# SVDetect - a tool to identify genomic structural variations from paired-end and mate-pair sequencing data

Bruno Zeitouni; Valentina Boeva; Isabelle Janoueix-Lerosey; Sophie Loeillet; Patricia Legoix-Né; Alain Nicolas; Olivier Delattre; Emmanuel Barillot.

# Supplementary data

**Supplementary table 1. Comparison of current tools for SV prediction from NGS data**

| Name | SVDetect | GASV | BreakDancer | VariationHunter |
|---|---|---|---|---|
| **Refs.** | | (1) | (2) | (3) |
| **Compatibility** | | | | |
| Paired-ends (FR orientation) | + | + | + | + |
| Mate-paired reads (RF or FF orientation) | + | +/-* | - | - |
| Front-end aligner input format | + | + | + | - |
| SAM/BAM input format | + | + | + | - |
| **Strategy** | | | | |
| Clustering | + | + | + | + |
| Coverage sliding window | + | - | - | - |
| **Signatures detected** | | | | |
| Deletion | + | + | + | + |
| Insertion | + | - | + | + |
| Inversion | + | + | + | + |
| Duplication (small/large/inverted) | + | - | - | - |
| Inter-chromosomal translocation | + | + | + | - |
| Balanced rearrangement | + | - | - | - |
| Gain/loss | + | - | - | - |
| **SV Comparison** | | | | |
| Across various samples | + | +/-** | - | - |
| aCGH data | + | + | - | - |
| **Output file formats** | | | | |
| Bed format | + | - | + | - |
| Circos format | + | - | - | - |
| **Computational requirements** | | | | |
| Memory usage | high | medium | medium | medium |
| CPU time | rapid | rapid | slow | rapid |

SVDetect manages both paired-end and mate-pair data, detects different types of duplications, discriminates balanced rearrangements from unbalanced events, allows for direct comparison between samples, calculates copy-number profiles, and provides output file formats for direct graphical SV view.

*Mate-pair data is not officially supported by GASV. ** GASV only filters a target set of pairs by using a reference file, no SV comparison is implemented. +: supported, -: not supported. F, Forward; R, Reverse.

**Supplementary table 2. Yeast mate-pair sequencing data used for tool comparison**

| *Saccharomyces cerevisiae* | | **Mutant** |
|---|---|---|
| **Genotype** | **WT** | ***pif1Δ* (4)** |
| **Strain** | **ORT2914** | **ORT4841** |
| **Known rearrangements** | | |
| A - Deletion chrXV: 721947-722609(W) - HIS3 gene | + | + |
| B - Insertion chrVIII:141539<br>Human CEB1 sequence (2350 bp) in the ARG4 promoter | + | + |
| C - Duplication-insertion of 2 intergenic sequences chrVIII:136581-139649, chrVIII:141708-148825(C) in chrV:116541-116601(W) - URA3 gene | + | + |
| D - Insertion chrXIII:151532-148953(C)<br>KanMX sequence in the PIF1 gene | - | + |
| E - Total loss of mitochondrial DNA | - | + |
| **Statistics** | | |
| SOLiDv2 mate-paired 25 bp reads (million) | 101 | 96 |
| Mapped* (million)/perc. | 54.6/54% | 51.6/53% |
| Paired (million) | 10.9 | 9.36 |
| Good pairs (million) | 7.08 | 5.93 |
| Mean insert size from good pairs (bp) | 1451 | 1584 |
| Insert size s.d. from good pairs (bp) | 375 | 425 |
| Avg. depth-of-coverage | 34X | 30X |
| Anomalously mapped in the same chromosome | 99 443 | 113 816 |
| Mapped to different chromosomes | 1 118 370 | 1 261 208 |

* Mate-paired reads mapped to the Sc288c reference sequence + CEB sequence + KanMX sequence with the ABI SOLiD corona-lite's mapping and pairing pipelines (2 mismatches maximum, "FF" normal strand orientation of reads). C, Crick strand; W, Watson strand.

**Supplementary table 3. Comparison of SVDetect with the variant detection tool GASV**

| **Yeast strains** | **WT** | | **Mutant *pif1Δ*** | |
|---|---|---|---|---|
| **Number of clusters/SVs** | **Intra-chr** | **Inter-chr** | **Intra-chr** | **Inter-chr** |
| **GASV analysis (1)** | | | | |
| Mate-pair clusters | 60 092 | 1 044 736 | 71 509 | 1 133 580 |
| Filtering of clusters | **31** | **76** | **129** | **60** |
| Intra-SV types (DEL,INV,DIV) | $30^{\mathbf{ABC}}$,1,0 | / | $128^{\mathbf{ABCD}}$,0,1 | / |
| Inter-SV types (NU-TR,RU-TR) | / | $39^{\mathbf{B}}$,$37^{\mathbf{C}}$ | / | $36^{\mathbf{B}}$,$24^{\mathbf{CD}}$ |
| **SVDetect analysis** | | | | |
| Mate-pair clusters | 68 506 | 893 034 | 75 909 | 886 457 |
| Filtering of clusters | **29** | **37** | **43** | **36** |
| Intra-SV types (DEL,INV,S-DUP,L-DUP) | $23^{\mathbf{ABC}}$,2,3,1 | / | $23^{\mathbf{ABCD}}$,2,18,0 | / |
| Inter-SV types (NU-TR,RU-TR,NB-TR) | / | $24^{\mathbf{B}}$,$12^{\mathbf{C}}$,1 | / | $22^{\mathbf{B}}$,$14^{\mathbf{CD}}$,0 |
| **Comparaison of clusters** | | | | |
| *GASV as reference* | | | | |
| Found with both tools | **27/31** | **30/76** | **39/129** | **28/60** |
| Found by GASV only | **4/31** | **46/76** | **90/129** | **32/60** |
| Why not found with SVDetect? | | | | |
| -"number of pairs below the cutoff (8-9)" | 1 | 17* | 52 | 15* |
| -"removed by the strand filter only" | 2 | 16* | 11 | 14* |
| -"removed by the order filter only" | 0 | 12* | 20 | 1 |
| -"removed by both strand & order filters" | 1 | 0 | 4 | 1 |
| -"removed by the insert size filter only" | 0 | / | 3 | / |
| -"unknown" | 0 | 1 | 0 | 1 |
| *SVDetect as reference* | | | | |
| Found with both tools | **27/29** | **30/37** | **39/42** | **28/36** |
| Found by SVDetect only | **2/29** | **7/37** | **3/42** | **8/36** |
| Why not found with GASV? | | | | |
| -"number of pairs below the cutoff (8-9)" | 0 | 7 | 2 | 8 |
| -"unknown" | 2 | 0 | 1 | 0 |

Mate-pair clusters (MPCs) predicted by SVDetect on the yeast sequencing data were compared to sets of clusters detected by GASV. MPCs from GASV were computed using Lmin=600 and Lmax=2600 as lower and upper bounds respectively (based on the insert size distribution) for both strains. MPCs from SVDetect were computed using a window size and a step size set to "w=3652, s=913" and "w=4018, s=1004" for the wild-type (WT) and the mutant strain, respectively (based on the mean μ and s.d. σ values from the insert size distribution). At the filtering step of both GASV and SVDetect analysis, only clusters harboring at least 10 read pairs were selected. From the remaining MPCs, those involving the mitochondrial chromosome have been removed. 72% of GASV clusters and 95% of SVDetect clusters in the WT strain were then filtered out. None of MPCs in the mutant was discarded, underlying the expected loss of the mitochondrial chromosome in the mutant strain. MPCs located at the chrV:541600-545180 coordinates (3580 bp) were filtered out due to library artifacts, representing ~5% of clusters in

both analysis. MPCs detected by SVDetect have been also filtered by specific filtering procedures not implemented in GASV. Strand and order filters of pairs were applied to the clusters (number of pairs in a sub-cluster≥2) together with the insert size filtering (indel σ cutoff=2, duplication σ cutoff=2), and only MPCs with a final confidence filtering score ≥ 0.8 (i.e. at least 80% of reads pairs passed through the strand/order filtering requirements of pairs) were kept and assigned as predicted clusters.

The overlap between the predicted MPCs from the GASV analysis was examined with MPCs obtained from the SVDetect analysis, and inversely. A minimum length overlap of 50% was required to define a common MPC between both tools. 58/73 (79%) inter-MPCs and 66/72 (93%) of intra-MPCs predicted by SVDetect were found with GASV. 17/20 (85%) of the unpredicted MPCs by GASV can be explained by a number of mate-pairs (8-9) close to the selected cutoff. Among the MPCs predicted by GASV, 27/31 (87%) and 39/129 (30%) of intra-chromosomal clusters, and 30/76 (30%) and 28/60 (40%) of inter-chromosomal clusters overlapped with those obtained with SVDetect, in the WT and in the mutant respectively. Half of MPCs (84/172, 49%) not predicted by SVDetect are due to a number of mate-pairs (8-9) just below the selected cutoff. 43/172 (25%) were filtered out by the strand filtering requirements, 33/172 (19%) were filtered out by the order filtering procedure and 5/172 (0.3%) by both filters. Among the 78 inter-chromosomal MPCs not detected by SVDetect, most of them (69/78, 96% - 40 in the WT and 29 in the mutant, indicated with "*") have one cluster of reads mapped onto a region (chrXII:462300-463000) belonging to the ribosomal DNA (rDNA). The rDNA (chrXII:458433-465071) is known to be highly repeated in the genome of *S. cerevisiae* (100-200 repeats of 9kb). As all MPCs detected by GASV and found in the rDNA were filtered out by SVDetect through the strand/order filters (and the number of pairs' cutoff), these MPCs probably rely to potential inconsistent mapping of mate-pairs in this complex genomic region.

Known rearrangements listed in the Supplementary table 2 are retrieved by both tools and are indicated in exponent by the corresponding ID letter.

DEL, deletion; INV, inversion; DIV, divergent; TR, Translocation; NU-TR, Unbalanced translocation; RU-TR, Inverted unbalanced translocation; NB-TR, Balanced translocation; S-DUP, Small duplication; L-DUP, Large duplication. Intra-chr, intra-chromosomal; Inter-chr, inter-chromosomal.

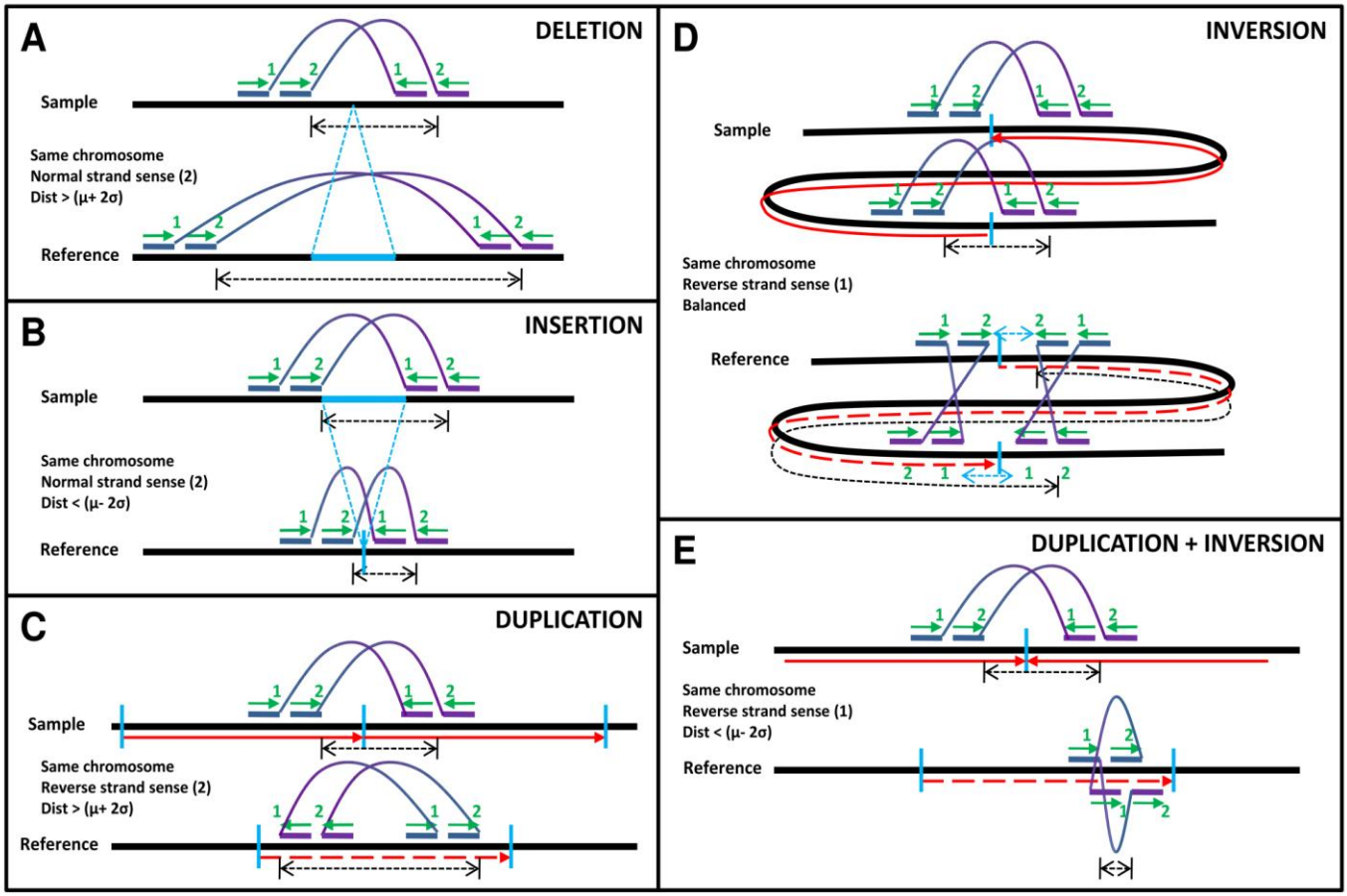
**Supplementary table 4. Computational efficiency**

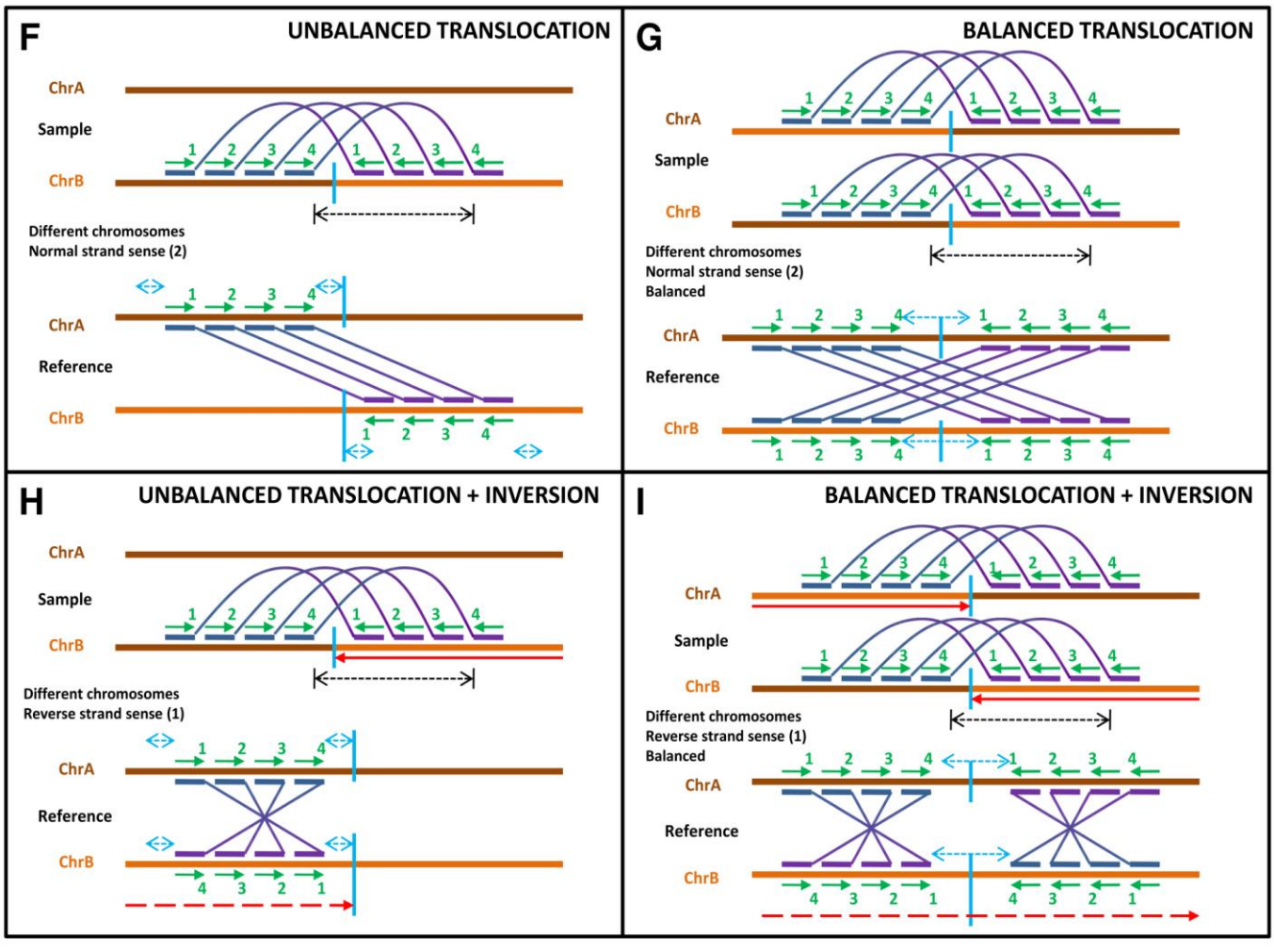
| | **Dataset 1** | **Dataset 2** |
|---|---|---|
| Organism | Yeast | Human |
| Data type | SOLiDv2 mate-paired reads | GAII mate-paired reads |
| Read length (bp) | 25 | 50 |
| Number of aberrant pairs | 1 375 024 | 599 905 |
| Window size, step size (bp) | 5000, 1000 | 6000, 1000 |
| Memory usage | 10.9 GB | 5.9 GB |
| Elapsed time | 40 min,13 sec | 13 min,46 sec |

The memory usage and the elapsed time for each dataset were measured on an eight CPU 3.16GHz Intel(R) Xeon(R) system with 16MB of system memory. Elapsed time represents execution on a single CPU.

## Supplementary table 5. Description of usage parameters in SVDetect

| Parameter | **Description** | **Type/Value/Example** |
|---|---|---|
| Input parameters : | | |
| input_format | Input file format | Text, value = eland, solid or sam |
| sv_type | Type of the structural variations to identify. This parameter can be used to analyze and generate distinct files for intra- and/or inter-chromosomal SVs | Text, value = intra, inter or all (default = all) |
| mates_orientation | Expected normal strand orientation of the paired-end/mate-pair reads | Text, value= FF, RR, FR or RF |
| Algorithm parameters : | | |
| tag_length | Length of mapped reads in base pairs (bp) | Integer, ex: 25 |
| window_size | Size of the sliding-window used for partioning the genome in bp. This parameter determines the range size of the predicted clusters. To identify balanced translocations, a window size equal to at least "2 μ +2 $\sqrt{2}$ σ" should be set | Integer, ex: 4000 |
| step_length | Length of the sliding-window step in bp. Set this parameter to get overlapped windows. Generally equal to ½ or ¼ of the window size | Integer, ex: 1000 |
| mates_file | Full path to the abnormal mate-pair/paired-end input data file | Text, ex: /align/bowtie/ sample_ab_mates.sam |
| mates_file_ref | Full path to the normal mate-pair/paired-end input data file of the reference (for cnv only). | Text, ex: /align/bowtie/ reference_norm_mates.sam |
| cmap_file | Full path to a file with chromosome lengths | Text, ex: /align/bowtie/hs18.len |
| strand_filtering | Flag to run the strand orientation filtering. Only pairs sharing the same read orientation in the link are kept for further analysis. If no main group of pairs can be found (ex: 50% of "FF", 50% of "RF"), the link is filter out | Boolean, value= 1 (run), 0 (skip) |
| order_filtering | Flag to run the order filtering (see Supplementary method 1 for details of the procedure). Also use it to detect balanced rearrangements | Boolean, value= 1 (run), 0 (skip) |
| insert_size_filtering | Flag to run the filtering on the separation distance between mate-pair/paired-end reads (for intra-chromosomal links only) | Boolean, value= 1 (run), 0 (skip) |
| chromosomes | List of chromosome names to keep or exclude | Text, ex: -chr1 |
| nb_pairs_threshold | Minimum number of pairs in a cluster | Integer, ex: 3 |
| nb_pairs_order_threshold | Minimal number of pairs in a subgroup of paired-end reads for balanced events (for order filtering only) | Integer, ex: 2 (default=1) |
| indel_sigma_threshold | Minimal number of sigma fold for the insert size filtering and to call insertions and deletions.<br>A deletion will be found if the mean insert size of the cluster > μ+threshold*σ.<br>An insertion will be found if the mean insert size of the cluster < μ-threshold*σ. | Integer, ex: 3 (default=2) |
| dup_sigma_threshold | Minimal number of sigma fold for the insert size filtering to call tandem duplications.<br>A large duplication will be found if the mean insert size of the cluster > μ+threshold*σ. A small/inverted duplication will be found if the mean insert size of the cluster < μ-threshold*σ. | Integer, ex: 3 (default=2) |
| mu_length (μ) | Mean insert size value of normally mapped paired-ends, in bp. This parameter becomes mandatory for order or insert size filtering. The mu_length parameter can be obtained by running the cnv program. | Real, ex: 1500 |
| sigma_length (σ) | Calculated s.d. value from the distribution of normally mapped paired-end reads, in bp. This parameter becomes mandatory for order or insert size filtering steps. The sigma_length parameter can be obtained by running the cnv program. | Real, ex: 400 |
| Output parameters : | | |
| organism_id | Symbol name of the organism (see the name in the karyotype file used for circos). | Text, ex: hs (for *Homo sapiens*) |
| colorcode | A color-code has to be associated to the number of pairs involved in links, parameters need to be set in the <colorcode> subblock. The format is as follows: color=minimum number,maximum number of pairs for the circos output or R,G,B=minimum number,maximum number of pairs for the bed output. | Text, ex: green= 5,10 (circos) or Text, ex: 0,255,0= 5,10 (bed) |
| list_samples | Name list of samples to compare (= list of file name prefixes) | Text, ex: sample, reference |
| file_suffix | Link file name suffixes. Need to be the same for all samples | Text, ex: _mates.links.filtered |
| circos_output | Creation of the circos converted link file after comparing links | Boolean, value= 1 (run), 0 (skip) |
| bed_output | Creation of the bed converted link file after comparing links | Boolean, value= 1 (run), 0 (skip) |
| sv_output | Creation of the SV converted file after comparing links | Boolean, value= 1 (run), 0 (skip) |


A
DELETION
Sample
Same chromosome
Normal strand sense (2)
Dist > (μ+ 2σ)
Reference
B
INSERTION
Sample
Same chromosome
Normal strand sense (2)
Dist < (μ- 2σ)
Reference
C
DUPLICATION
Sample
Same chromosome
Reverse strand sense (2)
Dist > (μ+ 2σ)
Reference
D
INVERSION
Sample
Same chromosome
Reverse strand sense (1)
Balanced
Reference
E
DUPLICATION + INVERSION
Sample
Same chromosome
Reverse strand sense (1)
Dist < (μ- 2σ)
Reference


F
UNBALANCED TRANSLOCATION
ChrA
Sample
ChrB
Different chromosomes
Normal strand sense (2)
ChrA
Reference
ChrB
G
BALANCED TRANSLOCATION
ChrA
Sample
ChrB
Different chromosomes
Normal strand sense (2)
Balanced
ChrA
Reference
ChrB
H
UNBALANCED TRANSLOCATION + INVERSION
ChrA
Sample
ChrB
Different chromosomes
Reverse strand sense (1)
ChrA
Reference
ChrB
I
BALANCED TRANSLOCATION + INVERSION
ChrA
Sample
ChrB
Different chromosomes
Reverse strand sense (1)
Balanced
ChrA
Reference
ChrB

**Supplementary figure 1. Illustrations of PEM signatures recognized by SVDetect for SV type prediction**

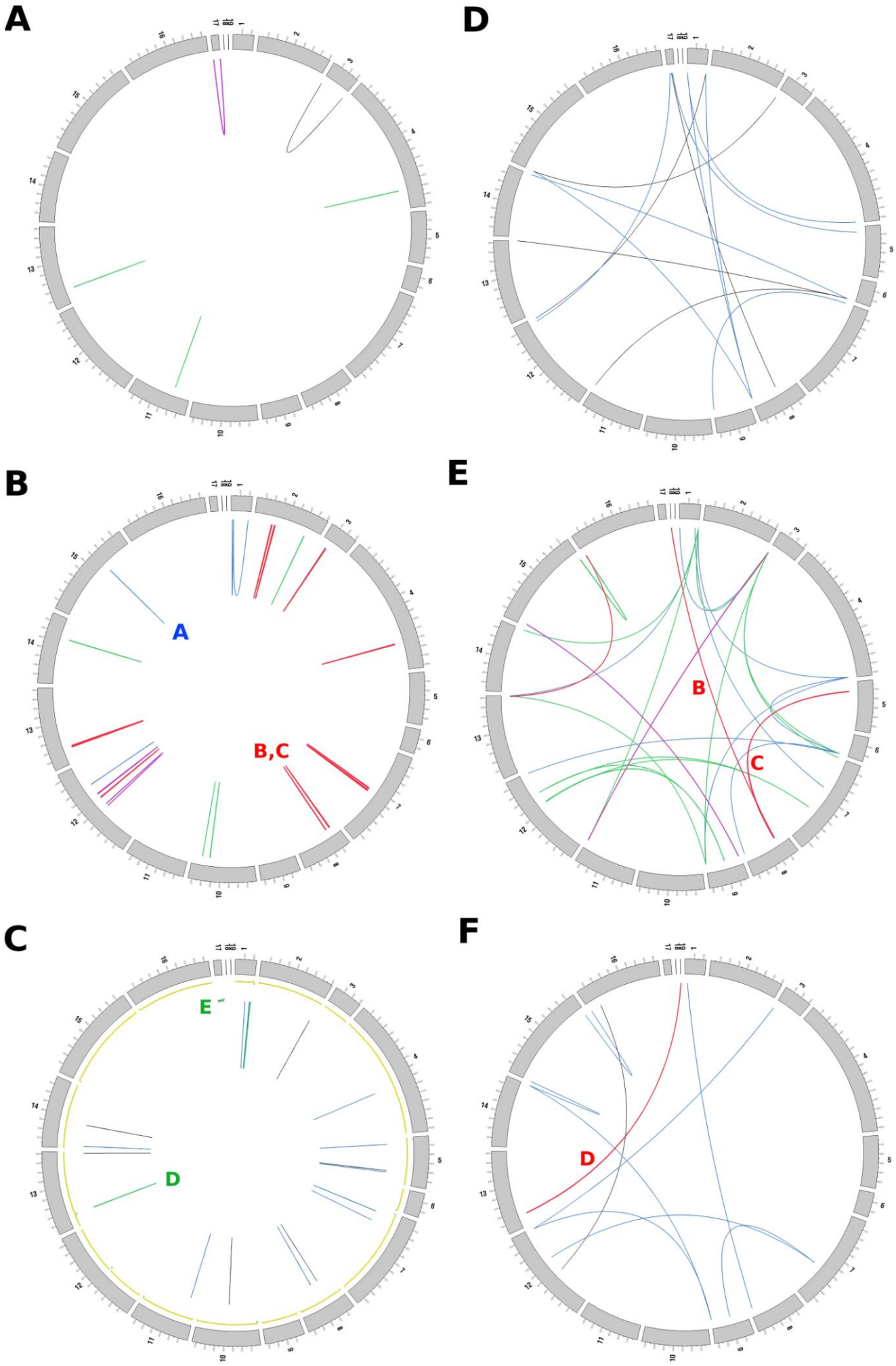
Sky-blue lines represent differences between the donor sample and the reference sample whose chromosomal coordinates are returned by SVDetect: the deleted or inserted regions (A, B) and the predicted location of breakpoints represented by double dotted arrows (D, F-I).

A pair of ticks represents the location of a read pair (the blue mate follows the purple) connected by a span line, the arrows represent their orientation and the associated number indicates the order. A black line represents a chromosome that is the same between the donor sample and the reference (A-E), whereas brown and orange lines represent two distinct chromosomes (F-I). Dotted black arrows indicate the span distance of paired ends. Red arrows indicate the orientation of chromosomal regions (these arrows are dotted in the reference).

**(A-E) Intra-chromosomal SVs.** A deletion is predicted from paired-end spans larger than a specific cutoff defined by the mean insert size $\mu$ plus a fold-number of the standard deviation $\sigma$ as a threshold value (here set to 2) (A). Similarly, a simple insertion has an insert distance < cutoff (B). Inversions are detected when one end changes orientation and the order of the two mates is reversed. We need a balanced PEM signature between paired-ends to detect the location of the two breakpoints underlying the inversion (D). Large tandem duplications are predicted when ends have both reversed orientations with an insert distance > cutoff (C), whereas inverted duplications are seen when only one of the two ends has preserved orientation with an insert distance < cutoff (E).

**(F-I) Inter-chromosomal SVs.** An inter-chromosomal translocation is predicted when paired-ends map to two different chromosomes. The identification of two ordered end subgroups with proper orientations in each mapped chromosome is necessary to discriminate between balanced and unbalanced chromosomal translocations (F,G). Under these conditions, if one of the two pairs has a different orientation combined with the expected reversed order of ends in the second group, inverted balanced or unbalanced translocations are predicted (H,I).

Dist, distance; ChrA, chromosome A; ChrB, chromosome B.


A
D
B
A
B,C
E
B
C
C
E
D
F
D

**Supplementary figure 2. Graphical visualization of predicted SVs from yeast data analysis with SVDetect**

Common/sample-specific predicted intra-chromosomal and inter-chromosomal SVs were detected between the WT and mutant strains with the SV comparison function of SVDetect.
(A) WT-specific intra-chromosomal SVs (n=5). (B) Common intra-chromosomal SVs (n=24). (C) Mutant-specific intra-chromosomal SVs (n=19). (D) WT-specific inter-chromosomal SVs (n=12). (E) Common inter-chromosomal SVs (n=25). (F) Mutant-specific inter-chromosomal SVs (n=11).

Genomic locations of inter-and intra-chromosomal links are shown using the *Circos* software (**5**). Starting from outside of the circle, the following features are displayed: chromosome ideograms with associated genomic coordinates, scatter plot of the copy-number profile (for C only), and spans of inter- and intra chromosomal links. The number of pairs involved in links is here color-coded with black (10), blue (11-20), green (21-50), purple (51-100), orange (101-200) and red (>200) pairs. Known structural variants are shown with the same ID letter used in the Supplementary table 2. Chromosomes 17, 18 and 19 correspond here to the mitochondrial chromosome, the CEB1 and KanMX sequences, respectively.

**Supplementary method 1. The order filtering procedure**

*Definition*. Direct order of read clusters.

A cluster of reads is in *direct order* when more than one half of its reads are mapped in the same orientation, and for these reads the following is true: “if one ranks paired-end reads according to their left end genomic positions, then the same ranking is observed for the right ends with possible permutations for *closely mapped* read pairs (as defined below)”.

*Definition*. Inverted order of read clusters.

A cluster of reads is in *inverted order* when more than one half of its reads are mapped in the same orientation, and for these reads the following is true: “if one ranks paired-end reads according to their left end genomic positions, then the opposite ranking is observed for the right ends with possible permutations for *closely mapped* read pairs”.

Using the information from the read pair orientation of clusters, we can determine the expected read order of pairs in a potential SV. For example, with paired-ends in which a correct read orientation of ends is defined by Forward-Reverse (FR), if the orientation of the majority of read pairs in a cluster is FR then the order of reads should be *direct*, otherwise the order should be *inverted*.

In the case of a deletion (see Supplementary figure 1A), an insertion (B), a duplication (C) or an unbalanced translocation without change of chromosome orientation (F), the order of reads should be *direct*. In case of an inverted duplication (E), an inversion (D) or an unbalanced translocation with a change of the chromosome orientation (H), the order of reads should be *inverted*.

We define *closely mapped* read pairs as pairs that have their left ends mapped within a certain distance from each other so that genomic positions of their right ends can be possibly inverted comparing to the positions of their left ends due to insert size variability. We apply a filtering procedure to discard all read pairs whose order disagrees with the order of the majority of pairs in a cluster (explained further at the end of this section). Since the order of closely mapped read pairs can be inverted by chance, in the filtering procedure we do not take into account changes of order for *closely mapped* pairs. With SVDetect, we select the cutoff based on the distance between left end positions of reads to which we apply filtering. The cutoff value M is selected so that for two left ends of pairs separated by a distance larger than M, the probability that the order between the two right ends is correct would be ≥95%.

*Definition*. Closely mapped read pairs.
Two pairs with a distance between their left end positions shorter than $\sqrt{2}\sigma \cdot x_{95\%}$, where σ is the standard deviation of insert size lengths and $x_{95\%}$ is a 0.95-quantile of the standard normal distribution, are considered to be closely mapped.

*Proposition*. If (*i*) the distance between left ends of two pairs is longer than $\sqrt{2}\sigma \cdot x_{1-\alpha}$, (*ii*) the insert size lengths follow a normal distribution $N(\mu,\sigma^2)$ and (*iii*) the cluster is identified in *direct order* (or *inverted order*), then the probability that the left end of the first pair maps upstream of the left end of the second pair, and the right end of the first pair maps downstream (or upstream) of the right end of the second pair, is less than or equal to α.

Proof (in the case of *direct order* clusters). Suppose that the insert size lengths of two read pairs A and B are normally distributed with known μ and σ, and *l* is the distance between left ends of pairs ($A_{start}$ - $B_{start}$ = *l*). Then, the difference between right ends ($A_{end}$ - $B_{end}$) follows the normal distribution $N(l, 2\sigma^2)$. Hence, the probability of getting a negative value of ($A_{end}$ - $B_{end}$) is equal to $\Phi(-l/(\sqrt{2}\,\sigma))$, where $\Phi(x)$ is the cumulative distribution function of the

standard normal distribution $N(0,1)$. Thus, if $l > \sqrt{2}\sigma \cdot x_{1-\alpha}$, the probability of ($A_{end}$ - $B_{end}$ < 0) is less than $\Phi(-x_{1-\alpha}) = \alpha$. □

In the case of clusters in *inverted order*, the proof is analogous.

Below, we describe the filtering algorithm that is used to filter out pairs whose order is inconsistent with the majority of pairs in a cluster.

**Order filtering algorithm**. When the order filtering algorithm is called, the information of the cluster order, *direct* or *inverted*, is known. For each pair *P*, we can find (*i*) pairs that are *closely mapped* to *P*, (*ii*) pairs which are located at a larger distance to *P* and whose order is in agreement with *P*, and (*iii*) pairs located at a larger distance to *P* and whose order disagrees with *P*. If there are no pairs in the cluster whose orders disagree with each other, the whole read cluster is conserved, and the order filtering is stopped. Otherwise, we continue the algorithm by removing the read pair showing the highest number of cases (*iii*). Then, we check again if all pair orders are in agreement and if it is still not the case, the next pair is removed following the same decision parameters. We iterate until the order of all pairs is correct.

In the output file of SVDetect, we indicate how many read pairs have been removed after running this procedure. The user can then define a cutoff based on the percentage of the pairs that have been filtered out. We suggest keeping clusters which had, at most, 1/3 of read pairs discarded.

**References**


1. Sindi, S. *et al.* (2009) A geometric approach for classification and comparison of structural variants. *Bioinformatics* 25(12):i222-230.
2. Chen, K. *et al.* (2009) BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* 6, 677-681.
3. Hormozdiari, F. *et al.* (2009) Combinatorial algorithms for structural variation detection in high-throughput sequenced genomes. *Genome Res.* 19(7):1270-8.
4. Ribeyre *et al.* (2009) The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming sequences in vivo. *PloS Genetics* 5, e1000475 (pp. 14)
5. Krzywinski, M. *et al.* (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639-1645.