



# Genome-Wide Analysis of Interchromosomal Interaction Probabilities Reveals Chained Translocations and Overrepresentation of Translocation Breakpoints in Genes in a Cutaneous T-Cell Lymphoma Cell Line

## OPEN ACCESS

### Edited by:

Michael Breitenbach,  
University of Salzburg, Austria

### Reviewed by:

Luisa Lanfrancione,  
Istituto Europeo di  
Oncologia s.r.l., Italy  
Wei Chen,  
Southern University of  
Science and Technology, China

### \*Correspondence:

Reinhard Ullmann  
reinhard1ullmann@bundeswehr.org

### \*Present address:

Anne Steinger,  
Steglitz, Berlin, Germany;  
Grit Ebert,  
TMF – Technology, Methods,  
and Infrastructure for Networked  
Medical Research, Berlin, Germany

### Specialty section:

This article was submitted  
to Molecular  
and Cellular Oncology,  
a section of the journal  
Frontiers in Oncology

**Received:** 14 November 2017

**Accepted:** 09 May 2018

**Published:** 30 May 2018

### Citation:

Steinger A, Ebert G, Becker BV,  
Assaf C, Möbs M, Schmidt CA,  
Grabarczyk P, Jensen LR,  
Przybylski GK, Port M, Kuss AW and  
Ullmann R (2018) Genome-Wide  
Analysis of Interchromosomal  
Interaction Probabilities Reveals  
Chained Translocations  
and Overrepresentation of  
Translocation Breakpoints in  
Genes in a Cutaneous T-Cell  
Lymphoma Cell Line.  
*Front. Oncol.* 8:183.  
doi: 10.3389/fonc.2018.00183

Anne Steinger<sup>1†</sup>, Grit Ebert<sup>1†</sup>, Benjamin V. Becker<sup>2</sup>, Chalid Assaf<sup>3</sup>, Markus Möbs<sup>4</sup>, Christian A. Schmidt<sup>5</sup>, Piotr Grabarczyk<sup>5</sup>, Lars R. Jensen<sup>6</sup>, Grzegorz K. Przybylski<sup>7</sup>, Matthias Port<sup>2</sup>, Andreas W. Kuss<sup>6</sup> and Reinhard Ullmann<sup>2\*</sup>

<sup>1</sup>Max Planck Institute for Molecular Genetics, Berlin, Germany, <sup>2</sup>Bundeswehr Institute of Radiobiology Affiliated to the University of Ulm, Munich, Germany, <sup>3</sup>Department of Dermatology and Venerology, Helios Klinikum Krefeld, Krefeld, Germany, <sup>4</sup>Berlin Institute of Health, Institute of Pathology, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin, Germany, <sup>5</sup>Clinic for Internal Medicine C, University Medicine Greifswald, Greifswald, Germany, <sup>6</sup>Human Molecular Genetics, Department of Functional Genomics, University Medicine Greifswald, Greifswald, Germany, <sup>7</sup>Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland

In classical models of tumorigenesis, the accumulation of tumor promoting chromosomal aberrations is described as a gradual process. Next-generation sequencing-based methods have recently revealed complex patterns of chromosomal aberrations, which are beyond explanation by these classical models of karyotypic evolution of tumor genomes. Thus, the term chromothripsis has been introduced to describe a phenomenon, where temporarily and spatially confined genomic instability results in dramatic chromosomal rearrangements limited to segments of one or a few chromosomes. Simultaneously arising and misrepaired DNA double-strand breaks are also the cause of another phenomenon called chromoplexy, which is characterized by the presence of chained translocations and interlinking deletion bridges involving several chromosomes. In this study, we demonstrate the genome-wide identification of chromosomal translocations based on the analysis of translocation-associated changes in spatial proximities of chromosome territories on the example of the cutaneous T-cell lymphoma cell line Se-Ax. We have used alterations of intra- and interchromosomal interaction probabilities as detected by genome-wide chromosome conformation capture (Hi-C) to infer the presence of translocations and to fine-map their breakpoints. The outcome of this analysis was subsequently compared to datasets on DNA copy number alterations and gene expression. The presence of chained translocations within the Se-Ax genome, partly connected by intervening deletion bridges, indicates a role of chromoplexy in the etiology of this cutaneous T-cell lymphoma. Notably, translocation breakpoints were significantly overrepresented in genes, which highlight gene-associated biological

processes like transcription or other gene characteristics as a possible cause of the observed complex rearrangements. Given the relevance of chromosomal aberrations for basic and translational research, genome-wide high-resolution analysis of structural chromosomal aberrations will gain increasing importance.

**Keywords:** chromosome conformation capture, chromoplexy, chromosomal translocations, deep sequencing, cutaneous T-cell lymphoma

## INTRODUCTION

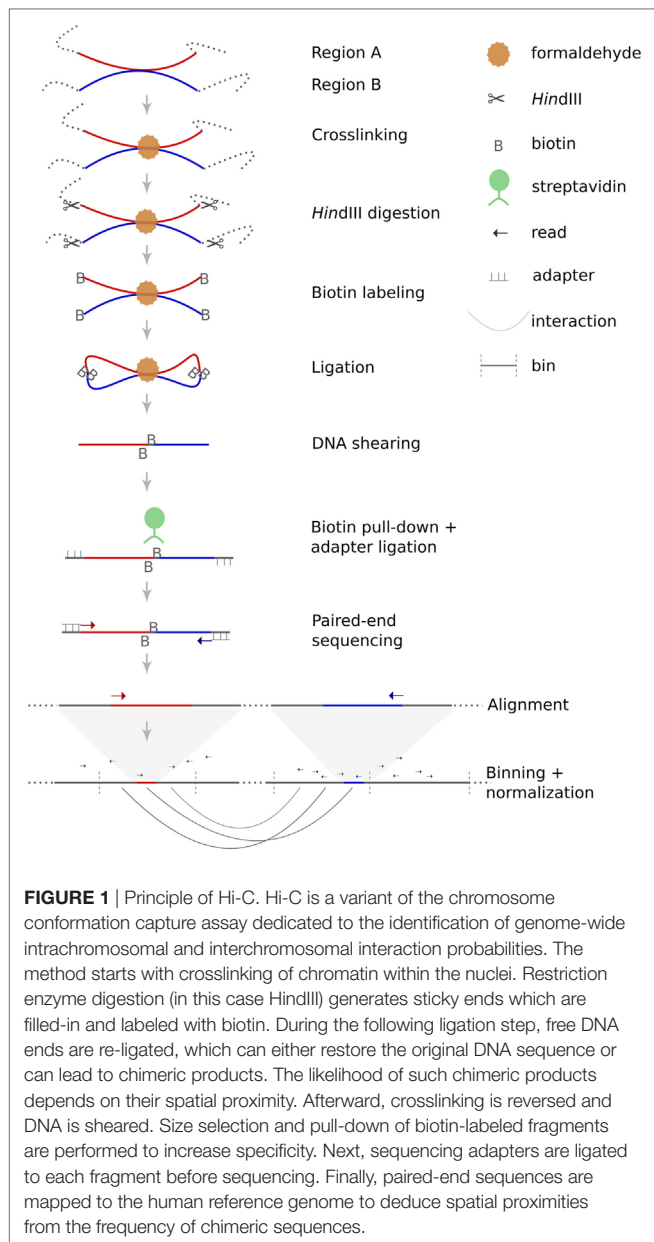
The analysis of structural chromosomal aberrations is of relevance for both basic and translational research. Several chromosomal markers are already routinely used in clinical tests for genotype-based sub-classification of tumors or to assist in therapeutic decisions. In addition, the identification of recurrent aberrations can highlight driver genes of tumorigenesis, which represent promising starting points for the development of targeted therapies. Apart from clinical applications, the characterization of chromosomal aberrations can shed light on the underlying mutational mechanisms and in this way contribute to a better understanding of the cause and course of intra-individual evolution of tumors.

According to classical models of tumorigenesis, complex abnormal karyotypes emerge through the stepwise acquisition of chromosomal rearrangements followed by expansion of those mutated clones with highest proliferative capacity (1). Yet, the conception of intra-individual karyotypic evolution has been biased by the limited perspective as provided by the low resolution of genome-wide datasets on structural chromosomal rearrangements for a long time. The lack of appropriate techniques capable of capturing karyotypic complexity both genome-wide and with high resolution has hampered the identification of mechanisms alternative to the well-documented gradual process. The introduction of array-based comparative genomic hybridization [(arrayCGH) (2, 3)] has mitigated this technical shortcoming for unbalanced structural chromosomal aberrations, but the situation has remained unsatisfactory for balanced chromosomal rearrangements. Until recently, their characterization required time-consuming cloning of breakpoints or, in case of translocations, depended on sophisticated sorting of derivative chromosomes followed by hybridization of sorted chromosomes on DNA microarrays (4–6). This situation has changed with the advent of next-generation sequencing (NGS), which has set the stage for the development of new protocols for the analysis of structural chromosome aberrations (7). Initially, these analyses have mainly focused on alterations of sequencing depth across the genome or along sorted chromosomes to define DNA copy number changes and translocation breakpoints (8), respectively. Later protocols have taken advantage of paired-end reads and used their mapping position and orientation with respect to the human reference genome to infer the presence and location of structural chromosome aberrations (9). Despite the development of paired-end NGS protocols, the identification of structural chromosomal aberrations such as balanced translocations has remained challenging. This is mainly due to the fact that strategies based on standard paired-end sequencing protocols have to rely on those

few sequenced chimeric fragments that span the chromosomal breakpoints. Hence, reliable detection of such rearrangements using standard NGS protocols requires considerable sequencing depth (10). Furthermore, even in case of sufficient sequencing depth, translocation breakpoints in the very vicinity of regions with low mappability, such as repetitive elements, segmental duplications or DNA segments with extreme bias of base composition might be missed or erroneously aligned (11). An alternative strategy capable to overcome these problems is based on the fact that chromosomal rearrangements such as translocations disrupt nuclear architecture and modify spatial proximities of chromosome territories. These modifications of nuclear organization can be monitored by chromosome conformation capture assays such as Hi-C (12–14). This technique combines proximity ligation and NGS to infer nuclear neighborhood of chromosomal regions (see **Figure 1** for explanation). The closer two chromosomal segments are within the nucleus, the more frequent Hi-C will detect interactions between them.

In general, interaction frequencies between two regions on the same chromosome decrease with linear distance and interactions between two chromosomes are considerably rare when compared to intra-chromosomal ones. In case of a translocation, regions of two or more chromosomes come into close contact. This results in an abrupt increase of interaction frequencies between the segments adjacent to the translocation breakpoints and makes Hi-C an ideal approach for the detection of balanced translocations (10, 15–17).

The application of array- and sequencing-based approaches as described above unveiled an unprecedented complexity of structural chromosomal aberrations in tumor genomes. In several cases, the observed mutational patterns were hardly compatible with a stepwise accumulation of chromosomal aberrations as described by the classical model of tumorigenesis (18). For example, regionally confined clusters of numerous chromosomal aberrations with limited DNA copy number states in the absence of general genomic instability have suggested a single catastrophic event as the underlying cause of this complex pattern of aberrations instead of a series of consecutive events. Meanwhile, it has been shown that this phenomenon, termed chromothripsis, can be encountered in a broad range of tumor types, where it can affect 2–3% of patients (19). In 2015, Zhang and colleagues succeeded to demonstrate that micronuclei formation and DNA damage confined to these structures can produce similar patterns of chromosomal changes as typical for chromothripsis (20). Multiple simultaneously arising DNA double-strand breaks also account for the emergence of chromoplexy, which is characterized by chained translocations and interlinking deletion bridges involving numerous chromosomes (21). The complex patterns of



chromosomal aberrations typical for chromoplexy are unlikely result of a stepwise process, as this would require the repeated use of the same chromosomal breakpoints (18, 21). It is still unclear what triggers these simultaneously arising DNA strand breaks in the context of chromoplexy (18).

In this study, we have employed chromosomal interaction probabilities to fine-map translocations in a cell line derived from a patient with Sézary syndrome. The etiology of this highly malignant cutaneous T-cell lymphoma is poorly understood (22) and thorough investigation of structural chromosomal aberrations promises more insights into its development and progression (23). We demonstrate the presence of chained translocations partly connected by interlinking deletion bridges, which suggests the manifestation of chromoplexy and argues against a gradual appearance of these chromosomal

aberrations. The overrepresentation of chromosomal translocation breakpoints within genes highlights the possible impact of spatial proximities of genes and biological processes associated with genes on the emergence of chromosomal translocations in this cutaneous T-cell lymphoma.

## MATERIALS AND METHODS

Hi-C is a high-throughput variant of the chromosome conformation capture assay and facilitates the genome-wide investigation of interaction probabilities of genomic segments within the nucleus. The technology is based on crosslinking of chromatin, fragmentation of DNA followed by re-ligation. Depending on their spatial distribution, not only the original DNA fragments but also fragments in spatial proximity will re-ligate. These chimeric fragments can be detected by paired-end sequencing and their frequency can be used to calculate the interaction probability of these fragments within the nucleus. The principle of Hi-C is schematically depicted in **Figure 1**. For this study, we have used a Hi-C protocol published in detail by Lieberman-Aiden and colleagues (13).

### Fixation, Cell Lysis, and Restriction Enzyme Digestion

In brief, 20–25 million Se-Ax cells (24) were cross-linked with formaldehyde (Thermo Fisher Scientific, Waltham, MA, USA). Crosslinking was stopped by addition of 125 mM glycine (Merck Millipore, Darmstadt, Germany). After washing cells in ice-cold DPBS buffer (Lonza, Basel, Switzerland), cell pellets were flash frozen and stored at  $-80^{\circ}\text{C}$ . For lysis, cells were resuspended in Hi-C lysis buffer and lysed using a Dounce homogenizer (Fisher Scientific GmbH, Schwerte, Germany). After centrifugation, pellets were washed twice in NEB buffer 2 (New England Biolabs, Ipswich, MA, USA), finally resuspended in 370  $\mu\text{l}$  NEB buffer 2 and 50  $\mu\text{l}$  were transferred to seven tubes each. In order to remove proteins not cross-linked to DNA, 38  $\mu\text{l}$  1% SDS (Sigma-Aldrich, St. Louis, MO, USA) was added to each tube and incubated for 10 min at  $65^{\circ}\text{C}$ . Afterward, SDS was inactivated by the addition of 44  $\mu\text{l}$  10% Triton X-100 (Sigma-Aldrich, St. Louis, MO, USA). In each but one tube 400 U HindIII (New England Biolabs, Ipswich, MA, USA) were added and DNA was digested overnight at  $37^{\circ}\text{C}$  with rotation. The next day, the tube with undigested DNA and one HindIII treated sample were removed to verify HindIII digestion efficiency.

### Endlabeling, Re-Ligation, Reversal of Crosslinking, and DNA Purification

For the remaining tubes a fill-in reaction was performed to blunt the sticky ends as generated by HindIII digestion. For later enrichment of re-ligated fragments, Biotin-dCTP (Invitrogen, Carlsbad, CA, USA) was incorporated during this fill-in reaction. Fragments were re-ligated with 15 U T4 DNA ligase per tube for 4 h at  $16^{\circ}\text{C}$ . Reversal of crosslinking was made by addition of 25  $\mu\text{l}$  Proteinase K (20 mg/ml, Thermo Fisher Scientific, Waltham, MA, USA) and incubation overnight at  $65^{\circ}\text{C}$ . After RNA digestion with 50  $\mu\text{l}$  RNase A (10 mg/ml, Thermo Fisher Scientific, Waltham, MA,

USA) for 45 min at 37°C, DNA was purified by means of standard phenol chloroform isoamylalcohol treatment (Sigma-Aldrich, St. Louis, MO, USA) and ethanol (Merck Millipore, Darmstadt, Germany) precipitation. DNA of the separate tubes was conflated and concentration measured with a Qubit fluorometric assay (Thermo Fisher Scientific, Waltham, MA, USA).

## Removal of Biotin From Unligated DNA Ends, Enrichment of Re-Ligated Fragments and Preparation of Sequencing Libraries

In order to remove biotin-labeling from unligated fragments, samples were treated for 2 h at 12°C with 5 U T4 DNA polymerase (New England Biolabs, Ipswich, MA, USA), whose exonuclease activity removed the biotin at the ends of the unligated fragments while keeping the centrally positioned biotin of ligated fragments untouched. Library generation was done according to the manufacturer's protocols, with minor adaptations concerning the pull-down of biotin-labeled fragments to eliminate unligated fragments. In brief, 5 µg DNA was sheared with the Covaris S2 system (Covaris, Woburn, MA, USA), DNA end-repaired and size selected by means of Agencourt AMPure XP Reagent beads (Beckman Coulter Genomics, Danvers, MA, USA). Afterward, size and quantity was verified employing a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Unligated fragments were depleted by pull-down of biotin-labeled fragments with 50 µl streptavidin beads (10 mg/ml, Life Technologies, Carlsbad, CA, USA). The resulting DNA-coated beads were resuspended in 34 µl TE buffer (Life Technologies, Carlsbad, CA, USA). After A-tailing, i.e., addition of a dATP to the repaired DNA ends, and ligation of sequencing adapters, 5 µl of DNA-coated beads were used for 10 cycles of PCR amplification with primers complementary to the ligated sequencing adapters. These amplicons were sequenced using the SOLiD 5500X1 Sequencing Instrument (Life Technologies, Carlsbad, CA, USA) using the paired-end protocol with 75 and 35 nucleotides for the forward and reverse strand, respectively.

## Processing and Quality Control of Hi-C Data

Forward and reverse sequence reads were separately aligned to the human reference genome (hg19) by means of LifeScope Genomic Analysis Software 2.5.1. The resulting file was imported into the software tool HOMER v.4.7 (25), where the dataset was filtered for possible PCR artifacts, reads with low mapping quality and reads derived from sites lacking HindIII motifs. For visualization of chromosomal interaction probabilities as heatmaps, reads that passed the above-mentioned filters were summarized to genomic bins of 100 and 250 kb in size and read counts were normalized. The normalization strategy as implemented in HOMER proceeds on the assumption that each region within the genome should have the same visibility and for that reason equalizes possible artifactual effects caused by differences in GC content, accessibility of DNA and unequal distribution of HindIII sites. Significantly interacting genomic

bins were determined employing HOMER's analyzeHiC module (FDR = 0.001; bin interaction distance >25 Mb). Visualization of interaction frequencies in heatmaps was done in JAVA Treeview (26). Translocations were preselected by visual inspection of the interaction heatmaps. Translocation breakpoints were fine-mapped by evaluating read distribution within a 2 Mb window surrounding the breakpoints in order to identify the HindIII fragment next to the breakpoint.

## Comparison of Translocation Breakpoint Regions With Data on Higher Order Chromatin Conformation, DNA Copy Number Alterations, and Search for Expressed Fusion Genes

Given the hypothesis that spatial proximity might promote the emergence of translocations, we have processed public Hi-C data on the B-lymphocyte cell line GM12878 (27) in the same ways as the data for Se-Ax and evaluated the presence of significant interchromosomal interactions as defined by HOMER (FDR = 0.001) connecting the translocation partner chromosomes by means of Circos (28). Additionally, we have visually inspected various public data on chromatin interaction deposited at the 4DGenome database (29) (<https://4dgenome.research.chop.edu/>) to identify possible interactions between our intervals of interest in other cell lines.

Data on DNA copy number alterations in Se-Ax that have been generated by means of arrayCGH in a previous study (30) were visualized for each translocation within a 2 Mb interval surrounding the breakpoint by means of R and the R packages reshape2 (31) and ggplot2 (32). The expression of fusion genes was tested using previously published RNA-Seq data (33, 34). Translocation breakpoints were verified by screening paired-end sequencing data for Se-Ax (33, 34) and the analysis of these data by Breakdancer (35).

## Analysis of Chromosomal Breakpoint Overrepresentation Within Genes

Overrepresentation of translocation breakpoints within genes was tested at the resolution of single HindIII fragments by calculating the likelihood that the same number of randomly distributed HindIII fragments map to genes as it has been observed for HindIII fragments located next to the translocation breakpoints. In a first step, we have cataloged all HindIII restriction sites within the human genome by means of Galaxy Emboss command *fuzznuc* (36, 37). Gaps in the human genome assembly (38) were subtracted with BEDtools (39). From the resulting dataset, the Unix command *shuf* was employed to generate 100,000 permutations of 32 HindIII fragments and the BEDtools command "intersectBed" was used to compute the frequency of overlap with RefSeq genes (40).

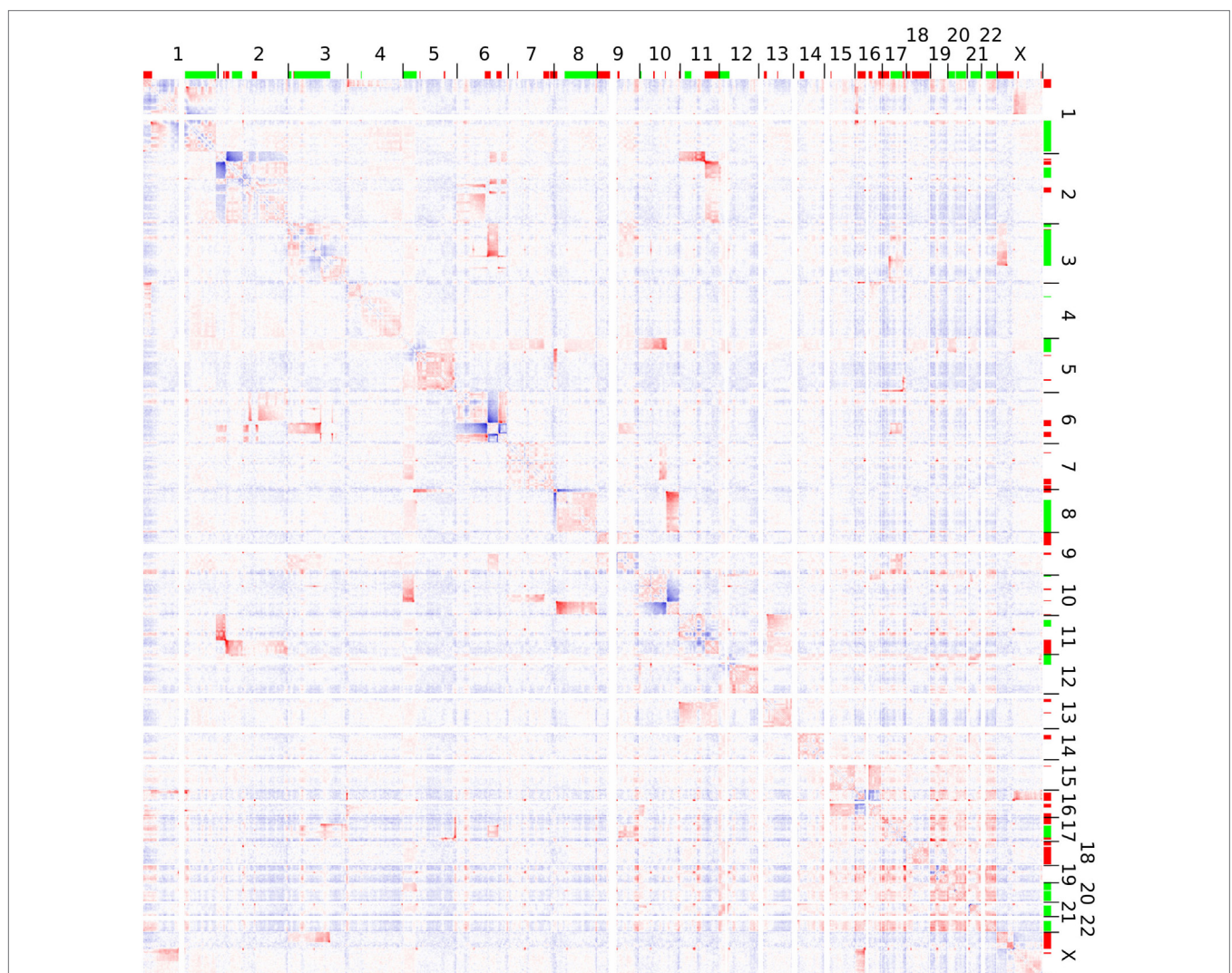
To calculate the *p*-value for Monte Carlo resampling according to Ref. (41), the number of permutation datasets that feature an equal or greater count of HindIII fragment regions with gene overlap as observed ( $\geq 24$ ) were used as the expected overlap.



## RESULTS

Hi-C analysis was based on 91.9 million read pairs that passed processing and quality filtering in HOMER. A genome-wide survey of structural aberrations is presented in **Figure 2**. This heatmap depicts the ratio of observed interaction frequencies and the expected frequencies based on a background model. Translocations are indicated by higher than expected frequencies of interchromosomal interactions (red color). Correspondingly, intrachromosomal interaction frequencies of the chromosomes involved in the translocation are decreased (blue color). The color gradient indicates the orientation of the breakpoint; i.e., interaction intensities decrease with distance from chromosomal breakpoints. In total, we identified 22 translocations, from which we were able to fine-map 32 breakpoints to a single HindIII

fragment (**Table 1**; note that only 32 of the 34 breakpoints as listed in **Table 1** were considered for the following analysis as in two cases breakpoints mapped to the same HindIII fragment). A comparison of Hi-C data with whole-genome sequencing data generated by a different laboratory using a different batch of Se-Ax cells (33, 34) revealed an overlap of 25 breakpoints. These have been highlighted in **Table 1**. A comparison of translocation breakpoints with array CGH data generated in a previous study by our laboratory with a resolution of ~100 kb (30) revealed that 11 of those breakpoints not identified by whole-genome sequencing were flanked by either deletions ( $n = 7$ ) or duplications ( $n = 4$ ). Other translocation breakpoints solely identified by Hi-C analysis were in close vicinity to other translocations, suggesting the presence of a complex rearrangement (t1/t10; t7/t8; t8/t15; and t13/t14). Yet, it has to be emphasized that non-overlapping



**FIGURE 2** | Genome-wide interaction frequencies in Se-Ax. Higher and lower than expected normalized interaction frequencies are shown with 2.5 Mb resolution in red and blue, respectively. The chromosome numbers are given at the top and to the right; together with information on DNA copy number losses (red) and gains (green) as detected by array comparative genomic hybridization. Translocations are characterized by interchromosomal interactions higher than expected, while their corresponding intrachromosomal interactions are decreased. A more detailed view of selected chromosomes is provided in **Figure 3**.

**TABLE 1** | Translocation breakpoints (hg19).

ID	Partner ID	Chromosome	Start	Stop	Genes
t1	t1_1p35.3	chr1	28238215	28248287	RPA2
	t1_4p16.3	chr4	4050000 <sup>b</sup>	4150000	BC042823
t2	t2_1q21.3	chr1	150769818	150781274	CTSK
	t2_16p13.13 <sup>a</sup>	chr16	11004475	11019710	CIITA
t3	t3_2q14.2 <sup>a</sup>	chr2	120270564	120285427	SCTR
	t3_6q16.1 <sup>a</sup>	chr6	98127299	98143928	LOC101927314
t4	t4_2q22.1 <sup>a</sup>	chr2	137227625	137228172	—
	t4_6q21 <sup>a</sup>	chr6	114395987	114399043	HDAC2-AS2
t5	t5_11q14.3 <sup>a</sup>	chr11	88350000 <sup>b</sup>	88450000	GRM5
	t5_2p22.3 <sup>a</sup>	chr2	32260000 <sup>b</sup>	32360000	Several genes
t6	t6_11q14.3 <sup>a</sup>	chr11	88350000 <sup>b</sup>	88450000	GRM5
	t6_2p22.3 <sup>a</sup>	chr2	32230000 <sup>b</sup>	32330000	Several genes
t7	t7_3q13.2	chr3	112240000 <sup>b</sup>	112340000	Several genes
	t7_6q21 <sup>a</sup>	chr6	106400000 <sup>b</sup>	106500000	—
t8	t8_3q13.2	chr3	113435703	113437137	NAA50
	t8_17q11.2	chr17	28807394	28808819	GOSR1
t9	t9_3q24 <sup>a</sup>	chr3	143214724	143222670	SLC9A9
	t9_Xp21.1 <sup>a</sup>	chrX	36597000	36600264	—
t10	t10_16p13.13	chr16	11954332	11955153	—
	t10_4p16.3	chr4	3315595	3322490	RGS12
t11	t11_5p13.2 <sup>a</sup>	chr5	37046678	37051545	NIPBL
	t11_8p23.1 <sup>a</sup>	chr8	10977335	10981039	XKR6
t12	t12_10q23.31 <sup>a</sup>	chr10	91799056	91801084	—
	t12_5q13.2 <sup>a</sup>	chr5	37046678	37051545	NIPBL
t13	t13_17q25.1	chr17	70991114	71006897	SLC39A11
	t13_5q31.3	chr5	143656972	143667413	KCTD16
t14	t14_5q35.2 <sup>a</sup>	chr5	176114621	176120288	—
	t14_17q24.3 <sup>a</sup>	chr17	70537801	70539618	LINC00673
t15	t15_6q24.2 <sup>a</sup>	chr6	143380000 <sup>b</sup>	143480000	AIG1
	t15_17q11.2 <sup>a</sup>	chr17	28770000 <sup>b</sup>	28870000	CPD, GOSR1
t16	t16_10q22.1	chr10	71038330	71050820	HK1
	t16_7q31.33	chr7	126890674	126899921	GRM8
t17	t17_10q23.33 <sup>a</sup>	chr10	95298787	95314132	—
	t17_8p23.1 <sup>a</sup>	chr8	10977335	10981039	XKR6
t18	t18_12p12.2	chr12	20833396	20843983	PDE3A
	t18_10p14	chr10	7970549	7976904	TAF3
t19	t19_13q12.3 <sup>a</sup>	chr13	31132320	31134497	HMGB1
	t19_11p15.5 <sup>a</sup>	chr11	1444072	1449221	BRSK2
t20	t20_12p12.3	chr12	18675961	18680225	PIK3C2G
	t20_Xq28	chrX	147292149	147300244	—
t21	t21_17q25.3 <sup>a</sup>	chr17	75511241	75516019	—
	t21_19p13.3	chr19	3737154	3750893	TJP3
t22	t22_17q24.3	chr17	69185554	69188239	CASC17
	t22_9q21.13	chr9	74350000 <sup>b</sup>	74600000	Several genes

<sup>a</sup>Identified in an independent Se-Ax cell batch by whole-genome sequencing.

<sup>b</sup>Not mapped to a single HindIII interval.

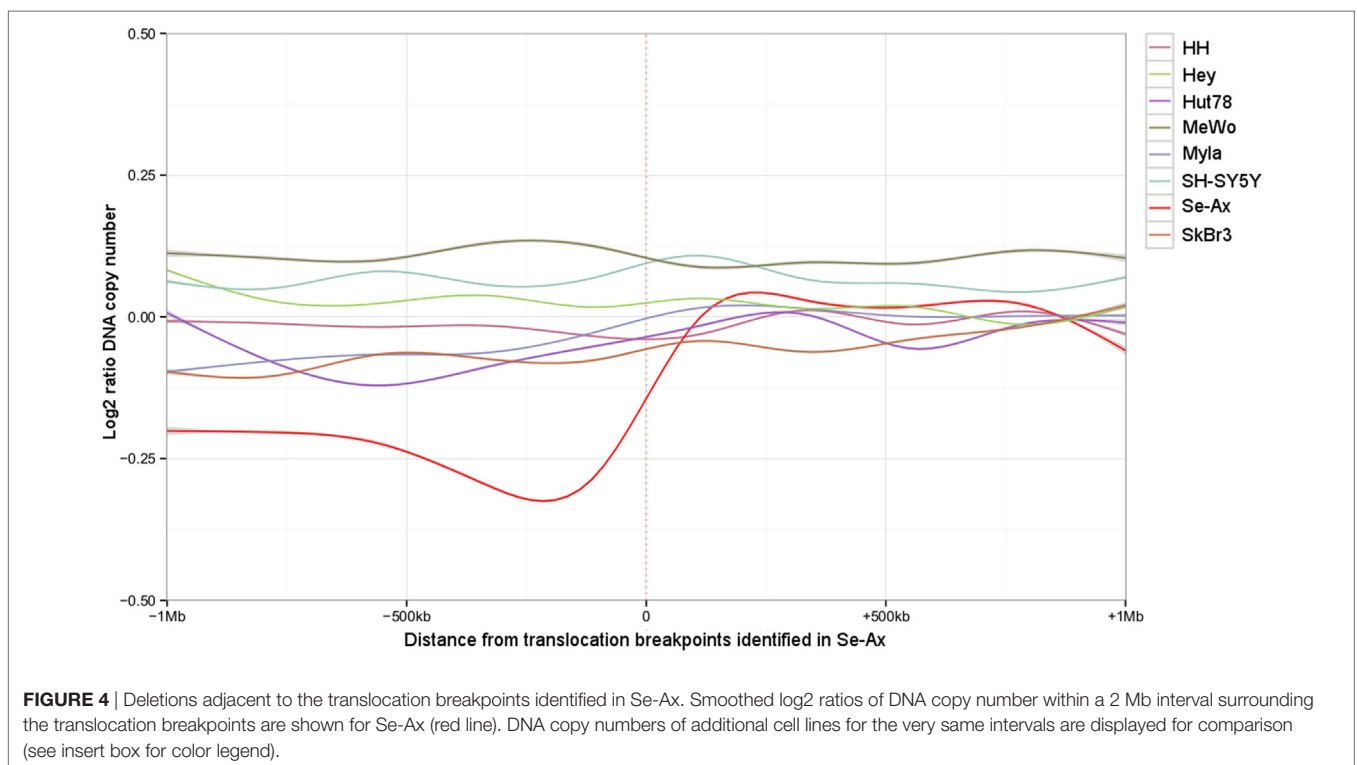
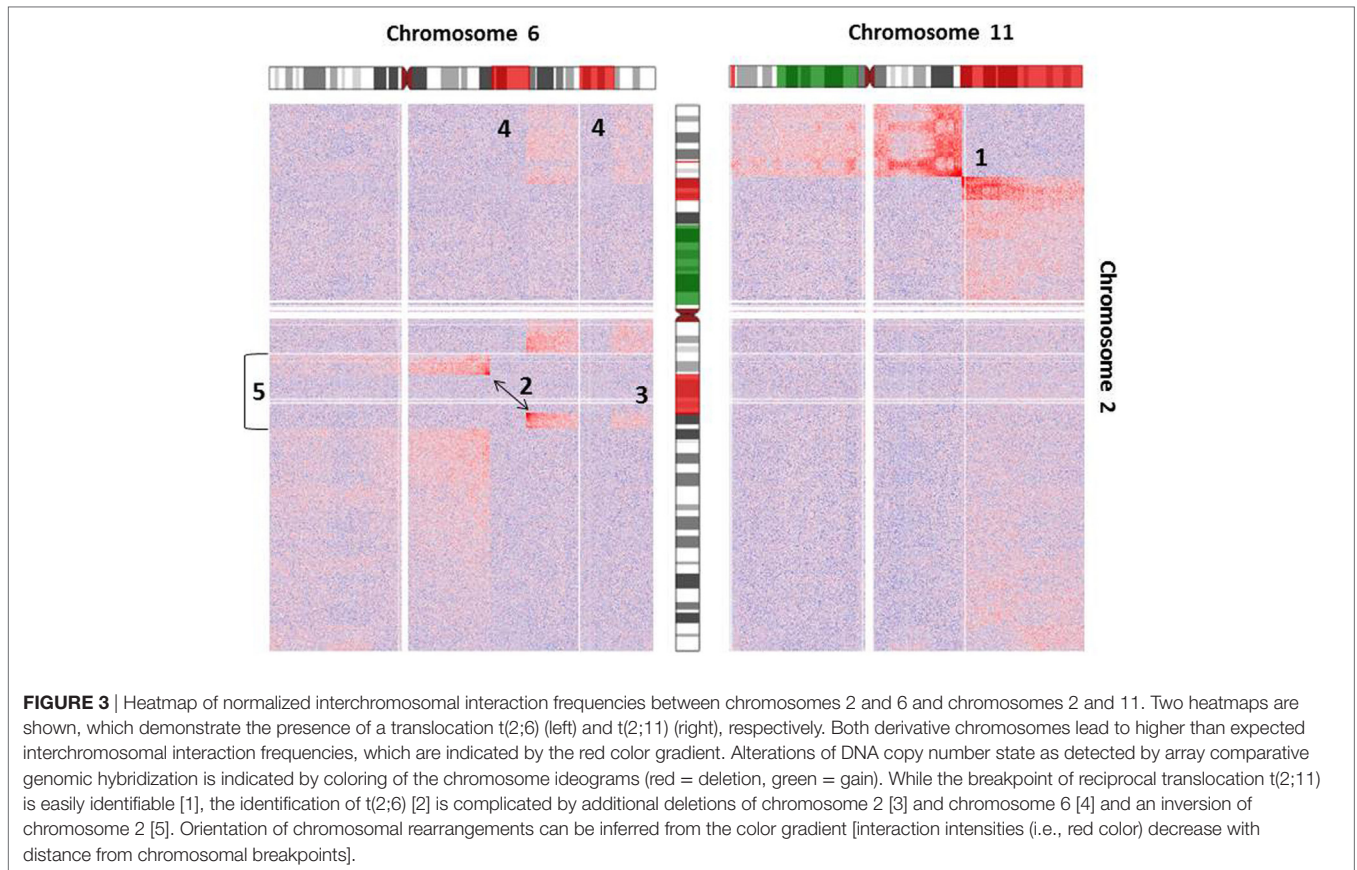
Start and stop define the HindIII interval covering the breakpoint.

breakpoints may also be owed to private mutations emerging during cultivation of Se-Ax cells in different laboratories over longer time or other technical reasons, in particular differences in resolution.

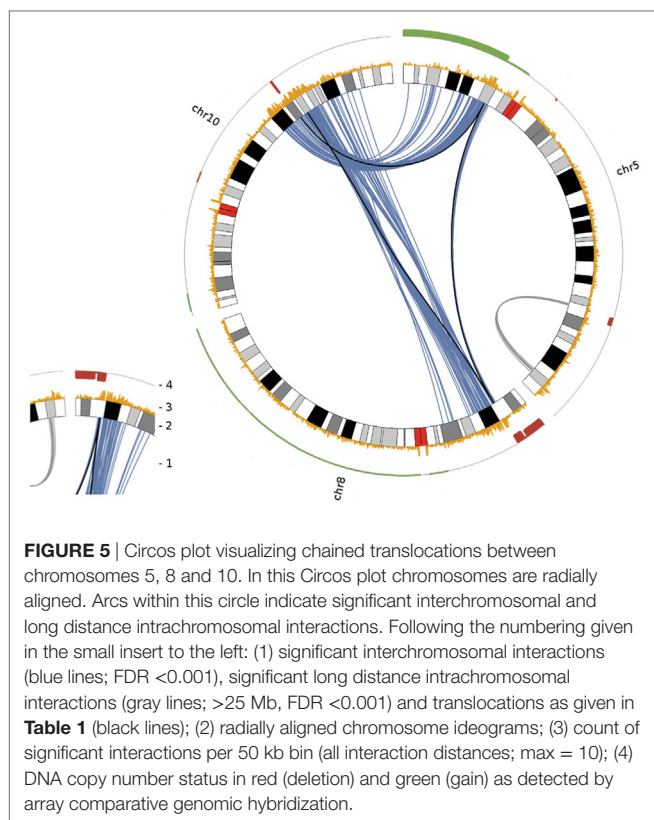
As an example for the complexity of chromosomal aberrations, a zoom-in depicting interchromosomal interactions for

chromosomes 2, 6, and 11 is given in **Figure 3**. Additionally, chromosomal deletions and duplications identified by arrayCGH analysis of Se-Ax are indicated in both heatmaps.

Deletions adjacent to translocation breakpoints have been encountered 12 times (out of 32 breakpoints; **Figure 4**). The Circos plot depicted in **Figure 5** demonstrates chained translocations







with shared breakpoints between several chromosomes on the example of chromosome 5, 8, and 10.

In order to evaluate the impact of spatial proximity of chromosomes on the emergence of translocations, we screened public Hi-C datasets for interactions between those chromosomal intervals affected by translocations in Se-Ax. We failed to get any clues on higher interaction probabilities between regions encompassing the translocation regions, neither in the data of the lymphoblastoid cell line GM12878, which we processed the same way as the Se-Ax data, nor in the datasets from the 4DGenome database. Permutation analysis revealed a significant overrepresentation of translocation breakpoint-associated HindIII fragments within genes ( $p = 0.00208$ , 100,000 permutations). For one of the possible fusion genes (*AIG1/GOSR1*), transcripts were identified in the corresponding published RNA-Seq data (33).

## DISCUSSION

The genome-wide identification and fine-mapping of chromosomal aberrations in tumor cells is instrumental in getting insights into the molecular mechanisms underlying their formation. On the example of a cutaneous T-cell lymphoma cell line we demonstrated the usefulness of Hi-C for the identification of balanced chromosomal translocations. We could show that the abrupt and prominent change of chromosomal interaction probabilities caused by derivative chromosomes facilitates the identification of translocation partners, their orientation to each other as well as their chromosomal breakpoints. Thereby, the

observed changes of interaction probabilities were not confined to the area surrounding the breakpoints, but have extended several megabases beyond (**Figures 2 and 3**). This makes the detection of translocations by Hi-C sensitive, robust, and less prone to artifacts, particularly if the chromosomal breakpoints are next to repetitive sequences, segmental duplications or DNA copy number aberrations. Consequently, the analysis of translocations by Hi-C overcomes some of the limitations of alternative deep sequencing approaches described above. Resolution and sensitivity of this approach might be further increased by recent modifications of the Hi-C protocol (27), by using more sequencing reads and more frequently cutting enzymes (e.g., 4 bp instead of 6 bp) (42) or DNase as an alternative (43). A current limitation of the presented approach is that the presence of translocations has been identified by visual inspection of interaction matrices and that their chromosomal breakpoints were pinned down to the level of single restriction enzyme recognition sites by scrutinizing read distribution within the preselected chromosomal intervals later on. There is need for automation and objectivation of this process and very recently, first software tools dedicated to this task have already been presented (44). Although the focus of this study is on balanced translocations, the observation of higher than expected interaction probabilities between chromosomal segments can also be employed for the detection of DNA copy number alterations and inversions, but particularly for small inversions higher sequencing depth is needed for their robust detection (10). Hi-C analysis with sufficient sequencing depth and read length would also facilitate the determination of haplotype phase (45, 46), which would allow the correct assignment of chromosomal breakpoints to either the maternally or paternally derived chromosomes.

In line with previous reports, the Hi-C data presented in this study have revealed a highly complex karyotype in the investigated cell line. Strikingly, several of the translocation breakpoints seem to be chained and associated with chromosomal deletions. Such patterns of rearrangements have already been observed in other tumors (47, 48), including one case of cutaneous T-cell lymphoma (49), and the term chromoplexy has been coined to describe this phenomenon (21). Simultaneously arising DNA double-strand breaks are a prerequisite for chromoplexy, but the triggers of the temporarily and spatially confined genomic instability have not been identified yet (18). Notably, the majority of breakpoints map within genes ( $p = 0.00208$ ), which suggests that transcription or other biological processes associated with genic sequences could be involved in this mutational event (50). Irrespective of the cause of genomic instability the fusion of DNA double-strand breaks requires their spatial proximity, either before DNA damage (contact first) or thereafter, when broken ends might migrate to some sort of repair center (breakage first) (51). Although our comparison of translocation breakpoints with public chromatin interaction data did not produce any conclusive results, there is evidence in the literature that nuclear neighborhood of chromosomes impacts the frequency of translocations (52, 53). Against this background, transcription factories could be the possible scene of the observed punctuated accumulation of chromosomal translocations, as



genes from different chromosomes cluster together in these nuclear structures (54, 55).

In summary, we have demonstrated the power of genome-wide chromosome conformation capture analysis to detect chromosomal translocations. We present evidence that several translocations identified in the investigated cutaneous T-cell lymphoma cell line likely emerged simultaneously leading to karyotypic features typical of chromoplexy. Overrepresentation of breakpoints within genic sequences highlights the role of transcription or gene-associated biological processes in the emergence of the observed pattern of structural chromosomal rearrangements.

## AUTHOR CONTRIBUTIONS

AS, GE, LRJ and AWK generated Hi-C data, which were analyzed by AS, GE and RU. GKP compared Hi-C results to whole genome

sequencing data. CA, MM, GE, MP, CAS, BVB, PG LRJ, GKP and AWK contributed to data interpretation and critically reviewed the manuscript. AS and RU designed the study and wrote the manuscript.

## ACKNOWLEDGMENTS

We thank Robert Weissmann for excellent bioinformatic support of Hi-C data.

## FUNDING

This project was funded by the German Ministry of Defence. CA was funded by the Wilhelm-Sander Stiftung (2011-066.1) and GKP by the National Science Centre (decision No 2013/08/M/NZ2/00962).

## REFERENCES

- Yates LR, Campbell PJ. Evolution of the cancer genome. *Nat Rev Genet* (2012) 13(11):795–806. doi:10.1038/nrg3317
- Solinas-Toldo S, Lampel S, Stilgenbauer S, Nicklenko J, Benner A, Dohner H, et al. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* (1997) 20(4):399–407. doi:10.1002/(SICI)1098-2264(199712)20:4<399::AID-GCC12>3.0.CO;2-I
- Pinkel D, Seagraves R, Sudar D, Clark S, Poole I, Kowbel D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* (1998) 20(2):207–11. doi:10.1038/2524
- Veltman IM, Veltman JA, Arkesteijn G, Janssen IM, Vissers LE, de Jong PJ, et al. Chromosomal breakpoint mapping by arrayCGH using flow-sorted chromosomes. *Biotechniques* (2003) 35(5):1066–70.
- Fiegler H, Gribble SM, Burford DC, Carr P, Prigmore E, Porter KM, et al. Array painting: a method for the rapid analysis of aberrant chromosomes using DNA microarrays. *J Med Genet* (2003) 40(9):664–70. doi:10.1136/jmg.40.9.664
- Kalscheuer VM, FitzPatrick D, Tommerup N, Bugge M, Niebuhr E, Neumann LM, et al. Mutations in autism susceptibility candidate 2 (AUTS2) in patients with mental retardation. *Hum Genet* (2007) 121(3–4):501–9. doi:10.1007/s00439-006-0284-0
- Abel HJ, Duncavage EJ. Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches. *Cancer Genet* (2013) 206(12):432–40. doi:10.1016/j.cancergen.2013.11.002
- Chen W, Kalscheuer V, Tzschach A, Menzel C, Ullmann R, Schulz MH, et al. Mapping translocation breakpoints by next-generation sequencing. *Genome Res* (2008) 18(7):1143–9. doi:10.1101/gr.076166.108
- Chen W, Ullmann R, Langnick C, Menzel C, Wotschofsky Z, Hu H, et al. Breakpoint analysis of balanced chromosome rearrangements by next-generation paired-end sequencing. *Eur J Hum Genet* (2010) 18(5):539–43. doi:10.1038/ejhg.2009.211
- Harewood L, Kishore K, Eldridge MD, Wingett S, Pearson D, Schoenfelder S, et al. Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. *Genome Biol* (2017) 18(1):125. doi:10.1186/s13059-017-1253-8
- Paterson AL, Weaver JM, Eldridge MD, Tavare S, Fitzgerald RC, Edwards PA, et al. Mobile element insertions are frequent in oesophageal adenocarcinomas and can mislead paired-end sequencing analysis. *BMC Genomics* (2015) 16:473. doi:10.1186/s12864-015-1685-z
- Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet* (2013) 14(6):390–403. doi:10.1038/nrg3454
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* (2009) 326(5950):289–93. doi:10.1126/science.1181369
- Belton J-M, McCord RP, Gibcus J, Naumova N, Zhan Y, Dekker J. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* (2012) 58(3):268–76. doi:10.1016/j.ymeth.2012.05.001
- Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat Biotechnol* (2013) 31(12):1119–25. doi:10.1038/nbt.2727
- Engreitz JM, Agarwala V, Mirny LA. Three-dimensional genome architecture influences partner selection for chromosomal translocations in human disease. *PLoS One* (2012) 7(9):e44196. doi:10.1371/journal.pone.0044196
- Barutcu AR, Lajoie BR, McCord RP, Tye CE, Hong D, Messier TL, et al. Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. *Genome Biol* (2015) 16:214. doi:10.1186/s13059-015-0768-0
- Zhang CZ, Leibowitz ML, Pellman D. Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. *Genes Dev* (2013) 27(23):2513–30. doi:10.1101/gad.229559.113
- Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* (2011) 144(1):27–40. doi:10.1016/j.cell.2010.11.055
- Zhang CZ, Spektor A, Cornils H, Francis JM, Jackson EK, Liu S, et al. Chromothripsis from DNA damage in micronuclei. *Nature* (2015) 522(7555):179–84. doi:10.1038/nature14493
- Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, et al. Punctuated evolution of prostate cancer genomes. *Cell* (2013) 153(3):666–77. doi:10.1016/j.cell.2013.03.021
- Bagherani N, Smoller BR. An overview of cutaneous T cell lymphomas. *F1000Res* (2016) 5. doi:10.12688/f1000research.8829.1
- Elenitoba-Johnson KS, Wilcox R. A new molecular paradigm in mycosis fungoides and Sezary syndrome. *Semin Diagn Pathol* (2017) 34(1):15–21. doi:10.1053/j.semdp.2016.11.002
- Kaltoft K, Bisballe S, Rasmussen HE, Thestrup-Pedersen K, Thomsen K, Sterry W. A continuous T-cell line from a patient with Sézary syndrome. *Arch Dermatol Res* (1987) 279(5):293–8. doi:10.1007/BF00431220
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* (2010) 38(4):576–89. doi:10.1016/j.molcel.2010.05.004
- Saldanha AJ. Java Treeview – extensible visualization of microarray data. *Bioinformatics* (2004) 20(17):3246–8. doi:10.1093/bioinformatics/bth349
- Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* (2014) 159(7):1665–80. doi:10.1016/j.cell.2014.11.021
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res* (2009) 19(9):1639–45. doi:10.1101/gr.092759.109

29. Teng L, He B, Wang J, Tan K. 4DGenome: a comprehensive database of chromatin interactions. *Bioinformatics* (2015) 31(15):2560–4. doi:10.1093/bioinformatics/btv158
30. Steininger A, Mobs M, Ullmann R, Kochert K, Kreher S, Lamprecht B, et al. Genomic loss of the putative tumor suppressor gene E2A in human lymphoma. *J Exp Med* (2011) 208(8):1585–93. doi:10.1084/jem.20101785
31. Wickham H. Reshaping data with the reshape package. *J Stat Softw* (2007) 21(12):1–20. doi:10.18637/jss.v021.i12
32. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag (2009).
33. Izykowska K, Zawada M, Nowicka K, Grabarczyk P, Braun FCM, Delin M, et al. Identification of multiple complex rearrangements associated with deletions in the 6q23–27 region in Sezary syndrome. *J Invest Dermatol* (2013) 133(11):2617–25. doi:10.1038/jid.2013.188
34. Izykowska K, Przybylski GK, Gand C, Braun FC, Grabarczyk P, Kuss AW, et al. Genetic rearrangements result in altered gene expression and novel fusion transcripts in Sezary syndrome. *Oncotarget* (2017) 8(24):39627–39. doi:10.18632/oncotarget.17383
35. Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* (2009) 6(9):677–81. doi:10.1038/nmeth.1363
36. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* (2016) 44(W1):W3–10. doi:10.1093/nar/gkw343
37. Rice P, Longden I, Bleasby A. EMBL: the European molecular biology open software suite. *Trends Genet* (2000) 16(6):276–7. doi:10.1016/S0168-9525(00)02024-2
38. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The UCSC table browser data retrieval tool. *Nucleic Acids Res* (2004) 32(Database issue):D493–6. doi:10.1093/nar/gkh103
39. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* (2010) 26(6):841–2. doi:10.1093/bioinformatics/btq033
40. O’Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* (2016) 44(D1):D733–45. doi:10.1093/nar/gkv1189
41. Phipson B, Smyth GK. Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. *Stat Appl Genet Mol Biol* (2010) 9:Article39. doi:10.2202/1544-6115.1585
42. Lajoie BR, Dekker J, Kaplan N. The Hitchhiker’s guide to Hi-C analysis: practical guidelines. *Methods* (2015) 72:65–75. doi:10.1016/j.ymeth.2014.10.031
43. Ramani V, Cusanovich DA, Hause RJ, Ma W, Qiu R, Deng X, et al. Mapping 3D genome architecture through in situ DNase Hi-C. *Nat Protoc* (2016) 11(11):2104–21. doi:10.1038/nprot.2016.126
44. Chakraborty A, Ay F. Identification of copy number variations and translocations in cancer cells from Hi-C data. *Bioinformatics* (2018) 34(2):338–45. doi:10.1093/bioinformatics/btx664
45. Edge P, Bafna V, Bansal V. HapCUT2: robust and accurate haplotype assembly for diverse sequencing technologies. *Genome Res* (2017) 27(5):801–12. doi:10.1101/gr.213462.116
46. Selvaraj S, Dixon JR, Bansal V, Ren B. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nat Biotechnol* (2013) 31(12):1111–8. doi:10.1038/nbt.2728
47. Dzamba M, Ramani AK, Buczkowicz P, Jiang Y, Yu M, Hawkins C, et al. Identification of complex genomic rearrangements in cancers using CouGaR. *Genome Res* (2017) 27(1):107–17. doi:10.1101/gr.211201.116
48. Mansfield AS, Murphy SJ, Harris FR, Robinson SI, Marks RS, Johnson SH, et al. Chromoplectic TPM3-ALK rearrangement in a patient with inflammatory myofibroblastic tumor who responded to ceritinib after progression on crizotinib. *Ann Oncol* (2016) 27(11):2111–7. doi:10.1093/annonc/mdw405
49. Choi J, Goh G, Walradt T, Hong BS, Bunick CG, Chen K, et al. Genomic landscape of cutaneous T cell lymphoma. *Nat Genet* (2015) 47(9):1011–9. doi:10.1038/ng.3356
50. Kim N, Jinks-Robertson S. Transcription as a source of genome instability. *Nat Rev Genet* (2012) 13(3):204–14. doi:10.1038/nrg3152
51. Meaburn KJ, Misteli T, Soutoglou E. Spatial genome organization in the formation of chromosomal translocations. *Semin Cancer Biol* (2007) 17(1):80–90. doi:10.1016/j.semcancer.2006.10.008
52. Zhang Y, McCord RP, Ho Y-J, Lajoie BR, Hildebrand DG, Simon AC, et al. Chromosomal translocations are guided by the spatial organization of the genome. *Cell* (2012) 148(5):908–21. doi:10.1016/j.cell.2012.02.002
53. Mathas S, Kreher S, Meaburn KJ, Johrens K, Lamprecht B, Assaf C, et al. Gene deregulation and spatial genome reorganization near breakpoints prior to formation of translocations in anaplastic large cell lymphoma. *Proc Natl Acad Sci U S A* (2009) 106(14):5831–6. doi:10.1073/pnas.0900912106
54. Iborra FJ, Pombo A, Jackson DA, Cook PR. Active RNA polymerases are localized within discrete transcription ‘factories’ in human nuclei. *J Cell Sci* (1996) 109(Pt 6):1427–36.
55. Osborne CS. Molecular pathways: transcription factories and chromosomal translocations. *Clin Cancer Res* (2014) 20(2):296–300. doi:10.1158/1078-0432.CCR-12-3667

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Steininger, Ebert, Becker, Assaf, Möbs, Schmidt, Grabarczyk, Jensen, Przybylski, Port, Kuss and Ullmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.