Multiplexed analysis of chromosome conformation at vastly improved sensitivity

James O J Davies1, Jelena M Telenius1, Simon J McGowan2, Nigel A Roberts1, Stephen Taylor2, Douglas R Higgs1 & Jim R Hughes1

Methods for analyzing chromosome conformation in mammalian cells are either low resolution or low throughput and are technically challenging. In next-generation (NG) Capture-C, we have redesigned the Capture-C method to achieve unprecedented levels of sensitivity and reproducibility. NG Capture-C can be used to analyze many genetic loci and samples simultaneously. High-resolution data can be produced with as few as 100,000 cells, and single-nucleotide polymorphisms can be used to generate allele-specific tracks. The method is straightforward to perform and should greatly facilitate the investigation of many questions related to gene regulation as well as the functional dissection of traits examined in genome-wide association studies.

The ability to annotate gene regulatory elements and investigate their function has been driven by technologies such as RNA-seq1, chromatin immunoprecipitation (ChIP)-seq2,3, DNase-seq4 and ATAC-seq (assay for transposase-accessible chromatin with high-throughput sequencing)5. An outstanding challenge is to understand the mechanisms by which regulatory elements control specific gene promoters at a distance (tens to thousands of kilobase pairs). Conventional chromosome-conformation capture (3C) allows for detailed analysis of the interactions between regulatory elements and promoters at individual loci6–11. Recently, we used a high-throughput approach (Capture-C) to investigate cis interactions at hundreds of loci at high resolution in a single experiment12. Such approaches are of immediate value in defining the regulatory landscapes of many loci and identifying the genes and the functional effects of single-nucleotide polymorphisms (SNPs) that are associated with complex diseases, the majority of which lie in intergenic cis-acting regulatory elements13–15.

The original Capture-C protocol12 uses oligos synthesized on a microarray with a design minimum of 40,000, irrespective of the number of desired viewpoints, so the cost per sample is very high for small designs. Experimental designs often require much smaller subsets of regions, but from multiple samples. Furthermore, Capture-C’s sensitivity does not readily allow for the analysis of very weak cis or trans interactions.

To address these limitations, we redesigned the Capture-C protocol to capture using DNA rather than RNA biotinylated oligos and to do so with vastly superior efficiency. Designs can be generated at any appropriate scale and can be easily expanded by the addition of new oligos to existing pools. Multiple independent 3C libraries from different samples can be processed in a single reaction, greatly increasing throughput, minimizing experimental variation and allowing for precise comparison of chromosome conformations in different cell types.

Here we used NG Capture-C to define the smallest number of cells required to identify robust interactions and show how SNP-specific interaction profiles can be generated.

RESULTS Overview and experimental workflow

Three factors influence the number of unique interactions that can be determined from each viewpoint in a 3C library. First, a maximum of only four interactions can be detected from each region per cell (one from each end of the captured viewpoint fragment on each allele), so the number of available cells and the complexity of the 3C library determine the maximum number of interactions that can be detected. Second, the hybridization efficiency of the capture probe is important, and is largely dictated by the underlying sequence. Third, the efficiency of the assay and depth of sequencing required are determined by the proportion of background fragments from noncaptured DNA contaminating the library.

We sonicated 3C libraries, made using standard methods similar to in situ Hi-C16 (Fig. 1a and Online Methods), to generate 200-bp fragments and then added Illumina paired-end sequencing adaptors. Sonication randomly generates unique fragments, which is an important advantage of Capture-C compared to 4C (circular chromosome-conformation capture) and 5C (chromosome-conformation capture carbon copy), as overamplified PCR duplicates can be removed bioinformatically, allowing the number of unique ligation junctions present in the 3C library to be quantified accurately (Fig. 1b).

We further optimized the protocol to minimize losses during the addition of sequencing adaptors and by pooling material from

1Medical Research Council, Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, Oxford University, Oxford, UK. 2Computational Biology Research Group, Weatherall Institute of Molecular Medicine, Oxford University, Oxford, UK. Correspondence should be addressed to J.R.H. (jim.hughes@imm.ox.ac.uk).

RECEIVED 22 APRIL; ACCEPTED 26 OCTOBER; PUBLISHED ONLINE 23 NOVEMBER 2015; DOI:10.1038/NMETH.3664
two parallel library preparations, which allowed us to use a total input of 10 µg of 3C library. This at least doubled the complexity of the material used for the hybridization reaction. Additionally, the amount of this material used in the hybridization reaction was increased fourfold (from 500 ng to 2 µg). These steps maximized the number of unique interactions defined, as the input 3C material contained eight times more ligation junctions than in the previous protocol.

To decrease the amount of background fragments, we implemented two changes. First, we simplified the library design to use single 120-bp biotinylated DNA oligonucleotides, which included the restriction sites, to capture each end of the target restriction fragment (rather than multiple overlapping oligos; Supplementary Fig. 1), maximizing the capture of informative junction fragments. This made additional steps, such as biotin fill-in, unnecessary and prevented losses in library complexity, which is the critical component of sensitivity, particularly at low cell numbers\(^{17,18}\). Crucially, we introduced a second, sequential round of capture, which markedly decreased the background of uncaptured material and reduced the need for prohibitively deep sequencing.

We tested a minimal design containing probes to only the Hba-\(a1\) and Hba-\(a2\) promoters, equivalent to a 4C-seq analysis, and found that a single oligonucleotide-capture step enriched the targets ~5–20,000-fold. Despite this, the captured DNA from this single region made up less than 1% of the sequenced reads, with the remainder being uncaptured background (Supplementary Fig. 1b).

In NG Capture-C, the use of two sequential oligonucleotide-capture steps resulted in up to 1,000,000-fold enrichment, so that the captured material made up approximately 50% of the

---

**Figure 1** Overview of the method. (a) Experimental workflow. (b) Data analysis.
sequenced material. This second step increased the number of PCR cycles (20–34) and the number of PCR duplicates sequenced because the library complexity (i.e., the number of interactions available to capture) limited the number of unique interactions that could be sequenced. The greatly improved enrichment meant that sequencing depth was no longer limiting and PCR duplicates could be easily excluded bioinformatically. This is demonstrated by the fact that we saw no differences in the local interaction profiles (Fig. 2) and little change in GC content and read length (Supplementary Fig. 1c) between single-captured and double-captured libraries.

To demonstrate the scalability of the approach, we combined the Hba-a1 and Hba-a2 capture probes with capture probes for Hbb-b1 and Hbb-b2 (encoding β-globin in adult mice) and for Slc25a37 (encoding Mitoferrin-1). The globins are among the most extensively characterized gene loci, and their regulatory interactions have been interrogated by almost every 3C-based method to date7,9,10,12,19–22, so they serve as important controls for the validation of any new methodology. The interaction profiles of all three control gene groups were almost identical in the biological replicates (Supplementary Fig. 2) and matched previously determined patterns of interaction (Fig. 2b and Supplementary Figs. 3–5). For the same depth of sequencing, the double capture increased the sensitivity of the profile 30-fold (Fig. 2b, Supplementary Figs. 6 and 7 and Supplementary Data). We next scaled up to a 35-gene design and increased the number of samples analyzed in a single experiment, capturing seven pooled indexed libraries in a single assay. The efficiency that resulted from the double-capture step allowed us to sequence these 245 interaction profiles (35 genes and seven samples) using a single Illumina HiSeq run (177 million reads).

After normalization of individual profiles for the total number of unique interactions across the genome from each viewpoint in each sample, the genome-wide correlation of the two replicates for all genes had an $R^2$ value of >0.97, showing exceptional correlation across biological replicates (Supplementary Fig. 1d). The coefficient of variation decreased substantially (to <50%) when more than ten normalized interactions mapped to any individual restriction fragment (Supplementary Fig. 1e). Thus ligation junctions present at 1 part in 10,000 in the 3C library could be detected reproducibly, as the data were normalized to a total of 100,000 unique interactions across the genome. Furthermore, the patterns of both short-range interactions (Supplementary Figs. 2 and 3) and long-range cis interactions (Supplementary Figs. 6 and 7) were highly reproducible. In addition, NG Capture-C produced a more comprehensive profile than 4C-seq10 (Supplementary Figs. 6 and 8) and did so regardless of the restriction enzyme fragment size (Supplementary Fig. 1f).

We developed a set of tools for the design and analysis of NG Capture-C experiments (Fig. 1b). An online tool for generating oligonucleotide designs for multiple targets can be found at http://apps.molbiol.ox.ac.uk/CaptureC/cgi-bin/CapSequm.cgi, and analysis scripts are available via GitHub (https://github.com/telenius/captureC/releases). The depth of NG Capture-C data allowed unique interactions to be reported per individual restriction fragment or half-fragment, the highest possible resolution for such experiments (Supplementary Fig. 8b); there was no requirement to integrate data by using a moving window.

Identification of regulatory elements using comparative analysis

Currently there is no ideal way to consistently call all likely important interactions from chromosome-conformation data. Sequences from any capture point will interact with the surrounding genome, in a distance-dependent manner, whether the targeted promoter or regulatory element is active or inactive. Therefore, current analysis of 3C data typically includes approaches to normalize interaction data, taking into account the distance from the viewpoint. In practice, the outputs from such approaches are highly dependent on the normalization model and input parameters used, and there is a tendency to undercall cis interactions with genuine regulatory
sequences lying close to the capture point, where normalization is most stringent (Supplementary Note).

The reproducibility of NG Capture-C profiles enabled us to test a complementary approach to identify regulatory interactions by comparing different cell types. Subtractive analysis of normalized data from erythroid and non-erythroid mouse embryonic stem (ES) cells successfully identified all known regulatory elements in well-characterized test loci (Figs. 3 and 4 and Supplementary Figs. 3–5). In the same data we identified similar interactions in other, less-well-characterized loci, which included clinically important genes (Cd47; Supplementary Fig. 9) and complete regulatory networks (Myc, Sox2, Pou5f1 (Oct4), Klf4 and Nanog) (Supplementary Figs. 10–14). We identified interactions with regulatory elements more than 1 Mb from the capture point, consistent with previously reported high-resolution Hi-C data (Supplementary Figs. 10 and 12).

Subtractive analysis also uncovered fine details of tissue-specific regulation of genes that are active in multiple cell types. For example, Pnpo encodes pyridoxal 5’-phosphate oxidase, which is a rate-limiting enzyme in vitamin B6 metabolism producing an essential cofactor in the heme synthetic pathway23. Pnpo is specifically upregulated in mouse erythroid cells by an erythroid-specific enhancer (~1 kb away from the capture point) (Supplementary Fig. 15).

Subtractive analyses not only showed new interactions in the specific cell type under investigation but also identified new patterns of interaction in the cell type used for comparison. For example, analysis of the Tall locus revealed one pattern of interaction in ES cells and another in erythroid cells; these cells act as reciprocal controls for each other (Supplementary Fig. 16).

It is important to note that because this approach relies on changes between active states, its goal is to find regulatory elements, rather than constitutive structural interactions.

The subtractive profiles can be additionally statistically interrogated using common approaches for the differential analysis of sequence count–based data, such as the Bioconductor package DESeq2 (ref. 24). We compared the effectiveness of this approach at identifying known regulatory elements with two tools commonly used for 3C analysis (Fig. 3, Supplementary Figs. 3–5 and Supplementary Note). FourCseq25 and r3C-seq26, which also use replicates and comparative analysis but additionally normalize for genomic distance using different models. We tested all approaches on the well-characterized α-globin, β-globin and Slc25a37 loci using default parameters to simulate the output at uncharacterized loci. Of these, the α- and β-globin loci are the gold standards in the 3C field because of the depth of the functional knowledge of their regulation. These tools called the known elements in the β-globin and Slc25a37 loci, but each variably missed the most proximal elements in the α-globin locus, unlike the comparative approach, which called all of the known elements in each locus (Supplementary Note).

Reproducible megabase-scale cis and trans interactions

The original Capture-C method does not readily detect weak long-range interactions. NG Capture-C enabled us to investigate such interactions and evaluate their strength relative to that of local interactions. Analysis of interaction frequencies across a whole chromosome containing a captured region showed that interactions with the entire chromosome were not easily seen.
when viewed on the same scale as interactions with the more local regulatory elements. However, reproducible, low-level (<100-fold) cis interactions were detected with other active regions of the chromosome (Supplementary Fig. 7). Similar patterns of general interactions could also been seen in trans, but these were a further tenfold weaker than the long-range cis interactions (Supplementary Figs. 17–19). The patterns of trans interactions became visible when the threshold for any interaction was decreased to fewer than 250 interactions per 100 kb; these interactions had similar distributions regardless of the gene promoter used as the viewpoint (Supplementary Figs. 18 and 19) and were correlated with gene density, the number of active promoters, enhancers and CTCF sites (Supplementary Fig. 20). This is of particular interest in the case of the Hba and Hbb genes, as they have been reported to interact with each other in erythroid cells. Some have suggested that these interactions are frequent\(^27\), whereas others have shown them to be rare\(^28\). The sensitivity and robust quantitation provided by the NG Capture-C approach showed that the trans interactions between these two genes were rare in the current study (~1,000-fold fewer than local cis interactions) and on the same scale as those with most other active regions of the genome. These weak interactions are unlikely to be functional, but this finding allows us to be confident that all functional interactions quantifiable by 3C approaches can be detected using NG Capture-C.

Robust interaction profiles from low numbers of cells
In human primary tissues, cell numbers are often limited, so we further adapted NG Capture-C to analyze small numbers of cells (Online Methods). The use of low cell numbers did not alter the digestion efficiency, and the amount of DNA extracted per cell was constant, so we optimized the preparation of hybridization-reaction material for the reduced DNA content of the 3C libraries. Using 100,000 cells, we generated ~19,000 interactions, compared to an average of 137,000 when cell number was not limiting; however, the interaction profiles at the α- and β-
β-globin loci remained virtually unchanged (Supplementary Fig. 21), although weak, long-range interactions became difficult to determine reproducibly (Supplementary Fig. 6) when we used only 100,000 cells.

**Generation of SNP-specific interaction profiles**

SNPs underlying the complex traits examined in genome-wide association studies are frequently heterozygous and may affect the regulatory interactions of the affected allele. Therefore, it would be of great value to generate allele-specific interaction tracks. The greatly improved depth of signal provided by the NG Capture-C protocol allowed us to distinguish between separate alleles when an SNP was included in a capture point and hence sequenced (Fig. 4a). These allele-specific interaction profiles showed that more than 95% of the strain-specific SNPs in *cis* were in phase with the captured strain-specific SNP, showing that interactions with the sister chromatid were relatively rare (Fig. 4b). For the examined genes, the interaction profiles were very similar, probably because none of the SNPs were of functional importance (Supplementary Fig. 22).

This type of analysis can also be applied to non-allelic SNPs. The paralogous genes *Hba-a1* and *Hba-a2* are almost identical, differing at only a few positions, one of which lies near the 5′ promoter capture point (Fig. 4c). This allowed us to generate separate interaction profiles for *Hba-a1* and *Hba-a2* (Fig. 4c). The 5′ *Hba-a1* interacted with the proximal regulatory elements (HS-12 and R4) more frequently than *Hba-a2* did. Interestingly, *Hba-a2* and *Hba-a1* had very similar interactions with the R1 and R2 regulatory elements, which are thought to have stronger enhancer functions than the other elements29,30.

**DISCUSSION**

NG Capture-C was able to reproducibly detect ligation junctions present at 1 part in 10,000 in the 3C library, potentially equivalent to the detection of interactions present in 1 in 10,000 cells at single-restriction-fragment resolution (~250 bp). This far exceeds the sensitivity and resolution of detection of current 3C methods through read counts per fragment, with no need to integrate interactions via a moving window10,12,16. Furthermore, the sensitivity provided by double capture together with the ability to remove PCR duplicates means that the interaction data faithfully represent all interactions in the library, which allows researchers to make estimates of relative quantitation between weak and strong interactions. The complete and detailed protocol presented here, with new, publicly available tools for library design and data analysis, is intended to allow any laboratory to perform chromatin-conformation capture analysis of the highest quality and at throughput levels that were previously impossible.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**Accession codes.** All of the Capture-C data sets are available at the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) with accession code GSE67959.

**Note:** Any Supplementary Information and Source Data files are available in the online version of the paper.

**ACKNOWLEDGMENTS**

J.O.J.D. thanks the Wellcome Trust for funding his work (Wellcome Trust Clinical Research Training Fellowship reference 098931/Z/12/Z). The work was also supported by a Wellcome Trust Strategic Award (reference 106130/Z/14/Z) and the Medical Research Council (MRC Core Funding and Centenary Award reference 405018/188). We thank E. Repapi for statistical advice and L. Hanssen, M. Oudelaar, D. Zeijlslorsa, B. Graham, M. Kassouf, M. Suciu, H. Long, S. Pasricha, V. Buckle, T. Milne, T. Fulga, T. Sauka-Spengler, D. Downes and A. Drakesmith for their critique of the manuscript. We thank S. Thongjuea for discussions on analysis.

**AUTHOR CONTRIBUTIONS**

J.O.J.D. performed the experiments, analyzed the data and wrote the manuscript. J.R.H. designed the experiments, assisted with the bioinformatic analysis and wrote the manuscript. N.A.R. assisted with the experiments. J.M.T. analyzed the data. S.J.M. and S.T. assisted with the bioinformatics analysis and prepared the software for public release. D.R.H. wrote the manuscript.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Wang, Z., Gerstein, M. & Snyder, M. RNA-seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).

2. Mikkelsen, T.S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).

3. Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods* **4**, 651–657 (2007).

4. Hesselberth, J.R. *et al.* Global mapping of protein–DNA interactions in vivo by digital genomic footprinting. *Nature Methods* **6**, 283–289 (2009).

5. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods* **10**, 1213–1218 (2013).

6. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
7. Tolhuis, B., Paclata, R.J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol. Cell 10, 1453–1465 (2002).
8. Noordermeer, D. et al. The dynamic architecture of Hox gene clusters. Science 334, 222–225 (2011).
9. Sanayi, A., Lajoie, B.R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. Nature 489, 109–113 (2012).
10. van de Werken, H.J. et al. Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat. Methods 9, 969–972 (2012).
11. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. Science 337, 1190–1195 (2012).
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, 205–212 (2014).
13. Pasquali, L. et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat. Genet. 46, 136–143 (2014).
14. Maurano, M.T. et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 337, 1190–1195 (2012).
15. Parker, S.C. et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc. Natl. Acad. Sci. USA 110, 17921–17926 (2013).
16. Rao, S.S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680 (2014).
17. Jäger, R. et al. Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. Nat. Commun. 6, 6178 (2015).
18. Schoenfelder, S. et al. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome Res. 25, 582–597 (2015).
19. Vernimmen, D., De Gobbi, M., Sloane-Stanley, J.A., Wood, W.G. & Higgs, D.R. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. EMBO J. 26, 2041–2051 (2007).
20. Hughes, J.R. et al. High-resolution analysis of cis-acting regulatory networks at the alpha-globin locus. Phil. Trans. R. Soc. Lond. B 368, 20120361 (2013).
21. Baü, D. et al. The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules. Nat. Struct. Mol. Biol. 18, 107–114 (2011).
22. 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).
23. Kang, J.H. et al. Genomic organization, tissue distribution and deletion mutation of human pyridoxine 5’-phosphate oxidase. Eur. J. Biochem. 271, 2452–2461 (2004).
24. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
25. Klein, F.A. et al. FourCseq: analysis of 4C sequencing data. Bioinformatics 31, 3085–3091 (2015).
26. Thongjuea, S., Stadhouders, R., Grosveld, F.G., Soler, E. & Lenhard, B. r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. Nucleic Acids Res. 41, e132 (2013).
27. Osborne, C.S. et al. Active genes dynamically colocalize to shared sites of ongoing transcription. Nat. Genet. 36, 1065–1071 (2004).
28. Noordermeer, D. et al. Variegated gene expression caused by cell-specific long-range DNA interactions. Nat. Cell Biol. 13, 944–951 (2011).
29. Bernet, A. et al. Targeted inactivation of the major positive regulatory element (HS-40) of the human alpha-globin gene locus. Blood 86, 1202–1211 (1995).
30. Anguita, E. et al. Deletion of the mouse alpha-globin regulatory element (HS-26) has an unexpectedly mild phenotype. Blood 100, 3450–3456 (2002).
31. de Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. Genes Dev. 26, 11–24 (2012).
32. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676 (2006).
33. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676 (2006).
ONLINE METHODS
Preparation of 3C libraries. We generated single-cell preparations of erythroid cells by gently dissociating cells from the spleens of 6–9-month-old female mice (F1 crosses between C57BL/6 and CBA/J and pure C57BL/6) treated with phenylhydrazine (40 mg/g body weight per dose, with three doses given 12 h apart; mice were killed on day 5). Mice were maintained in specific pathogen–free facilities at the Biomedical Services Unit of Oxford University. Protocols were approved through the Oxford University Local Ethical Review process. Experimental procedures were performed in accordance with European Union Directive 2010/63/EU and/or the UK Animals (Scientific Procedures) Act, 1986. Phenylhydrazine causes hemolytic anemia and marked erythroid expansion in the spleen so that 80% or more of cells are erythroid cells (defined as CD71+ter119+). We passed cells through a 40-μm CellTrix strainer (Sysmex) to remove clumps. For ter119+ cell selection, we stained cells with ter119-phycocerythrin and positively selected using anti-phycocerythrin MACS beads (Miltenyi Biotec) before fixation with formaldehyde.

We prepared mouse ES cells (from mice at embryonic day 14) by trypsinization and washing once before fixation.

We resuspended each aliquot of 10^7 cells in 10 ml of RPMI medium with 10% fetal calf serum in a 15-ml conical centrifuge tube, added 549 μl of 37% (vol/vol) formaldehyde to each for a final concentration of 2% (vol/vol), incubated the tubes for 10 min at room temperature on a roller mixer, quenched the cross-linking reaction with 1.5 ml of cold 1 M glycine, and centrifuged samples immediately for 5 min at 220 g in a precooled centrifuge at 4 °C. The supernatant was discarded and the pellet was gently resuspended in 10 ml of cold phosphate-buffered saline (PBS). The cells were centrifuged again (5 min, 220 g, 4 °C), and the supernatant was discarded. We resuspended the pellet in 5 ml of cold lysis buffer (10 mM Tris-HCl, pH 8, 10 mM NaCl, 0.2% Igepal NP-40 (Sigma), 1× cOmplete protease inhibitor (Roche)), incubated the suspension on ice for 20 min, centrifuged it to pellet nuclei (5 min, 500 g, 4 °C), and carefully discarded the supernatant. We snap-froze multiple aliquots using liquid N2 or dry ice with ethanol, which allows for storage for up to several months at −80 °C. We resuspended cells in 1 ml of water (Milli-Q or Sigma), homogenized them (45 strokes in a 5-ml Dounce homogenizer) on ice to break open intact cells, pelleted them by centrifugation (5 min, 22000 g at 4 °C) and resuspended them in 650 μl of water (Milli-Q or Sigma). We set up three parallel digestion reactions to check for nonspecific degradation with 50 μl of restriction enzyme buffer, 10 μl of 20% SDS (vol/vol) and 414 μl of water for a final volume of 800 μl. We added 4 ml of chloroform (25:24:1) and reconstituted them with 10 ml of water (final volume of 200 μl) by pipetting. We also determined digestion efficiency using quantitative PCR with primers designed across one of the restriction enzyme gene promoters (DpnII digestion control) and another primer set that lay close to the other end of the same restriction fragment (Hba-a1/2 control primer). Digestion efficiencies were always in excess of 70% for libraries used for analysis. We used a Qubit dsDNA BR assay kit (Life Technologies) to determine DNA concentration.

Real-time primers for assessing digestion were as follows:

- **DpnII digestion control forward primer**, GTGTCAACAAACACAGCTCA
- **DpnII digestion control reverse primer**, CCTGGAATCCTTTTG
- **GCTCAAG**
- **DpnII digestion control Taqman probe**, CGGCGCCTAGATGCAAGTC
- **Hba-a1/2 control forward primer**, TGGAGGGCATATAAGTC
- **Hba-a1/2 control reverse primer**, TGCTTTTGTCTTC
- **CCACAGA**
- **Hba-a1/2 control Taqman probe**, TGCAGGTCAGACACT
- **TCTGATTCTGACA**

Addition of sequencing adaptors. We sonicated two aliquots of 5 μg of 3C library to 200 bp using a Covaris S220 Focused ultrasonicator (six cycles of 60 s; duty cycle, 10%; intensity, 5; cycles per burst, 200) and assessed the degree of sonication using an Agilent
Bioanalyzer or Tapestation (DNA 1000). We indexed libraries with Illumina Truseq indexed sequencing adaptors using NEBnext reagents (E6000, E6040, E7335 or E7500) for end repair, dA labeling, adaptor ligation and PCR indexing, primarily following the manufacturer’s instructions. We performed DNA clean-up steps with Ampure XP beads at a 1:1.8 ratio to minimize the selection of larger fragments and loss of material and used six to eight cycles of PCR when adding the Truseq indices using the Herculase II PCR kit (Agilent). Generally we generated 1.5–2 µg of adaptor-ligated material per 3C library; however, to maximize library complexity, we prepared libraries in duplicate (to use 10 µg of input material) and pooled after indexing. We assessed libraries using an Agilent Bioanalyzer or D1000 Tapestation both before and after the PCR and addition of sequencing adaptors, as this allowed us to determine the DNA losses (and library complexity) before amplification.

We reconstituted biotinylated DNA oligonucleotides (IDT ultramers or Sigma long synthesis) to a concentration of 2.9 µM. This allowed us to mix different oligonucleotides in equimolar quantities so that 4.5 µl of the resulting library would always contain a total of 13 pmol of oligonucleotide pool. These oligonucleotides are vastly in excess in the hybridization reaction such that contamination with very small quantities can result in substantial capture, which can lead to spurious results. We recommend that oligonucleotides for different experiments be ordered from the manufacturer and handled separately, as contamination can occur during the manufacturing process.

**Oligonucleotide capture.** We placed 1.5–2 µg of adaptor-ligated material in a 1.5-ml microcentrifuge tube with 5 µg of COT DNA from the appropriate species, 1,000 pmol of Nimblegen HE universal blocking oligo and 1,000 pmol of Nimblegen HE index-specific blocking oligo (corresponding to the Illumina TS index used) and thoroughly dried the contents of the tube using a vacuum centrifuge (50–60 °C). We resuspended DNA in 7.5 µl of Nimblegen hybridization buffer and 3 µl of Nimblegen hybridization component A and then denatured this mix at 95 °C for 10 min. Concurrently, we heated 4.5 µl of the biotinylated capture oligonucleotide library (total: 13 pmol) in a 0.2-ml PCR tube to 47 °C in a PCR thermocycler. After denaturation we added the 3C library and blocking oligonucleotides to the biotinylated oligonucleotides without removing them from the heating block and incubated this hybridization reaction in the thermocycler at 47 °C for 64–72 h (with a heated lid at 57 °C).

We prepared the Nimblegen SeqCap EZ wash buffers (I, II, III, stringent and bead wash buffers), preheated to 47 °C in a thermomixer where necessary, according to the instructions. We aliquoted 100 µl of M270 streptavidin beads per library into a 1.5-ml microcentrifuge tube and incubated them at room temperature for 30 min. We washed beads twice with 200 µl of bead wash buffer, using a DynaMag device to capture the beads and allow the supernatants to be discarded. Immediately after the final wash, we added the hybridization reaction mixture directly to the beads and mixed it thoroughly by pipetting up and down and then vortexing. We allowed captured material to bind to the beads by incubating samples on a thermomixer (47 °C, 500 r.p.m., 45 min). To wash beads, we added 100 µl of warmed wash buffer I (47 °C), mixed samples by vortexing for 10 s and placed them in a DynaMag device, discarding the liquid once it was clear. We immediately added 200 µl of warmed stringent wash buffer (47 °C), incubated samples at 47 °C for 5 min and placed them in a DynaMag device, again discarding the liquid once it was clear. We repeated this step so that two washes were performed with stringent wash buffer. We added 200 µl of wash buffer I, mixed samples by vortexing for 1 min, and placed them in a DynaMag device, discarding the liquid once it was clear. We added 200 µl of wash buffer III, mixed samples by vortexing for 30 s, and placed them in a DynaMag device, discarding the liquid once it was clear. We resuspended beads in 40 µl of PCR-grade water and either proceeded immediately to amplification or stored the beads at −15 °C to −25 °C for up to 3 weeks.

We amplified the captured material directly from the beads using either the SeqCap EZ post-capture LM PCR master mix and post–LM-PCR oligos (18 cycles) or the newer KAPA master mix supplied in the SeqCap EZ accessory kit v2 (14 cycles) in two separate reactions. We used Ampure-XP beads to clean up the amplification reaction and eluted DNA in 30 µl of PCR-grade water (Sigma). We used either an Agilent Bioanalyzer or a D1000 Tapestation to assess the quality of captured material.

We used 75% of the amplified captured material (up to a maximum of 2 µg) for the second round of oligonucleotide capture, following the protocol outlined above, except that material was hybridized for 24 h rather than 64–72 h.

We determined the mass of captured material using the Agilent Bioanalyzer or D1000 Tapestation and Qubit dsDNA BR assay kit. We diluted libraries to 4 nM on the basis of size determination from the Bioanalyzer or D1000 Tapestation and the concentration measured by Qubit. We determined oligonucleotide capture enrichment by real-time PCR, using the Hba-a1/2 control primers described above and a standard curve of genomic DNA for comparison to the concentration of the input material determined by Qubit.

**Multiplexed library capture.** Multiple samples can be captured simultaneously if they are labeled with different index adaptors and mixed before the oligonucleotide hybridization. To maintain library complexity, for the first capture we pooled 1–2 µg from each sample in an exact 1:1 stoichiometry. It is important to do this precisely, as the percentage of reads obtained from each sample will be directly related to the amount of DNA mixed. We then added 5 µg of COT DNA and 1,000 pmol of universal TS HE blocking oligonucleotides for every sample and 1,000 pmol of the index-specific blocking oligonucleotide for each relevant sample. We then either split the mixture into multiple identical hybridization reactions of the same volume of a single sample or performed one large hybridization reaction. We performed the hybridization, streptavidin bead capture and wash protocols as outlined above, except that the volumes were adjusted appropriately when larger-volume captures were undertaken. We performed the PCR reactions using the same volumes as for a single capture (multiple reactions were performed in parallel). For the second capture, we pooled material from all of the PCRs and performed a single second capture on this material. It is possible to use a single volume capture at this point in the protocol because the library should contain thousands of copies of each
captured read, making it unlikely that substantial complexity will be lost during the second capture.

**Sequencing.** We determined the mass of captured material using the Agilent Bioanalyzer or D1000 Tapestation and Qubit dsDNA BR assay kit. We diluted libraries to 4 nM on the basis of size determination from the Bioanalyzer or D1000 Tapestation and the concentration measured by Qubit. We confirmed this final concentration using real-time PCR (SYBR green) with the P5 and P7 sequences on the adaptors or Qubit dsDNA HS assay kit (Life Technologies). We sequenced the majority of material using the Illumina MiSeq (300 bp, V2 chemistry), which produced 10–20 million 150 bp paired-end reads, depending on the cluster density. We also performed one larger sequencing run on the Illumina HiSeq, which produced 100–bp paired-end sequences.

**Adaptations for low cell numbers.** We performed 3C library preparation as above with the following adaptations: we decreased the volume of the digestion reaction to 200 μl for between 500,000 and 3 million cells and 50 μl for 500,000 cells or fewer. When fewer than 1 million cells were used, to save material, we omitted the two control samples and assessed digestion efficiency on the basis of the ligation reaction using real-time PCR. The Ct value for the ligation reaction is nearly identical to that for the digestion control because the probability of the fragment ligating back to its original position is negligible given the proportion of undigested material. We performed the entire library preparation in a single 1.5-ml Eppendorf tube to minimize losses. We performed phenol-chloroform extraction as described above, except that the DNA precipitation was performed in a smaller volume (3× volume of chloroform extraction as described above, except that the DNA present at a specific position in the capture fragment.

**Data analysis.** We removed adapter sequences from the reads in the raw FASTQ files using Trim_galore (a wrapper tool around Cutadapt and FastQC; Babraham Institute, http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). This was particularly necessary when using 150-bp paired-end sequencing, as some of the sequenced fragments were shorter than 150 bp and the sequence extended into the adaptor. We reconstructed the paired-end reads into single reads where possible using FLASH with interleaved output settings34. These two steps were omitted when shorter reads were used that did not have an area of central overlap; in such cases a file of the reads was generated with the paired-end reads interleaved in strict order (read 1 FASTQ followed by read 2). We performed an in silico restriction enzyme digestion of the reads using the DpnII2E.pl script (https://github.com/telenius/captureC/releases), with the name of the read used to keep a record of each subfragment, and aligned the resulting FASTQ file of subfragments using Bowtie1 (using p1, m2, best and strata settings). Fragments that resulted from nonspecific ligation and did not contain the restriction cut sequence were unmappable with Bowtie1 and were therefore discarded. It is important that the reads are in strict order for the subsequent analysis; we achieved this by either sorting by read name or using one processor (p1) for the alignment.

We analyzed the resulting SAM file with the main script CCanalyser2.pl (https://github.com/telenius/captureC/releases). This allowed us to classify the subfragments as (a) ‘capture’ if they were contained within the capture fragment, (b) ‘proximity exclusion’ if they were inside the defined proximity-exclusion coordinates (usually 1 kb on either side of the capture fragment) or (c) ‘reporter’ if they were outside all of the capture and proximity-exclusion regions in the entire experiment. We removed PCR duplicates in CCanalyser2.pl by excluding reads that had the same start and end coordinates of each subfragment. For long-range cis and trans analysis, the start and stop coordinates of the interacting read itself also had to be unique. We used this stringent filter to remove PCR duplicates, because occasional sequencing errors in the captured restriction fragment allowed PCR duplicates to seem unique. Unique interactions were reported only when the read was unique and a single capture and one or more reporter subfragments were defined from a single read.

We designed CCanalyser2.pl to either map reads to the whole restriction enzyme fragment or, to give the maximum resolution possible, map reads to the half-fragment on the basis of the midpoint of the read and restriction fragment.

We also designed CCanalyser2.pl with the facility to generate SNP-specific tracks. This analysis requires a specific base to be present at a specific position in the capture fragment.

After using CCanalyser2.pl, we filtered the data to remove regions with problematic mappability due to copy-number differences and mismapped reads from the proximity-exclusion region. For the latter, we mapped the sequence of the proximity-exclusion zone back to the genome using BLAT, excluding restriction fragments more than 2 Mb from the viewpoint if the proximity-exclusion zone mismapped to them (the 2-Mb distance was chosen so that gene paralogs such as Hba-a1 and Hba-a2 were not excluded). We normalized the read count per fragment to the total number of reads in the track to give the number of interactions per 100,000 interactions in the whole track. Finally, we converted these data to a format (e.g., BigBed or GFF) suitable for viewing in the UCSC Genome Browser35,36 (http://genome.ucsc.edu/) and performed statistical analysis using DESeq2 (ref. 24) using un-normalized raw counts per restriction fragment and excluding restriction fragments with no mapped reads. To allow for correct statistical interrogation of each interaction profile, we generated a minimum of three biological replicates for each cell type.

34. Magoč, T. & Salzberg, S.L. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963 (2011).
35. Kent, W.J. et al. The human genome browser at UCSC. Genome Res. 12, 996–1006 (2002).
36. Raney, B.J. et al. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. Bioinformatics 30, 1003–1005 (2014).