

ShatterSeek: an R package for the detection of chromothripsis from Next-Generation Sequencing (NGS) data

Isidro Cortes-Ciriano^{1,2,3}, Ruibin Xi⁴, and Peter J. Park^{*1,2}

¹*Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, USA*

²*Ludwig Center at Harvard, Boston, MA 02115, USA*

³*Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom*

⁴*School of Mathematical Sciences and Center for Statistical Science, Peking University, Beijing 100871, China*

April 9, 2018

Contents

1	Introduction	2
2	Workflow implemented in ShatterSeek	3
2.1	Detection of clusters of interleaved SVs	4
2.2	Statistical criteria implemented in ShatterSeek	5
2.2.1	Equal distribution of SV types	5
2.2.2	Chromosomal enrichment of breakpoints	5
2.2.3	Random distribution of breakpoints	5
2.2.4	Number of oscillating copy number segments	6

*peter_park@hms.harvard.edu

2.2.5	Interspersed loss of heterozygosity (LOH)	6
2.3	Detection of multichromosomal chromothripsis events	6
3	Recommended cut-off values to interpret the output of ShatterSeek	7
4	How to use ShatterSeek	8
4.1	Installation	8
4.2	Load SV and CNV data into R	8
5	Visualization of chromothripsis regions	16
6	Bibliography	19

1 Introduction

Chromothripsis refers to the genomic alterations characterized by massive de novo rearrangements, often generated in a single catastrophic event, where the DNA is shattered into a number of fragments that are subsequently stitched together in random order and orientation. Chromothripsis can be confined to small chromosomal regions or can affect multiple chromosomes, and can involve from tens to hundreds of rearrangements [1]. Therefore, chromothripsis represents a mechanism for the accrual of tens to hundreds of rearrangements in a few cell divisions.

Chromothripsis regions are characterized by copy number (CN) profiles oscillating between two or three states, (in cases where partial duplications follow chromothripsis), interspersed loss of heterozygosity (LOH), and clusters of interleaved structural variations (SVs), where the proportions of fragment joins (i.e., duplication-like, deletion-like, head-to-head and tail-to-tail inversions) are roughly equal, consistent with the random stitching of genomic fragments through mostly non-homologous end-joining (NHEJ) DNA repair (Figure 1).

Initial studies used array-derived copy number profiles to detect chromothripsis, as the amount of whole-genome sequencing data sets was limited [2], [3]. SNP arrays do not permit to fully characterize the genome-wide landscape of structural rearrangements at single-base resolution, notably interchromosomal events. Hence, the detection of chromothripsis is more accurate if structural

rearrangements and copy number data are integrated [4]–[7].

Korbel and Campbell [6] proposed a set of statistical criteria for the detection of chromothripsis that have been widely used in the literature. To date, there exist two publicly available packages for the detection of chromothripsis, namely: CTLP scanner [3], [8] and Shatterproof [7]. The former uses SNP array data, whereas the latter both copy number variation (CNV) and SV data.

We have developed ShatterSeek, an R package that integrates copy number and SV data for the detection and visualization of chromothripsis events from NGS data. ShatterSeek implements a custom graph-based approach to identify candidate chromothripsis regions, and then applies a set of statistical criteria based on Korbel and Campbell [6] to detect both single and multichromosomal chromothripsis events. In addition, ShatterSeek provides functionalities for the easy visualization of SVs, as well as CN and LOH profiles. Visual inspection of candidate chromothripsis regions is still required in a number of cases due to the complexity of the observed events, and the overlapping features of chromothripsis and other complex events. We have recently validated ShatterSeek in a large-scale study of ca. 2,600 cancer genomes and shown its higher sensitivity and specificity with respect to Shatterproof. We refer the reader to this work for further details about the rates and characteristics of chromothripsis across diverse human cancers. The chromothripsis calls for all these tumors generated using ShatterSeek can be accessed at <http://compbio.med.harvard.edu/chromothripsis/>

In the next sections we explain in detail the statistical criteria implemented in ShatterSeek to detect chromothripsis events, and illustrate its functionalities using data from a kidney renal cell carcinoma patient.

2 Workflow implemented in ShatterSeek

To identify chromothripsis-like patterns in cancer genomes, we implemented and extended the set of criteria proposed by Campbell and Korbel [6]. The pipeline to detect chromothripsis consists of three major steps, namely: (i) discovery of clusters of interleaved SVs using a graph-based approach, (ii) evaluation of a set of statistical criteria in the genomic regions spanned by these clusters, and (iii) evaluation of whether chromothripsis is confined to a single or to multiple chromosomes.

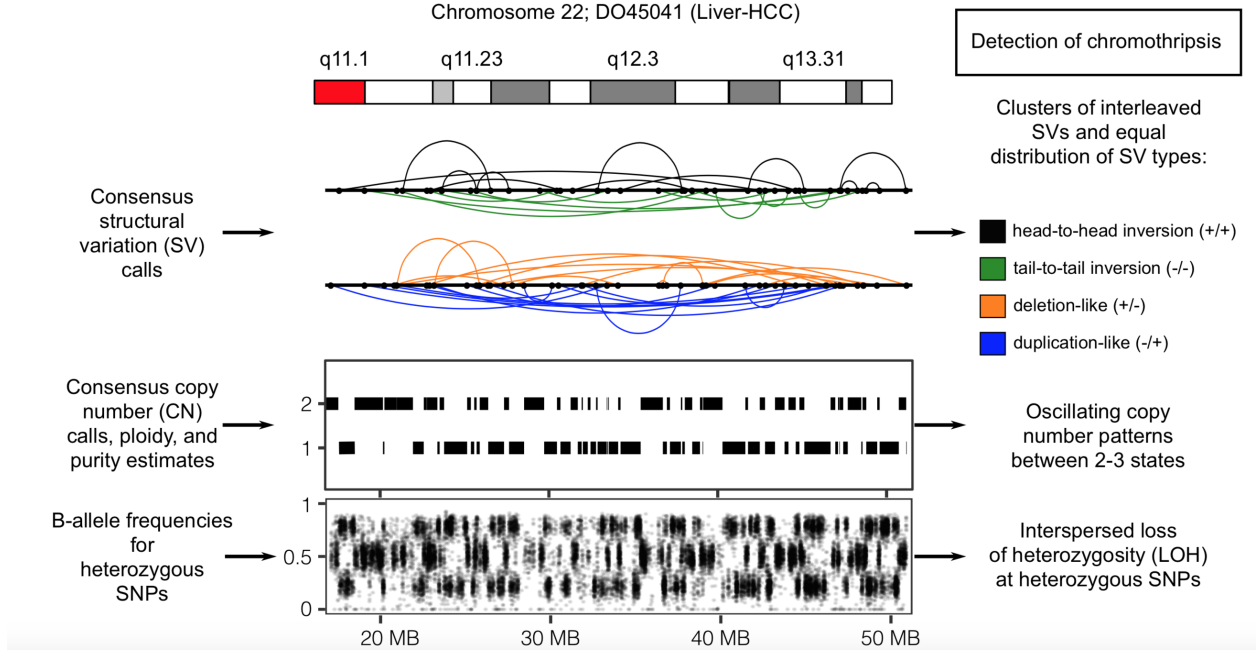


Figure 1: Illustrative example of chromothripsis and workflow and criteria used to detect chromothripsis events by ShatterSeek

2.1 Detection of clusters of interleaved SVs

ShatterSeek firstly scans the cancer genomes for the presence of clusters of interleaved SVs, which are expected to be generated by the random fragmentation and stitching of the genome. We consider that two SVs are interleaved if the genomic regions bridged by their breakpoints overlap but are not nested. To find clusters within a given chromosome, ShatterSeek constructs an undirected graph using intrachromosomal SVs whose nodes correspond to SVs and whose edges connect interleaved SVs. Thus, clusters of SVs can be detected by finding the connected components in the graph. The connected component in each chromosome with the highest number of SVs is considered for further analysis. By default, ShatterSeek only considers chromosomes 1-22 and X. In the following, we use the term *chromothripsis region* to refer to the genomic regions affected by chromothripsis within one chromosome, whereas we term *chromothripsis event* the set of chromothripsis regions involved in a single catastrophic event. Thus, a chromothripsis event can affect a single chromosome, multiple regions within a chromosome, or comprise genomic regions from multiple chromosomes.

2.2 Statistical criteria implemented in ShatterSeek

Once the clusters of interleaved SVs are detected, ShatterSeek evaluates the following statistical criteria following the work by Korbelt and Campbell [6].

2.2.1 Equal distribution of SV types

Once the SV clusters are detected, ShatterSeek evaluates whether the distribution of DNA fragment joins, i.e., deletion-like (+/-; or "head"/"tail"), duplication-like (-/+), head-to-head (+/+), and tail-to-tail (-/-) inversions, diverges from a multinomial distribution with equal probabilities for each SV category using the goodness-of-fit test for the multinomial distribution from the R package stats (chisq.test function). The types of SVs are determined by the orientation of the reads mapped to the breakpoints (see [9] for further details). We term this test 'fragment joins test'.

2.2.2 Chromosomal enrichment of breakpoints

Massive chromothripsis events, which involve hundreds of SVs, generally affect large genomic regions in one or multiple chromosomes against the background of quiescent genomes. In these cases the chromosomes harboring chromothripsis are highly enriched for breakpoints and can be detected by comparing the SV rate in a given chromosome (or region) against the genome-wide (or chromosome-wide) SV rate.

To test for the enrichment of breakpoints in each chromosome, ShatterSeek uses the binomial test corrected for mappability. We note that this test might be misleading for focal chromothripsis events appearing in highly rearranged genomes, where other chromosomes not displaying chromothripsis harbor a number of SVs (interleaved or not) comparable to that detected in the chromosomes displaying chromothripsis. Therefore, this criterion needs to be interpreted carefully on a per-case basis (please see below). We term this test 'chromosomal breakpoint enrichment test'.

2.2.3 Random distribution of breakpoints

ShatterSeek also evaluates whether the distribution of the breakpoints comprised in the SV clusters differs from an exponential distribution, as described by Korbelt and Campbell [6]. We term this test 'random distribution of breakpoints test'.

2.2.4 Number of oscillating copy number segments

A hallmark of chromothripsis is the presence of copy number profiles oscillating between two or three copy number states. This feature has been widely used before to detect chromothripsis from genome-wide copy number profiles using operational definitions, e.g., at least 10 contiguous segments oscillating between two copy number states [10].

To uncover oscillating patterns in the genomic regions delimited by the distal breakpoints composing the clusters of interleaved SVs, ShatterSeek reports the number of uninterrupted CN segments oscillating across 2 and 3 states.

2.2.5 Interspersed loss of heterozygosity (LOH)

We decided not to use the criterion of loss of heterozygosity proposed by Campbell and Korbel [6], as evaluating the presence of LOH in some bona fide chromothripsis cases identified in low-purity samples and in aneuploid tumors was not straightforward. In LOH regions, the allelic ratios for heterozygous SNPs (i.e., B allele frequencies or BAF) are either 1 or 0. However, in low-purity tumors the infiltration of normal tissue heavily distorts the BAF profile. Hence, this limits the power to detect significant changes due to LOH in the BAF profile. In such cases, the BAF values would not divert significantly from 0.5 even in LOH regions, and thus, would hamper the observation of alternating LOH patterns associated to copy losses [11]. In addition, assessing LOH in aneuploid tumors is often difficult, where the BAF values for oscillating segments between high CN levels do not vary strongly.

However, given that the loss of heterozygosity profiles are very useful to determine the temporal profile of chromothripsis events, and to distinguish chromothripsis from chromoanasythesis in nearly-diploid cases, ShatterSeek provides capabilities to visualize LOH/CN minor profiles at chromothripsis regions.

2.3 Detection of multichromosomal chromothripsis events

ShatterSeek considers that two or more chromothripsis regions belong to the same catastrophic event if these regions $\pm 10\text{Kb}$ (default value) are linked by at least two interchromosomal SVs. Given that the sensitivity of SV detection algorithms is still limited, it is not always possible

to detect rearrangements between all regions belonging to the same chromothripsis event. Thus, in cases where at least three chromosomes are involved, ShatterSeek applies transitive reasoning to identify the full extent of the events. For instance, if the chromothripsis regions detected in chromosomes 1 and 2 are linked, and the event detected in chromosome 2 is also linked to a chromothripsis region in chromosome 3, ShatterSeek considers that the chromothripsis regions detected in these three chromosomes were generated as a result of the same catastrophic event.

3 Recommended cut-off values to interpret the output of ShatterSeek

After manual curation of hundreds of massive and focal chromothripsis calls, we derived the following guidelines to detect chromothripsis using ShatterSeek.

We assign two levels of confidence depending on the set of statistical tests satisfied by a candidate chromothripsis region.

- High confidence: at least 6 interleaved intrachromosomal SVs, 7 contiguous segments oscillating between 2 CN states, the fragment joins test, and either the chromosomal enrichment or the exponential distribution of breakpoints test.
- High confidence: at least 3 interleaved intrachromosomal SVs and 4 or more interchromosomal SVs, 7 contiguous segments oscillating between 2 CN states and the fragment joins test.
- Low confidence: at least 6 interleaved intrachromosomal SVs, 4, 5 or 6 adjacent segments oscillating between 2 CN states, the fragment joins test, and either the chromosomal enrichment or the exponential distribution of breakpoints test.

Application of these criteria to 2,600 whole genomes still led to a small set of false positives that were removed by visual inspection. This is mostly due to the fact that multiple layers of rearrangements other than chromothripsis might coexist, and the aggregate of these might satisfy the statistical tests described above. In addition, chromothripsis regions are highly heterogeneous, running the gamut from massive chromothripsis cases involving hundreds of SVs in an otherwise quiescent genome, to focal events in the context of genomes highly rearranged by other mutational processes. This heterogeneity makes it almost impossible to define a set of criteria with perfect discriminative power for chromothripsis. Therefore, we advocate, as far as possible, for the manual curation of

candidate chromothripsis regions. To facilitate this process, ShatterSeek provides functionalities to visually depict the candidate chromothripsis regions (see below).

4 How to use ShatterSeek

In the following sections, we illustrate how to install and use ShatterSeek to detect and visualize chromothripsis using SV and CNV data. This tutorial assumes minimal knowledge of the R programming language.

4.1 Installation

ShatterSeek is entirely written in R. To install ShatterSeek type the following in R:

```
require(devtools)
install_github("parklab/ShatterSeek")
```

Alternatively, please download the latest release of ShatterSeek by running in a bash terminal:

```
git clone git@github.com:parklab/ShatterSeek.git
unzip ShatterSeek-master.zip
R CMD INSTALL ShatterSeek-master
```

4.2 Load SV and CNV data into R

We first load ShatterSeek and the test data that is provided with the package. This corresponds to the SV and CN data for a kidney renal cell carcinoma tumor (ICGC ID: DO17373). The chromothripsis events detected in this tumor sample are depicted in Figure 2. The other chromosomes in this tumor showed unaltered CN profiles with respect to the normal tissue sample.

```
library(ShatterSeek)
data(DO17373)
```


Running this command loads to the workspace two R dataframes, corresponding to the somatic SVs (SV_DO17373) and CNV (SCNA_DO17373) calls for this tumor. These calls are consensus calls generated by the Pan-Cancer Analysis of Whole Genomes (PCAWG) project. ShatterSeek accepts CN and SV calls from any caller, provided they are encoded in the format required by ShatterSeek (please see below). However, we note that the sensitivity of ShatterSeek depends on the quality of the SV and CN calls used. Hence, it is the responsibility of the user to guarantee that the SV and CN calls are of sufficient quality.

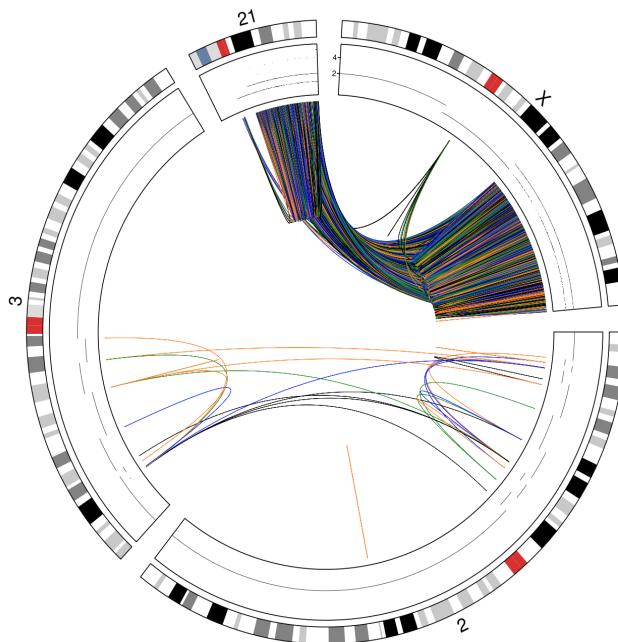


Figure 2: Evidence of two massive multichromosomal chromothripsis events involving chromosomes 21-X and 2-3 detected in a kidney renal cell carcinoma tumor in patient DO17373 (TCGA-CJ-5681).

ShatterSeek requires the SV data to be stored in a dataframe with the following columns:

- chrom1 (character): chromosome for the first breakpoint
- pos1 (character): position for the first breakpoint
- chrom2 (character): chromosome for the second breakpoint
- pos2 (character): position for the second breakpoint
- SVtype (character): type of SV, encoded as: DEL (deletion-like; +/-), DUP (duplication-like; -/+), h2hINV (head-to-head inversion; +/-), and t2tINV (tail-to-tail inversion; -/-).

- strand1 (e.g. + for DEL)
- strand2 (e.g. - for DEL)

Chromosomes are expected to be in Ensembl chromosome notation, i.e. NOT contain the prefix "chr".

```
head(SV_D017373)
```

##	chrom1	start1	end1	chrom2	start2	end2	sv_id
## 1	1	142618685	142618686	21	21708092	21708093	SVMERGE1159
## 2	1	142623413	142623414	21	25233137	25233138	SVMERGE269
## 3	1	142639910	142639911	21	33703275	33703276	SVMERGE127
## 4	1	143536371	143536372	21	29698088	29698089	SVMERGE1343
## 5	11	8540384	8540385	11	8541772	8541773	SVMERGE55
## 6	14	61896353	61896354	14	62061595	62061596	SVMERGE1006
##	pe_support	strand1	strand2	svclass	svmethod		
## 1	18	+	-	TRA	SNOWMAN_DELLY		
## 2	50	-	-	TRA	SNOWMAN_dRANGER_DELLY		
## 3	34	+	-	TRA	SNOWMAN_dRANGER		
## 4	39	-	+	TRA	SNOWMAN_dRANGER		
## 5	37	+	-	DEL	SNOWMAN_BRASS_DELLY		
## 6	111	-	+	DUP	SNOWMAN_BRASS_dRANGER_DELLY		

Please remember that ShatterSeek only considers chromosomes 1-22 and X. Thus, make sure that the SVs comprised in your input data only correspond to these chromosomes. The SV data is loaded into an object of class "SV", using the function *SVs*:

```
SV_data <- SVs(chrom1=as.character(SV_D017373$chrom1),
               pos1=as.numeric(SV_D017373$start1),
               chrom2=as.character(SV_D017373$chrom2),
               pos2=as.numeric(SV_D017373$end2),
               SVtype=as.character(SV_D017373$svclass),
               strand1=as.character(SV_D017373$strand1),
               strand2=as.character(SV_D017373$strand2))
```

ShatterSeek requires the CNV data to be in the following format:

- chrom (character): chromosome (also in Ensembl notation)
- start (numeric): start position for the CN segment
- end (numeric): end position for the CN segment
- CN (numeric): integer total copy number (e.g. 2 for unaltered chromosomal regions)

```
head(SCNA_D017373)
```

##	chromosome	start	end	total_cn
## 1	1	1 249250620		2
## 3	2	20016 11969465		2
## 4	2	11969466 14420187		1
## 5	2	14420188 16916033		2
## 6	2	16916034 16937471		1
## 7	2	16937472 17054487		2

The CNV data is loaded into an object of class "CNVsegs", using the function *CNVsegs*:

```
CN_data <- CNVsegs(chrom=as.character(SCNA_D017373$chromosome),  
                  start=SCNA_D017373$start,  
                  end=SCNA_D017373$end,  
                  total_cn=SCNA_D017373$total_cn)
```

Once the input data has been loaded, we proceed to run the main function of the package, namely *shatterseek*. This function runs the code to detect clusters of interleaved SVs, and subsequently evaluates each candidate chromothripsis region for the statistical criteria described above.

```
library(ShatterSeek)  
start_time <- Sys.time()  
chromothripsis <- shatterseek(SV.sample=SV_data, seg.sample=CN_data)  
  
## Running..  
##
```

```
##
## Evaluating the statistical criteria
## Successfully finished!

end_time <- Sys.time()
print(paste0("Running time (s): ",round(end_time - start_time,digits=2)))

## [1] "Running time (s): 17.78"

print(head(chromothripsis@chromSummary))
```

##	chrom	start	end	number_DEL	number_DUP	number_h2hINV
## 1	1	NA	NA	0	0	0
## 2	2	11969466	75460870	3	3	2
## 3	3	20704347	85709275	2	1	1
## 4	4	NA	NA	0	0	0
## 5	5	140129029	171935828	0	2	0
## 6	6	NA	NA	0	0	0

##	number_t2tINV	number_TRA	clusterSize_including_TRA	number_SVs_sample
## 1	0	0	0	1426
## 2	0	7	15	1426
## 3	0	6	10	1426
## 4	0	0	0	1426
## 5	0	0	2	1426
## 6	0	0	0	1426

##	number_CNV_segments	pval_fragment_joins	chr_breakpoint_enrichment
## 1	NA	NA	8.612611e-17
## 2	13	0.8653370	2.949495e-08
## 3	14	0.5724067	1.185413e-07
## 4	NA	NA	1.852862e-17
## 5	1	0.1116102	6.729654e-15
## 6	NA	NA	3.277719e-14

##	pval_exp_chr	pval_exp_cluster
## 1	NA	NA
## 2	8.656372e-06	0.000000e+00

```

## 3 0.000000e+00      4.440892e-16
## 4      NA      NA
## 5      NA      NA
## 6      NA      NA
##   max_number_oscillating_CN_segments_2_states
## 1      NA
## 2      11
## 3      12
## 4      NA
## 5      NA
## 6      NA
##   max_number_oscillating_CN_segments_3_states number_CN_segments_chr
## 1      NA      NA
## 2      11      13
## 3      12      13
## 4      NA      NA
## 5      NA      NA
## 6      NA      NA
##   max_number_oscillating_CN_segments_2_states_chr
## 1      NA
## 2      13
## 3      13
## 4      NA
## 5      NA
## 6      NA
##   max_number_oscillating_CN_segments_3_states_chr inter_number_DEL
## 1      NA      0
## 2      13      4
## 3      13      4
## 4      NA      0
## 5      NA      0
## 6      NA      0
##   inter_number_h2hINV inter_number_t2tINV inter_number_DUP
## 1      0      0      0
## 2      2      3      2
## 3      3      4      3
## 4      0      0      0

```

```
## 5          0          0          0
## 6          0          0          0
##   inter_pval_fragment_joins inter_other_chroms
## 1          NA
## 2      0.8012520          3
## 3      0.9626925          2
## 4          NA
## 5          NA
## 6          NA
##   inter_other_chroms_coords_all
## 1
## 2      3:20704347-85709275;
## 3      2:11969466-75460870;
## 4
## 5
## 6
```

The function *ShatterSeek* returns an instance of the 'chromoth' class that contains two slots, namely: *detail* and *chromSummary*. The slot *detail* contains a list containing the input data (SV and CNV calls), as well as intermediate results obtained by running the graph-based approach implemented to discover clusters of interleaved SVs:

```
names(chromothripsis@detail)

## [1] "SV"          "graph"        "connComp"     "num.chromth"
## [5] "maxSVs"      "degree"       "numSVByChrom" "maxClusterSize"
## [9] "SVinter"     "CNV"
```

The slot *chromothripsis@chromSummary* is a data.frame where each row corresponds to a chromosome, and the columns to the values of the statistical criteria and additional information. The columns are:

- chrom: chromosome
- start: start position for the cluster of interleaved SVs detected in chromosome 'chrom'
- end: end position for the cluster of interleaved SVs detected in chromosome 'chrom'

- `number_DEL`: number of intrachromosomal deletion-like SVs (+/-) mapped within the cluster region
- `number_DUP`: number of intrachromosomal duplication-like SVs (-/+) mapped within the cluster region
- `number_h2hINV`: number of intrachromosomal head-to-head SVs (+/+) mapped within the cluster region
- `number_t2tINV`: number of intrachromosomal tail-to-tail SVs (-/-) mapped within the cluster region
- `number_TRA`: number of interchromosomal SVs mapped within the cluster
- `clusterSize_including_TRA`: total number of SVs (including inter- and intrachromosomal SVs) mapped to the cluster region
- `number_SVs_sample`: total number of SVs (including inter- and intrachromosomal SVs) detected in the sample
- `number_CNV_segments`: number of CN segments located within the cluster boundaries
- `pval_fragment_joins`: P value for the fragment joins test considering only the intrachromosomal SVs mapped to the cluster region
- `chr_breakpoint_enrichment`: P value for the "chromosomal breakpoint enrichment" test
- `pval_exp_chr`: P value for the 'random distribution of breakpoints' test considering all breakpoints in the chromosome
- `pval_exp_cluster`: P value for the 'random distribution of breakpoints' test considering only the breakpoints mapped to the cluster region
- `max_number_oscillating_CNV_segments_2_states`: Maximum number of uninterrupted oscillations between 2 CN states in the cluster region
- `max_number_oscillating_CNV_segments_3_states`: Maximum number of uninterrupted oscillations across 3 CN states in the cluster region
- `number_CNV_segments_chr`: number of CN segments in the chromosome
- `max_number_oscillating_CNV_segments_2_states_chr`: Maximum number of uninterrupted oscillations between 2 CN states in the chromosome
- `max_number_oscillating_CNV_segments_3_states_chr`: Maximum number of uninterrupted oscillations across 3 CN states in the chromosome

- `inter_number_DEL`: number of interchromosomal deletion-like SVs (+/-) with one breakpoint mapped within the cluster region
- `inter_number_DUP`: number of interchromosomal duplication-like SVs (-/+) with one breakpoint mapped within the cluster region
- `inter_number_h2hINV`: number of interchromosomal head-to-head SVs (+/+) with one breakpoint mapped within the cluster region
- `inter_number_t2tINV`: number of interchromosomal tail-to-tail SVs (-/-) with one breakpoint mapped within the cluster region
- `inter_pval_fragment_joins`: P value for the fragment joins test considering the inter- and intrachromosomal SVs mapped to the cluster region
- `inter_other_chroms`: chromosomes linked by at least 2 SVs with the cluster region detected in chromosome 'chrom'
- `inter_other_chroms_coords_all`: chromosomal coordinates for the clusters of SVs detected in other chromosomes linked by at least 2 SVs with the cluster region detected in the chromosome under consideration (i.e. the one specified in 'chrom')

5 Visualization of chromothripsis regions

ShatterSeek provides functionalities to inspect the detected chromothripsis regions. The function `plot_chromothripsis` takes as input the output of the function `ShatterSeek` and a chromosome number. It returns a list containing 4 ggplot objects:

- ideogram of the affected region (only hg19 is supported at the moment).
- Representatin of the SVs. SVs are depicted as arcs with the breakpoints represented by black points. The breakpoints corresponding to interchromosomal SVs are depicted as colored points. Duplication-like SVs, deletion-like SVs, head-to-head and tail-to-tail inversions are depicted by default in blue, orange, black, and green, respectively.
- Total CN profile. Each CN segment is represented by a black rectangle.
- Information for the depicted region is given as a table.

These four ggplot objects, which can be modified to tailor the user's needs, can be easily combined using the function *arrangeGrob* from the R package *gridExtra*, and visualized using e.g. the function *plot_grid* from the R package *cowplot*.

```
library(gridExtra)

plots_chr3 = plot_chromothripsis(ShatterSeek_output = chromothripsis,chr = "3")
plot_chr3 = arrangeGrob(plots_chr3[[1]],
                        plots_chr3[[2]],
                        plots_chr3[[3]],
                        plots_chr3[[4]],
                        nrow=4,ncol=1,heights=c(0.2,.4,.4,.4))

plots_chr2 = plot_chromothripsis(ShatterSeek_output = chromothripsis,chr = "2")
plot_chr2 = arrangeGrob(plots_chr2[[1]],
                        plots_chr2[[2]],
                        plots_chr2[[3]],
                        plots_chr2[[4]],
                        nrow=4,ncol=1,heights=c(0.2,.4,.4,.4))

plots_chr21 = plot_chromothripsis(ShatterSeek_output = chromothripsis,chr = "21")
plot_chr21 = arrangeGrob(plots_chr21[[1]],
                        plots_chr21[[2]],
                        plots_chr21[[3]],
                        plots_chr21[[4]],
                        nrow=4,ncol=1,heights=c(0.2,.4,.4,.4))

plots_chrX = plot_chromothripsis(ShatterSeek_output = chromothripsis,chr = "X")
plot_chrX = arrangeGrob(plots_chrX[[1]],
                        plots_chrX[[2]],
                        plots_chrX[[3]],
                        plots_chrX[[4]],
                        nrow=4,ncol=1,heights=c(0.2,.4,.4,.4))

library(cowplot)
plot_grid(plot_chr3,plot_chr2)
```

The plots can also be combined into a grid of e.g. four plots and saved to a file:

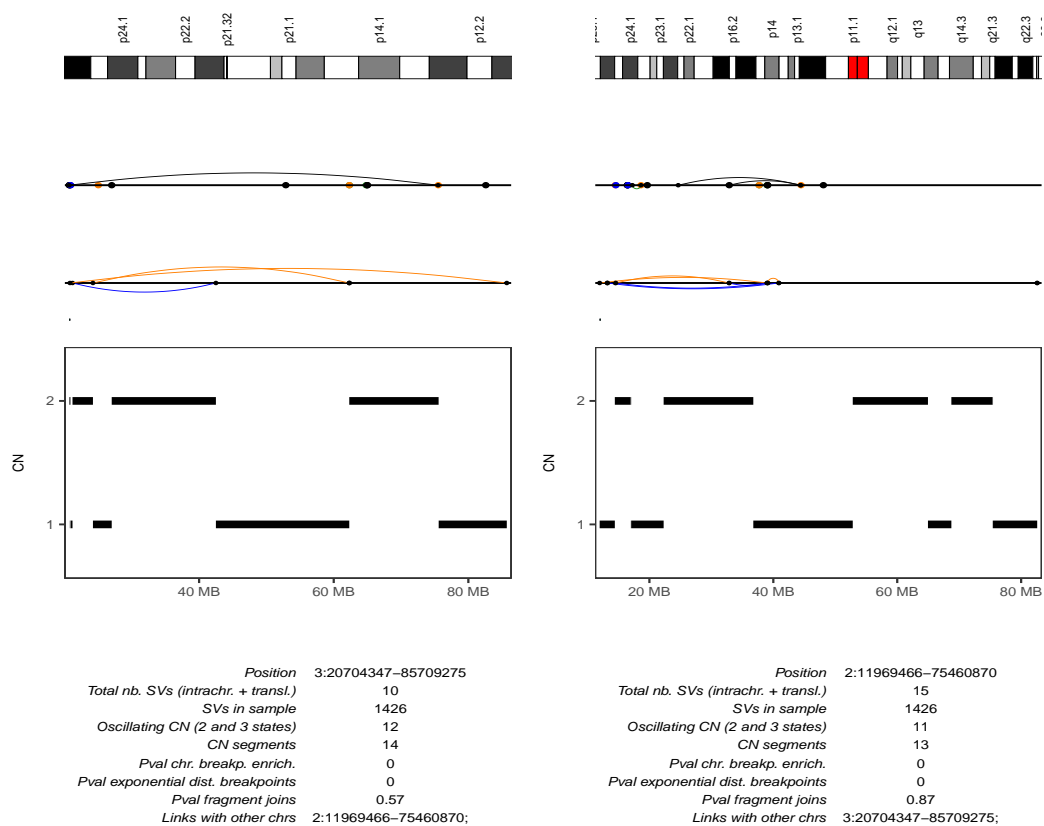


Figure 3: Chromothripsis regions detected in patient DO17373 in chromosomes 2 and 3.

6 Bibliography

References

- [1] P. J. Stephens, C. D. Greenman, B. Fu, F. Yang, G. R. Bignell, L. J. Mudie, E. D. Pleasance, K. W. Lau, D. Beare, L. A. Stebbings, S. McLaren, M.-L. Lin, D. J. McBride, I. Varela, S. Nik-Zainal, C. Leroy, M. Jia, A. Menzies, A. P. Butler, J. W. Teague, M. A. Quail, J. Burton, H. Swerdlow, N. P. Carter, L. A. Morsberger, C. Iacobuzio-Donahue, G. A. Follows, A. R. Green, A. M. Flanagan, M. R. Stratton, P. A. Futreal, and P. J. Campbell, “Massive genomic rearrangement acquired in a single catastrophic event during cancer development.,” *Cell*, vol. 144, no. 1, pp. 27–40, 2011, ISSN: 1097-4172. DOI: 10.1016/j.cell.2010.11.055. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21215367><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3065307>.
- [2] T.-M. Kim, R. Xi, L. J. Luquette, R. W. Park, M. D. Johnson, and P. J. Park, “Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes.,” *Genome research*, vol. 23, no. 2, pp. 217–27, 2013, ISSN: 1549-5469. DOI: 10.1101/gr.140301.112. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23132910><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3561863>.
- [3] H. Cai, N. Kumar, H. C. Bagheri, C. von Mering, M. D. Robinson, and M. Baudis, “Chromothripsis-like patterns are recurring but heterogeneously distributed features in a survey of 22,347 cancer genome screens,” *BMC Genomics*, vol. 15, no. 1, p. 82, 2014, ISSN: 1471-2164. DOI: 10.1186/1471-2164-15-82. [Online]. Available: <http://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-82>.
- [4] F. Notta, M. Chan-Seng-Yue, M. Lemire, Y. Li, G. W. Wilson, A. A. Connor, R. E. Denroche, S.-B. Liang, A. M. K. Brown, J. C. Kim, T. Wang, J. T. Simpson, T. Beck, A. Borgida, N. Buchner, D. Chadwick, S. Hafezi-Bakhtiari, J. E. Dick, L. Heisler, M. A. Hollingsworth, E. Ibrahimov, G. H. Jang, J. Johns, L. G. T. Jorgensen, C. Law, O. Ludkovski, I. Lungu, K. Ng, D. Pasternack, G. M. Petersen, L. I. Shlush, L. Timms, M.-S. Tsao, J. M. Wilson, C. K. Yung, G. Zogopoulos, J. M. S. Bartlett, L. B. Alexandrov, F. X. Real, S. P. Cleary, M. H. Roehrl, J. D. McPherson, L. D. Stein, T. J. Hudson, P. J. Campbell, and S. Gallinger, “A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns,” *Nature*, vol. 538, no. 7625, pp. 378–382, 2016, ISSN: 0028-0836. DOI: 10.1038/nature19823. [Online]. Available: <http://www.nature.com/doifinder/10.1038/nature19823>.
- [5] Y. Li, C. Schwab, S. L. Ryan, E. Papaemmanuil, H. M. Robinson, P. Jacobs, A. V. Moorman, S. Dyer, J. Borrow, M. Griffiths, N. A. Heerema, A. J. Carroll, P. Talley, N. Bown,

- N. Telford, F. M. Ross, L. Gaunt, R. J. Q. McNally, B. D. Young, P. Sinclair, V. Rand, M. R. Teixeira, O. Joseph, B. Robinson, M. Maddison, N. Dastugue, P. Vandenberghe, C. Haferlach, P. J. Stephens, J. Cheng, P. Van Loo, M. R. Stratton, P. J. Campbell, and C. J. Harrison, "Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia.," *Nature*, vol. 508, no. 7494, pp. 98–102, 2014, ISSN: 1476-4687. DOI: 10.1038/nature13115. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24670643><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3976272>.
- [6] J. O. Korbel and P. J. Campbell, "Criteria for inference of chromothripsis in cancer genomes.," *Cell*, vol. 152, no. 6, pp. 1226–36, 2013, ISSN: 1097-4172. DOI: 10.1016/j.cell.2013.02.023. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23498933>.
- [7] S. K. Govind, A. Zia, P. H. Hennings-Yeomans, J. D. Watson, M. Fraser, C. Anghel, A. W. Wyatt, T. van der Kwast, C. C. Collins, J. D. McPherson, R. G. Bristow, and P. C. Boutros, "ShatterProof: operational detection and quantification of chromothripsis," *BMC Bioinformatics*, vol. 15, no. 1, p. 78, 2014, ISSN: 1471-2105. DOI: 10.1186/1471-2105-15-78. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24646301><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3999944><http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-15-78>.
- [8] J. Yang, J. Liu, L. Ouyang, Y. Chen, B. Liu, and H. Cai, "CTLPScanner: a web server for chromothripsis-like pattern detection," *Nucleic Acids Research*, vol. 44, no. W1, W252–W258, 2016, ISSN: 0305-1048. DOI: 10.1093/nar/gkw434. [Online]. Available: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw434>.
- [9] C.-Z. Zhang, M. L. Leibowitz, and D. Pellman, "Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements.," *Genes & development*, vol. 27, no. 23, pp. 2513–30, 2013, ISSN: 1549-5477. DOI: 10.1101/gad.229559.113. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24298051><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3861665>.
- [10] T. Rausch, D. Jones, M. Zapatka, A. Stütz, T. Zichner, J. Weischenfeldt, N. Jäger, M. Remke, D. Shih, P. Northcott, E. Pfaff, J. Tica, Q. Wang, L. Massimi, H. Witt, S. Bender, S. Pleier, H. Cin, C. Hawkins, C. Beck, A. vonDeimling, V. Hans, B. Brors, R. Eils, W. Scheurlen, J. Blake, V. Benes, A. Kulozik, O. Witt, D. Martin, C. Zhang, R. Porat, D. M. Merino, J. Wasserman, N. Jabado, A. Fontebasso, L. Bullinger, F. G. Rücker, K. Döhner, H. Döhner, J. Koster, J. Molenaar, R. Versteeg, M. Kool, U. Tabori, D. Malkin, A. Korshunov, M. Taylor, P. Lichter, S. Pfister, and J. Korbel, "Genome Sequencing of Pediatric Medulloblastoma Links Catastrophic DNA Rearrangements with TP53 Mutations," *Cell*, vol. 148, no. 1-2, pp. 59–71, 2012, ISSN: 00928674. DOI: 10.1016/j.cell.2011.12.013. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0092867411015169>.

- [11] S. Song, K. Nones, D. Miller, I. Harliwong, K. S. Kassahn, M. Pinese, M. Pajic, A. J. Gill, A. L. Johns, M. Anderson, O. Holmes, C. Leonard, D. Taylor, S. Wood, Q. Xu, F. Newell, M. J. Cowley, J. Wu, P. Wilson, L. Fink, A. V. Biankin, N. Waddell, S. M. Grimmond, and J. V. Pearson, “qpure: A Tool to Estimate Tumor Cellularity from Genome-Wide Single-Nucleotide Polymorphism Profiles,” *PLoS ONE*, vol. 7, no. 9, A. H. Ting, Ed., e45835, 2012, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0045835. [Online]. Available: <http://dx.plos.org/10.1371/journal.pone.0045835>.