline for Hi-C or HiChIP data analysis, as outlined in the Procedure. To obtain differences in human and mouse ratios for species-specific interactive PETs across experiments, the same raw data can be run through the HiC-Pro pipeline twice, once with a human genome reference (we used hg19 here) and a second time with a mouse genome build (we used mm10 here). AQuA-HiChIP with robust signal (e.g., >100-fold enrichment at ChIP target sites over background) for species-specific PETs can be obtained from active histone modification marks (H3K27ac and H3K4me3). Such samples need relatively small read quantities (100–200 million total PETs per experimental condition) to accurately define architecture in both the primary and spiked-in genomes.

We also suggest, as a quality-control metric, to run AQuA-HiChIP data through a standard ChIP-seq pipeline (e.g., mapping single-end sequencing reads to the human genome using Bowtie2, peak calling with MACS2 and generation of .bed files encoding non-redundant mapped reads[25,26]) to visualize the quality of the ChIP enrichment in 2D, compared to ChIP-seq of the same target. In this way, it is also possible to validate the quality of the ChIP step of AQuA-HiChIP using very low read depth sequencing, before sequencing deeply. In our experience, failed ChIP steps still generate AQuA-HiChIP data that have recognizable Hi-C structures and appear very similar to Hi-C itself (no ChIP step).

AQuA normalization is accomplished by calculating the AQuA factor as the ratio of human to mouse reads and applying this sample-specific scaling factor to its valid contact matrix at a locus of interest. Currently, the available visualization tools for Hi-C and HiChIP data types are not able to use an external quantification metric such as this. Therefore, we have utilized R-based graphics to generate contact matrices and APA (Aggregate Peak Analysis, a matrix summarizing a set of contact pairs genome wide[11,27]) plots with and without AQuA normalization (details for usage are in the protocol below, and full scripts are available at https://github.com/GryderArt/AQuA-HiChIP).

Materials

Biological materials

- NIH-3T3 mouse fibroblast cells (NCI CCR, Oncogenomics Section; RRID: CVCL_0594)
- RH4 cells, derived from PAX3-FOXO1 positive rhabdomyosarcoma (NCI CCR, Oncogenomics Section; RRID: CVCL_5916) **! CAUTION** The cell lines used in your research should be regularly checked to ensure they are authentic and are not infected with mycoplasma.

Reagents

- DMSO, molecular biology grade (Sigma, cat. no. D8418)
- Formaldehyde, molecular biology grade (ThermoFisher Scientific, cat. no. 28906) **! CAUTION** Formaldehyde is toxic and should be handled with protective equipment (safety data sheet: http://www.labchem.com/tools/msds/msds/VT310.pdf).
- Entinostat (Selleckchem, cat. no. S1053)
- Trypsin EDTA (ThermoFisher Scientific, cat. no. 25300-120)
- Protease Inhibitor cocktail (Active Motif, cat. no. 37490)
- E-Gel EX Agarose Gels, 2% (ThermoFisher Scientific, cat. no. G401002)
- DMEM (Quality Biological, cat. no. 112-013-101)
- FBS (ThermoFisher Scientific, cat. no. A3160502)
- PBS (VWR, cat. no. 114-057-101 CS)
- Glycine (Sigma, cat. no. G7126)
- Pen-Strep (penicillin–streptomycin solution, Fisher Scientific, cat. no. 15140163)
- MboI, (New England BioLabs, cat. no. R0147M)
- CutSmart Buffer (New England BioLabs, cat. no. B7204S)
- T4 DNA ligase (New England BioLabs, cat. no. M0202L)
- T4 DNA Ligase Buffer (10×) (New England BioLabs, cat. no. B0202S)
- DNA Polymerase I, Large (Klenow) Fragment (New England BioLabs, cat. no. M0210S)
- Biotin-14-dATP 0.4 mM (ThermoFisher Scientific, cat. no. 19524016)
- dCTP 10 mM (ThermoFisher Scientific, cat. no. 18253013)
- dGTP 10 mM (ThermoFisher Scientific, cat. no. 18254011)
- dTTP 10 mM (ThermoFisher Scientific, cat. no. 18255018)
- dATP 100 mM (ThermoFisher Scientific, cat. no. R0141)
- BSA, lyophilized (Sigma, cat. no. A9418)
- Proteinase K (Active Motif, cat. no. 53001)
- DNA Clean & Concentrator (Zymo Research, cat. no. D5205)
- Dynabeads M-280 Streptavidin (ThermoFisher Scientific, cat. no. 11205D)
- End-It DNA End-Repair Kit (Lucigen, cat. no. ER81050)
- Klenow Fragment (3’→5’ exo-) (New England BioLabs, cat. no. M0212S)
- NEBuffer2 (Klenow Buffer; New England BioLabs, cat. no. B7002S)

- Primers and adaptors (made by IDT; Table 1) ▲ **CRITICAL** Primers are reconstituted in water to 100 μM and stored at −20 °C.
- Phusion High-Fidelity PCR Master Mix with HF Buffer (New England BioLabs, cat. no. M0531S)
- Agencourt AMPure XP beads (Beckman Coulter, cat. no. A63880)
- E-Gel EX Agarose Gels (ThermoFisher Scientific, cat. no. G401002)
- KAPA Library Quantification kit (Roche, cat. no. KK4835)
- Anti-H3K27ac antibody (Active Motif, cat. no. 39133)
- Dynabeads Protein A (ThermoFisher Scientific, cat. no. 10001D)

Equipment

- E-Gel Power Snap electrophoresis device (ThermoFisher Scientific, cat. no. G8100) **! CAUTION** Utilize eye protection or a Safe Imager Blue-Light Transilluminator for safety during visual inspection of bands on E-Gel EX using blue-light transillumination.
- Nexellom Cellometer (Nexcelom, cat. no. T4-IQOQ)
- 96R Ring Magnet Plate (ThermoFisher Scientific, cat. no. NC0280470)
- DynaMag Magnet (ThermoFisher Scientific, cat. no. 12-321-D)
- mySPIN 6 Mini Centrifuge (ThermoFisher Scientific, cat. no. 75004061)
- EpiShear Probe Sonicator (Active Motif, cat. no. 53051)
- EpiShear Cooled Sonication Platform (Active Motif, cat. no. 53080)
- Eppendorf Thermomixer (ThermoFisher Scientific, cat. no. 05-412-500)
- Roto-Therm (VWR, cat. no. 470313-920)
- 2100 Bioanalyzer Instrument (Agilent, cat. no. G2939BA)
- Viia 7 (ThermoFisher Scientific, cat. no. 4453545)
- Hardware specifications used for data analysis: four parallel cluster nodes, each with 64 GB RAM, 16 CPU cores, and 400 GB hard drive space (https://hpc.nih.gov/systems/)

Software

- bcl2fastq v2.20.0 (https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html)
- Bowtie 2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml)[28]
- MACS v2.1.1.20160309 (https://github.com/taoliu/MACS)[29]
- RStudio (https://www.rstudio.com/products/RStudio/)
- HOMER v4.9.1 (http://homer.ucsd.edu/homer/motif/)[30]

- AQuA (https://github.com/GryderArt/AQuA-HiChIP)
- HiC-Pro v2.11.1 (https://github.com/nservant/HiC-Pro)[24]
- Juicebox v1.11.08 (https://github.com/aidenlab/Juicebox)[31]

Reagent setup

PE Adapter Oligo Mix.—Make the adaptor by annealing multiplexing adapter-top with multiplexing adapter-bottom (Table 1) for 5 min at 95 °C followed by slowly cooling to room temperature over 120 min. The final concentration of the annealed PE Adapter Oligo Mix is 15 μM. This annealed adaptor should be prepared in advance and stored at −20 °C for ≤6 months.

Hi-C lysis buffer

Prepare as outlined below and store at room temperature for ≤1 year.

Stock concentration	Final concentration	Volume (ml) for 50 ml final	Fold dilution
1 M Tris-HCl pH 7.5	10 mM Tris-HCl pH 7.5	0.5	100
5 M NaCl	10 mM NaCl	0.1	500
10% (vol/vol) NP-40	0.2% (vol/vol) NP-40	1.0	50

Nuclear lysis buffer

Prepare as outlined below and store at room temperature for ≤1 year.

Stock concentration	Final concentration	Volume (ml) for 50 ml final	Fold dilution
1 M Tris-HCl pH 7.5	50 mM Tris-HCl pH 7.5	2.5	20
0.5 M EDTA	10 mM EDTA	1.0	50
10% (wt/vol) SDS	1% (wt/vol) SDS	5.0	10

ChIP dilution buffer

Prepare as outlined below and store at room temperature for ≤1 year.

Stock concentration	Final concentration	Volume (ml) for 50 ml final	Fold dilution
1 M Tris-HCl pH 7.5	16.7 mM Tris-HCl pH 7.5	0.8	59.9
0.5 M EDTA	1.2 mM EDTA	0.1	416.7
10% (wt/vol) SDS	0.01% (wt/vol) SDS	0.05	1,000.0
10% (vol/vol) Triton X-100	1.1% (vol/vol) Triton X-100	5.5	9.1
5 M NaCl	167 mM NaCl	1.7	29.9

Low-salt wash buffer

Prepare as outlined below and store at room temperature for ≤1 year.

Stock concentration	Final concentration	Volume (ml) for 50 ml final	Fold dilution
1 M Tris-HCl pH 7.5	20 mM Tris-HCl pH 7.5	1.0	50.0
0.5 M EDTA	2 mM EDTA	0.2	250.0
10% SDS (wt/vol)	0.1% (wt/vol) SDS	0.5	100.0
10% (vol/vol) Triton X-100	1% (vol/vol) Triton X-100	5.0	10.0
5 M NaCl	150 mM NaCl	1.5	33.3

High-salt wash buffer

Prepare as outlined below and store at room temperature for ≤1 year.

Stock concentration	Final concentration	Volume (ml) for 50 ml final	Fold dilution
1 M Tris-HCl pH 7.5	20 mM Tris-HCl pH 7.5	1.0	50
0.5 M EDTA	2 mM EDTA	0.2	250
10% SDS (wt/vol)	0.1% (wt/vol) SDS	0.5	100
10% (vol/vol) Triton X-100	1% (vol/vol) Triton X-100	5.0	10
5 M NaCl	500 mM NaCl	5.0	10

LiCl wash buffer

Prepare as outlined below and store at 4° C for ≤2 months.

Stock concentration	Final concentration	Volume (ml) for 50 ml final	Fold dilution
1 M Tris-HCl pH 7.5	10 mM Tris-HCl pH 7.5	0.5	100
2.5 M LiCl	250 mM LiCl	5.0	10
10% (vol/vol) NP-40	1% (vol/vol) NP-40	5	10
10% (wt/vol) sodium deoxycholate	1% (wt/vol) sodium deoxycholate	5	10
0.5 M EDTA	1 mM EDTA	0.1	500

Tween wash buffer

Prepare as outlined below and store at room temperature for ≤1 year.

Stock concentration	Final concentration	Volume (ml) for 50 ml final	Fold dilution
1 M Tris-HCl pH 7.5	5 mM Tris-HCl pH 7.5	0.3	200
0.5 M EDTA	0.5 mM EDTA	0.1	1000

Stock concentration	Final concentration	Volume (ml) for 50 ml final	Fold dilution
5 M NaCl	1 M NaCl	10.0	5
10% (vol/vol) Tween-20	0.05% (vol/vol) Tween-20	0.3	200

Biotin binding buffer (2×)

Prepare as outlined below and store at room temperature for ≤1 year.

Stock concentration	Final concentration	Volume (ml) for 50 ml final	Fold dilution
1 M Tris-HCl pH 7.5	10 mM Tris-HCl pH 7.5	0.5	100
0.5 M EDTA	1 mM EDTA	0.1	500
5 M NaCl	2 M NaCl	20.0	2.5

Elution buffer

Prepare as outlined below and store at room temperature for ≤1 d.

Stock concentration	Final concentration	Volume (ml) for 10 ml final	Fold dilution
0.5 M NaHCO4 pH 8.0	50 mM NaHCO3	1	10
10% SDS (wt/vol)	1% SDS (wt/vol)	1	10

Procedure

Growing control (untreated) cells, treated cells and (control) species-orthogonal cells ● Timing 18–24 h

1. Plate between 10 and 12 million cells (in our experiments, adherent human cancer cell lines) in a 15-cm dish for each condition and let adhere overnight at 37 °C. According to the biological question being asked, treat one dish of cells with control (e.g., 20 μl of 0.1% (vol/vol) DMSO in 20 ml of culture medium) and a second dish with perturbing agent (e.g., 1 μM entinostat by adding 20 μl of a 1 mM stock solution into 20 ml of culture medium) for the desired time (e.g., 6 h).

2. In parallel, and not necessarily on the same day, grow cells from an orthogonal species. Seed 10 million cells (in our case, NIH-3T3 mouse cells) and grow until the next day in an incubator set to 37 °C.

Cell harvesting, quantification and formaldehyde fixation ● Timing 2 h

3. Harvest and count treated and control cells by aspirating medium, washing with PBS gently and applying 4 ml of trypsin (0.05%) and incubating for 2 min until cells are detached and clumps are evenly resuspended. Quench trypsin with 9 ml of complete culture medium (containing 10% (vol/vol) FBS) and achieve

close to a single-cell suspension by pipetting cells between 5 and 10 times once transferred into a 15-ml conical tube.

4. Count cells (expected: 1–3 million cells/ml) automatically in a Nexcelom Cellometer with at least four replicates per condition achieving low standard deviation (we aim for <5% coefficient of variation).

5. Calculate the total number of cells in each condition. For the condition with more cells, discard volume to achieve an identical total number of cells per condition and refill with complete medium to create equivalent volumes (13 ml each).

6. Add formaldehyde to achieve a final concentration of 1% (wt/vol) and mix gently using RotoTherm (see Equipment) for 10 min at room temperature. For 13 ml of counted cells in culture medium from Step 5, add 360 µl of 37% (wt/vol) formaldehyde. Quench the formaldehyde with a final concentration of 125 mM glycine at 4 °C for 5 min.

   ! CAUTION Formaldehyde is toxic and should be handled with protective equipment.

   ▲**CRITICAL STEP** Allowing the fixation to proceed for >10–12 min may result in over-fixing and an overall signal loss in the immunoprecipitation steps.

7. Pellet cells at 1,250*g* for 3 min at 4 °C. Remove and discard the supernatant. Resuspend in cold PBS (5 ml) on ice, pellet again at 1,250*g* for 3 min at 4 °C and discard the supernatant.

8. Flash-freeze pellets at −80 °C or move ahead to Step 10.

9. Prepare exogenous spiked-in cells from an orthogonal species in the same manner as described in Steps 3–8, but before freezing, divide into aliquots 2 million cells per 1.5-ml tube (approximately six aliquots for a 15-cm dish of NIH-3T3 cells).

   ■ **PAUSE POINT** Accurately counted and fixed cell aliquots for paired treated samples and/or exogenous chromatin can be stored at −80 °C for ≤1 year.

Lysis and restriction digest ● Timing 3.5 h

10. Thaw the cells gently on ice for 8–10 min to recover from snap-freeze. Resuspend ~6–10 million crosslinked human cells mixed with 2 million crosslinked mouse cells (~20% orthologous chromatin, by cell equivalency for AQuA) in 500 µl of ice-cold Hi-C Lysis Buffer (with freshly added protease inhibitor cocktail) and rotate at 4 °C for 30 min.

11. Centrifuge at 2,500*g* for 5 min at 4 °C, discard the supernatant and proceed to the next step with the pellet.

12. Wash pellet from Step 11 with 500 µl of ice-cold Hi-C Lysis Buffer (without resuspending it). Centrifuge again at 2,500*g* for 5 min at 4 °C.

13. Remove and discard the supernatant and resuspend in 100 µl of 0.5% (wt/vol) SDS (in TE pH 7.4) to permeabilize nuclei in preparation for in situ enzymatic digestion steps.

14. Incubate mixture for 10 min at 62 °C and then add 285 µl of water and 50 µl of 10% (vol/vol) Triton X-100 to quench the SDS.

15. Mix by pipetting, spin down the sample and incubate for 15 min at 37 °C.

16. Add 50 µl of 10× NEBuffer 2, mix well and then add 200 U (8 µl) of MboI restriction enzyme, mix and digest chromatin for 2 h at 37 C with rotation or agitation.

17. Heat inactivate MboI for 20 min at 62 °C. Then, cool the sample at 4 °C for 5 min.

Biotin incorporation and proximity ligation ● Timing ~5 h

18. After in situ enzymatic digestion with MboI, overhangs will be blunted and biotinylated. To the heat-inactivated mixture from Step 17, add 52 µl of the following master mix:

Reagent	Amount (µl)	Final concentration
0.4 mM biotin-dATP	37.5	288 µM
10 mM dCTP	1.5	288 µM
10 mM dGTP	1.5	288 µM
10 mM dTTP	1.5	288 µM
5U/µl DNA Polymerase I Large (Klenow) Fragment	10	0.96 U/µL
Total	52	

19. Mix gently by pipetting and incubate at 37 °C for 1 h with gentle agitation in a ThermoMixer.

20. Prepare the following ligation master mix (948 µl per sample). Add to samples from Step 19, mix well and proceed to the next step.

Reagent	Amount (µl)	Final concentration
10× NEB T4 DNA ligase buffer	150	1.58×
10% (vol/vol) Triton X-100	125	1.3%
50 mg/ml BSA	3	158 µg/ml
Water	660	–
400 U/µl T4 DNA Ligase	10	4.2 U/µl
Total	948	

21. Incubate at room temperature for 4 h with gentle agitation in a ThermoMixer (300 r.p.m). Alternatively, the sample can be divided into eight PCR strip tubes

and temperature controlled with a thermocycler to leave overnight (set to 23 °C for 4 h, then 4 °C overnight).

22. Centrifuge reaction mixture at 2,500*g* for 5 min at 4 °C and remove and discard the supernatant. Use pelleted nuclei in subsequent steps.

Sonication ● Timing 1 h (overnight step when optimizing shearing)

23. Bring pellet up to 700 μl in Nuclear Lysis Buffer supplemented with freshly added protease inhibitor cocktail.

24. Use 10 min of total shearing 'on' time with 30 s 'on' and 30 s 'rest' with an Active Motif Epi-shear probe sonicator (see Equipment) at 30% power. While checking the shearing efficiency, keep samples at 4 °C.

! CAUTION Sonication can cause ear damage and hearing loss and must be carried out either in a sound-proof container (such as available here: https://www.activemotif.com/catalog/785/epishear-cooled-sonication-platform) or with the use of sound-blocking ear muffs.

? TROUBLESHOOTING

25. Check shearing: take 12.5 μl of sonicated lysate and add to 12.5 μl of Nuclear Lysis Buffer. Add 1 μl of Proteinase K and incubate at 65 °C overnight after mixing well. Purify DNA with a MinElute PCR purification kit. Run E-Gel EX with standards to get shearing fragment length to ~300–1200 bp. If run on an Agilent Bioanalyzer, most fragment lengths can appear slightly shorter, in a range from 100 to 700 bp (see Box 1).

Immunoprecipitation ● Timing overnight

26. Spin the sample from Step 24 for 15 min at 16,000*g* at 4 °C. This step adheres to the first-generation HiChIP protocol and aims to homogenize the solution.

27. Add 2× volume of ChIP Dilution Buffer (split into two tubes of ~400 μl each and add 750 μl of Dilution Buffer for 1:3 dilution). Then, dilute again 1:1 with ChIP Dilution Buffer (with 1× protease inhibitors). It will be ~1,100 μl per tube in four final tubes with ~0.16% (wt/vol) SDS (total 1:6 dilution from 1% SDS).

? TROUBLESHOOTING

28. Add 5 μl of Active Motif #39133 anti-H3K27ac antibody per tube and incubate at 4 °C overnight with overhead rotation.

29. Resuspend 100 μl of Protein A beads in an equivalent volume of Dilution Buffer, add resuspended beads to the samples (25 μl of beads per tube, four tubes per ChIP target) and incubate at 4 °C for 2 h with rotation.

▲**CRITICAL STEP** Steps 29–32 equilibrate Dynabeads Protein A into ChIP Dilution Buffer. Approximately 100 μl of Dynabeads Protein A are used per AQuA-HiChIP sample (25 μl of beads per tube, four tubes with 1.1 ml each).

▲**CRITICAL STEP** Protein A and Protein G have different affinities for certain species' IgG. Generally, Protein A binds well to rabbit IgG antibodies, while Protein G binds well to rabbit, goat and mouse IgG. A useful compatibility reference can be found on the NEB website: https://www.neb.com/tools-and-resources/selection-charts/affinity-of-protein-ag-for-igg-types-from-different-species).

? TROUBLESHOOTING

30. Settle beads on the side of the tubes at room temperature using a magnetic stand.

31. Wash beads twice with low-salt wash buffer, high-salt wash buffer, and LiCl wash buffer each. Washing can be performed at room temperature on a magnet by adding 500 µl of a wash buffer, swishing the beads back and forth twice by moving the sample relative to the magnet and then removing and discarding the supernatant.

ChIP DNA reverse crosslinking ● Timing 3 h

32. Resuspend beads in 100 µl of DNA Elution Buffer.

33. Incubate at 23 °C for 10 min with rotation, followed by shaking at 500–700 r.p.m. for 3 min at 37 °C.

34. Place samples on the magnet and remove the supernatant to a fresh tube. Add another 100 µl of DNA Elution Buffer to the ChIP samples and repeat incubation as described in Step 33.

35. Remove the ChIP eluent again and combine for 200 µl total per sample.

36. Add 10 µl of Proteinase K per eluate and incubate at 55 °C for 45 min with shaking in a benchtop ThermoMixer at 500–700 r.p.m.

37. Adjust to 67 °C and incubate for ≥1.5 h with shaking at 500–700 r.p.m.

38. Purify the DNA from the samples with ChIP DNA Clean & Concentrator (Zymo Research) and elute in 10 µl per AQuA-HiChIP immunoprecipitation, with elution buffer provided in the Zymo kit.

■ **PAUSE POINT** While we recommend continuing immediately to biotin capture, it is possible to store eluted AQuA-HiChIP DNA at −20 °C for a few days before continuing to Step 42.

Biotin capture ● Timing 40 min

39. Prepare for biotin pulldown by washing 5 µl of M-280 Streptavidin Dynabeads twice with 10 µl of Tween Wash Buffer.

40. Resuspend the beads in 10 µl of 2× Biotin Binding Buffer and add to the 10-µl sample from Step 38. Incubate the mixture at room temperature for 15 min with rotation.

41. Let stand on the magnet and remove and discard the buffer.

42. Resuspend the beads in 500 µl of Tween Wash Buffer and incubate at 55 °C for 2 min with shaking at 300–400 r.p.m.; then, wash with 500 µl of TE pH 7.4 and leave at room temperature for on-bead library preparation. Proceed to on-bead End-Repair.

End-Repair ● Timing 1 h 10 min

43. Resuspend the bead-bound sample from Step 42 with 34 µl of TE pH 7.4 and add the following (from the End-It DNA End-Repair Kit) to make the End-Repair reaction mix:

Reagent	Amount (µl)	Final concentration
10× *End-Repair Buffer*	5	1×
dNTPs (2.5 mM each)	5	250 µM
ATP (10 mM)	5	1 mM
End-Repair Enzyme mix	1	–
Total	50 (with sample)	

44. Mix well by pipetting (do not vortex) and incubate for 1 h at room temperature with slow agitation (350 r.p.m.) in a Thermomixer.

45. Resuspend to wash once with 500 µl of Tween Wash Buffer and then once with 500 µl of TE pH 7.4, keeping beads on the magnet throughout washes.

A-tailing to 3′ ends of DNA fragments ● Timing 50 min

Resuspend the bead-bound sample from Step 45 with 32 µl of TE pH 7.4 and add the following to make the A-tailing reaction mix:

Reagent	Amount (µl)	Final concentration
10× Klenow buffer (NEBuffer 2)	5	1×
1 mM dATP	10	200 µM
5 U/µl Klenow Fragment (3’→5’ exo-)	3	0.3 U/µl
Total	50 (with sample)	

46. Incubate for 45 min at 37 °C in ThermoMixer with gentle rotation (300 r.p.m.).

Resuspend to wash once with 500 µl of Tween Wash Buffer and then once with 500 µl of TE pH 7.4, keeping beads on the magnet throughout washes.

Linker ligation ● Timing 1 h

47. Resuspend the bead-bound sample from Step 46 with 22 µl of TE pH 7.4 and add the following to make the Adaptor ligation reaction mix:

Reagent	Amount (μl)	Final concentration
10× T4 DNA ligase buffer	3	1×
Index PE Adapter Oligo Mix	2	1 μM
T4 DNA ligase (400 U/μl)	3	40 U/μl
Total	30 (with sample)	

48. Incubate for a minimum of 1 h at room temperature with gentle rotation.

49. Resuspend to wash once with 500 μl of Tween Wash Buffer and then once with 500 μl of TE pH 7.4, holding beads magnetically.

Amplify AQuA-HiChIP sequencing libraries with unique indexes ● Timing 1–2 h

50. Resuspend beads from Step 49 in 23 μl of water. Add 1 μl of Multiplexing PCR FWD (forward) Primer (Box 2) and 1 μl of a REV (reverse) Primer M1–12, with unique barcode (Box 2). Mix well with quick vortex and table-top centrifugation for several seconds.

51. Add 25 μl of Phusion High-Fidelity PCR Master Mix with HF Buffer and pipette to mix. Remove 25 μl of sample and place in a 200-μl PCR strip tube to prepare for amplification.

52. Run the following program for on-bead PCR amplification with 25 μl (half of the total material, after thorough mixing). Check library amplicon fragment lengths with aliquots of 2 μl diluted in 10 μl of water and run with E-Gel EX after 10, 12 and 14 cycles. Desired fragment length should be in the range of 300–800 bp.

Cycle number	Denature	Anneal	Extend
1	98 °C, 30 s		
2 – 14	98 °C, 10 s	65 °C, 30 s	72 °C, 30 s
Hold	4 °C, ∞		

? TROUBLESHOOTING

53. After PCR, place libraries on a magnet, remove the supernatant and transfer into new tubes. Save the unused 25-μl PCR reaction mix at 4 °C to amplify separately after the desired cycle number from Step 52 has been defined with the initial on-bead amplification reactions.

54. Add 25-μl qPCR reaction mix to 25 μl of AMPure beads (1:1 ratio to purify sample from unreacted PCR primers).

55. Mix well and let stand 5 min with beads. Put on a magnet for 5 min. Then, remove the supernatant and wash quickly two times with 80% ethanol with gentle pipette aspiration.

56. Let stand to dry on a magnet for 10–15 min. Add 20 μl of ultrapure water. Mix well and incubate for 5 min. Put on a magnet for 3 min.

57. Remove the 18 μl of purified indexed AQuA-HiChIP DNA.

■**PAUSE POINT** Indexed AQuA-HiChIP DNA can be stored at −20 °C until ready for sequencing.

Sequencing and bioinformatic analysis

58. Quantify the libraries using a qPCR-based approach (we use the KAPA Library Quantification kit) and pool to multiplex (2.6 fmol total library material; thus, if multiplexing two separate experiments together, each sample will require 1.3 fmol) according to standard Illumina protocols (www.illumina.com/documents/products/datasheets/datasheet_sequencing_multiplex.pdf).

59. Sequence the libraries in 150-bp paired-end mode. A typical run can be performed on an Illumina NextSeq High Output flow cell in paired-end mode with four samples multiplexed to achieve ~100–120 million reads per sample.

60. The output of Illumina sequencing is a BCL (binary base call) file. From this, create paired fastq (R1 and R2) files using Illumina's bcl2fastq software version 2.17 or greater.

61. To perform dual human and mouse mapping of fastq files, use HiC-Pro (https://github.com/nservant/HiC-Pro). First, establish appropriate reference digested genomes for human and mouse using the following example code.

    Command line example code for digesting hg19:

```
/usr/local/hicpro/2.10.0/HiC-Pro_2.10.0/bin/utils/digest_genome.
py -r mboi -o MboI.ucsc.hg19.bed /HiC/reference_files/
bowtie2_index/
ucsc.hg19.fasta
```

    Command line example code for digesting mm10:

```
/usr/local/hicpro/2.10.0/HiC-Pro_2.10.0/bin/utils/digest_genome.
py -r mboi -o MboI.ucsc.mm10.bed /UCSC/mm10/Sequence/Bowtie2Index/
genome.fa
```

62. Link or copy fastq files from Step 60 into a /SampleSet_hg19/DATA/ folder, where 'SampleSet' represents the set of samples to be operated on for a given project or experiment set. Create a mirrored copy, /SampleSet_mm10/DATA/, with the same fastq files. An example initial folder structure is as follows:

```
projects/
├── RH4_H3K27ac_HiChIP_hg19
│   └── DATA
```

```
│  ├── Sample_RH4_D6_H3K27ac_HiChIP_HKJ22BGX7
│  └── Sample_RH4_Ent6_H3K27ac_HiChIP_HKJ22BGX7
└── RH4_H3K27ac_HiChIP_mm10
└── DATA
├── Sample_RH4_D6_H3K27ac_HiChIP_HKJ22BGX7
└── Sample_RH4_Ent6_H3K27ac_HiChIP_HKJ22BGX7
```

where each /Sample_*/ folder contains both Sample_*R1.fastq.gz and Sample_*R2.fastq.gz (paired-end read mates).

63. Next, run HiC-Pro pipeline twice, separately, once per reference genome.

? TROUBLESHOOTING

hg19 example code to run HiC-Pro:

```
/hicpro/2.10.0/HiC-Pro_2.10.0/bin/HiC-Pro -i /projects/RH4_H3K27ac_
HiChIP_hg19/DATA/ -o /projects/RH4_H3K27ac_HiChIP_hg19/
HiCpro_OUTPUT/
-c /reference_files/config_hg19.txt -p
```

64. When running on a cluster, HiC-Pro will generate individual shell scripts for part 1 on the pipeline execution (mapping, filtering for valid pairs; this step often runs overnight) and part 2 (summary statistics, ICE (iterative correction) normalization outputs, etc.) and place them in the /HiCpro_OUTPUT/ folder. Navigate on the command line to the /HiCpro_OUTPUT/ folder and execute, once per hg19 and again for mm10.

hg19 example code to run HiC-Pro pipeline part 1:

```
sbatch -J HiCstep1 --time=24:00:00 --mem=121g --cpus-per-task=4
--gres=lscratch:200 HiCPro_step1_TestHiCpro.sh sbatch -J HiCstep2
--time=24:00:00 --mem=121g --cpus-per-task=4 --gres=lscratch:200
HiCPro_step2_TestHiCpro.sh
```

Once part 1 is complete (if your email was placed in the configuration file, you should be notified of this), then run HiC-Pro pipeline part 2:

```
sbatch -J HiCstep2 --time=24:00:00 --mem=121g --cpus-per-task=4
--gres=lscratch:200 HiCPro_step2_TestHiCpro.sh
```

Once complete, you can check the fragment size distribution by comparing bioanalyzer results across samples, and also by size distribution from the results of the sequencing as processed by HiC-Pro (Fig. 3). These results

can be found in the standard subdirectory: HiCpro_OUTPUT/hic_results/pic/Sample_**/ plotHiCFragmentSize_Sample_**.pdf.

65. Calculate AQuA factors and sample statistics

- HiC-Pro generates summary statistics files for each sample, including the total numbers of valid contacts, and these can be gathered together as human and mouse pairs.
- AQuA factors can be calculated from merged statistics (dataframe `mergestat.all` in example R code below; output as mergestat.HiChIP.all.txt), where the AQuA factor is the ratio of human:mouse unique and valid contact pairs for a given sample.

```
### Set up project and samples
setwd("K:/projects/HiC/projects/")
project.title = "RH4_H3K27ac_HiChIP"
project.folder = paste(project.title,"/HiCpro_OUTPUT/
hic_results/
data/",sep="")
sample.list = list.dirs(path = project.folder, full.names =
F, recursive = F)
project.title.mm10 = paste(project.title,"_mm10",sep="")
project.folder.mm10 = paste(project.title.mm10,"/
HiCpro_OUTPUT/
hic_results/data",sep="")
sample.list.mm10 = list.dirs(path = project.folder.mm10,
full.names
= F, recursive = F)
### Obtain human read counts and spike in mouse read counts
from HiC-pro
output
mergestat.all =
as.data.frame(read.table(paste(project.folder,"/",
sample.list[1],
"/",sample.list[1],"_allValidPairs.mergestat",sep=""),
sep="\t",
header=F))
mergestat.all = as.data.frame(mergestat.all$V1)
##load and merge hg19 sample data
lapply(sample.list, function(x) {
mergestat <- read.table(paste(project.folder,"/",x,"/",x,
"_allValidPairs.mergestat",sep=""), sep="\t", header=F)
mergestat.sample = as.data.frame(mergestat[,2])
removable.string = "Sample_" ; sample.name = gsub(removable.
```

```
string,"",x)
colnames(mergestat.sample) = c(sample.name)
mergestat.all <<- cbind(mergestat.all,mergestat.sample)
})
##load and merge mm10 sample data
lapply(sample.list.mm10, function(x) {
mergestat <-
read.table(paste(project.folder.mm10,"/",x,"/",x,
"_allValidPairs.mergestat",sep=""), sep="\t", header=F)
mergestat.sample = as.data.frame(mergestat[,2])
removable.string = "Sample_" ; sample.name = gsub(removable.
string,"mm10_",x)
colnames(mergestat.sample) = c(sample.name)
mergestat.all <<- cbind(mergestat.all,mergestat.sample)
})
write.table(mergestat.all,"mergestat.HiChIP.all.txt",sep="\t
",
col.names=T,row.names=F)
```

66. Convert HiC-Pro allValidPairs file type to the Juicebox compatible .hic file type. HiC-Pro comes equipped with a command to create a .hic filetype suitable for the Juicebox[31] and Juicer[27] tool box developed by the Aiden laboratory (http://aidenlab.org/software.html). These tools present one of the most advanced tool suites for visualizing Hi-C–type interaction data, as well as rapidly extracting data and also plotting data as APA plots.

Example code to run hicpro2juicebox.sh:

```
/hicpro/2.10.0/HiC-Pro_2.10.0/bin/utils/hicpro2juicebox.sh -i /
projects/
RH4_H3K27ac_HiChIP_hg19/HiCpro_OUTPUT/hic_results/data/
Sample_RH4_D6_
H3K27ac_HiChIP_HKJ22BGX7/Sample_RH4_D6_H3K27ac_HiChIP_HKJ22BGX7_all
ValidPairs -g hg19 -j /local/apps/juicer/juicer-1.5.6/
scripts/juicer_
tools.jar
```

67. Generate a matrix for downstream analysis (heatmaps and virtual 4C plots) of AQuA-HiChIP data at a locus of interest. The .hic file is a genome-wide interaction file. To import a region of interest into R (scripts designed for interactive R-studio sessions) to create AQuA-normalized plots (contact frequency heatmaps and virtual 4C normalized by the AQuA factor), first use the juicer ‘dump’ command to create a sparse matrix.

Extract a sparse formatted contact matrix for a region of interest from the .hic file using the juicer dump tool, as follows:

```
java -jar /juicer/juicer-1.5.6/scripts/juicer_tools.jar dump
observed
NONE /HiC/projects/RH4_H3K27ac_HiChIP/HiCpro_OUTPUT/hic_results/
data/
Sample_RH4_Ent6_H3K27ac_HiChIP_HKJ22BGX7/Sample_RH4_Ent6_H3K27ac_
HiChIP_HKJ22BGX7_allValidPairs.hic 11:17600000:17800000
11:17600000:
17800000 BP 5000 Sample_RH4_Ent6_H3K27ac_HiChIP_HKJ22BGX7.chr11.
176MB-178MB.5KB.matrix.txt
```

68. Once the sparse matrix is created, import into R-Studio and modify the following R code to match the output file name for the juicer dump tool. The AQuA factor calculations are performed from the `mergestat.all` data frame gathered from the HiC-Pro statistical output, generated in Step 65. Use HiCcompare[32] to convert from sparse to dense format (`sparMat` to `denMat`; see code below).

```
### Load and transform sparse matrix
source("https://bioconductor.org/
biocLite.R");biocLite("HiCcompare")
library(HiCcompare)
sample1 = "Sample_RH4_D6_H3K27ac_HiChIP_HKJ22BGX7"
sample2 = "Sample_RH4_Ent6_H3K27ac_HiChIP_HKJ22BGX7"
removable.string = "Sample_"
name1=gsub(removable.string,"",sample1)
name2=gsub(removable.string,"",sample2)
mm10_name1 = gsub(removable.string,"mm10_",sample1)
mm10_name2 = gsub(removable.string,"mm10_",sample2)
matrix.suffix = ".chr11.176MB-178MB.5KB.matrix.txt"
sparMat1 = read.table(paste(project.folder,sample1,"/",
sample1,matrix.suffix,sep=""),header=F, sep="\t")
sparMat2 = read.table(paste(project.folder,sample2,"/",
sample2,matrix.suffix,sep=""), header=F, sep="\t")
denMat1 <- sparse2full(sparMat1,
hic.table = FALSE, column.name = NA)
denMat2 <- sparse2full(sparMat2,
hic.table = FALSE, column.name = NA)
denMat1.CPM = denMat1*1000000/
(mergestat.all[2,name1]+mergestat.all
[2,mm10_name1])
denMat2.CPM = denMat2*1000000/
(mergestat.all[2,name2]+mergestat.all
[2,mm10_name2])
AQuAfactor1 = mergestat.all[2,name1]/mergestat.all[2,mm10_name1]
```

```
AQuAfactor2 = mergestat.all[2,name2]/mergestat.all[2,mm10_name2]
denAQuA1 = denMat1.CPM*AQuAfactor1
denAQuA2 = denMat2.CPM*AQuAfactor2
```

69. Visualize the AQuA-normalized matrices using pheatmap in R. To set the scale to a maximum contact matrix value, pheatmap requires a maximum be applied outside the generation of the heatmap; otherwise, out-of-scale values will be plotted as gray boxes. To maintain the structure and coordinates of the contact matrix, pheatmap clustering is turned off. An example output from the code below is shown in Fig. 4a.

```
### Build matrices with AQuA maximums, plot HEATMAPS
library(pheatmap)
quant_cut = 0.95 #caps the contact map plot values at a given
percentile
quantile(denAQuA1, probs = c(quant_cut))
quantile(denAQuA2, probs = c(quant_cut))
AQuAmax = max(quantile(denAQuA1, probs = c(quant_cut)),
quantile(denAQuA2, probs = c(quant_cut)))
denAQuA1max = denAQuA1; denAQuA1max[denAQuA1max>AQuAmax] <-
AQuAmax
denAQuA2max = denAQuA2; denAQuA2max[denAQuA2max>AQuAmax] <-
AQuAmax
### plot control and treated samples
pheatmap(((denAQuA1max)), cluster_rows = F, cluster_cols = F,
color = colorRampPalette(c("white", "red"))(50))
pheatmap(((denAQuA2max)), cluster_rows = F, cluster_cols = F,
color = colorRampPalette(c("white", "red"))(50))
### plot delta samples
denDeltaAQuA = denAQuA2 - denAQuA1
denDeltaAQuAmax = quantile(denDeltaAQuA, probs = c(quant_cut))
denDeltaAQuA[denDeltaAQuA>denDeltaAQuAmax] = denDeltaAQuAmax
pheatmap((denDeltaAQuA), cluster_rows=F, cluster_cols=F,
color = colorRampPalette(c("dodgerblue", "white",
"mediumvioletred"))
(33))
```

70. Virtual AQuA-4C (a 2D slice from HiChIP or HiC data, which shows the contacts from a single genomic location to the rest of the genome) can be extracted for a given matrix by isolating a single locus, at resolution equal to the binsize. Use the code below to generate splined bedgraphs of Virtual AQuA-4C at a single region of interest within the matrix. This generates interaction profiles that can be overlaid conveniently with 2D genome browser views (Fig. 4b).

```
### Extract a Virtual AQuA-4C viewpoint, make bedgraphs, plot
VIRTUAL4C
virt4C_viewpoint_chr = ‘chr11’; virt4C_viewpoint = ‘17670000’
df_AQuA1 = as.data.frame(denAQuA1, row.names = row.names(denAQuA1),
col.names=col.names(denAQuA1))
df_AQuA2 = as.data.frame(denAQuA2, row.names = row.names(denAQuA2),
col.names = col.names(denAQuA2))
virt4C.bedgraph1 = as.data.frame(df_AQuA1[,virt4C_viewpoint])
colnames(virt4C.bedgraph1) = "AQuA_contact_freq"
virt4C.bedgraph1$chr = virt4C_viewpoint_chr
virt4C.bedgraph1$start = as.numeric(rownames(df_AQuA1))
virt4C_binsize = virt4C.bedgraph1[2,c(‘start’)]-virt4C.bedgraph1[1,
c(‘start’)]
virt4C.bedgraph1$stop = virt4C.bedgraph1$start + virt4C_binsize
virt4C.bedgraph1 = virt4C.bedgraph1[,c(2,3,4,1)]
virt4C.bedgraph2 = as.data.frame(df_AQuA2[,virt4C_viewpoint])
colnames(virt4C.bedgraph2) = "AQuA_contact_freq"
virt4C.bedgraph2$chr = virt4C_viewpoint_chr
virt4C.bedgraph2$start = as.numeric(rownames(df_AQuA2))
virt4C_binsize = virt4C.bedgraph2[2,c(‘start’)]-virt4C.bedgraph2[1,
c(‘start’)]
virt4C.bedgraph2$stop = virt4C.bedgraph2$start + virt4C_binsize
virt4C.bedgraph2 = virt4C.bedgraph2[,c(2,3,4,1)]
window_coordinates=paste(virt4C_viewpoint_chr,":",
virt4C.bedgraph1[1,c(‘start’)],"-",max(virt4C.bedgraph1$stop),
sep="")
## spline to smooth
spline1 <- as.data.frame(spline(virt4C.bedgraph1$start,
virt4C.bedgraph1$AQuA_contact_freq))
spline2 <- as.data.frame(spline(virt4C.bedgraph2$start,
virt4C.bedgraph2$AQuA_contact_freq))
spline_delta = spline1; spline_delta$y = (spline2$y-spline1$y)
## plots
library(ggplot2)
ggplot(spline1, aes(x=(x+virt4C_binsize/2),y=y))+ theme_bw()+
geom_line(color = ‘red’)+ geom_line(data=spline_delta, color =
‘pur-
ple’)+
geom_vline(xintercept=as.numeric(virt4C_viewpoint))+
geom_vline(xintercept=(as.numeric(virt4C_viewpoint)
+virt4C_binsize))
```

71. APA is a convenient and rapid way to determine the enrichment of contacts across a series of paired coordinates (such as loop anchors). There is not a

mechanism to incorporate spike-in to the Juicebox APA map generator currently, but the non-normalized output APA matrix can be loaded into R and replotted after AQuA factor incorporation.

Run APA with the example code (command line):

```
juicer_tools apa -r 10000,5000 -k NONE -n 0 -u /HiC/projects/
RH4_H3K27ac_
HiChIP/HiCpro_OUTPUT/hic_results/data/
Sample_RH4_D6_H3K27ac_HiChIP_
HKJ22BGX7/Sample_RH4_D6_H3K27ac_HiChIP_HKJ22BGX7_allValidPairs.hic
RH4_SEtoSE.2D.txt RH4_D6_H3K27ac_SEtoSE_APA
```

where the *.2D.txt is a tab-delimited coordinate file with the format:

```
chr1x1x2chr2y1y2colorcomment
chr15275254950chr179244802460,0,0SEtoSE
```

Example code (R-studio) for plotting APA matrices after AQuA Factor normalization is given below.

```
### 1. Set up project and samples
project.folder = "VASE/"
sample.list = list.dirs(path=project.folder,full.names=F,
recursive=F)
sample.list = sample.list[grep("RH4",sample.list)]
### 2. Extract APA matrices
APAmatrix.scale = 10000
APA.all =
as.data.frame(read.table(paste(project.folder,sample.list[1],
"/",APAmatrix.scale,"/gw/APA.txt",sep=""),sep=",",header=F))
APA.all$comparison.name = sample.list[1] ; APA.all = APA.all[0,]
#initiate
## load and merge sample data
lapply(sample.list, function(x) {
APA = as.data.frame(read.table(paste(project.folder,x,"/",
APAmatrix.scale,"/gw/APA.txt",sep=""), sep=",", header=F))
APA$V1 = as.numeric(gsub("\\[|\\]", "", APA$V1)) #cleanup juicer
output
APA$V21 = as.numeric(gsub("\\[|\\]", "", APA$V21)) #cleanup juicer
output
APA$comparison.name = x
APA.all <<- rbind(APA.all,APA)
```

```
})
APA.DMSO = APA.all[grep("D6", APA.all$comparison.name),] #subset
control
APA.Ent6 = APA.all[grep("Ent6", APA.all$comparison.name),] #subset
treat-
ment
```

It detects all directories (one per APA output), which means that multiple region sets can be calculated by APA (repeat for both control and treatment samples for each 2D coordinate set to compare).

72. Once the APA matrices are built for each condition, perform AQuA factor adjustments and then plot individual APA sets. An example APA plot, comparing raw and AQuA-normalized APA plots, is shown in Fig. 4c.

Timing

Steps 1–9, grow, treat, harvest, count and fix the cells: 1–2 d

Steps 10–22, nuclei isolation, DNA digestion, DNA-biotinylation and proximity ligation: 8–10 h

Steps 23–38, shearing, ChIP, washing, elution and reversal of crosslinks: 24–28 h

Steps 39–49, biotin capture, end repair, A-tailing and linker ligation: 4 h

Steps 50–57, amplify on-bead to produce sequencing libraries and purify: 2–4 h

Steps 58–60, multiplex samples and sequence 200 million PETs: 1 d

Steps 61–72, bioinformatic analysis: 2 d to 1 week

Troubleshooting

Troubleshooting advice can be found in Table 2.

Anticipated results

HiChIP experiments are expected to yield improved signal-to-noise ratios over Hi-C and also contain a higher percentage of informative contacts (unique PETs)[12]. AQuA-HiChIP sequenced at a depth of 100 million paired-end reads should contain ~70–80 million mapped read pairs to the human genome and 20–30 million to the mouse genome. In a successful experiment, between 20% and 40% of reads will be duplicates, presumed as artifacts from PCR and discarded. Then, low-quality mapping reads and pairs lacking a predicted digestion site will be discarded. Of the human reads, one should anticipate ~25–40 million unique PETs (valid interaction pairs) to pass all quality-filtering steps, and unlike Hi-C or chromatin interaction analysis by paired-end tag sequencing, HiChIP contains a high percentage of *cis* (not *trans*) interactions (Fig. 5a). An absolute increase in contacts may be expected to also create a more diverse and higher starting quantity pool of biotinylated DNA fragments,

which would result in effectively lowering the duplicate read count and increasing the proportion of valid human reads (as seen for H3K27ac AQuA-HiChIP under HDAC inhibition conditions; Fig. 5b). The relative amount of PETs for human and mouse should trend with the expected total protein-mediated contacts in the treated human sample: for a degron[20] experiment with 80% reduction of the IP target, one would expect an increase from 10% mouse PETs to upwards of 50%; for experiments with an upregulation of a protein target by twofold, one would expect a drop in mouse PETs from 10% to 5% (similar to the decrease seen in Fig. 5c). If the same experimental conditions also impact global 3D contact frequencies among all contacts regardless of mediating IP target proteins, one might expect a change in the ratio of human:mouse PETs in the input Hi-C DNA, although we expect the ratio of human:mouse reads (and the AQuA factor) to be less variable than experiments involving a ChIP step (Fig. 5d,e) because Hi-C data have much more background data from close-range PETs than HiChIP. A successful experiment should produce high-quality enrichment (>30-fold enrichment of contacts over interactions expected from random polymer movements, a calculation provided automatically in Juicebox[31]), as assessed by eye in a Hi-C browser such as Juicebox[31] (https://www.aidenlab.org/juicebox/).

Preparing libraries from HiChIP experiments involves amplification that can introduce duplicate reads. We found, however, that variability in the AQuA normalization factor was not altered by duplication. Among two technical replicate library preparations from the same ChIP eluent DNA, the second replicate had greatly increased duplication rates but did not suffer substantial changes in AQuA normalization factors (Box 3). We also note that the percentage of *cis* and *trans* contacts was invariant for the spiked-in mouse chromatin across all conditions and replicates (88–89% *cis* contacts; Box 3), increasing confidence.

For a given experiment where one protein or histone mark of interest is globally increasing or decreasing (i.e., histone acetylation increasing after treatment with an HDAC inhibitor), many other targets are not expected to change globally (i.e., promoter-associated histone H3K4 trimethylation after treatment with an HDAC inhibitor). To ensure that in such cases the AQuA normalization method did not introduce spurious artifacts, we performed AQuA-HiChIP for H3K27ac and H3K4me3 from human RH4 cells treated with DMSO or the HDAC inhibitor entinostat for 6 h. Mouse cells (equal numbers of fixed NIH3T3 cells) were spiked in to each condition as described in the protocol. After performing AQuA-HiChIP, we normalized the data both as normal HiChIP and with the AQuA factor (Fig. 6). Whereas the global increase in H3K27ac was not visible without AQuA normalization, we found that AQuA normalization did not greatly change the interaction frequencies with H3K4me3 HiChIP (Fig. 6). H3K4me3 was not directly modulated by the drug used (entinostat) and had similar human-to-mouse PET ratios and AQuA factors across samples, and overall a small AQuA impact ratio (1.2×; Table 3). In contrast, the histone mark globally modulated by entinostat (H3K27ac) had an increase in percentage of human reads, more than double the AQuA factor and a strong AQuA impact ratio (2.7×; Table 3). For regions with virtually no interaction signal, multiplying by the AQuA factor did not introduce false signal ($0 \times 15 = 0$). In addition, while the AQuA normalization reveals critical changes at sites most enriched for H3K27ac (super-enhancer peak pairs), this same normalization does not influence the signal at repressed H3K27me3-rich domains (Fig. 7).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We gratefully acknowledge Y. Song and the NCI Sequencing Core for assistance. We are grateful to Tom Misteli and Justin Demmerle for helpful conversations relating to this work. We thank Emma Chory for technical advice. We wish to honor the lasting memory of Joseph P. Calarco. This work was facilitated by Biowulf High Performance Computing Systems and enabled by funding from the Division of Intramural Research from NIH NCI CCR.

Data availability

Data generated by this protocol, and visualized herein, is available through Gene Expression Omnibus, accession number GSE120770.

References

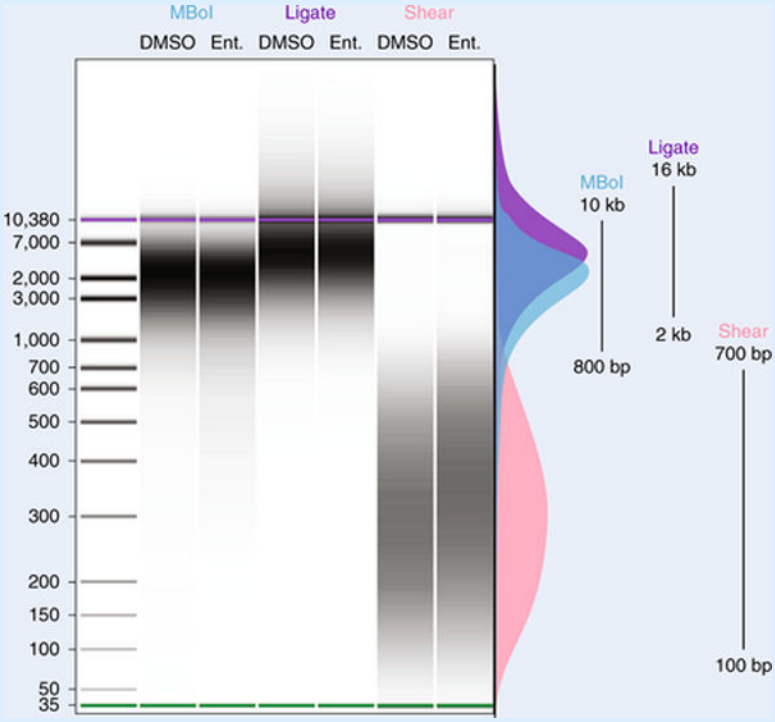
1. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 409, 860 (2001). [PubMed: 11237011]
2. Roh T.-y, Ngau WC, Cui K, Landsman D & Zhao K High-resolution genome-wide mapping of histone modifications. Nat. Biotechnol 22, 1013 (2004). [PubMed: 15235610]
3. Roh T-Y, Cuddapah S & Zhao K Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes Dev 19, 542–552 (2005). [PubMed: 15706033]
4. Bernstein BE et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125, 315–326 (2006). [PubMed: 16630819]
5. Barski A et al. High-resolution profiling of histone methylations in the human genome. Cell 129, 823–837 (2007). [PubMed: 17512414]
6. ENCODE Project Consortium. The ENCODE (ENCyclopedia of DNA elements) project. Science 306, 636–640 (2004). [PubMed: 15499007]
7. Birney E et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799–816 (2007). [PubMed: 17571346]
8. Pennisi E ENCODE Project writes eulogy for junk DNA. Science 337, 1159–1161 (2012). [PubMed: 22955811]
9. Lieberman-Aiden E et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289–293 (2009). [PubMed: 19815776]
10. Fullwood MJ et al. An oestrogen-receptor-α-bound human chromatin interactome. Nature 462, 58 (2009). [PubMed: 19890323]
11. Rao SS et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680 (2014). [PubMed: 25497547]
12. Mumbach MR et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat. Methods 13, 919–922 (2016). [PubMed: 27643841]
13. Mumbach MR et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat. Genet 49, 1602 (2017). [PubMed: 28945252]
14. Goldberg AD, Allis CD & Bernstein E Epigenetics: a landscape takes shape. Cell 128, 635–638 (2007). [PubMed: 17320500]
15. Orlando DA et al. Quantitative ChIP-seq normalization reveals global modulation of the epigenome. Cell Rep 9, 1163–1170 (2014). [PubMed: 25437568]
16. Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330 (2015). [PubMed: 25693563]
17. Lai B et al. Trac-looping measures genome structure and chromatin accessibility. Nat. Methods 15, 741–747 (2018). [PubMed: 30150754]

18. Nishimura K, Fukagawa T, Takisawa H, Kakimoto T & Kanemaki M An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat. Methods 6, 917 (2009). [PubMed: 19915560]
19. Tan X et al. Mechanism of auxin perception by the TIR1 ubiquitin ligase. Nature 446, 640 (2007). [PubMed: 17410169]
20. Nabet B et al. The dTAG system for immediate and target-specific protein degradation. Nat. Chem. Biol 14, 431 (2018). [PubMed: 29581585]
21. Pogo B, Allfrey V & Mirsky A RNA synthesis and histone acetylation during the course of gene activation in lymphocytes. Proc. Natl Acad. Sci. USA 55, 805–812 (1966). [PubMed: 5219687]
22. Gryder BE et al. Histone hyperacetylation disrupts core gene regulatory architecture in rhabdomyosarcoma. Nat. Genet 51, 1714–1722 (2019). [PubMed: 31784732]
23. Yohe ME et al. MEK inhibition induces MYOG and remodels super-enhancers in RAS-driven rhabdomyosarcoma. Sci. Transl. Med 10, eaan4470 (2018). [PubMed: 29973406]
24. Servant N et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol 16, 259 (2015). [PubMed: 26619908]
25. Gryder BE et al. PAX3-FOXO1 establishes myogenic super enhancers and confers BET bromodomain vulnerability. Cancer Disco 7, 884–899 (2017).
26. Stanton BZ et al. Smarca4 ATPase mutations disrupt direct eviction of PRC1 from chromatin. Nat. Genet 49, 282–288 (2017). [PubMed: 27941795]
27. Durand NC et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst 3, 95–98 (2016). [PubMed: 27467249]
28. Langmead B, Trapnell C, Pop M & Salzberg S Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10, R25 (2009). [PubMed: 19261174]
29. Zhang Y et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol 9, R137 (2008). [PubMed: 18798982]
30. Heinz S et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589 (2010). [PubMed: 20513432]
31. Durand NC et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst 3, 99–101 (2016). [PubMed: 27467250]
32. Stansfield JC, Cresswell KG, Vladimirov VI & Dozmorov MG HiCcompare: an R-package for joint normalization and comparison of HI-C datasets. BMC Bioinforma 19, 279 (2018).

Box 1 |

Quality control for digestion, ligation and shearing

Small amounts of sample can be extracted throughout the protocol at key steps, including MBoI digestion, ligation with T4 DNA ligase and after shearing chromatin before the ChIP step. AQuA-HiChIP experiments were performed in RH4 cells treated with DMSO or Entinostat (Ent.) and were monitored using an Agilent Bioanalyzer to reveal DNA fragment size distribution changes upon digestion (800 bp to 10 kb range), ligation (2–16 kb) and shearing (100–700 bp). Between treatment conditions, the DNA showed very little difference in distribution at steps prior to chromatin immunoprecipitation.



Box 2 |

Illumina's multiplexing indexes

Index sequence

M1 ATCACG

M2 CGATGT

M3 TTAGGC

M4 TGACCA

M5 ACAGTG

M6 GCCAAT

M7 CAGATC

M8 ACTTGA

M9 GATCAG

M10 TAGCTT

M11 GGCTAC

M12 CTTGTA

Illumina's multiplexing primer sequences (* denotes phosphorothioate)

Multiplexing PCR FWD Primer:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T

(P5 sequence is underlined)

Use REV Primers M1–12, with sequencing barcodes for indexed library amplification (begins with P7 sequence, underlined in first primer M1, and the reverse complimented index sequence is in italics).

Primer M1

CAAGCAGAAGACGGCATACGAGAT*CGTGAT*GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T

Primer M2

CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T

Primer M3

CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T

Primer M4

CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGC TCTTCCGATC*T

Primer M5

CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGC TCTTCCGATC*T

Primer M6

CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGC TCTTCCGATC*T

Primer M7

CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGC TCTTCCGATC*T

Primer M8

CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGC TCTTCCGATC*T

Primer M9

CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGC TCTTCCGATC*T

Primer M10

CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGC TCTTCCGATC*T

Primer M11

CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGC TCTTCCGATC*T

Primer M12

CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTGC TCTTCCGATC*T

Box 3 |

AQuA-HiChIP normalization is robust against excessive library duplication

The table displays the number of valid contacts (numbers are ×1,000), non-duplicate contacts, percentage of *cis* and *trans* contacts, percentage of human and mouse and AQuA factors from technical replicate sample preparations of H3K27ac AQuA-HiChIP. The technical replicates have the same cell preparation and the same ChIP experiment, but biotin capture and library preparation were performed twice (a month apart) from the same ChIP eluent DNA. The second library preparation had less diverse fragments and more duplicated contacts; yet, this amplification-induced duplication did not change the AQuA factors.

	Replicate 1				Replicate 2			
	Human		Mouse		Human		Mouse	
	DMSO	Ent.	DMSO	Ent.	DMSO	Ent.	DMSO	Ent.
Valid contacts (×1,000)	45,470	46,919	8,177	3,200	58,700	81,636	10,397	5,460
Valid contacts, non-duplicate	25,367	36,367	4,580	2,479	18,953	38,895	3,358	2,550
Trans contacts	3,865	6,917	525	292	2,986	7,660	404	315
Cis contacts	21,502	29,449	4,056	2,187	15,967	31,235	2,954	2,234
% non-duplicate	56%	78%	56%	77%	32%	48%	32%	47%
% trans	15%	19%	11%	12%	16%	20%	12%	12%
% cis	85%	81%	89%	88%	84%	80%	88%	88%
% human or mouse	84.8%	93.6%	15.2%	6.4%	85.0%	93.7%	15.0%	6.3%

	Replicate 1		Replicate 2	
	DMSO	Ent.	DMSO	Ent.
AQuA factor	5.6	14.7	5.6	15.0

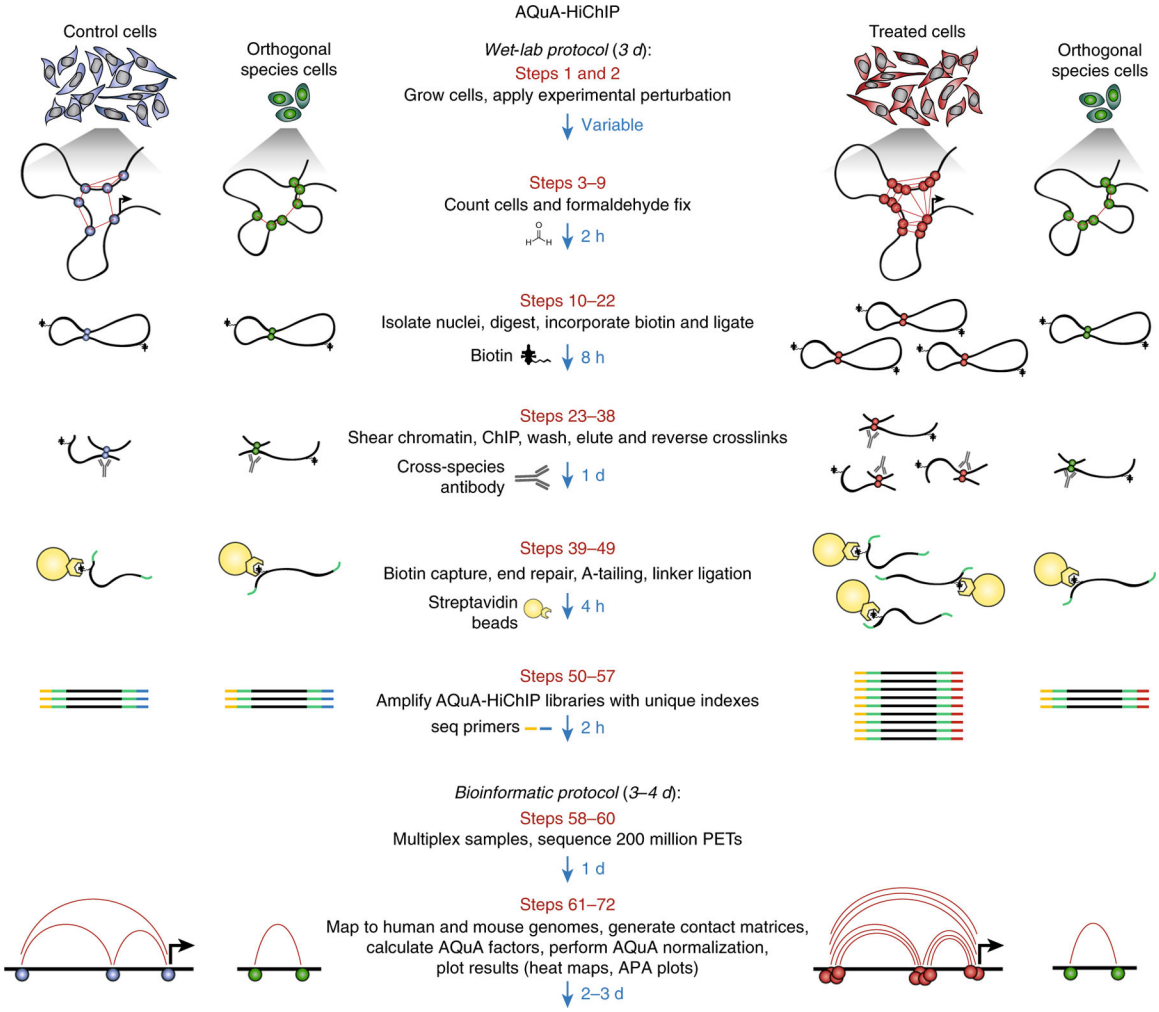


Fig. 1 |. AQuA-HiChIP experimental schematic overview.

AQuA-HiChIP performed on carefully counted cells from different treatment conditions, with orthogonal species cells spiked in prior to in situ crosslinking, ChIP, biotin capture and library preparation. Protein-mediated chromatin interactions induced by the treatment are readily differentiated using AQuA-HiChIP, whereas normalizing the control and treated sample by total PETs may not reveal differential interaction frequencies between the two conditions.

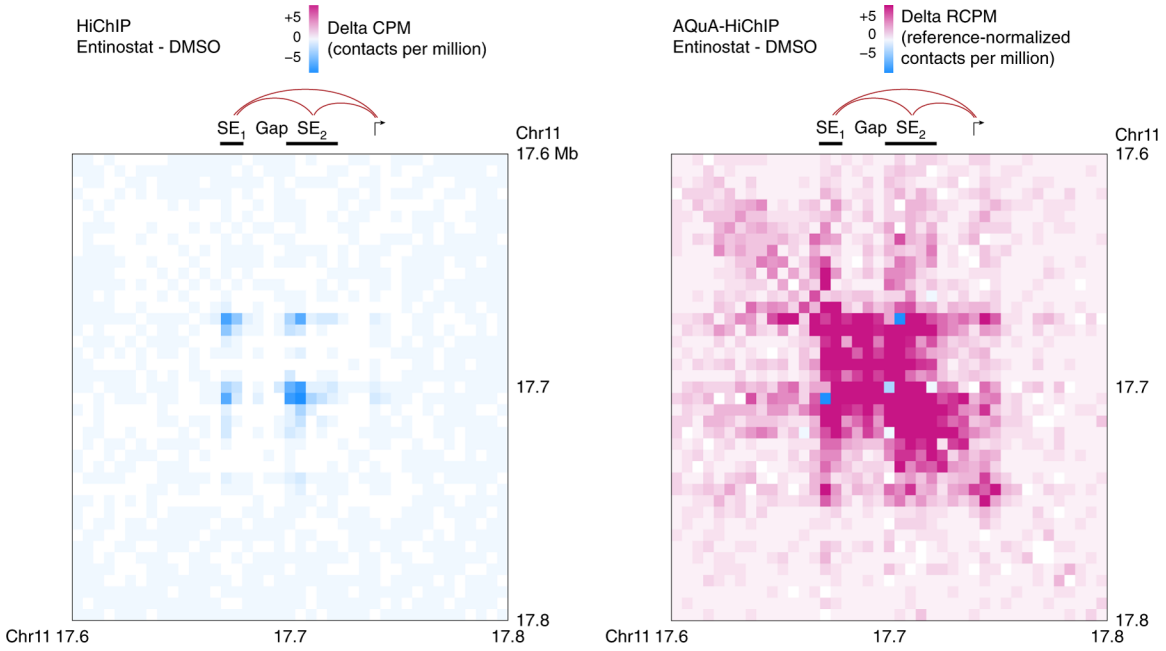


Fig. 2 |. Comparison of HiChIP and AQuA-HiChIP between treated and untreated cells reveals the difference between apparent and absolute contact changes.

Direct comparison of AQuA-HiChIP and non-reference normalized HiChIP samples indicates changes in apparent chromatin interactions anchored in H3-acetylation marks after rapid HDAC inhibition with 1 μM entinostat (subtracting the matrix on contacts found in cells treated with entinostat minus DMSO). These changes in architecture are focused near super enhancers (SEs) proximal to a TSS on chromosome 11 (chr11). Data were generated from RH4, a rhabdomyosarcoma cell line. Maps are binned at 5-kb resolution. SE1 and SE2 are the first and second super enhancers found near the *MYOD1* gene in RH4 cells (defined by H3K27ac ChIP-seq). Red lines are heuristic indicators of the major native connections.

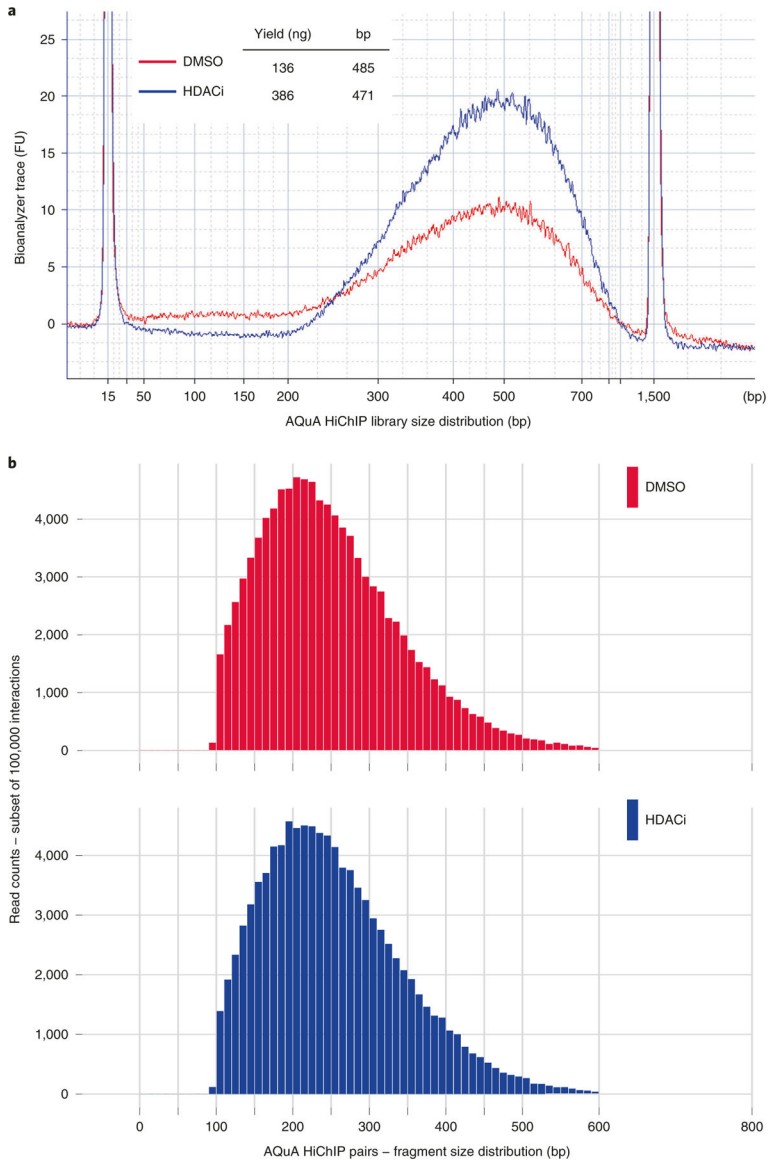


Fig. 3 |. AQuA-HiChIP size distribution.

a, Size of libraries, measured using Agilent Bioanalyzer. **b**, Size distribution of read-pairs after sequencing, mapping to hg19 and filtering for correct restriction digest at contact junction. HDACi, histone deacetylase inhibitor (1 μM entinostat).

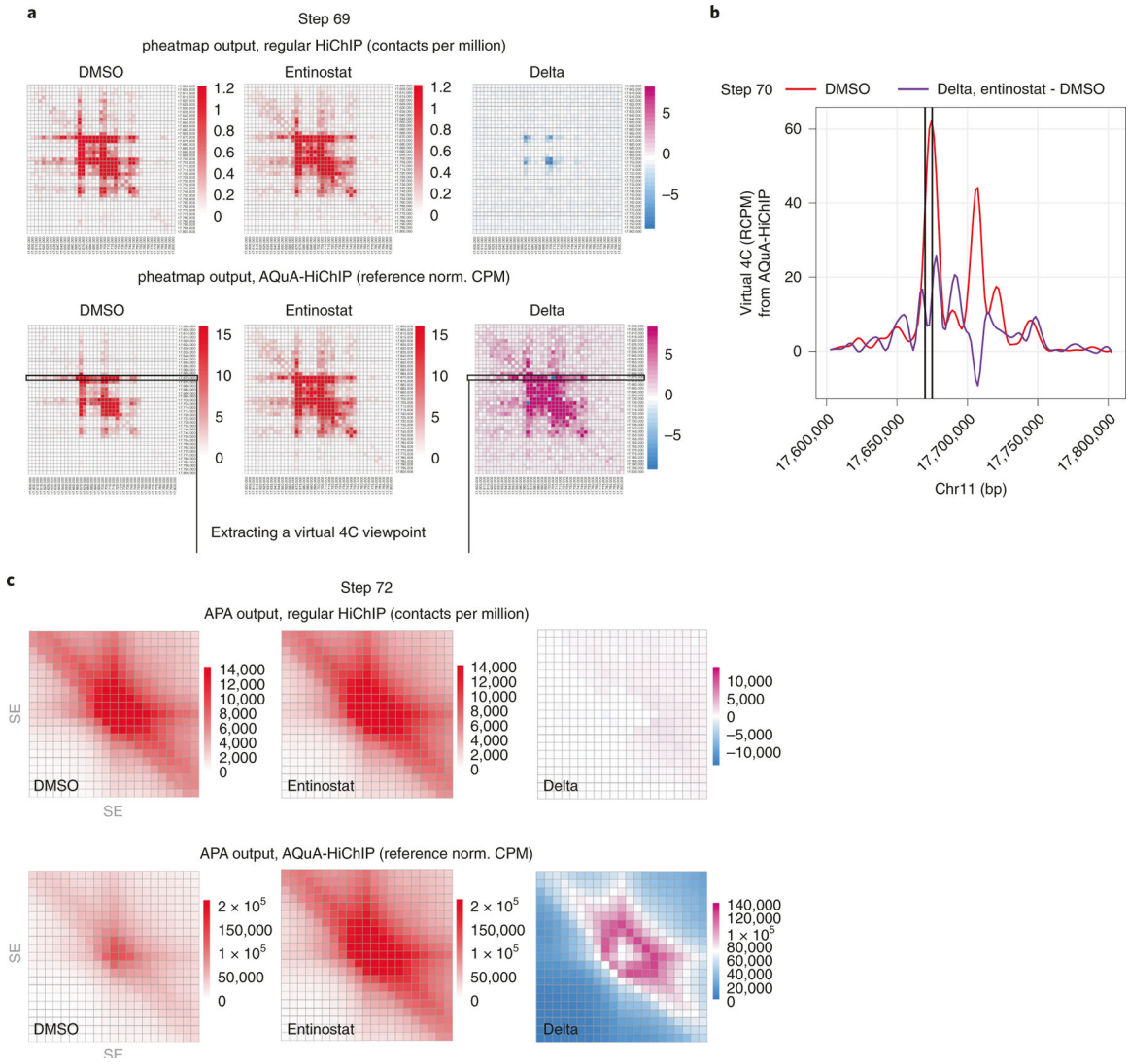


Fig. 4 |. Bioinformatic analysis example plots.

a, Matrix plots from pheatmap (in R Studio) of HiChIP and AQuA-HiChIP data as resulting from code in Step 69. **b**, Virtual 4C plot from ggplot2 (used from R Studio) as resulting from code in Step 70. Lines to the matrix in a from which the virtual 4C data are extracted are shown in black. **c**, APA plots from SE pairs and plots made in pheatmap (used from R Studio) as described in code from Step 72. Norm., normalized.

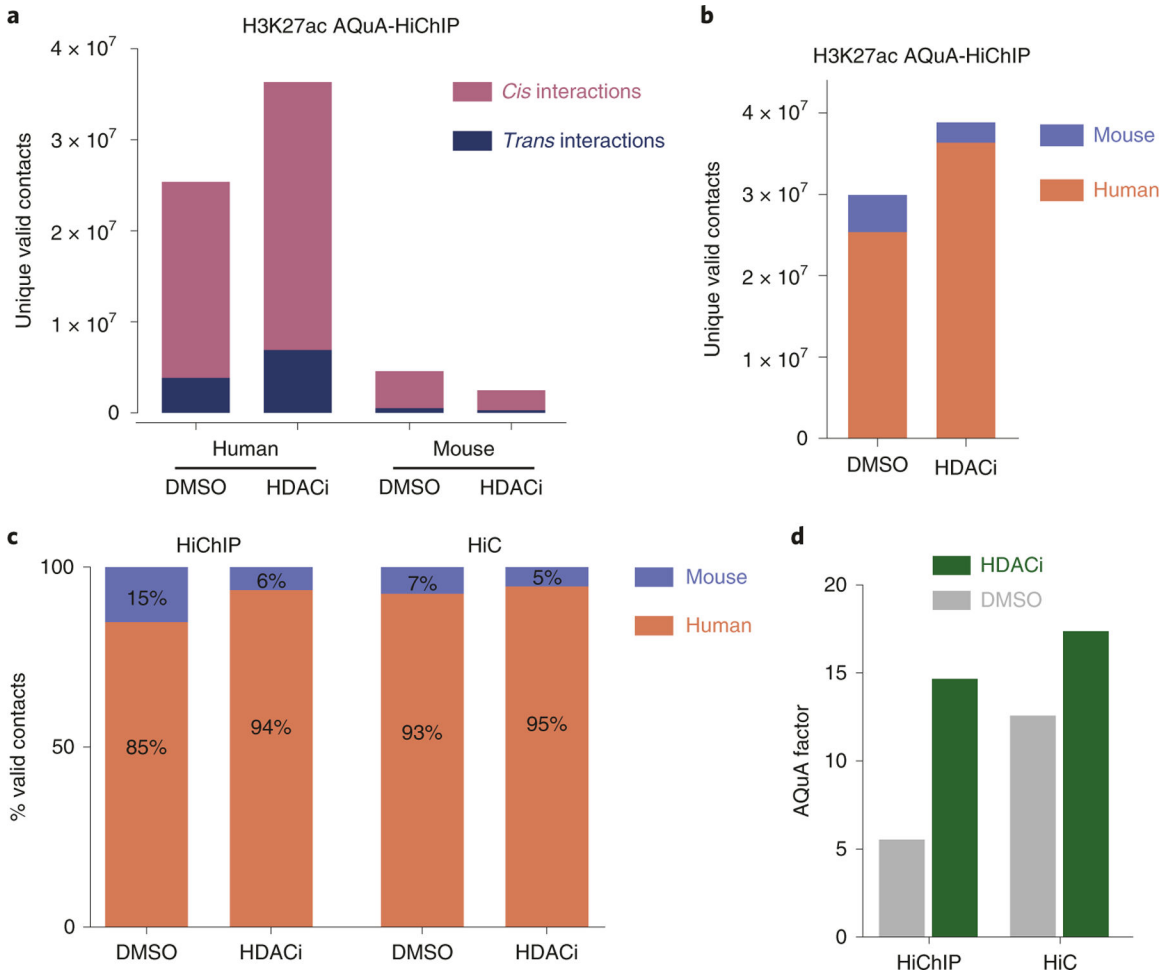


Fig. 5 |. AQuA-HiChIP contact frequency calculations.

a, Valid H3K27ac-mediated contacts (unique: duplicated read-pairs removed) were tabulated after separate mapping of the same fastq file pairs to the human (hg19, from RH4 cells) and mouse (mm10, from NIH3T3 cells) genomes. *trans* (multi-chromosome) and *cis* (intra-chromasome) contact frequencies among samples (treated with DMSO or 1 μM entinostat) and across species are presented. **b**, Direct comparison of mouse to human contacts in DMSO- and HDACi (1 μM entinostat)-treated HiChIP experiments. **c**, Percentage of mouse and human contacts compared across AQuA-HiChIP and AQuA-input (HiC, material processed from the same cells but without a ChIP step). **d**, AQuA Factor for AQuA-HiChIP and AQuA-input (HiC), both DMSO and HDACi (1 μM entinostat) treated. Results from a single experiment are shown.

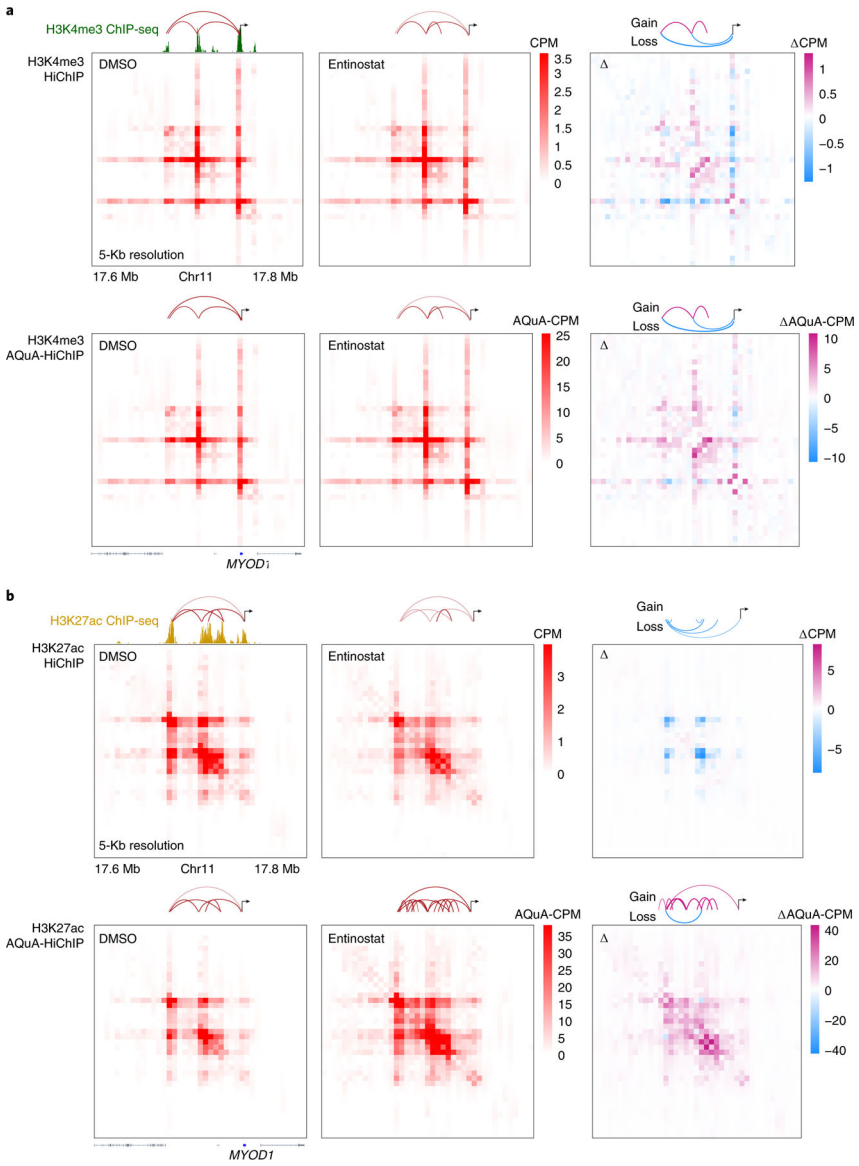


Fig. 6 |. AQuA-HiChIP reveals target-dependent architectural changes.

a, Comparison of HiChIP and AQuA-HiChIP for H3K4me3 in RH4 cancer cells, treated with either DMSO or HDACi (1 μM entinostat) for 6 h. ChIP-seq of the respective marks in RH4 cells is shown overlapping the HiChIP data[25]. Maps are shown at 5-kb resolution. CPM, contacts per million; AQuA-CPM, absolute quantification of architecture (spike-in reference-normalized) contacts per million. **b**, Same as **a** but for H3K27ac.

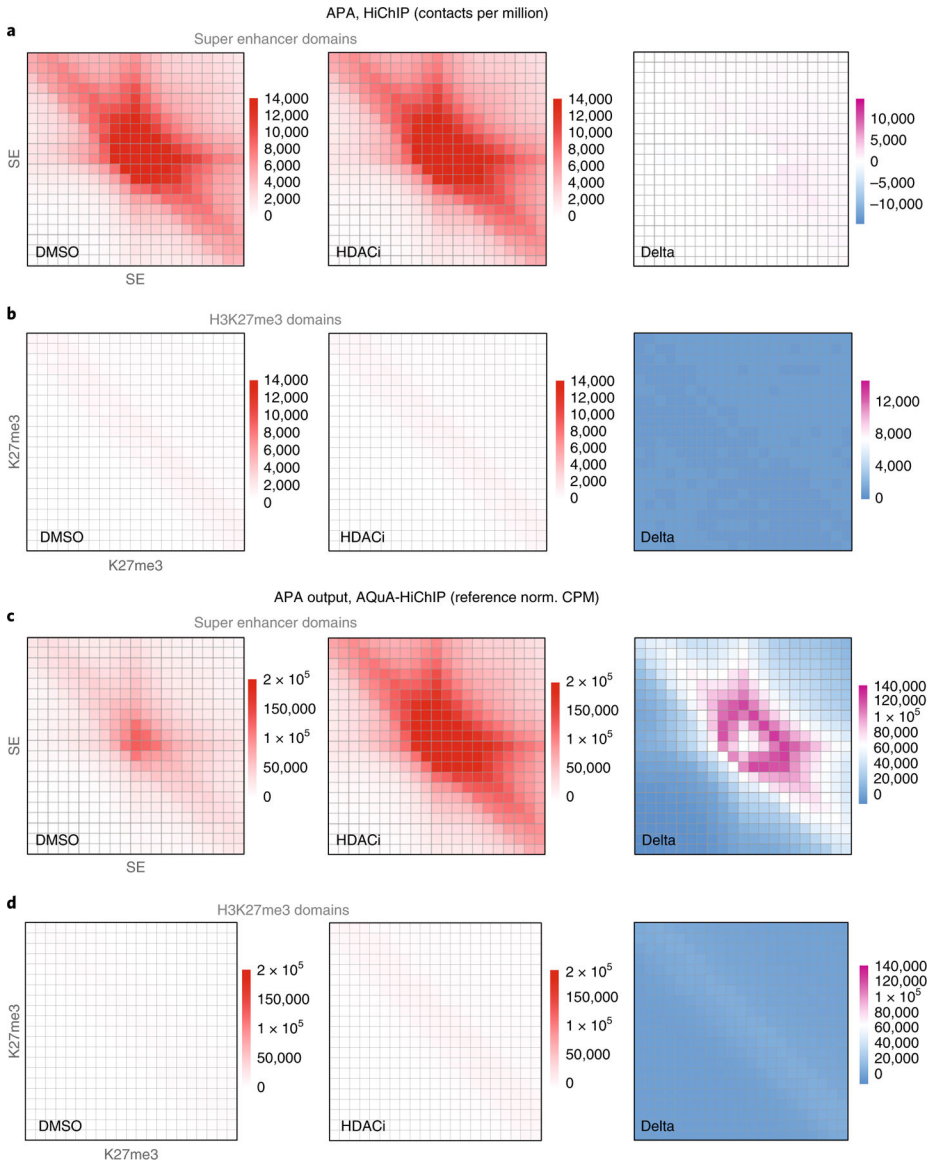


Fig. 7 |. Global AQuA normalization does not influence H3K27ac HiChIP contact frequencies at H3K27me3 domains.

a, APA plot at SE domains of HiChIP for H3K27ac in RH4 cells (treated with DMSO or HDACi entinostat) and the delta (matrix obtained by subtracting ‘treated’ from ‘untreated’ matrix). Boxes are 10 kb, with total APA area covering +/− 100 kb surrounding the center of each pair of peaks. **b**, APA plot at H3K27me3 domains of HiChIP for H3K27ac in RH4 cells (DMSO of HDACi entinostat) and the delta CPM. **c**, Same as **a** but with AQuA normalization; heatmaps plot the reference-normalized CPM. **d**, Same as **b** but with AQuA normalization.

Table 1 |

Oligonucleotide sequences

Name	Sequence (5′–3′)	Purpose
Primer M1	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	Index primer
Primer M3	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	Index primer
Primer M4	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	Index primer
Primer M5	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	Index primer
Primer M6	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	Index primer
Primer M7	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	Index primer
Primer M8	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	Index primer
Primer M9	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	Index primer
Primer M10	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	Index primer
Primer M11	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	Index primer
Primer M12	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	Index primer
Multiplexing adapter-top	5’ Phosphorylation-GATCGGAAGAGCACACGTCT	Sequencing adaptor
Multiplexing adapter-bottom	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	Sequencing adaptor

*T denotes phosphorothioate.

Table 2 |

Troubleshooting table

Step	Problem	Possible reason	Solution
24	Loss of efficiency in the ChIP, as detected by overall loss of signal over background	Over-sheared chromatin may result in elimination of epitopes	Perform reverse crosslinking after chromatin fragmentation (Step 24). Check that chromatin fragments range between 200 and 2,000 bp. Prepare samples again from Step 10 and shear with less cycles
	Loss of efficiency in the ChIP; overall loss of signal.	Under-sheared chromatin may result in inefficient pulldown	Perform reverse crosslinking after chromatin fragmentation (Step 24). Check for target shearing from 200 to 2,000 bp. Add additional sonication cycles if necessary
27	Loss of signal enrichment over background	Performing ChIPs with SDS concentrations >0.2%	Dilute ChIP reactions in Step 27 to a final concentration of SDS of ~0.1–0.2%. Alternatively, reduce SDS from 1% to 0.1% in the Nuclear Lysis Buffer used in Step 23. This allows ChIP reactions to be performed at higher concentrations of chromatin in a single tube, eliminating the need for dilution to 4 ml and four separate tubes to achieve <0.2% SDS
29	Some loss of sample during ChIP washes	Can occur if Dynabeads are retained in tube caps during or after washes	Briefly rotate tubes while on the magnet so that the caps are oriented toward the floor and repeat several times before removing wash buffer
52	AQuA-HiChIP libraries are saturated or have low complexity after sequencing	PCR over-amplification: E-Gel indicates primer signal is no longer present after on-bead PCR reaction(s). This indicates that AQuA-HiChIP libraries may be saturated. It is optimal if strong/visible bands can be seen at 300–800 bp, but unreacted primer signal is still detectable	Split the post-ligation sample (Step 49) in half before PCR amplification. It is better to try the amplification reaction again with the other half of the sample and check that some residual primer signal is present still after amplification
63	HiC-Pro pipeline fails to convert fastq to BAMs (aligned reads file) and make contact maps	Reference genome fasta files are incomplete or do not share the same name	Unify the name of the bowtie2 indexes, the genome name and the restriction-digested genome. For instance: REFERENCE_GENOME = ucsc.hg19, FASTA file in bowtie2 folder = ucsc.hg19.fasta and GENOME_FRAGMENT = /reference_files/MboI.ucsc.hg19. bed To build bowtie2 index files from a genome fasta file, see: http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml#the-bowtie2-build-indexer
	HiC-Pro pipeline fails to convert fastq to BAMs and make contact maps	Configuration file(s) incomplete or in error	(1) The restriction-digested BED (genomic coordinates) file must be present and must match the restriction digestion enzyme employed in Step 16 (2) The chromosome size file must be in correct format and present in a location properly referred to by the configuration file (3) The configuration file must follow the instructions here: http://nservant.github.io/HiC-Pro/MANUAL.html#setting-the-configuration-file

Table 3 |

AQuA factor and AQuA impact ratio

	H3K4me3		H3K27ac	
Condition	**DMSO**	**Entinostat**	**DMSO**	**Entinostat**
Human contacts, %	86	88	85	94
Mouse contacts, %	14	12	15	6
AQuA factor	6	7	6	15
AQuA impact ratio (Entinostat/DMSO)	1.2	—	2.7	

In summary, when a HiChIP target does not change globally, AQuA-ChIP is just as reliable as HiChIP, but when a HiChIP target does change globally, AQuA-HiChIP provides the internal normalization needed to properly interpret the data and see the correct changes.