bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

CaptureCompendium: a comprehensive toolkit for 3C analysis.

Jelena M. Telenius[1,*], Damien J. Downes[2,*], Martin Sergeant[1,*], A. Marieke Oudelaar[1,2], Simon McGowan[1], Jon Kerry[1], Lars L.P. Hanssen[2], Ron Schwessinger[1], Chris Q. Eijsbouts[3,4], James O.J. Davies[2], Stephen Taylor[1], Jim R. Hughes[1,2,†].

* These authors contributed equally: Jelena M. Telenius, Damien J. Downes, Martin Sergeant.

† Corresponding author: jim.hughes@imm.ox.ac.uk

1. MRC WIMM Centre for Computational Biology, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK
2. MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK
3. Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK
4. Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK

Keywords: 3C, chromosome conformation capture, Capture-C, software.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ABSTRACT

DNA folding within nuclei is a highly ordered process, with implications for gene regulation and development. An array of chromosome conformation capture (3C) methods have been developed to investigate how DNA is packaged within nuclei and to interrogate specific interactions. While these methods use different approaches to examine target loci (many-versus-all) or the entire genome (all-versus-all), they all rely on the core principle of endonuclease digestion and proximity-based ligation to re-arrange genomic order to reflect the three-dimensional nuclear conformation. This sequence reorganization creates novel chimeric DNA fragments which require specialist bioinformatic tools to analyze and visualize. Despite this need for specialist bioinformatic skills, the core biological importance of genome folding has seen widespread methodological uptake. To service the needs of experimentalists using the many-versus-all Capture-C family of methods we have developed CaptureCompendium; a toolkit of software to simplify the design, analysis and presentation of 3C experiments.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

The arrangement of the genome into the nucleus is highly regulated to facilitate correct interactions between promoters and their requisite enhancers. Chromosome conformation capture (3C) has emerged as a highly malleable method for studying both local and global trends in genome folding[1]. All 3C methods rely on the underlying principle that following DNA digestion, generally with a restriction endonuclease (RE), fragment ends in proximity will ligate together at high frequency. This digestion and proximity driven ligation functions to re-organize the linear genome sequence to reflect its 3D structure[2]. This re-organized sequence is then examined primarily by targeted high-resolution many-versus-all approaches, e.g. NG Capture-C[3], 4C-seq[4], UMI-4C[5], i4C-seq[6], and Tri-C [7]; or lower resolution approaches across the genome, e.g. HiC[8] and Capture-HiC[9].

The importance of genome folding to transcriptional regulation has brought about the application of 3C methods to investigate a range of questions, including basic biological problems and the identification of disease relevant gene-to-polymorphism interactions[10,11]. Owing to the enhanced signal provided by targeted sequencing, oligonucleotide capture methods, based on Capture-C[3,7,9,12–14], are emerging as a preferred method to look at loci of interest. A standard Capture-C experiment consists of three main bioinformatic components: target selection and oligonucleotide probe design; sequence mapping with read classification; and statistical comparison with data presentation. For each of these processes an array of software tools is available, including HiCapTools[15], GOPHER[16], Capsequm[12], capC-MAP[17], FourCSeq[18], peakC[19], CHiCAGO[20], peaky[21], ChiCMaxima[22], HiC-Pro[23], and 4See[24]; yet, to date no comprehensive pipeline for the complete design, mapping and interpretation of 3C experiments has been made available. This lack of an easily implemented, automated workflow for Capture-C based experiments forms a barrier to the general use of such methods, particularly in non-specialist groups. To address this need we have developed the CaptureCompendium toolkit (Fig. 1) which combines oligonucleotide design (Capsequm2), sequence mapping and data extraction (CCseqBasic), and statistical analyses with data presentation and distribution (CaptureCompare, Tri-C, CaptureSee). This workflow streamlines the bioinformatic process required to generate high quality 3C data using Capture-C experiments.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

RESULTS

Capsequm2: A user-friendly interface for probe design.

Targeted enrichment with biotinylated DNA oligonucleotides forms the basic tenet of Capture-C methods[3]. Probe design is essential to a successful experiment; probes must be situated near to a RE cut site, be proximal to the object of interest, and not be located on a highly repetitive element. These parameters allow for enrichment of informative on-target chimeric sequencing fragments, while simultaneously avoiding off-target sequencing. Currently, we know of three tools are available for probe design: HiCapTools - a command line-based method[15], GOPHER – an installable JAVA package[16], and Capsequm – an online interface[12]. As a webtool, Capsequm provides the most user-friendly option for initial probe design but previously required integration with MIG[25], a registration-based service, for probe filtering.

We redeveloped Capsequm (http://apps.molbiol.ox.ac.uk/CaptureC/cgi-bin/CapSequm.cgi) with a web-based filtering portal (AltSort) to generate Capsequm2 (Fig. 2). Capsequm2 allows researchers to select from a range of 11 genomes covering seven species and the five commonly used RE, as well as allowing a choice in probe length for design to targets of interest. Following fragment identification and BLAT processing, probes can be filtered using AltSort. In AltSort probes filtering are automatically categorised as good or bad using recommended settings for duplication number, BLAT density score (measuring likely off-target capture) (Kent 2002), G/C content, CpG content and, repeat content (Table 1). Probes can be visually inspected with integrated UCSC browser tracks, and the default parameters adjusted for all or specific regions of interest. Upon selection of desired probes, users can then download the oligonucleotide sequences, ready for ordering. This download also provides correctly formatted viewpoint input files required for subsequent analysis with the downstream tools CCseqBasic and CaptureCompare.

Table 1. Default AltSort parameters for probe design.

Filter	"Good"	"Bad"
Duplicates	≤2*	≥3
Density	≤40	>40
% G/C content	≤60	≥60
Repeat	False	True

*Often interactions at duplicated genes, e.g. *Hba-1*, *Hba-2*, can still be understood.

CCseqBasic: Data extraction and quality control.

Analysis of sequencing data with the CCseqBasic command line tool (Fig. 3a) is launched using the probe design parameter file provided by AltSort. Enriched chimeric fragments from the

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Capture-C family of methods, NG Capture-C, LI-Capture-C, Tag Capture-C, Tri-C, are generally sequenced using 150-bp paired-end reads[3,7,13,26]; this gives 300-bp of total sequence, therefore the majority of read-pairs overlap and the chimeric fragment can be reconstructed through flashing[27]. RE sites in chimeric fragments generated from 3C digestion and ligation are then identified in both flashed and non-flashed reads and digested-*in silico*. These digested reads are then mapped and categorised as capture, occurring within a targeted RE fragment, or reporter, occurring in any other fragment. Because Capture-C methods use sonication or tagmentation to generate random ends for reads, they can then be PCR duplicate filtered. All of these steps are carried out automatically by CCseqBasic with minimal user input. The main output of CCseqBasic is a convenient UCSC-loadable data hub which contains reporter tracks for each viewpoint (Fig. 4), complete with sample quality control metrics for 3C library quality and enrichment efficiency (Fig. 3b), as well as sequence quality control in the form of fastQC reports. Additionally, all generated data is stored in an easy-to-follow hierarchical folder structure allowing re-analysis, or immediate further processing with CaptureCompare or Tri-C.

Tri-C: Investigation of 3-way interaction hubs.

Most 3C methods interrogate single, pairwise interactions to generate a population view of interaction frequencies. Recently, interest has been given to the formation of multi-way interaction hubs[28–30]. By enriching for chimeric fragments containing multiple ligation junctions, Tri-C allows investigation of these multi-way interactions from single-alleles[7]. Tri-C interaction matrices are generated using an optional flag, which can be implemented during either the initial run or re-analysis with CCseqBasic. When using Tri-C analysis, users can specify the genomic regions to analyse, sonication fragment size, and bin size to match their specific experiment. The output of this analysis is a pdf of an interaction matrix for the region surrounding the viewpoint of interest (Fig 5.).

CaptureCompare: Statistical analysis and sample comparison.

A major strength of Capture-C methods is their ability to confidently distinguish PCR duplicates, while simultaneously generating data from multiple replicates and cell types by multiplexing samples[3]. This allows experiments to be designed with advanced statistical analysis in mind, such as the comparison between cell types, differentiation states, genetic models, drug treatment or any other biological question of interest. As with analysis by CCseqBasic, CaptureCompare is implemented using a command line interface and AltSort-generated parameter file. CaptureCompare extracts the raw interaction count data for each viewpoint from replicates of multiple conditions and generates a UCSC loadable hub containing normalised mean interactions, standard deviation, and difference in interaction level for different conditions on a per RE fragment basis (Fig. 4). Additionally, CaptureCompare generates and implements

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

a modifiable R script for DESeq2 analysis[31], which compares RE fragments to determine if observed differences between conditions reach significance. CaptureCompare also automates the preparation of the necessary files for more complex interaction frequency analysis via the Bayesian modelling tool peaky[21]. Finally, to facilitate presentation, a pdf is generated for each viewpoint with publication-style windowed interaction profiles (Fig. 4).

CaptureSee: A simple interface for navigating and sharing experiments.

On-going methodological improvements have increased the ability of Capture-C methods to target thousands rather than tens to hundreds of loci. Exploration of these large datasets is difficult when using the UCSC browser. To facilitate easy interaction with 3C data, and provide a platform for sharing results, we developed the web-based platform CaptureSee. CaptureSee takes the output hub from CaptureCompare to generate a browsable repository of viewpoints in both per RE fragment and windowed formats (Fig 6.). Within CaptureSee users can quickly jump between capture viewpoints in the genome, to see both the total number and percent of *cis* reporter counts for each viewpoint on a per sample basis, and filter DESeq2 results to find significantly interacting RE fragments.

DISCUSSION

The importance of genome folding for gene regulation has come to the fore in recent years. An extensive array of techniques in fixed and native cells has been developed to examine this folding based on the premise of proximity ligation. Despite being based on a simple premise, 3C sequencing generates a complex bioinformatic problem of mapping and interpreting chimeric DNA fragments. To meet this need, an equally wide range of tools have been developed to either design experiments[12,15,16], map sequence reads[3,17], or interpret interaction profiles[18–22]. While some of these tools have been developed with a non-bioinformatically trained user in mind, many are complex to implement. The absence of a coherent analysis workflow necessitates the definition and testing of appropriate solutions and tools for each step of design, analysis, interpretation and visualisation which must then be assembled into a functioning bioinformatics pipeline, before such experiments can be productively undertaken. This has slowed the general uptake into other fields, such as genetics, clinical studies and developmental biology, though Capture-C methods have proven utility in these areas[10,32–41].

We have generated a systematic, semi-automated workflow for the design and analysis of Capture-C based 3C experiments. This pipeline is constructed with the end-user in mind, including easy to navigate web tools where computationally possible. For the command line tools the default settings will suffice for general use, and user input is only required at three stages. More advanced users will find their needs met, with options and flags available at all

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

stages to customise analysis – such as the choice in mapping algorithm or using SNPs to produce allele-specific profiles[3]. Furthermore, the availability of well-documented open source software in commonly used programming languages (Unix, perl, R) means further modifications can be easily incorporated. Therefore, CaptureCompendium represents a significant advance for the implementation of capture-based 3C methodologies that will enable widespread use of these powerful techniques to explore genomes structure.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

METHODS

Capture-C data presented in this work are available on GEO (GSE129378) and can be visualised at (http://capturesee.molbiol.ox.ac.uk/projects/capture_compare/1387). The Tri-C interaction profile was generated from published data[7] (GSE107755). CaptureCompendium with links to software, installation guide, usage manual and tutorial vignettes is available online (http://userweb.molbiol.ox.ac.uk/public/telenius/CaptureCompendium/). Capsequm2, AltSort and CaptureSee are available as web tools with full descriptions on the websites (http://apps.molbiol.ox.ac.uk/CaptureC/cgi-bin/CapSequm.cgi and https://capturesee.molbiol.ox.ac.uk/). To facilitate saving and sharing data, CaptureSee requires users to register for a free account. CCseqBasic is modified from previous published versions[3], and now contains Tri-C. CCseqBasic is available on GitHub (https://github.com/Hughes-Genome-Group/CCseqBasicS) and implements a suite of tools to analyse data. Briefly adaptor sequences a trimmed with TrimGalore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), reads are flashed using FLASH[27], then *in silico* digested, and mapped with either bowtie or bowtie2[42,43]. Read quality is measured at multiple stages of processing using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads are classified as capture or reporter depending on mapping location and PCR duplicates are filtered based on duplication of random end pairs generated by sonication. Skew arising from co-capture from oligonucleotides at two independent viewpoints is excluded by masking all viewpoints included within a Capture-C experiment. CaptureCompare is available on github (https://github.com/Hughes-Genome-Group/CaptureCompare), and uses custom perl and R scripts to analyse each viewpoint independently by normalising reads per 100,000 *cis* interactions, combining replicates using BEDtools[44], generating mean and subtraction data and inputting read counts into DESeq2[31] treating each restriction fragment as an individual object. Bedgraphs are converted to bigwigs with deepTools2[45]. Windowed plots are generated with adjustable bin and window sizes using custom R scripts.

Acknowledgements

This work was carried out as part of the WIGWAM Consortium (Wellcome Investigation of Genome Wide Association Mechanisms) funded by a Wellcome Trust Strategic Award (106130/Z/14/Z) and Medical Research Council (MRC) Core Funding (MC_UU_12009). We wish to thank all the testers and users of the tools, particularly Duantida Songdej and Nigel Roberts, for dedicated and detailed error reporting and alpha testing, and Jason Torres and Gabriele Mingiolini for portability testing. We acknowledge the CCB computational cluster and the AVI research group for providing the computational resources and system administration for the software development and production runs. Wellcome Trust Doctoral Programmes

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

212 supported R.S. (203728/Z/16/Z), C.Q.E. (203141/Z/16/Z), A.M.O. (105281/Z/14/Z) and L.L.P.H
213 (099684/Z/12/Z). A.M.O was also supported by the Stevenson Junior Research Fellowship
214 (University College, Oxford). J.O.J.D. is funded by an MRC Clinician Scientist Award
215 (MR/R008108) and received Wellcome Trust Support (098931/Z/12/Z).
216

217 **Competing Interests**

218 J.R.H and J.O.J.D are founders and shareholders of Nucleome Therapeutics.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

1. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing Chromosome Conformation. *Science (80-.).* **295**, 1306–1311 (2002).
2. Davies, J. O. J., Oudelaar, A. M., Higgs, D. R. & Hughes, J. R. How best to identify chromosomal interactions: a comparison of approaches. *Nat. Methods* **14**, 125–134 (2017).
3. Davies, J. O. J. *et al.* Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat. Methods* **13**, 74–80 (2016).
4. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). *Nat. Genet.* **38**, 1348–1354 (2006).
5. Schwartzman, O. *et al.* UMI-4C for quantitative and targeted chromosomal contact profiling. *Nat. Methods* **13**, 685–91 (2016).
6. Brant, L. *et al.* Exploiting native forces to capture chromosome conformation in mammalian cell nuclei. *Mol Syst Biol* **12**, 1–8 (2016).
7. Oudelaar, A. M. *et al.* Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat. Genet.* **50**, 1744–1751 (2018).
8. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science (80-.).* **326**, 289–293 (2009).
9. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
10. Downes, D. J. *et al.* An integrated platform to systematically identify causal variants and genes for polygenic human traits. *bioRxiv* 813618 (2019). doi:10.1101/813618
11. Baxter, J. S. *et al.* Capture Hi-C identifies putative target genes at 33 breast cancer risk loci. *Nat. Commun.* **9**, (2018).
12. Hughes, J. R. *et al.* Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* **46**, 205–12 (2014).
13. Oudelaar, A. M., Davies, J. O. J., Downes, D. J., Higgs, D. R. & Hughes, J. R. Robust detection of chromosomal interactions from small numbers of cells using low-input Capture-C. *Nucleic Acids Res.* **45**, (2017).
14. Oudelaar, A. M. *et al.* Dissection of the 4D chromatin structure of the α-globin locus through in vivo erythroid differentiation with extreme spatial and temporal resolution. *bioRxiv* 763763 (2019). doi:10.1101/763763
15. Anil, A., Spalinskas, R., Åkerborg, Ö. & Sahlén, P. HiCapTools: a software suite for probe design and proximity detection for targeted chromosome conformation capture applications. *Bioinformatics* **34**, 675–677 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

16. Hansen, P. *et al.* GOPHER: Generator Of Probes for capture Hi-C Experiments at high Resolution. *BMC Genomics* **20**, 40 (2019).
17. Buckle, A., Gilbert, N., Marenduzzo, D. & Brackley, C. A. capC-MAP: software for analysis of Capture-C data. *Bioinformatics* (2019). doi:10.1093/bioinformatics/btz480
18. Klein, F. A. *et al.* FourCSeq: analysis of 4C sequencing data. *Bioinformatics* **31**, 3085–3091 (2015).
19. Geeven, G., Teunissen, H., de Laat, W. & de Wit, E. peakC: a flexible, non-parametric peak calling package for 4C and Capture-C data. *Nucleic Acids Res.* **46**, e91–e91 (2018).
20. Cairns, J. *et al.* CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* **17**, 127 (2016).
21. Eijsbouts, C. Q., Burren, O. S., Newcombe, P. J. & Wallace, C. Fine mapping chromatin contacts in capture Hi-C data. *BMC Genomics* **20**, 77 (2019).
22. Ben Zouari, Y., Molitor, A. M., Sikorska, N., Pancaldi, V. & Sexton, T. ChiCMaxima: a robust and simple pipeline for detection and visualization of chromatin looping in Capture Hi-C. *Genome Biol.* **20**, 102 (2019).
23. Servant, N. *et al.* HiC-Pro : an optimized and flexible pipeline for Hi-C data processing. *Genome* 1–11 (2015). doi:10.1186/s13059-015-0831-x
24. Ben Zouari, Y., Platania, A., Molitor, A. M. & Sexton, T. 4See : A Flexible Browser to Explore 4C Data. *Front. Genet.* **10**, 1–13 (2020).
25. McGowan, S. J., Hughes, J. R., Han, Z. P. & Taylor, S. MIG: Multi-Image Genome viewer. *Bioinformatics* **29**, 2477–2478 (2013).
26. Oudelaar, A. M., Downes, D., Davies, J. & Hughes, J. Low-input Capture-C: A Chromosome Conformation Capture Assay to Analyze Chromatin Architecture in Small Numbers of Cells. *BIO-PROTOCOL* **7**, (2017).
27. Magoc, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
28. Zheng, M. *et al.* Multiplex chromatin interactions with single-molecule precision. *Nature* **566**, 558–562 (2019).
29. Allahyar, A. *et al.* Enhancer hubs and loop collisions identified from single-allele topologies. *Nat. Genet.* **50**, 1151–1160 (2018).
30. Oudelaar, A. M. *et al.* A revised model for promoter competition based on multi-way chromatin interactions at the Î±-globin locus. *Nat. Commun.* (2019). doi:10.1038/s41467-019-13404-x
31. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
32. Simon, C. S. *et al.* Functional characterisation of cis -regulatory elements governing

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

dynamic Eomes expression in the early mouse embryo. *Development* **144**, 1249–1260 (2017).

33. Thurner, M. *et al.* Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci. *Elife* **7**, 1–30 (2018).
34. Brown, J. M. *et al.* A tissue-specific self-interacting chromatin domain forms independently of enhancer-promoter interactions. *Nat. Commun.* **9**, 1–15 (2018).
35. Schäfer, A. *et al.* Impaired DNA demethylation of C / EBP sites causes premature aging. *Genes Dev.* **32**, 742–762 (2018).
36. Godfrey, L. *et al.* DOT1L inhibition reveals a distinct subset of enhancers dependent on H3K79 methylation. *Nat. Commun.* **10**, 1–15 (2019).
37. Blackledge, N. P. *et al.* PRC1 Catalytic Activity Is Central to Polycomb System Function. *Mol. Cell* **77**, 1–18 (2020).
38. Rhodes, J. D. P. *et al.* Cohesin Disrupts Polycomb-Dependent Chromosome Interactions in Embryonic Stem Cells. *Cell Rep.* **30**, 820–835 (2020).
39. Furlan, G. *et al.* The Ftx Noncoding Locus Controls X Chromosome Inactivation Independently of Its RNA Products Article. *Mol. Cell* **70**, 462–472 (2018).
40. van Bemmel, J. G. *et al.* The bipartite TAD organization of the X-inactivation center ensures opposing developmental regulation of Tsix and Xist. *Nat. Genet.* **51**, (2019).
41. Hyle, J. *et al.* Acute depletion of CTCF directly affects MYC regulation through loss of enhancer–promoter looping. *Nucleic Acids Res.* **47**, 6699–6713 (2019).
42. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
43. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
44. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
45. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

FIGURES

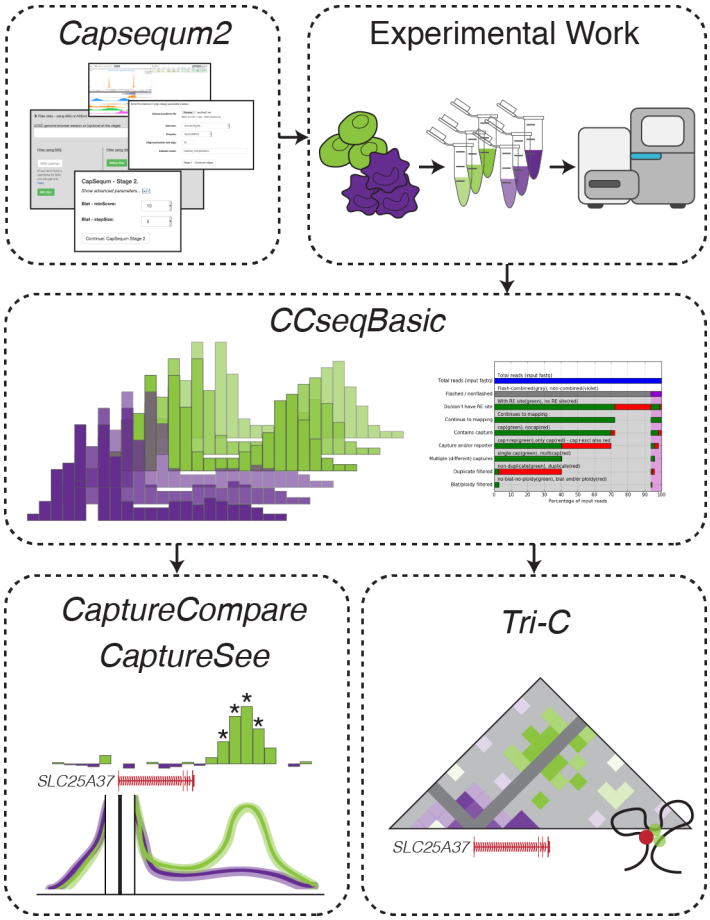

Figure 1 | Overview of CaptureCompendium. Biotinylated oligonucleotides for 3C experiments can be designed using Capsequm2. Following the experimental work, individual samples are analysed and QC is performed in CCseqBasic. Whole experiments, with replicates and comparators are analysed in CaptureCompare and visualized in CaptureSee. Multi-interaction hubs in Tri-C can also be detected using CCseqBasic.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

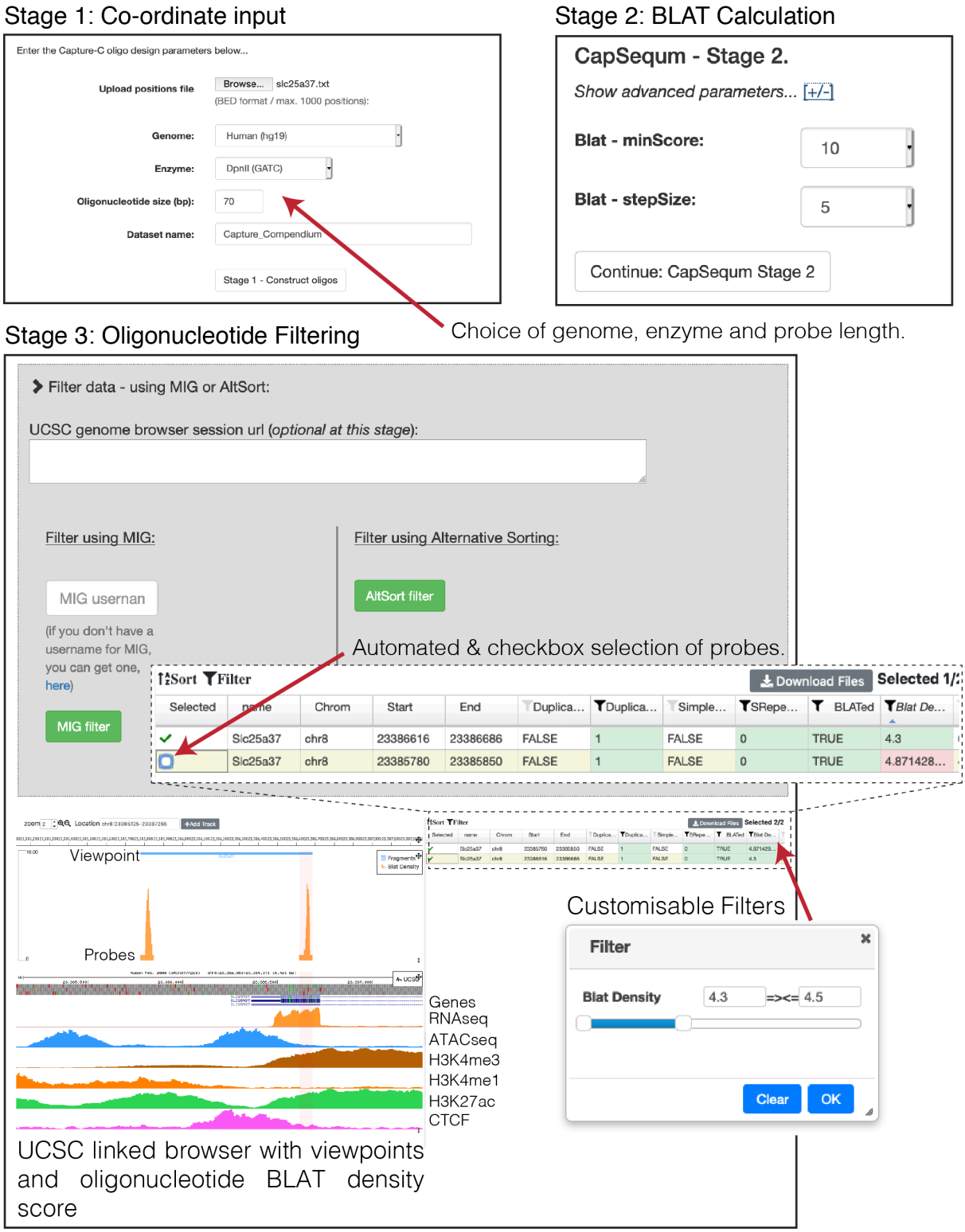

Figure 2 | CapSequm2 provides a user-friendly web interface for oligonucleotide design. The online tool allows design to desired capture locations with selection of genome from Human (hg18/19/38), Mouse (mm9, mm10), Chimpanzee (panTro3), Zebrafish (danRer7, danRer10), *Drosophila melanogaster* (dm3), Chicken (galGal4, galGal5), & Sickleback (BROADS1), of enzyme (*Dpn*II, *Nla*III, *Bam*HI, *Bgl*II, *Hind*III), and of oligonucleotide size. Designed oligonucleotides are then quality filtered using MIG or AltSort which allows integration with chromatin data. AltSort automatically selects oligonucleotides on recommended BLAT density, duplication, and repeat values which can be manually adjusted.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

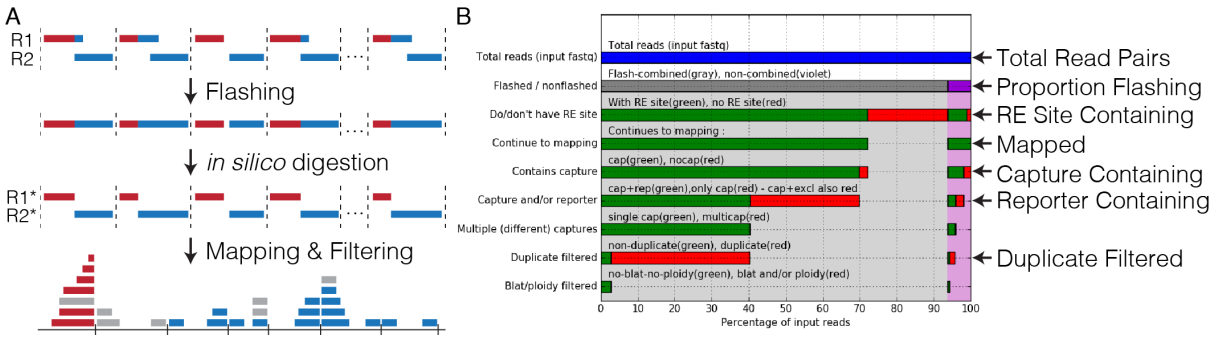

Figure 3 | CCseqBasic extracts informative data from 3C sequence reads. (a) 3C libraries are analysed by reconstructing original fragments by flashing. Subsequent *in silico* digestion of reads resolves chimeric fragments, and allows mapping with bowtie. Mapped reads are identified as “Capture” (red): coming from a targeted viewpoint, or “Reporter” (blue): coming from an interacting fragment. Capture-reporter read pairs are then filtered for PCR duplicates and off target capture sites (grey). **(b)** 3C library and capture quality metrics are reportable as fractions of total reads to allow easy tracking of experimental success, or where necessary, facilitate troubleshooting.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

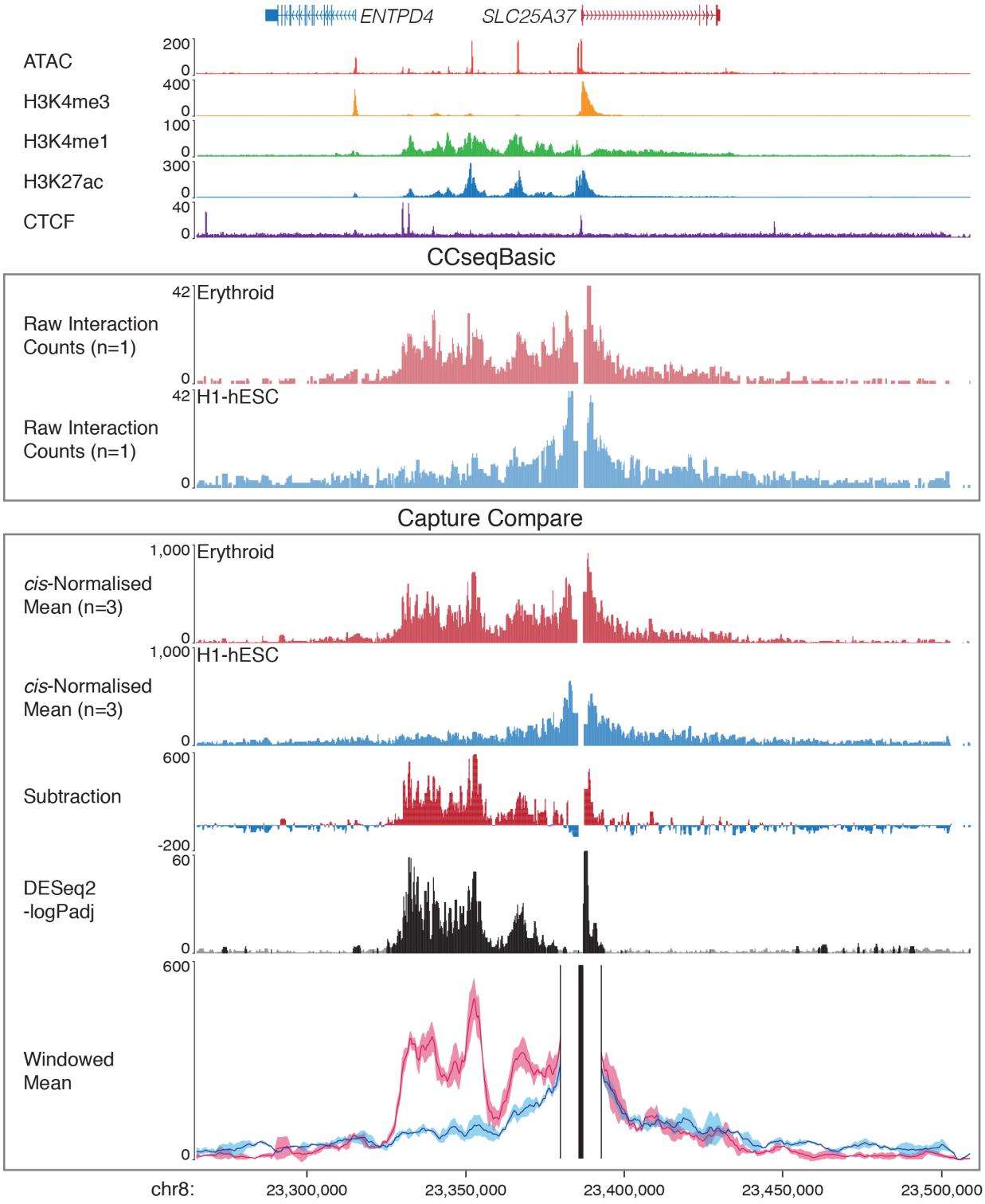

Figure 4 | CaptureCompare allows statistical analysis of experiments. The outputs of CCseqBasic and CaptureComparecan be directly loaded into UCSC for interpretation with chromatin epigenetic data. Example ouput is a Capture-C experiment from the viewpoint of the human mitoferrin gene *SLC25A37*. Output tracks include raw reporter counts, *cis*-normalised reporter counts, subtraction between comparison samples, and -logPadj values from DESeq2 on a per fragment basis. A pdf of windowed mean interactions with standard deviations is also provided. For DESeq2, significantly different fragments have been coloured black (FDR q<0.05).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

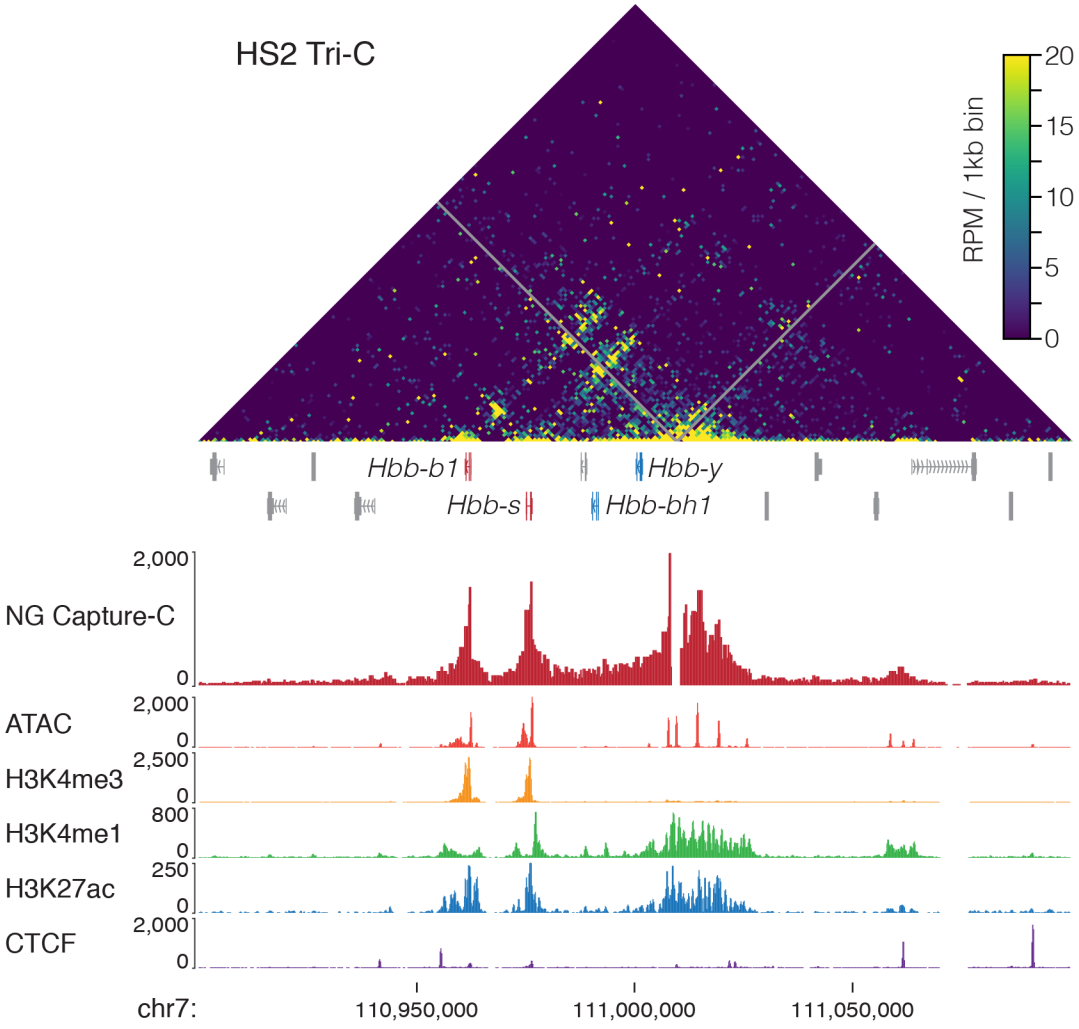

Figure 5 I Detection of multi-way interaction hubs with Tri-C analysis. The output interaction map from Tri-C analysis of the viewpoint from the HS2 element of the mouse beta-globin Locus Control Region (LCR) shows concurrent interaction between HS2, the active beta-globin genes, and other elements within the LCR.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952572; this version posted February 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 6 | Data exploration and sharing with CaptureSee. Data is loaded into CaptureSee using the CaptureCompare generated public hub. Within CaptureSee users can quickly move between viewpoints; with both per fragment and widowed views. In addition to 3C profiles, reporter counts and percent *cis* interactions are provided for each sample 3C library. DESeq2 logPadj scores are provided on a per fragment basis and can be filtered for level of significance.