High-resolution identification of balanced and complex chromosomal rearrangements by 4C technology

Marieke Simonis[1,5], Petra Klous[1,5], Irene Homminga[2], Robert-Jan Galjaard[3], Erik-Jan Rijkers[4], Frank Grosveld[5], Jules P P Meijerink[2] & Wouter de Laat[1,5]

Balanced chromosomal rearrangements can cause disease, but techniques for their rapid and accurate identification are missing. Here we demonstrate that chromatin conformation capture on chip (4C) technology can be used to screen large genomic regions for balanced and complex inversions and translocations at high resolution. The 4C technique can be used to detect breakpoints also in repetitive DNA sequences as it uniquely relies on capturing genomic fragments across the breakpoint. Using 4C, we uncovered *LMO3* as a potentially leukemogenic translocation partner of *TRB@*. We developed multiplex 4C to simultaneously screen for translocation partners of multiple selected loci. We identified unsuspected translocations and complex rearrangements. Furthermore, using 4C we detected translocations even in small subpopulations of cells. This strategy opens avenues for the rapid fine-mapping of cytogenetically identified translocations and inversions, and the efficient screening for balanced rearrangements near candidate loci, even when rearrangements exist only in subpopulations of cells.

Chromosomal rearrangements (deletions, amplifications, inversions and translocations) occur naturally in the genome[1–6] and often cause disease, particularly when they affect gene expression. This can happen when the rearrangement leads to changes in gene copy number, creates fusion genes or results in a repositioning of regulatory elements such as enhancers. Accurate mapping of chromosomal rearrangements is required to find disease-associated genes and is important both for understanding the mechanism of disease and for optimal diagnosis and treatment decisions. Deletions and amplifications causing copy-number variation can be detected at high resolution using microarray-based comparative genomic hybridization (array-CGH). High-resolution mapping of translocations and inversions not accompanied by loss or gain of DNA can be done at a genome-wide scale by massive parallel paired-end sequencing[7,8]. Given the repetitive nature of the human genome, this is not trivial and massive genome-wide sequencing currently cannot be done routinely for every patient. Detection of balanced rearrangements still mostly relies on molecular-cytogenetic techniques such as chromosomal karyotyping. However, its limited resolution (5–50 megabases (Mb)) necessitates additional labor-intensive experiments to validate genomic breakpoints.

Here we demonstrate that the recently developed chromatin conformation capture on chip (4C) technology[9] can be used to quickly fine-map chromosomal breakpoints in large selected genomic regions that span at least 3 Mb on each side of a selected chromosomal location. Unlike other genomics strategies for mapping chromosomal rearrangements, 4C does not rely on finding the one fragment that carries the breakpoint; rather, it identifies rearrangements based on the capture and identification of many fragments across the breakpoints. As a consequence, 4C can even be used to detect chromosomal breakpoints when they are located within repetitive DNA sequences. Using a single microarray, 4C can be used to analyze multiple selected chromosomal regions simultaneously for breakpoints and rearranged partners throughout the genome, even when the rearrangements are balanced or complex. We used the 4C technology to uncover new rearrangements in cell lines and samples from individuals with T cell–derived acute lymphocytic leukemia (T-ALL) and identified *LMO3* as a potential leukemogenic translocation partner of T cell receptor beta locus (*TRB@*). Finally, we found that 4C, can be used to identify balanced rearrangements even when they are present in small subpopulations of cells. The fact that large megabase-sized regions around target sites are captured efficiently by 4C, no matter the three-dimensional (3D) structure of the DNA, has consequences also for the interpretation of results obtained by other chromatin conformation capture (3C)-based methods, which we discuss here.

RESULTS

4C identifies balanced chromosomal rearrangements

The 4C technology[10] (**Supplementary Fig. 1a**) is based on 3C technology[11]. It involves treatment of cells with formaldehyde to cross-link parts of the genome that are physically close in the nucleus. The DNA is then digested with a restriction enzyme (such as HindIII used here) and cross-linked DNA fragments are ligated. Inverse PCR with primers specific to a selected locus (the 'viewpoint') subsequently allows amplification of its

[1]Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and University Medical Center Utrecht, Utrecht, The Netherlands. [2]Department of Pediatric Oncology, [3]Department of Clinical Genetics, [4]Department of Biochemistry and [5]Department of Cell Biology, Erasmus Medical Center, Rotterdam, The Netherlands. Correspondence should be addressed to W.d.L. (w.delaat@hubrecht.eu).

RECEIVED 21 JULY; ACCEPTED 16 SEPTEMBER; PUBLISHED ONLINE 11 OCTOBER 2009; DOI:10.1038/NMETH.1391

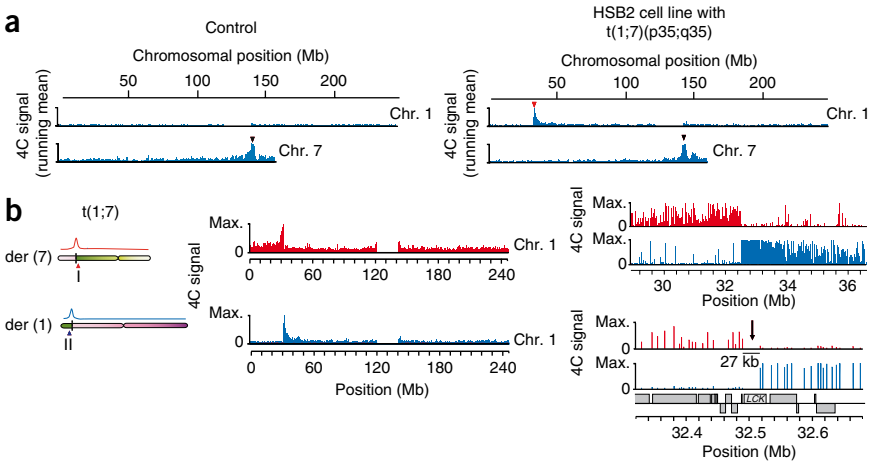

Figure 1 | 4C accurately detects a balanced translocation. (**a**) The 4C signals across chromosomes (chr.) 1 and 7 in cells from a healthy individual and the HSB-2 cell line carrying t(1;7)(p35;q35). The black and red arrowheads indicate positions of viewpoint sequences and translocation site, respectively. Running mean data were plotted, using a window size of ~60 kb. Scale on *y* axes (arbitrary units) is identical for all chromosomes, with the highest mean value set to maximum. (**b**) The 4C signals on chromosome 1, captured by viewpoint fragments I (red) and II (blue) located at opposite sides of the *TRB@* locus on chromosome 7. The regions on chromosome 1 captured by viewpoint fragments on chromosome 7 directly neighbor each other and flank the previously cloned breakpoint (arrow) located in the *LCK* gene. The highest signal in each sample was set to maximum (max.).

interacting partners. When analyzed on a 4C-tailored microarray (385,000 probes) that analyzes the entire human genome at an average resolution of 7 kb (ref. 9), the highest hybridization signals always map to a region of several megabases surrounding the viewpoint sequence (**Supplementary Fig. 1b**). Thus, 4C technology predominantly identifies flanking sequences of the viewpoint, and we reasoned it should therefore detect genomic rearrangements present in this chromosomal area.

To test this hypothesis, we applied the 4C technology to the T-ALL cell line HSB-2. This cell line contains a reciprocal translocation between the *TRB@* locus on band 3, sub-band 5 on the q arm of chromosome 7 (7q35) and the lymphocyte cell–specific protein-tyrosine kinase (*LCK*) locus on 1p35, t(1;7)(p35;q35) (ref. 12). We performed two independent 4C experiments, analyzing DNA interactions with viewpoint sequences located 462 kb centromeric and 239 kb telomeric of the breakpoint in *TRB@*. With both viewpoint sequences, we observed strong hybridization signals not only around the *TRB@* locus on chromosome 7 but also across a megabase-sized region on 1p35, specifically in HSB-2 cells (**Fig. 1**). These signals represented restriction fragments captured by the viewpoint sequences on chromosome 7 as an effect of their close physical proximity. Notably, the first restriction fragments captured on chromosome 1 in both experiments directly flanked the previously identified chromosomal breakpoint. Thus, 4C identifies translocation partners based on the appearance of unusually large clusters of signals on unrelated chromosomes and maps the translocation breakpoint to a position just upstream of the first captured restriction fragment on this chromosome.

Next, we tested whether 4C can be used to identify inversions by applying it to a sample from an individual with T-ALL that, based on a fluorescence *in situ* hybridization (FISH) analysis, carries an inversion on chromosome 7, inv(7)(p15q35). This abnormality leads to the rearrangement of the *TRB@* locus into the *HOXA* gene cluster and activation of the *HOXA* genes[13,14]. The same set of *TRB@* viewpoint sequences described above was used. The telomeric *TRB@* viewpoint sequence captured centromeric *HOXA* fragments and the centromeric *TRB@* viewpoint fragment captured telomeric *HOXA* fragments, thus revealing an inversion between the loci in the sample from an individual with T-ALL (**Fig. 2a**). The two captured regions directly neighbor each other and locate the breakpoint to a 6-kb region. Restriction-fragment paired-end sequencing of the breakpoint (**Supplementary Fig. 2**) confirmed the location of this breakpoint (**Fig. 2a**). Thus 4C technology can detect balanced translocations and inversions at high resolution.

4C identifies unbalanced chromosomal rearrangements

We explored the potential of 4C technology by applying it to an Epstein-Barr virus (EBV) transformed cell line derived from an individual with postaxial polydactyly (PAP). PAP is an autosomal-dominant heritable disorder characterized by extra ulnar or fibular digits. The cells had been previously characterized by karyotyping and FISH analyses and found to contain an unbalanced translocation between chromosomes 4 and 7, t(4;7)(p15.2;q35), with a micro-deletion of unknown size[15]. We performed two 4C experiments, analyzing DNA interactions with two viewpoint fragments located on either side of the rearranged part of chromosome 7. In contrast to what had been found for the balanced translocation, the chromosome 4 fragments captured by the two viewpoint sequences on chromosome 7 did not directly flank each other but were 2.8 Mb apart (**Fig. 2b** and **Supplementary Fig. 3**). We cloned one of the breakpoints and sequenced it, confirming the breakpoint location at 20.08 Mb (**Fig. 2b**). The sequence revealed that the breakpoint on chromosome 7 was more than 3 Mb away from the viewpoint sequence. Thus, 4C viewpoint sequences can be used to capture DNA fragments and characterize rearrangements even when the breakpoints are several megabases away. The data also showed that 4C technology is very suitable to fine-map poorly characterized rearrangements identified by chromosomal karyotyping. When directed to both sides of a genomic breakpoint, 4C can be used to immediately identify whether a translocation or inversion is balanced or accompanied by additional rearrangements such as a deletion (that is, unbalanced).

We next investigated whether 4C technology can be used to identify a deletion not associated with a translocation. For this, we applied 4C to a sample from an individual with T-ALL that was previously characterized by array-CGH to contain a homozygous deletion of the *p15-p16* loci on chromosome 9p21 (J.P.P.M.; unpublished data). Using a viewpoint fragment located ~2 Mb away from the predicted, but not fine-mapped, deletion, we observed a region lacking probe signals, demarcating the deleted area (**Supplementary Fig. 4a**). Notably, we observed increased hybridization signals for the region beyond the deletion. We

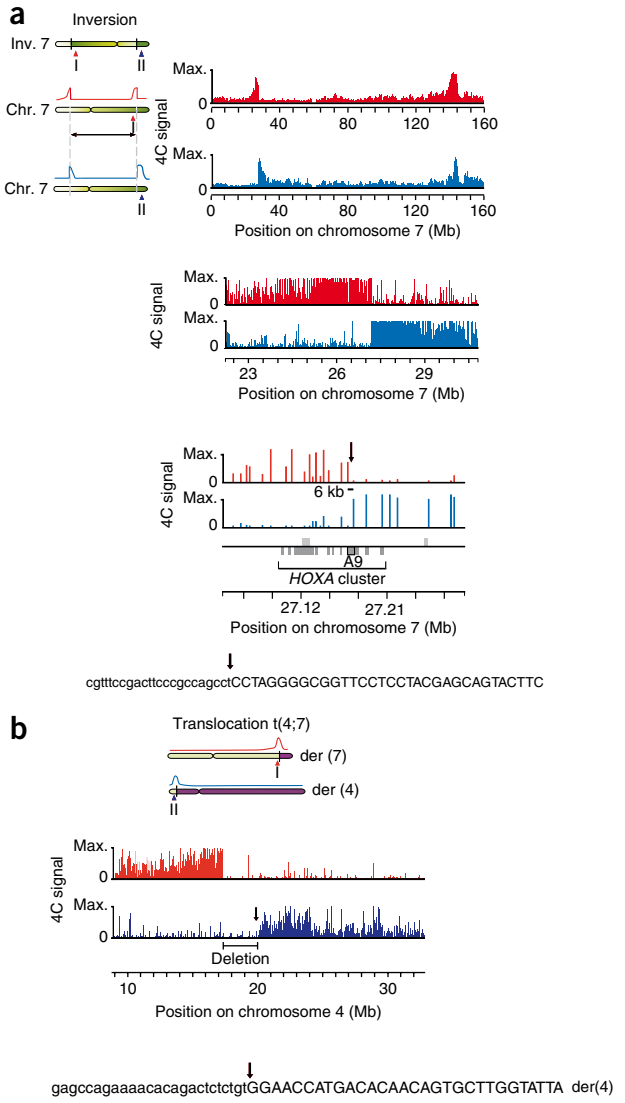

Figure 2 | 4C accurately detects a balanced inversion and an unbalanced translocation. (**a**) Balanced inversion detected in a sample from an individual with T-ALL. The 4C signals on chromosome 7 are fragments captured by viewpoint fragments I (red) and II (blue) located at opposite sides of the *TRB@* locus on chromosome 7. The breakpoint (arrow) was sequenced and confirmed to locate immediately adjacent to the captured sequences. (**b**) The 4C technology accurately detects an unbalanced translocation t(4;7). The 4C data show fragments captured on chromosome 4 by viewpoint fragments I (red) and II (blue) located on opposite sides of the breakpoints on chromosome 7. The two captured regions do not directly neighbor each other, demonstrating a deletion on chromosome 4. The breakpoint of derivate chromosome 4 (arrow) was sequenced. *Y* axes represent raw intensities of the microarray signals; maximum (max.) = 65,535 counts. For whole-chromosome views a running mean with a window size of 29 probes was applied.

expected this because the deletion brings the region in closer physical proximity to the viewpoint fragment. We used the 4C data to predict the breakpoints and design primers for PCR amplification across the ~2-Mb deleted region, which allowed us to map the two breakpoints at the base-pair resolution (**Supplementary Fig. 4b**). This indicates that 4C technology can be used to identify homozygous deletions as regions containing reduced hybridization signals across the deleted area in combination with increased hybridization signals on the other side of the deletion.

The 4C technology identifies translocations in cell mixtures

Tumors and tumor samples are often mosaic, and current high-throughput techniques cannot detect rearrangements present in small subpopulations of cells. We applied 4C to cell mixtures containing control cells (K562) and various amounts of HSB-2 cells carrying the t(1;7) described above. We found that even if only 5% of the analyzed pool of cells carried the translocation, this rearrangement could still be detected efficiently (**Fig. 3** and **Supplementary Fig. 5**). Thus, 4C can be used to identify balanced rearrangements in nonhomogeneous samples, enabling early detection of rearrangements in small tumors.

The 4C identifies new chromosomal rearrangements

We next asked whether 4C can be used to easily identify new rearrangements. *TCR* loci are frequently involved in chromosomal rearrangements in T-ALL cells because translocations can arise during the process of variable-diversity-joining (VDJ) recombination. We screened samples from five individuals with T-ALL for new genetic rearrangements associated with the *TRB@* locus. One sample had a translocation between *TRB@* and the p arm of chromosome 12, plus an additional deletion ~3 Mb away from the translocation breakpoint on chromosome 12 (**Fig. 4** and see **Supplementary Fig. 6** for other chromosomes). The translocation t(7;12)(q35;p12.3) is new in T-ALL. We mapped the breakpoint on chromosome 12 at 6-kb resolution, and cloned and sequenced it (**Fig. 4b**). The translocation positioned the enhancer of *TRB@* 70 kb downstream of the still intact LIM domain only gene *LMO3* (**Fig. 4c**). Microarray expression data showed that *LMO3*, normally silenced in T cells, was highly active in the corresponding T-ALL sample (**Supplementary Fig. 7**). The protein family members *LMO1* and *LMO2*, but not *LMO3*, have previously been found as oncogenic translocation partners of the *TCR* loci in T-ALL. Notably, *LMO3* was recently found to act as an oncogene in neuroblastoma[16]. Thus, 4C technology can be used to discover new oncogenes rearranged with frequently modified loci.

Multiplex 4C technology

Finally, we aimed to develop a 4C strategy that would simultaneously identify multiple recurrent rearrangements associated with a given disease using a single microarray (**Fig. 5** and **Supplementary Figs. 8**,**9**). In T-ALL, a set of loci, in particular *TRA@-TRD@*, *TRB@*, *BCL11B* and *MLL*, frequently recombine with various other genes[17,18]. We included these four loci, together with nine other loci that have been described either as their translocation partner or to be rearranged otherwise in T-ALL[17,18], in a 4C-multiplex strategy. The genomic sites interacting with each of the 13 viewpoints were PCR amplified separately and pooled in two mixes representing the chromosomal neighborhoods of 6 and 7 viewpoints, respectively. These mixes were differentially labeled and hybridized to the same microarray. The data show that multi-view 4C accurately identifies rearrangements in each of 10 T-ALL samples analyzed. We identified translocations between *TRD@-HOX11* and *TRD@-LMO2*, translocations between *BCL11B-Nkx-2.5* and *BCL11B-TLX3*, the translocation between *CALM-AF10*, the common SET-Nup214 deletion, a *TRA@–c-myc* translocation and an *MLL-AF-6* translocation

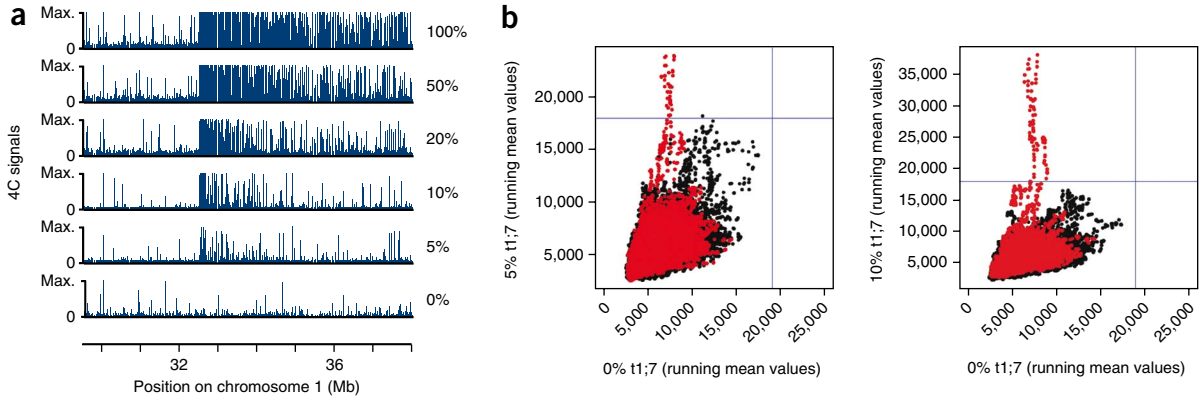

Figure 3 | 4C detects translocations in small subpopulations of cells. (**a**) The 4C technology was applied to mixtures of K562 cells and various amounts of HSB-2 cells, the latter containing translocation t(1;7). Percentages of cells carrying the translocation are indicated on the right. Using a viewpoint neighboring the breakpoint on chromosome 7, sequences on chromosome 1 were captured efficiently in all mixtures, even when only 5% of the cells carried the translocation. Scale on *y* axes are hybridization signal intensities (arbitrary units), with the highest value set to maximum (max.). (**b**) The 4C values from the translocation site clearly separate from other 4C data on *trans* chromosomes. The amount of captured fragments per chromosomal region was determined by calculating running mean values (window size 49 probes, ~300 kb). In the samples containing 5% (left) or 10% (right) HSB-2 cells, higher running mean values were found than in the sample that did not contain cells carrying the t(1;7) translocation (distance to median >3× median, blue line). The high 4C values are located on chromosome 1 (red), next to the breakpoint (**Supplementary Fig. 5**).

(**Fig. 5** and **Supplementary Figs. 8**,**9**). In one sample we found a novel *LMO1-TRB@* translocation (**Supplementary Fig. 9**). Analogous to *HOX11* (ref. 19,20) and *LMO2* (ref. 21), *LMO1* has now been found to translocate to both *TRB@* and *TRA@-TRD@*[17,18]. Retrospectively, we could confirm all these rearrangements by chromosomal karyotyping and/or FISH analyses (J.P.P.M.; data not shown). In another sample (10110) from an individual with T-ALL, we initially failed to identify the deletion previously mapped by array-CGH to locate between the *LMO2* gene and the locus containing the *RAG1*and *RAG2* genes on chromosome 11 (ref. 22). Instead, using a target sequence ~500 kb telomeric of *LMO2*, we found the region to be fused to an area on chromosome 1 that contained the *STIL* and *TAL1* genes, both also implicated in T-ALL (**Supplementary Fig. 10a**). To further map this rearrangement, we applied 4C to sequences on either side of the breakpoints of the two chromosomes. Together, the data revealed a complex chromosomal rearrangement involving a translocation, t(1;11) (**Supplementary Fig. 10a**). The breakpoint on derivate chromosome 11 was flanked by two small (100–200 kb) regions, one of which carried the T-ALL gene *LMO2*, and the other carrying the T-ALL genes *STIL* and *TAL1*. On either side this was followed by a large deleted region spanning 1–3 Mb (**Supplementary Fig. 10b**). Thus, the deletion identified by array-CGH[22] was shown by 4C to be part of a more complex chromosomal rearrangement involving a translocation. Notably, oncogenes like *LMO2* and *TAL1* are known to translocate to the *TCR* loci in T-ALL, but have not been documented previously to rearrange with each other. The fact that these oncogenes also recombine may support the idea that nuclear co-localization of all these loci at some stage of T-cell development is an important mechanism behind translocation partner selection in T-ALL. Collectively, the data showed that multiview 4C identifies many known and new translocations as well as complex chromosomal rearrangements associated with T-ALL with a single microarray. Clearly, this strategy can also be adapted to detect such rearrangements in samples from patients with other diseases.

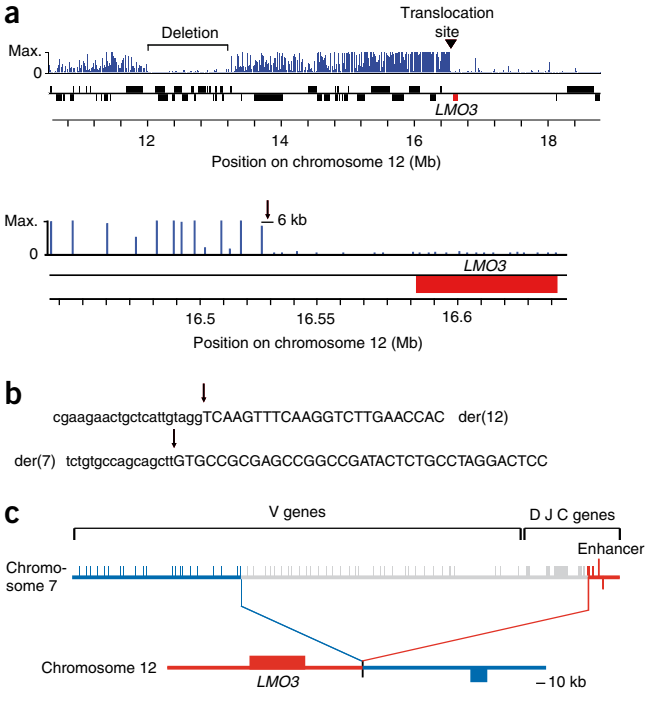

Figure 4 | 4C identifies novel translocation partners. (**a**) Uncharacterized samples from five individuals with T-ALL were screened with 4C, using a viewpoint fragment near the *TRB@* locus on chromosome 7. Shown 4C signals represent fragments on chromosome 12 captured by the viewpoint. Scale on *y* axes are hybridization signal intensities (arbitrary units), with the highest value set to maximum (max.). In one sample, many high 4C signals appeared specifically on chromosome 12, revealing a translocation, t(7;12)(q35;p12.3). A deletion is present several megabases from the translocation site (arrow) on chromosome 12 (top). The translocation site is present in a 6 kb region close to *LMO3* (bottom). (**b**) Sequences of both breakpoints of t(7;12)(q35;p12.3); nucleotides in upper case are from chromosome 12, in lower case from chromosome 7 and in italics are from unknown origin. (**c**) Schematic representation of the translocation site of t(7;12)(q35;p12.3). The enhancer of *TRB@* is positioned 70 kb downstream of the *LMO3* gene.

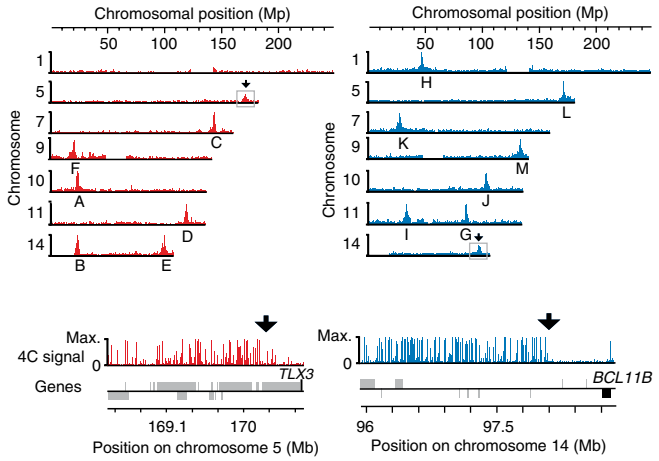

Figure 5 | 4C can be used to analyze multiple sites frequently involved in rearrangements in T-ALL cells on one microarray. The 4C analysis was performed on 13 viewpoints separately: A, *AF10*; B, TRA@-TRD@; C, TRB@; D, *MLL*; E, *BCL11B*; F, *CDKN2A-CDKN2B*; G, *CALM*; H, *TAL1*; I, *LMO2*; J, *HOX11*; K, *HOXA*; L, *TLX3*; and M, *NUP214*. PCR products are combined in two pools. A–F were labeled with Cy-5 (left) and G–M were labeled with Cy-3 (right), and both pools were hybridized to a microarray. The 4C data show high signals on all the viewpoints included in the pools. Additional signals in the Cy-5 pool (boxed) on chromosome 5 demonstrate *TLX3* is a translocation partner of one of the Cy-5 viewpoints. In the Cy-3 pool *BCL11B* was found as a translocation partner (square). Together the data demonstrate the tested sample carries a *BCL11B-TLX3* translocation.

DISCUSSION

The idea to use 4C technology to map genomic rearrangements is based on our observation that 4C technology predominantly identifies large genomic regions flanking the viewpoint[9,10]. Other groups have published similar strategies to uncover long-range DNA interactions in the nucleus[23–26], but these studies did not clearly reveal this important feature of 4C, probably because not enough captured sequences had been analyzed. To prevent the description of an anecdotal collection of interacting DNA elements, we recommend that for each viewpoint sufficient numbers of captured sequences are analyzed by 4C.

Folding of chromosomes is indeed not random, and some regions of the genome are captured and identified by 4C owing to the 3D structure of the chromosomes[9]. When interpreting 4C data, chromosomal rearrangements can easily be discerned from looped chromatin structures. First, signal intensities seen for genomic regions that are physically close on the linear chromosome template are much higher than those observed for distant regions that loop in 3D to the genomic viewpoint. Second, signal profiles from translocations, inversions and deletions each have a very distinct shape that is different from the more Gaussian curves seen for looped regions. Most distinctly, chromosomal rearrangements will yield a sharp transition in signal intensities at the probes that surround the breakpoint. Third and most importantly, the number of probes with positive signals near genomic breakpoints is usually much higher than observed for a region which loops toward the viewpoint. In one example, we found this to be true even in a cell mixture with only 5% of the cells carrying the translocation. Thus, it is not difficult to discriminate structural variation in the genome from 3D configurations detected by 4C. It is important that sufficient numbers of ligation products are analyzed simultaneously. Each diploid cell donates a maximum of two ligation products per viewpoint. For singleplex 4C, we therefore routinely pool at least 15 PCRs, each performed on 200 ng DNA template, for hybridization on a microarray, meaning that we simultaneously analyze ~0.5 million cells, or ~1 million interactions with the viewpoint. Under such conditions, translocations and other rearrangements can be readily identified, even when the breakpoint is 3 Mb away from the viewpoint or present in small subpopulations of cells. In the 4C-multiplex strategy, aimed to simultaneously screen at several sites for structural rearrangements, we performed fewer PCRs per viewpoint and PCR products from different viewpoints were mixed, reducing the net amount of PCR material per viewpoint hybridized to the array. As a result, signal intensities and the number of positive probes identified at the breakpoint dropped, but in each case we could still identify the underlying rearrangement, as confirmed by FISH analysis and by additional 4C experiments. In addition to the complexity of the sample, the distance between the breakpoint and the viewpoint is important. We recommend defining a viewpoint every 3 Mb when screening a large genomic region for breakpoints. The resolution provided by 4C technology is limited by the density of restriction enzyme digestion sites. Here average resolution was 7 kb, but with the same restriction enzyme (HindIII) this can be improved to 4–5 kb when using a higher-resolution array. Cross-hybridization to probes on a microarray always causes undesired background. However, this usually occurs at probes randomly distributed over the genome. It therefore has little impact on the detection of rearrangements by 4C, which depends on chromosomal clustering of probes with strong hybridization signals.

As 4C identifies rearrangements based on the capture of many genomic fragments across the breakpoint, it can be used to uncover balanced rearrangements when breakpoints are present in, or are surrounded by, repetitive sequences. Deletions are currently better detected by, for example, array-CGH, but if deletions are associated with (unbalanced) translocations, they are readily identified by 4C as well and accurately mapped to one of the two derivative chromosomes (**Figs. 2b**, **4** and **Supplementary Fig. 10**). The same is true for large (balanced) inversions, but smaller inversions will be more difficult to identify. Future systematic analyses of samples with inversions varying in size should reveal the detection limit, but we currently estimate that inversions smaller than ~1 Mb cannot be detected by 4C. Possibly the future use of next-generation sequencing instead of microarrays will provide a more quantitative analysis of captured sequences that may improve the detection limit of small rearrangements.

Paired-end sequencing is extremely powerful for genome-wide analysis of both unbalanced and balanced chromosomal rearrangements[7,8]. However, the repetitive nature of the human genome makes the detection of balanced rearrangements by paired-end mapping not trivial, particularly when they are present only in a subpopulation of cells. The 4C technology enables a more focused approach, screening for rearrangements in large genomic regions around candidate loci. Such candidate loci can be regions suspected to carry a rearrangement based on low-resolution techniques such as FISH and chromosomal karyotyping. They may be loci recurrently involved in rearrangements such as the T-cell receptor loci in leukemia samples and the immunoglobulin loci in lymphoma samples. These are large

loci that can carry a break anywhere in a region of up to several megabases in size, making it very difficult to design ligation-mediated PCR strategies for the identification of rearrangement partners. Finally, candidate loci may represent genes that are aberrantly expressed without apparent variation in their DNA copy number[27]. We expect that a systematic analysis by 4C focusing on such genes in many samples will lead to the identification of new balanced rearrangements. In this respect, it is relevant to mention that 4C can also be applied to analyze solid tumor material (data not shown).

Array painting is another recently developed technique for fine-mapping translocation breakpoints. It involves isolating chromosomes based on size using flow-sorting and characterization of a selected (derivative) chromosome by hybridization to a microarray or by large-scale sequencing[28,29]. However, not all chromosomes and chromosome derivatives can be isolated based on size, and inversions cannot be detected using this technique. Finally, the fact that 4C can be used to detect translocations present in small subpopulations of cells makes it a potent technique to study rearrangements in mixtures of cells and mosaic tumors and creates the prospect of identifying tumor cells in early stages of metastasis.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemethods/.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank A. de Klein, W. van Ijcken, J. Gladdines, J. Veltman, A. Hoischen and E. Splinter for assistance. This work was supported by grants from the Dutch Cancer Society (KWF-EMCR 2006-3500) to J.P.P.M., grants from the European Union and the Cancer Genomics Centre to F.G., grants from the Dutch Scientific Organization (NWO) (912-04-082), Netherlands Genomics Initiative (050-71-324) and the Erasmus MC to W.d.L.

AUTHOR CONTRIBUTIONS

M.S. designed, performed and analyzed experiments and wrote the manuscript; P.K. performed and analyzed experiments; I.H. helped perform experiments; R.-J.G.: provided PAP cell line and helped design the PAP experiment; E.-J.R. designed the microarray. F.G. helped design experiments; J.P.P.M. provided leukemia samples, helped design T-ALL experiments and helped write the manuscript; and W.d.L. designed experiments, supervised the project and wrote the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturemethods/.

Published online at http://www.nature.com/naturemethods/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Iafrate, A.J. *et al.* Detection of large-scale variation in the human genome. *Nat. Genet.* **36**, 949–951 (2004).
2. Mehan, M.R., Freimer, N.B. & Ophoff, R.A. A genome-wide survey of segmental duplications that mediate common human genetic variation of chromosomal architecture. *Hum. Genomics* **1**, 335–344 (2004).
3. Sharp, A.J., Cheng, Z. & Eichler, E.E. Structural variation of the human genome. *Annu. Rev. Genomics Hum. Genet.* **7**, 407–442 (2006).
4. Easton, D.F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087–1093 (2007).
5. Eichler, E.E. *et al.* Completing the map of human genetic variation. *Nature* **447**, 161–165 (2007).
6. Feuk, L., Carson, A.R. & Scherer, S.W. Structural variation in the human genome. *Nat. Rev. Genet.* **7**, 85–97 (2006).
7. Campbell, P.J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet.* **40**, 722–729 (2008).
8. Korbel, J.O. *et al.* Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**, 420–426 (2007).
9. 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).
10. Simonis, M., Kooren, J. & de Laat, W. An evaluation of 3C-based methods to capture DNA interactions. *Nat. Methods* **4**, 895–901 (2007).
11. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
12. Burnett, R.C., Thirman, M.J., Rowley, J.D. & Diaz, M.O. Molecular analysis of the T-cell acute lymphoblastic leukemia-associated t(1;7)(p34;q34) that fuses LCK and TCRB. *Blood* **84**, 1232–1236 (1994).
13. Soulier, J. *et al.* HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). *Blood* **106**, 274–286 (2005).
14. Speleman, F. *et al.* A new recurrent inversion, inv(7)(p15q34), leads to transcriptional activation of HOXA10 and HOXA11 in a subset of T-cell acute lymphoblastic leukemias. *Leukemia* **19**, 358–366 (2005).
15. Galjaard, R.J. *et al.* Isolated postaxial polydactyly type B with mosaicism of a submicroscopic unbalanced translocation leading to an extended phenotype in offspring. *Am. J. Med. Genet. A.* **121**, 168–173 (2003).
16. Aoyama, M. *et al.* LMO3 interacts with neuronal transcription factor, HEN2, and acts as an oncogene in neuroblastoma. *Cancer Res.* **65**, 4587–4597 (2005).
17. Armstrong, S.A. & Look, A.T. Molecular genetics of acute lymphoblastic leukemia. *J. Clin. Oncol.* **23**, 6306–6315 (2005).
18. Graux, C., Cools, J., Michaux, L., Vandenberghe, P. & Hagemeijer, A. Cytogenetics and molecular genetics of T-cell acute lymphoblastic leukemia: from thymocyte to lymphoblast. *Leukemia* **20**, 1496–1510 (2006).
19. Hatano, M., Roberts, C.W. & Minden, M. Crist. W.M. and Korsmeyer, S.J. Deregulation of a homeobox gene, HOX11, by the t(10;14) in T cell leukemia. *Science* **253**, 79–82 (1991).
20. Kennedy, M.A. *et al.* HOX11, a homeobox-containing T-cell oncogene on human chromosome 10q24. *Proc. Natl. Acad. Sci. USA* **88**, 8900–8904 (1991).
21. Boehm, T. Foroni, L., Kaneko, Y., Perutz, M.F. & Rabbitts, T.H. The rhombotin family of cysteine-rich LIM-domain oncogenes: distinct members are involved in T-cell translocations to human chromosomes 11p15 and 11p13. *Proc. Natl. Acad. Sci. USA* **88**, 4367–4371 (1991).
22. Van Vlierberghe, P. *et al.* The cryptic chromosomal deletion del(11)(p12p13) as a new activation mechanism of LMO2 in pediatric T-cell acute lymphoblastic leukemia. *Blood* **108**, 3520–3529 (2006).
23. Ling, J.Q. *et al.* CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science* **312**, 269–272 (2006).
24. Lomvardas, S. *et al.* Interchromosomal interactions and olfactory receptor choice. *Cell* **126**, 403–413 (2006).
25. Wurtele, H. & Chartrand, P. Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. *Chromosome Res.* **14**, 477–495 (2006).
26. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* **38**, 1341–1347 (2006).
27. Tomlins, S.A. *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644–648 (2005).
28. Chen, W. *et al.* Mapping translocation breakpoints by next-generation sequencing. *Genome Res.* **18**, 1143–1149 (2008).
29. Fiegler, H. *et al.* Array painting: a method for the rapid analysis of aberrant chromosomes using DNA microarrays. *J. Med. Genet.* **40**, 664–670 (2003).

ONLINE METHODS

Restriction-fragment paired-end sequencing. First, 10 μg of genomic DNA was digested in 500 μl with 10 U of an enzyme that recognizes 6 bases (HindIII, BglII or EcoRI) (37 °C for 2 h). Samples were purified by phenol-chloroform extraction and ethanol precipitation. Then samples were ligated in 2 ml with 40 U of ligase (Roche) for 4 h at 16 °C and 30 min at 20 °C. Ligated samples were purified by phenol-chloroform extraction and ethanol precipitation. A second digestion was performed with a restriction enzyme that recognizes 4 bases (for example, NlaIII or DpnII) under the same conditions as described for the 6-base–recognizing enzyme. Subsequent ligation was also as described above. Samples were purified by phenol-chloroform extraction and ethanol precipitation. Selected fragments were PCR amplified from 50–100 ng of DNA, using the following conditions: 94 °C for 3 min, followed by 30 cycles of 15 s at 94 °C, 1 min at 55 °C and 2 min at 72 °C, and one final step of 7 min at 72 °C.

General guidelines for the design of 4C experiments. It is crucial to process a sufficient number of cells, at least 1×10^7 cells. The complexity of the sample needs to be high, meaning that sufficient numbers of ligation products need to be analyzed. We routinely analyzed the equivalent of 1×10^6 ligation events per 4C experiment. For this, 16 PCRs were performed, each on 200 ng of 4C template (total amount of template, 3.2 mg). These PCRs were pooled, labeled and subsequently hybridized to one microarray.

If chromosomal rearrangement partners and breakpoints are to be identified near candidate genes, the target sequence (viewpoint) should ideally designed within 200 kb from the gene. For the mapping of chromosomal breakpoints in large, cytogenetically defined, chromosomal regions, target sequences (viewpoints) preferably should be defined every 3 Mb (or closer together), such that together they span the entire region. The 4C inverse PCR primers should be designed according to standard rules for PCR primer design. Primers used in this study are listed in **Supplementary Table 1**.

The 4C array design. The 60 bp probes were designed within 100 bp from a HindIII site, using criteria described previously, for example, the selection of only unique DNA sequences[9]. To cover the entire genome with the 385,000 probes that fit on the Nimblegen microarray, probe numbers were first reduced by keeping only one probe per HindIII fragment, instead of one on each side; second, probes were selected such that the spacing of probes was as equal as possible across the genome.

Sample preparation. Samples were processed as described[9,30]. The EBV-transformed cell line derived from the individual with PAP was grown and handled as described previously[9,15]. The patients' parents or their legal guardians provided informed consent to use leftover material for reserarch purposes in accordance with the rules of the review board of Erasmus MC and the Declaration of Helsinki.

Formaldehyde cross-linking and cell lysis. The procedure used was essentially as described previously[9,10]. We cross-linked 1×10^7 cells for 10 min at room temperature (18–20 °C) in 9.5 ml phosphate-buffered saline (PBS) (pH7.4) containing 2% formaldehyde and 10% FCS. Tubes were transferred to ice and 1.425 ml of 1 M ice-cold glycine was added to quench the cross-linking reaction, followed immediately by centrifugation for 8 min at 225*g* at 4 °C. Supernatant was carefully removed and cell pellet was dissolved in 5 ml cold lysis buffer (10 mM Tris-HCl (pH 7.5); 10 mM NaCl; 5 mM $MgCl_2$; 0.1 mM EGTA; 1× complete protease inhibitor (Roche) and incubated for 10–30 min on ice. Depending on the cell type, more stringent lysis buffers may be needed to prepare nuclei, which were collected by centrifugation for 5 min at 400*g* at 4 °C. The pelleted nuclei, without supernatant, can be frozen with liquid nitrogen and stored at −80 °C for several months.

Digestion and ligation. Nuclei were washed once in restriction buffer and taken up in 0.5 ml of 1.2× restriction buffer. Tubes were placed at 37 °C, and 7.5 μl 20% SDS (final, 0.3% SDS) was added and the tubes were incubated for 1 h at 37 °C while shaking gently, followed by the addition of 50 μl 20% Triton X-100 (final, 2% Triton X-100) and incubation for another 1 h at 37 °C while shaking gently. A 5-μl aliquot of the sample was taken as the 'undigested control'. This sample may be stored at −20 °C until it is needed to determine the digestion efficiency. We added 400 U of the selected restriction enzyme to the remaining sample and incubated it overnight at 37 °C with gentle shaking. The next day, a 5 μl aliquot of the sample was taken as the 'digested control', de-cross-linked by incubation with 10 μl Proteinase K in 90 μl of 10 mM Tris (pH 7.5) at 65 °C for 1 h. Digestion efficiency was roughly determined by running the sample on a 0.6% agarose gel and checking for a smear of DNA fragments with the majority of the fragments between 5–10 kb. If digestion was sufficient, 40 μl of 20% SDS (final, 1.6%) was added to the remaining sample for 20 to 25 min at 65 °C (shake gently). The digested nuclei were transferred to a 50 ml falcon tube and 6.125 ml of 1.15× ligation buffer together with 375 μl of 20% Triton X-100 (final, 1% Triton X-100) was added. After incubation for 1 h at 37 °C while shaking gently, 5 μl ligase HC 20 U μl^{-1} (100 U total) was added and ligation proceeded for 4 h at 16 °C followed by 30 min at room temperature.

DNA purification. DNA cross-links were reversed by the addition of 30 μl of 10 mg ml^{-1} Proteinase K (300 μg final) and incubation at 65 °C overnight. The next morning 30 μl of 10 mg ml^{-1} RNAse (300 μg final) was added and the reactions were incubated for 30–45 min at 37 °C, followed by phenol extraction and DNA purification as described[10]. The DNA pellet, referred to as 3C template, was dissolved in 150 μl of 10 mM Tris (pH 7.5).

Preparation of 4C template. DNA concentration of the 3C template was determined as described[10]. To prepare the 4C template, the 3C template was digested overnight with a frequent-cutter enzyme (for example, DpnII or NlaIII), at a DNA concentration of 100 ng μl^{-1} (total amount of DNA from a 1×10^7 cells is estimated at 60 μg) using 1 unit enzyme per microgram of DNA followed by heat-inactivation of the enzyme. DNA was purified by phenol extraction and DNA precipitation as described[10], and pellet was dissolved in 100 μl milli-Q. Digestion efficiency was checked by taking an aliquot (5 μl) and separating it on a 1.5% agarose gel. Typically, this yielded a smear of products with the majority around 300 bp. After digestion, ligation was carried out in 14 ml total volume with 200 U of ligase, for 4 h at 16 °C,

plus 30 min at room temperature. DNA was purified by phenol extraction and DNA precipitation as described previously[10]. Pellets were dissolved in 75 µl of 10 mM Tris and incubated at 37 °C for at least 30 min to dissolve.

Testing the 4C PCR and template. To check the linearity of the 4C PCR, a titration series was done. Inputs of 25, 50, 100 and 200 ng of 4C sample were tested. Reaction conditions were as follows: 5 µl 10× Buffer 1 (Expand Long Template system, Roche), 1 µl dNTPs (10 mM each), 10 µl of primer pair (25 pmol each), 0.75 µl DNA polymerase (Expand Long Template system), *x* µl sample (volume representing the required amount of template DNA) and H_2O to have a total volume of 50 µl. PCR parameters are as follows: 1 cycle 94 °C for 2 min, followed by 30 cycles of 94 °C for 15 s, 55 °C for 1 min, 68 °C for 3 min, followed by 1 cycle of 68 °C for 7 min.

PCR products are analyzed on a 1.5% agarose gel and three things should be checked (see ref. 10): (i) the 4C PCR product should appear as a smear on gel, typically with two prominent bands: one representing the 10–20% of the investigated restriction site that was not digested and one representing the events where the investigated fragment was ligated to its own end. (ii) Lighter bands should give a reproducible pattern in separate PCR. (iii) With a linear increase in amount of input, the increase in PCR product should be linear. The product can be quantified, by quantifying one of the prominent bands you see on gel, using ImageQuant software.

4C PCR. The highest amount of 4C template (being input DNA) for which the PCR is still linear is used, allowing a maximum of 200 ng template per reaction. As many PCRs were done as necessary to amplify a total amount of input of 3.2 µg. For example, if the reaction is linear up to 200 ng input, we performed 3.2 µg / 200 ng = 16 PCRs. Reaction conditions and PCR program are as described above. PCR products are pooled and purified using Qiagen PCR cleanup kit (2 columns per pooled sample, maximum binding capacity is 10 µg). DNA is eluted in H_2O (pH 7.5–8.0). Samples are checked on 1.5% agarose gel. Labeling and hybridization are done according to the standard ChIP-chip procedures of Nimblegen.

Data analysis. The 4C data were visualized using SignalMap software (Roche Nimblegen). To create whole chromosome view pictures of the 4C data and identify chromosomal rearrangement partners, unprocessed signal intensities were taken and a running mean with a window size of 29 probes was calculated using the R package (http://www.r-project.org/). Data were plotted along the chromosome axis and visually inspected. Unusual clustering of 4C positive signals identified the chromosomal rearrangement partner. Subsequent visual inspection of the unprocessed 4C signals at these chromosomal sites predicted the position of the breakpoint at a resolution of 7 kb, being the average distance between probes on the microarray. The 4C data were processed further for samples in which clustering of 4C positive signals was less apparent on visual inspection. The 4C datasets were normalized by quantile normalization. A running mean with a window size of 49 probes was applied to the normalized data. A median of the running mean values was calculated for each sample. A distance of three times the median from the median was set as threshold. Probes with running mean values above the threshold were found at the translocation site.

30. van Vlierberghe, P. *et al.* A new recurrent 9q34 duplication in pediatric T-cell acute lymphoblastic leukemia. *Leukemia* **20**, 1245–1253 (2006).

doi:10.1038/nmeth.1391