Structural Variants

Structural variations

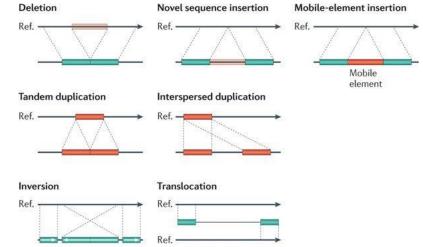
Structural genomic events > 50 bp

CNVs, but also structural rearrangements

Common in human genomes in normal population

Major cause of phenotypic variation

Common in some diseases, cancer +++

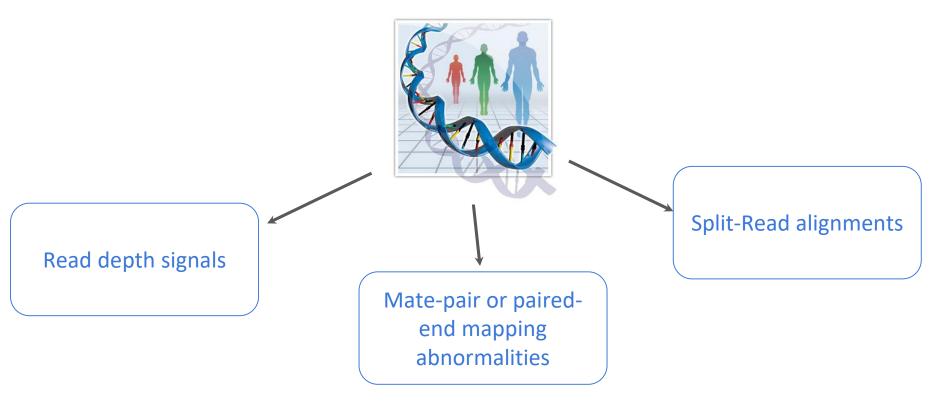


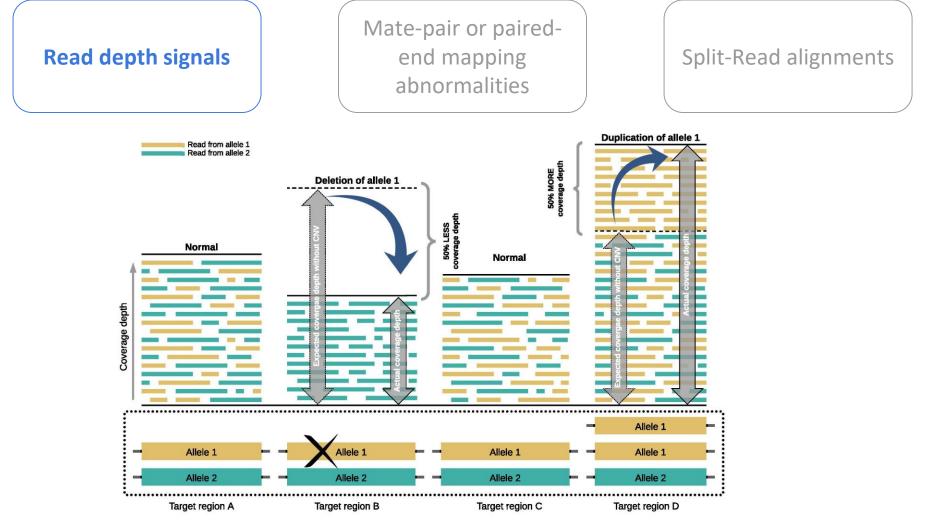
Nature Reviews | Genetics

Growing effects on rare disease, autism, schizophrenia

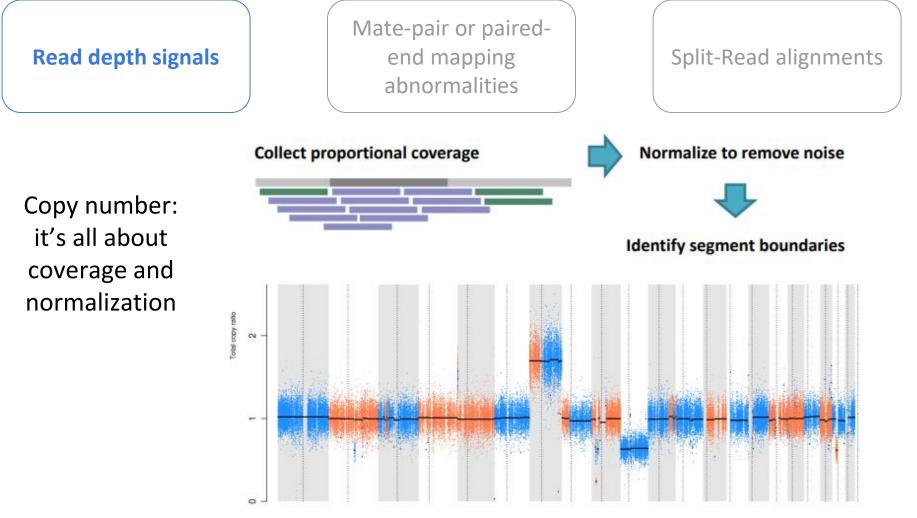
How to detect Structural variations ?

Detection of genomic rearrangements





Singh, A.K., Olsen, M.F., Lavik, L.A.S. et al. Detecting copy number variation in next generation sequencing data from diagnostic gene panels. BMC Med Genomics



from Introduction to Somatic Variant Discovery, GATK Best Practices for Variant Discovery

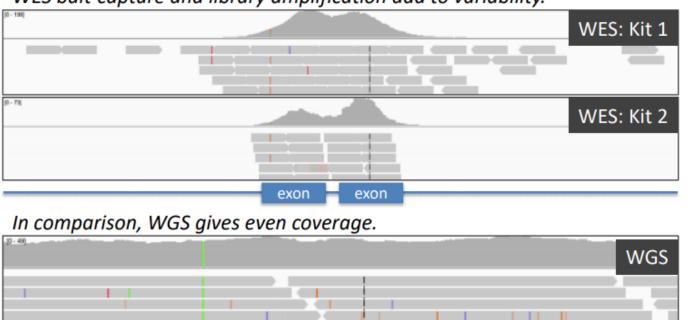
Mate-pair or pairedend mapping abnormalities

Split-Read alignments

WES bait-capture and library amplification add to variability.



Coverage is variable across WES targets and kits



from Somatic Copy Number Alterations, GATK Best Practices for Variant Discovery

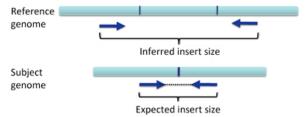
Mate-pair or pairedend mapping abnormalities

Split-Read alignments

Paired reads can yield evidence of structural events.

Alignment coloring options help highlight these events based on:

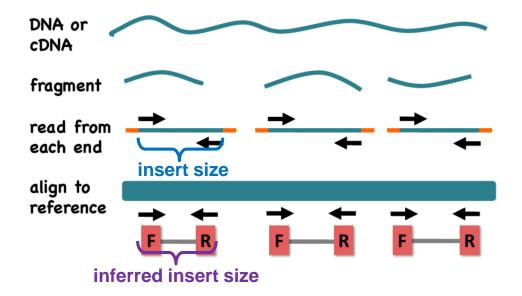
- Inferred insert size (template length)
- Pair orientation (relative strand of pair)

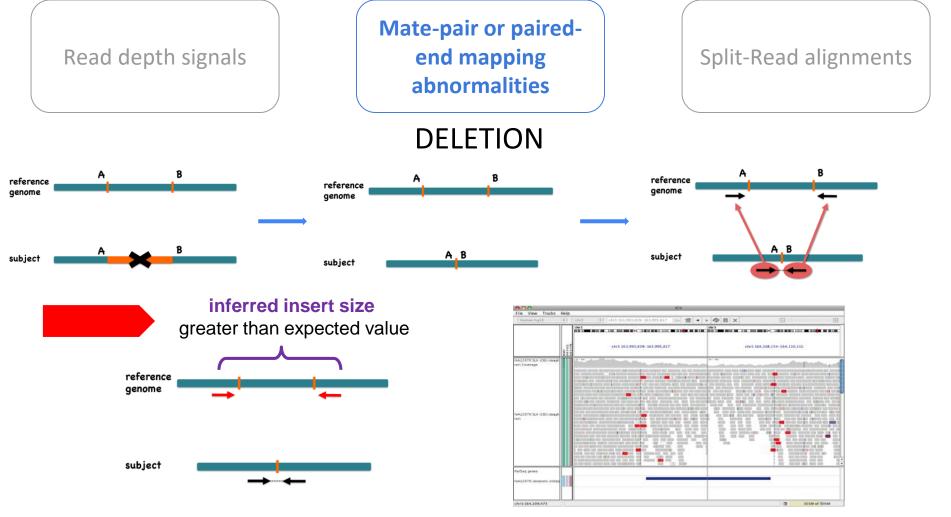




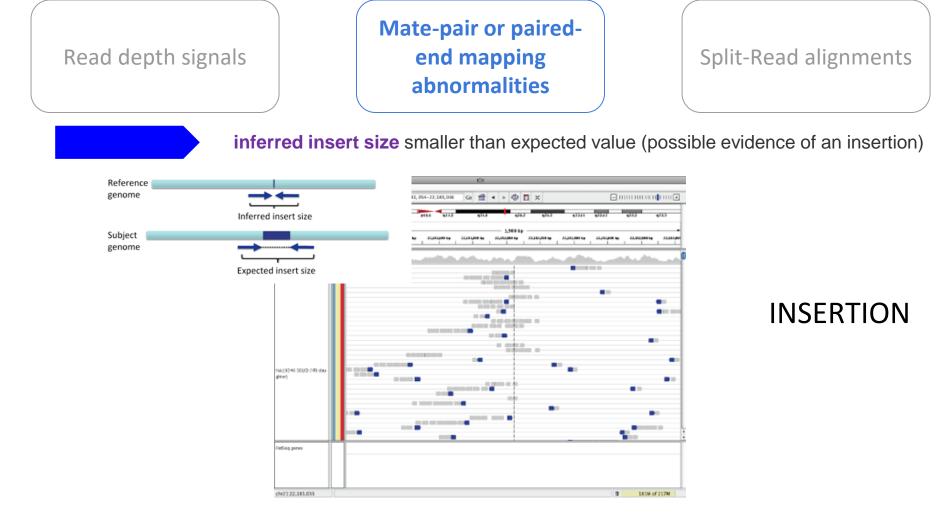
Differences between **insert size** and **inferred insert size** can be used to detect evidence of structural variants, including :

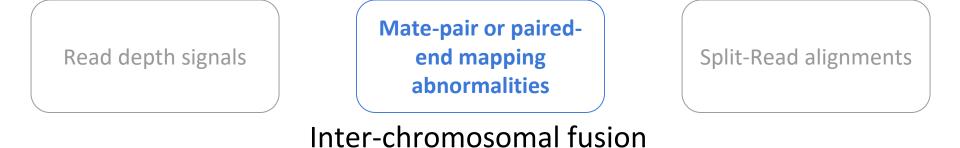
- Deletions
- Insertions
- Inter-chromosomal rearrangements

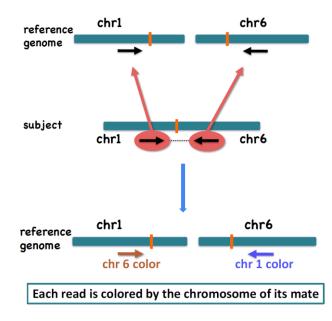




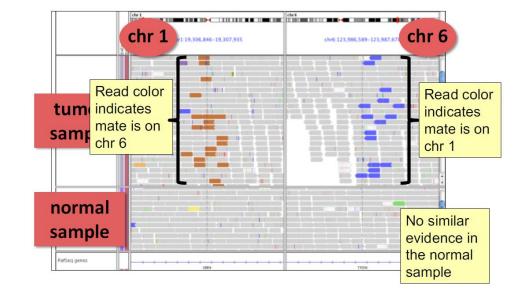
http://software.broadinstitute.org/software/igv/interpreting_insert_size







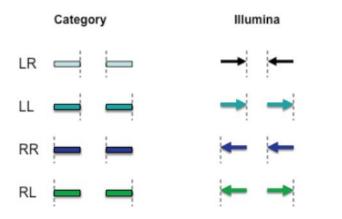
Chromosome colors:



Mate-pair or pairedend mapping abnormalities

Split-Read alignments

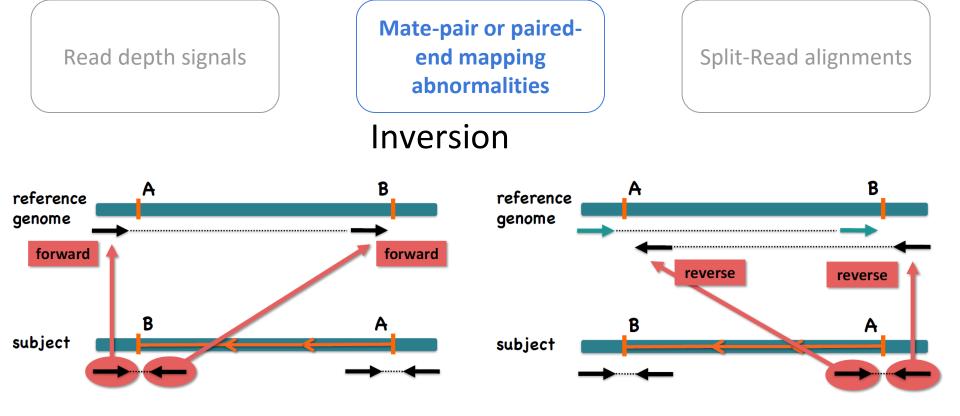
IGV Interpreting Color by Pair Orientation

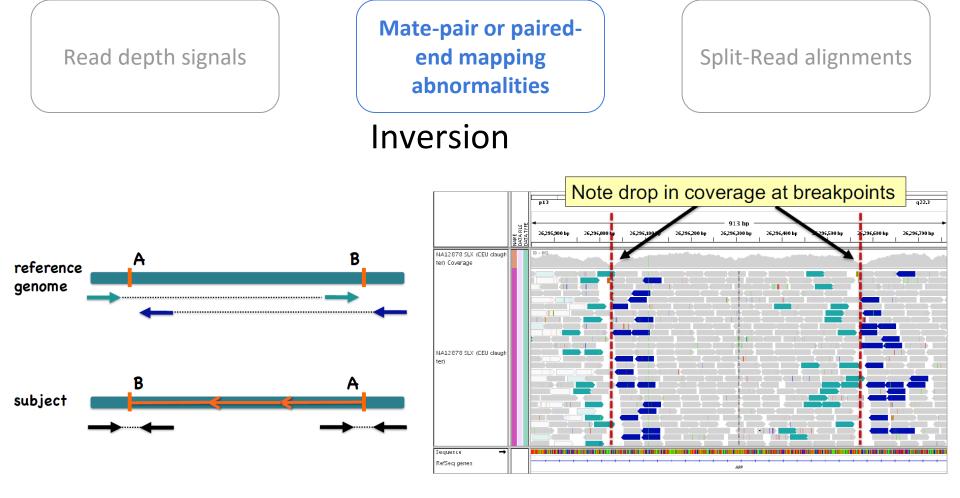


Orientation of paired reads can reveal

evidence of structural events, including:

- Inversions
- Duplications
- Translocations
- LR Normal reads. The reads are left and right (respectively) of the unsequenced part of the sequenced DNA fragment when aligned back to the reference genome.
- LL,RR Implies inversion in sequenced DNA with respect to reference.
- RL Implies duplication or translocation with respect to reference.

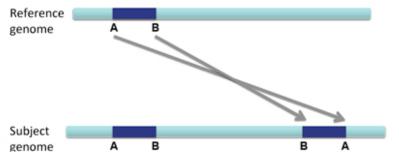




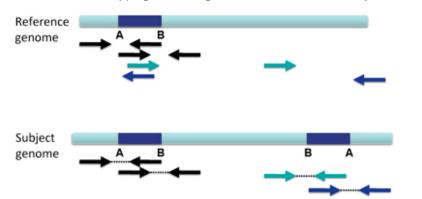
Mate-pair or pairedend mapping abnormalities

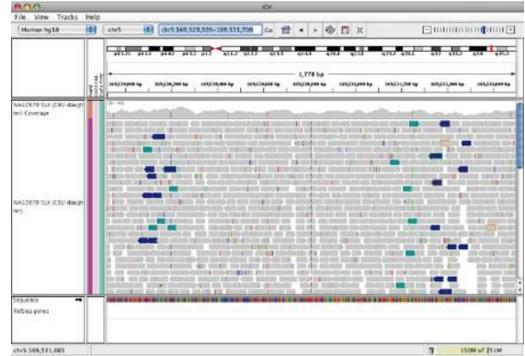
Split-Read alignments

Inverted Duplication



There will be overlapping left and right reads, and there will likely be alter



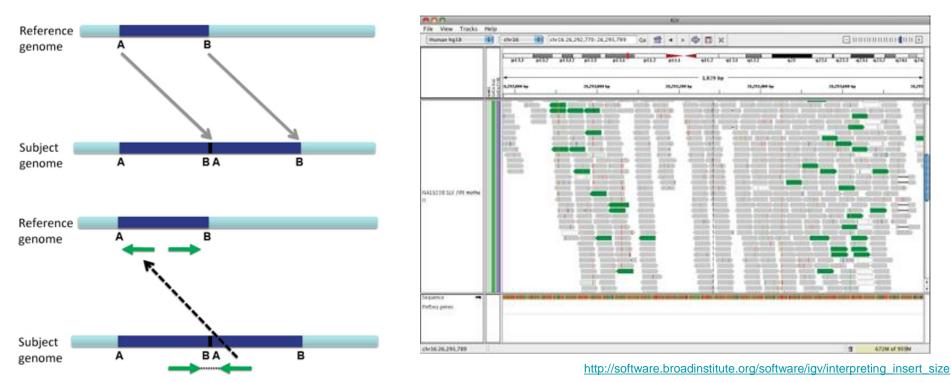


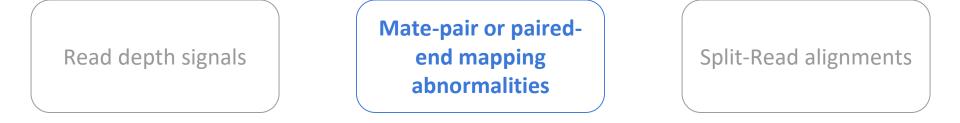
http://software.broadinstitute.org/software/igv/interpreting_insert_size

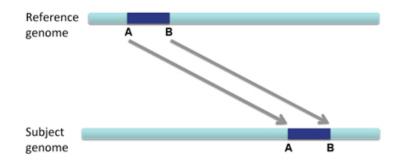
Mate-pair or pairedend mapping abnormalities

Split-Read alignments

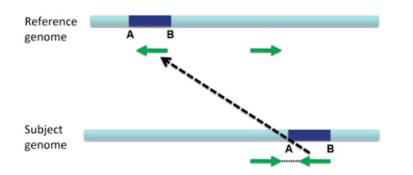
Tandem Duplication







Translocation on the Same Chromosome

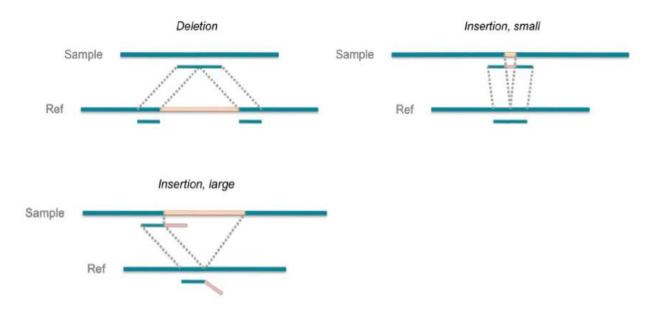


Mate-pair or pairedend mapping abnormalities

Read depth signals

Split-Read alignments

Reads spanning the exact breakpoint of a structural variation

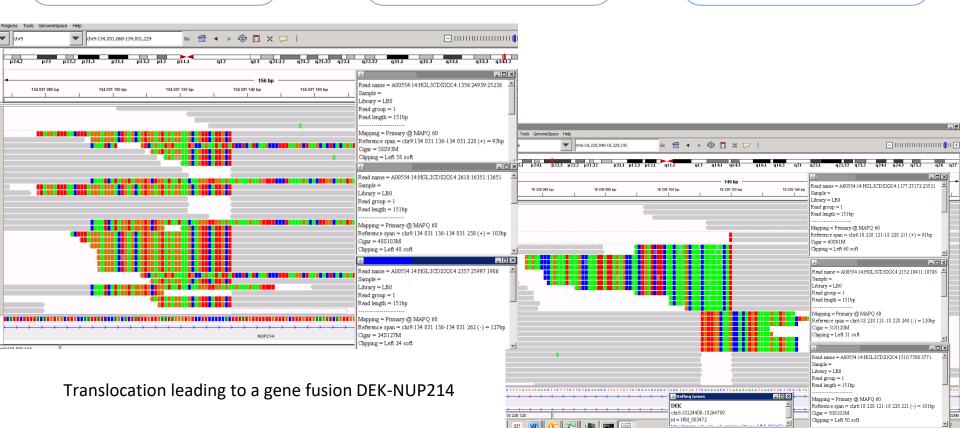


Zhang et al. Identification of genomic indels and structural variations using split reads. BMC Genomics 12, 375 (2011)

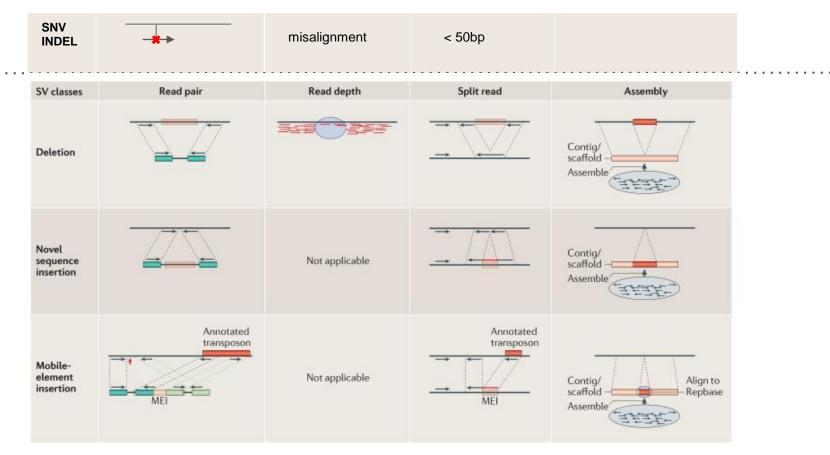
Mate-pair or pairedend mapping abnormalities

Read depth signals

Split-Read alignments

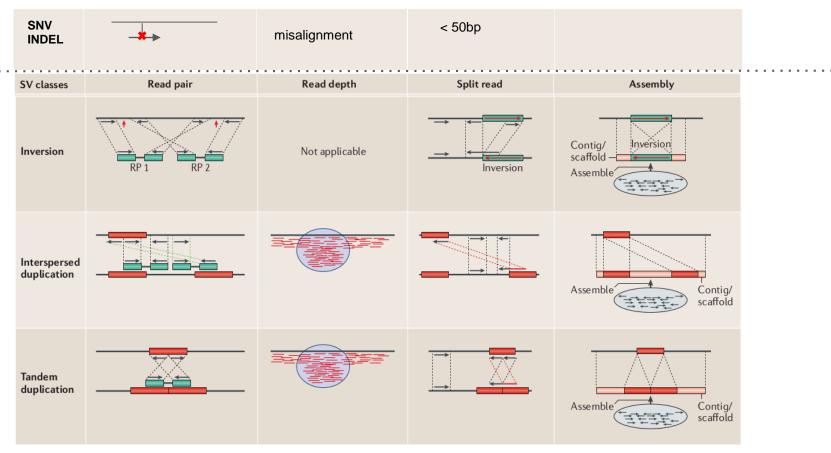


Structural Variation Calling



Genome structural variation discovery and genotyping. Nature Review Genetics 2014

Structural Variation Calling



Genome structural variation discovery and genotyping. Nature Review Genetics 2014

How to represent Structural Variations in VCF ?

<pre>##INF0=<id=ciend,number=2,type=integer,desc ##INF0=<id=cipos,number=2,type=integer,desc ##INF0=<id=end,number=1,type=integer,descri ##INF0=<id=homlen,number=.,type=integer,descri< pre=""></id=homlen,number=.,type=integer,descri<></id=end,number=1,type=integer,descri </id=cipos,number=2,type=integer,desc </id=ciend,number=2,type=integer,desc </pre>	cription="ID of t cription="Confide cription="Confide iption="End posit scription="Length cription="Sequence cription="Differe	the assess ance interaction of the of base the of base ance in the	abled alternate allele in the assembly file"> erval around END for imprecise variants"> erval around POS for imprecise variants"> the variant described in this record"> e pair identical micro-homology at event breakpoints"> se pair identical micro-homology at event breakpoints"> length between REF and ALT alleles">		
##ALT= <id=del,description="deletion"></id=del,description="deletion">					
##ALT= <id=del:me:alu,description="deletion< td=""><td>of ALU element"></td><td>•</td><td></td><td></td><td></td></id=del:me:alu,description="deletion<>	of ALU element">	•			
##ALT= <id=del:me:l1,description="deletion o<="" td=""><td>of L1 element"></td><td></td><td></td><td></td><td></td></id=del:me:l1,description="deletion>	of L1 element">				
##ALT= <id=dup,description="duplication"></id=dup,description="duplication">					
##ALT= <id=dup:tandem,description="tandem du<="" td=""><td>uplication"></td><td></td><td></td><td></td><td></td></id=dup:tandem,description="tandem>	uplication">				
##ALT= <id=ins,description="insertion nov<="" of="" td=""><td>vel sequence"></td><td></td><td></td><td></td><td></td></id=ins,description="insertion>	vel sequence">				
##ALT= <id=ins:me:alu,description="insertion< td=""><td>n of ALU element"</td><td>></td><td></td><td></td><td></td></id=ins:me:alu,description="insertion<>	n of ALU element"	>			
##ALT= <id=ins:me:l1,description="insertion< td=""><td>of L1 element"></td><td></td><td></td><td></td><td></td></id=ins:me:l1,description="insertion<>	of L1 element">				
##ALT= <id=inv,description="inversion"></id=inv,description="inversion">					
##ALT= <id=cnv,description="copy number="" td="" vari<=""><td>iable region"></td><td></td><td></td><td></td><td></td></id=cnv,description="copy>	iable region">				
##FORMAT= <id=gt,number=1,type=string,descri< td=""><td>iption="Genotype"</td><td>></td><td></td><td></td><td></td></id=gt,number=1,type=string,descri<>	iption="Genotype"	>			
##FORMAT= <id=gq,number=1,type=integer,descr< td=""><td></td><td></td><td></td><td></td><td></td></id=gq,number=1,type=integer,descr<>					
##FORMAT= <id=cn,number=1,type=integer,descr< td=""><td></td><td></td><td></td><td></td><td></td></id=cn,number=1,type=integer,descr<>					
##FORMAT= <id=cnq,number=1,type=float,descri< td=""><td>iption="Copy numb</td><td>per geno</td><td>type quality for imprecise events"></td><td></td><td></td></id=cnq,number=1,type=float,descri<>	iption="Copy numb	per geno	type quality for imprecise events">		
		. FILTER		FORMAT	NA00001
1 2827694 rs2376870 CGTGGATGCGGGGAC C		PASS	SVTYPE=DEL; END=2827708; HOMLEN=1; HOMSEQ=G; SVLEN=-14	GT : GQ	1/1:14
	 6	PASS	SVTYPE=DEL; END=321887; SVLEN=-205; CIPOS=-56, 20; CIEND=-10, 62		0/1:12
	<pre><del:me:alu> 12</del:me:alu></pre>	PASS	SVTYPE=DEL; END=14477381; SVLEN=-297; CIPOS=-22, 18; CIEND=-12, 32		0/1:12
	<ins:me:l1> 23</ins:me:l1>	PASS	SVTYPE=INS;END=9425916;SVLEN=6027;CIPOS=-16,22	GT:GQ	1/1:15
	<dup> 14</dup>	PASS			
4 18665128 . T <	<pre><dup:tandem> 11</dup:tandem></pre>	PASS	SVTYPE=DUP;END=18665204;SVLEN=76;CIPOS=-10,10;CIEND=-10,10	GT:GQ:CN:CNQ	./.:0:5:8.3

##fileformat=VCFv4.3		
##fileDate=20100501		
##reference=1000GenomesPilot-NCBI36		
##assembly=ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/release/sv/breakpoint_assemblies.fasta		
##INFO= <id=bkptid,number=.,type=string,description="id allele="" alternate="" assembled="" assembly="" file"="" in="" of="" the=""></id=bkptid,number=.,type=string,description="id>		
##INFO= <id=ciend,number=2,type=integer,description="confidence around="" end="" for="" imprecise="" interval="" variants"=""></id=ciend,number=2,type=integer,description="confidence>		
##INFO= <id=cipos,number=2,type=integer,description="confidence around="" for="" imprecise="" interval="" pos="" variants"=""></id=cipos,number=2,type=integer,description="confidence>		
##INFO= <id=end,number=1,type=integer,description="end described="" in="" of="" position="" record"="" the="" this="" variant=""></id=end,number=1,type=integer,description="end>		
##INFO= <id=homlen,number=.,type=integer,description="length at="" base="" breakpoints"="" event="" identical="" micro-homology="" of="" pair=""></id=homlen,number=.,type=integer,description="length>		
##INFO= <id=homseq,number=.,type=string,description="sequence at="" base="" breakpoints"="" event="" identical="" micro-homology="" of="" pair=""></id=homseq,number=.,type=string,description="sequence>		
##INFO= <id=svlen,number=.,type=integer,description="difference alleles"="" alt="" and="" between="" in="" length="" ref=""></id=svlen,number=.,type=integer,description="difference>		
##INF0= <id=svtype,number=1,type=string,description="type of="" structural="" variant"=""></id=svtype,number=1,type=string,description="type>		
##ALT= <id=del,description="deletion"></id=del,description="deletion">		
##ALT= <id=del:me:alu,description="deletion alu="" element"="" of=""></id=del:me:alu,description="deletion>		
##ALT= <id=del:me:l1,description="deletion element"="" l1="" of=""></id=del:me:l1,description="deletion>		
##ALT= <id=dup,description="duplication"></id=dup,description="duplication">		
##ALT= <id=dup:tandem,description="tandem duplication"=""></id=dup:tandem,description="tandem>		
##ALT= <id=ins,description="insertion novel="" of="" sequence"=""></id=ins,description="insertion>		
##ALT= <id=ins:me:alu,description="insertion alu="" element"="" of=""></id=ins:me:alu,description="insertion>		
##ALT= <id=ins:me:l1,description="insertion element"="" l1="" of=""></id=ins:me:l1,description="insertion>		
##ALT= <id=inv,description="inversion"></id=inv,description="inversion">		
##ALT= <id=cnv,description="copy number="" region"="" variable=""></id=cnv,description="copy>		
##FORMAT= <id=gt,number=1,type=string,description="genotype"> precise deletion with kn</id=gt,number=1,type=string,description="genotype">	lown breakpoint	
##FORMAT= <id=gq,number=1,type=integer,description="genotype quality"=""></id=gq,number=1,type=integer,description="genotype>		
##FORMAT= <id=cn,number=1,type=integer,description="copy events"="" for="" genotype="" imprecise="" number=""></id=cn,number=1,type=integer,description="copy>		
##FORMAT= <id=cnq,number=1,type=float,description="copy events"="" for="" genotype="" imprecise="" number="" quality=""></id=cnq,number=1,type=float,description="copy>		
	FORMAT NA00001	
	GT:GQ 1/1:14	
	GT:GQ 0/1:12	
	GT:GQ 0/1:12	
	GT:GQ 1/1:15	
	GT:GQ:CN:CNQ ./.:0:3:1	
4 18665128 . T <pre>CDUP:TANDEM> 11 PASS SVTYPE=DUP;END=18665204;SVLEN=76;CIPOS=-10,10;CIEND=-10,10</pre>	GT:GQ:CN:CNQ ./.:0:5:8	.3

##f	ileformat=VCFv4	4.3							
##1	ileDate=2010050	01							
##r	eference=1000Ge	enomesPilo	t-NCBI36						
##8	ssembly=ftp://f	ftp-trace.	ncbi.nih.gov/1	000genomes/ftp	/rel	ease/sv,	/breakpoint_assemblies.fasta		
##I	NFO= <id=bkptid,< td=""><td>Number=.</td><td>Type=String, Dea</td><td>scription="ID</td><td>of the</td><td>he asser</td><td>mbled alternate allele in the assembly file"></td><td></td><td></td></id=bkptid,<>	Number=.	Type=String, Dea	scription="ID	of the	he asser	mbled alternate allele in the assembly file">		
##I	NFO= <id=ciend.n< td=""><td>Number=2.T</td><td>vpe=Integer.De</td><td>scription="Cor</td><td>fide</td><td>nce inte</td><td>erval around END for imprecise variants"></td><td></td><td></td></id=ciend.n<>	Number=2.T	vpe=Integer.De	scription="Cor	fide	nce inte	erval around END for imprecise variants">		
							erval around POS for imprecise variants">		
							the variant described in this record">		
							e pair identical micro-homology at event breakpoints">		
							se pair identical micro-homology at event breakpoints">		
							length between REF and ALT alleles">		
	NFO= <id=svtype< td=""><td></td><td>· · · · · · · · · · · · · · · · · · ·</td><td></td><td></td><td></td><td>0</td><td></td><td></td></id=svtype<>		· · · · · · · · · · · · · · · · · · ·				0		
	LT= <id=del,desc< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></id=del,desc<>								
	LT= <id=del:me:< td=""><td></td><td></td><td>n of ALU eleme</td><td>nt"></td><td></td><td></td><td></td><td></td></id=del:me:<>			n of ALU eleme	nt">				
	LT= <id=del:me:i< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></id=del:me:i<>								
	LT= <id=dup,desc< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></id=dup,desc<>								
	LT= <id=dup:tane< td=""><td></td><td></td><td>Duplication"></td><td></td><td></td><td></td><td></td><td></td></id=dup:tane<>			Duplication">					
	LT= <id=ins,desc< td=""><td></td><td></td><td></td><td>></td><td></td><td></td><td></td><td></td></id=ins,desc<>				>				
	LT= <id=ins:me:< td=""><td></td><td></td><td></td><td></td><td>></td><td></td><td></td><td></td></id=ins:me:<>					>			
	LT= <id=ins:me:i< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></id=ins:me:i<>								
	LT= <id=inv,desc< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></id=inv,desc<>								
	LT= <id=cnv,desc< td=""><td></td><td></td><td>riable region'</td><td>></td><td></td><td></td><td></td><td></td></id=cnv,desc<>			riable region'	>				
	ORMAT= <id=gt, no<="" td=""><td></td><td></td><td>0</td><td></td><td>></td><td>imprecise deletion of a</td><td>pproximat</td><td>elv 205 bp</td></id=gt,>			0		>	imprecise deletion of a	pproximat	elv 205 bp
	ORMAT= <id=gq, nu<="" td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></id=gq,>								
							otype for imprecise events">		
##F	ORMAT= <id=cnq, n<="" td=""><td>Number=1,T</td><td>ype=Float,Desc</td><td>ription="Copy</td><td>numb</td><td>er geno</td><td>type quality for imprecise events"></td><td></td><td></td></id=cnq,>	Number=1,T	ype=Float,Desc	ription="Copy	numb	er geno	type quality for imprecise events">		
#CH	ROM POS ID	RE	F	ALT	QUAL	FILTER	INFO	FORMAT	NA00001
1	2827694 rs2	2376870 CG	TGGATGCGGGGGAC	с		PASS	SVTYPE=DEL:END=2827708:HOMLEN=1:HOMSEQ=G:SVLEN=-14	GT: GQ	1/1:14
2	321682 .	т			6	PASS	SVTYPE=DEL; END=321887; SVLEN=-205; CIPOS=-56, 20; CIEND=-10, 62	GT: GQ	0/1:12
2	14477084 .	С		<del:me:alu></del:me:alu>	12	PASS	SVTYPE=DEL; END=14477381; SVLEN=-297; CIPUS=-22, 18; CIEND=-12, 32	GT:GQ	0/1:12
3	9425916 .	С		<ins:me:l1></ins:me:l1>	23	PASS	SVTYPE=INS; END=9425916; SVLEN=6027; CIPOS=-16, 22	GT: GQ	1/1:15
3	12665100 .	A		<dup></dup>	14	PASS	SVTYPE=DUP; END=12686200; SVLEN=21100; CIPOS=-500, 500; CIEND=-500, 500	GT: GQ: CN: CNQ	./.:0:3:16.2
4	18665128 .	т		<dup:tandem></dup:tandem>	11	PASS	SVTYPE=DUP;END=18665204;SVLEN=76;CIPOS=-10,10;CIEND=-10,10	GT: GQ: CN: CNQ	./.:0:5:8.3

	##fileformat=VCFv4.3					
	##fileDate=20100501					
	##reference=1000GenomesPilot-NCBI36					
	##assembly=ftp://ftp-trace.ncbi.nih.gov	v/1000genomes/ftp/re	lease/sv/	breakpoint_assemblies.fasta		
	##INFO= <id=bkptid,number=.,type=string< td=""><td>Description="ID of</td><td>the asser</td><td>abled alternate allele in the assembly file"></td><td></td><td></td></id=bkptid,number=.,type=string<>	Description="ID of	the asser	abled alternate allele in the assembly file">		
	##INFO= <id=ciend,number=2,type=integer< td=""><td>Description="Confid</td><td>lence inte</td><td>erval around END for imprecise variants"></td><td></td><td></td></id=ciend,number=2,type=integer<>	Description="Confid	lence inte	erval around END for imprecise variants">		
	##INFO= <id=cipos,number=2,type=integer< td=""><td>Description="Confid</td><td>lence inte</td><td>erval around POS for imprecise variants"></td><td></td><td></td></id=cipos,number=2,type=integer<>	Description="Confid	lence inte	erval around POS for imprecise variants">		
	##INFO= <id=end,number=1,type=integer,de< td=""><td>escription="End posi</td><td>tion of t</td><td>the variant described in this record"></td><td></td><td></td></id=end,number=1,type=integer,de<>	escription="End posi	tion of t	the variant described in this record">		
	##INFO= <id=homlen,number=.,type=integer< td=""><td>r,Description="Lengt</td><td>h of base</td><td>e pair identical micro-homology at event breakpoints"></td><td></td><td></td></id=homlen,number=.,type=integer<>	r,Description="Lengt	h of base	e pair identical micro-homology at event breakpoints">		
	##INFO= <id=homseq.number=type=string< td=""><td>Description="Sequen</td><td>ice of bas</td><td>se pair identical micro-homology at event breakpoints"></td><td></td><td></td></id=homseq.number=type=string<>	Description="Sequen	ice of bas	se pair identical micro-homology at event breakpoints">		
	##INFO= <id=svlen,number=.,type=integer< td=""><td>Description="Differ</td><td>ence in 1</td><td>ength between REF and ALT alleles"></td><td></td><td></td></id=svlen,number=.,type=integer<>	Description="Differ	ence in 1	ength between REF and ALT alleles">		
	##INFO= <id=svtype,number=1,type=string< td=""><td>Description="Type o</td><td>f structu</td><td>ural variant"></td><td></td><td></td></id=svtype,number=1,type=string<>	Description="Type o	f structu	ural variant">		
	##ALT= <id=del.description="deletion"></id=del.description="deletion">					
	##ALT= <id=del:me:alu,description="delet< td=""><td>tion of ALU element"</td><td>></td><td></td><td></td><td></td></id=del:me:alu,description="delet<>	tion of ALU element"	>			
	##ALT= <id=del:me:l1,description="delet:< td=""><td></td><td></td><td></td><td></td><td></td></id=del:me:l1,description="delet:<>					
	##ALT= <id=dup.description="duplication"< td=""><td></td><td></td><td></td><td></td><td></td></id=dup.description="duplication"<>					
	##ALT= <id=dup:tandem,description="tande< td=""><td></td><td></td><td></td><td></td><td></td></id=dup:tandem,description="tande<>					
	##ALT= <id=ins,description="insertion of<="" td=""><td></td><td></td><td></td><td></td><td></td></id=ins,description="insertion>					
	##ALT= <id=ins:me:alu,description="inse< td=""><td></td><td>"></td><td></td><td></td><td></td></id=ins:me:alu,description="inse<>		">			
	##ALT= <id=ins:me:l1,description="insert< td=""><td></td><td></td><td></td><td></td><td></td></id=ins:me:l1,description="insert<>					
	##ALT= <id=inv,description="inversion"></id=inv,description="inversion">					
	##ALT= <id=cnv,description="copy number<="" td=""><td></td><td></td><td></td><td></td><td></td></id=cnv,description="copy>					
	##FORMAT= <id=gt,number=1,type=string,de< td=""><td>0</td><td>"></td><td>imprecise deletion of an</td><td>ALU elem</td><td>ent</td></id=gt,number=1,type=string,de<>	0	">	imprecise deletion of an	ALU elem	ent
	##FORMAT= <id=gq,number=1,type=integer,i< td=""><td></td><td></td><td></td><td></td><td></td></id=gq,number=1,type=integer,i<>					
	##FORMAT= <id=cn,number=1,type=integer,i< td=""><td></td><td></td><td></td><td></td><td></td></id=cn,number=1,type=integer,i<>					
	##FORMAT= <id=cnq,number=1,type=float,d< td=""><td></td><td></td><td></td><td></td><td></td></id=cnq,number=1,type=float,d<>					
	#CHROM POS ID REF		L FILTER		ORMAT N	A00001
	1 2827694 rs2376870 CGTGGATGCGGGGG				T:GQ 1	1/1:14
	2 321682 T	(DEL) 6				0/1:12
ſ	2 14477084 . C	<del:me:alu> 12</del:me:alu>	PASS			0/1:12
5	3 9425916 . C	<ins:me:l1> 23</ins:me:l1>	PASS		T:GQ 1	1/1:15
	3 12665100 . A	<dup> 14</dup>			T:GQ:CN:CNQ .	/.:0:3:16.2
	4 18665128 . T	<dup:tandem> 11</dup:tandem>			T:GQ:CN:CNQ .	

##1	ileformat=VCFv4.3							
##1	ileDate=20100501							
##r	eference=1000Genome	sPilot-NCBI36						
##a	ssembly=ftp://ftp-t	race.ncbi.nih.gov/1	000genomes/ft	p/rel	ease/sv	/breakpoint_assemblies.fasta		
##I	NFO= <id=bkptid,numb< td=""><td>er=.,Type=String,De</td><td>scription="ID</td><td>of t</td><td>he asse</td><td>mbled alternate allele in the assembly file"></td><td></td><td></td></id=bkptid,numb<>	er=.,Type=String,De	scription="ID	of t	he asse	mbled alternate allele in the assembly file">		
##I	NFO= <id=ciend.numbe< td=""><td>r=2.Type=Integer.De</td><td>scription="Con</td><td>nfide</td><td>nce int</td><td>erval around END for imprecise variants"></td><td></td><td></td></id=ciend.numbe<>	r=2.Type=Integer.De	scription="Con	nfide	nce int	erval around END for imprecise variants">		
						erval around POS for imprecise variants">		
						the variant described in this record">		
						e pair identical micro-homology at event breakpoints">		
						se pair identical micro-homology at event breakpoints">		
						length between REF and ALT alleles">		
	NFO= <id=svtype, numb<="" td=""><td>, ,,</td><td></td><td></td><td></td><td>0</td><td></td><td></td></id=svtype,>	, ,,				0		
	LT= <id=del,descript< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td></id=del,descript<>							
	LT= <id=del:me:alu,d< td=""><td></td><td>on of ALU elem</td><td>ent"></td><td></td><td></td><td></td><td></td></id=del:me:alu,d<>		on of ALU elem	ent">				
	LT= <id=del:me:l1,de< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td></id=del:me:l1,de<>							
	LT= <id=dup,descript< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td></id=dup,descript<>							
	LT= <id=dup:tandem,d< td=""><td></td><td>Duplication"></td><td></td><td></td><td></td><td></td><td></td></id=dup:tandem,d<>		Duplication">					
	LT= <id=ins,descript< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td></id=ins,descript<>							
	LT= <id=ins:me:alu,d< td=""><td></td><td></td><td></td><td>></td><td></td><td></td><td></td></id=ins:me:alu,d<>				>			
	LT= <id=ins:me:l1,de< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td></id=ins:me:l1,de<>							
##A	LT= <id=inv,descript< td=""><td>ion="Inversion"></td><td></td><td></td><td></td><td></td><td></td><td></td></id=inv,descript<>	ion="Inversion">						
##A	LT= <id=cnv,descript< td=""><td>ion="Copy number va</td><td>riable region</td><td>"></td><td></td><td></td><td></td><td></td></id=cnv,descript<>	ion="Copy number va	riable region	">				
	ORMAT= <id=gt, number<="" td=""><td></td><td></td><td></td><td>></td><td>imprecise insertion of a</td><td>an L1 eleme</td><td>ent</td></id=gt,>				>	imprecise insertion of a	an L1 eleme	ent
##F	ORMAT= <id=gq, number<="" td=""><td>=1,Type=Integer,Des</td><td>cription="Gen</td><td>otype</td><td>qualit</td><td></td><td></td><td></td></id=gq,>	=1,Type=Integer,Des	cription="Gen	otype	qualit			
##F	ORMAT= <id=cn, number<="" td=""><td>=1, Type=Integer, Des</td><td>cription="Copy</td><td>y num</td><td>ber gen</td><td>otype for imprecise events"></td><td></td><td></td></id=cn,>	=1, Type=Integer, Des	cription="Copy	y num	ber gen	otype for imprecise events">		
##F	ORMAT= <id=cnq, numbe<="" td=""><td>r=1,Type=Float,Desc</td><td>ription="Copy</td><td>numb</td><td>er geno</td><td>type quality for imprecise events"></td><td></td><td></td></id=cnq,>	r=1,Type=Float,Desc	ription="Copy	numb	er geno	type quality for imprecise events">		
#CH	ROM POS ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	NA00001
1	2827694 rs23768	70 CGTGGATGCGGGGAC	С		PASS	SVTYPE=DEL; END=2827708; HOMLEN=1; HOMSEQ=G; SVLEN=-14	GT: GQ	1/1:14
2	321682 .	т		6	PASS	SVTYPE=DEL; END=321887; SVLEN=-205; CIPOS=-56, 20; CIEND= 10, 62	GT: GQ	0/1:12
2	14477084 .	С	<del:me:alu></del:me:alu>	12	PASS	SVTYPE=DEL:END=14477381:SVLEN=-297:CIPOS=-22.18:CIEND=-12.32	GT: GQ	0/1:12
3	9425916 .	С	<ins:me:l1></ins:me:l1>	23	PASS	SVTYPE=INS; END=9425916; SVLEN=6027; CIPOS=-16, 22	GT: GQ	1/1:15
3	12665100 .	A	<dup></dup>	14	PASS	SVTYPE=DUP; END=12686200; SVLEN=21100; CIPOS=-500, 500; CIEND=-500, 500	GT: GQ: CN: CNQ	./.:0:3:16.2
4	18665128 .	Т	<dup: tandem=""></dup:>	11	PASS	SVTYPE=DUP; END=18665204; SVLEN=76; CIPOS=-10, 10; CIEND=-10, 10	GT: GQ: CN: CNQ	./.:0:5:8.3

##fileformat=VCFv4.3		
##fileDate=20100501		
##reference=1000GenomesPilot-NCBI36		
##assembly=ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/release/sv/breakpoint_assemblies.fasta		
##INFO= <id=bkptid,number=.,type=string,description="id allele="" alternate="" assembled="" assembly="" file"="" in="" of="" the=""></id=bkptid,number=.,type=string,description="id>		
##INFO= <id=ciend_number=2.type=integer.description="confidence around="" end="" for="" imprecise="" interval="" variants"=""></id=ciend_number=2.type=integer.description="confidence>		
##INFO= <id=cipos,number=2,type=integer,description="confidence around="" for="" imprecise="" interval="" pos="" variants"=""></id=cipos,number=2,type=integer,description="confidence>		
##INFO= <id=end.number=1.type=integer.description="end described="" in="" of="" position="" record"="" the="" this="" variant=""></id=end.number=1.type=integer.description="end>		
##INFO= <id=homlen,number=.,type=integer,description="length at="" base="" breakpoints"="" event="" identical="" micro-homology="" of="" pair=""></id=homlen,number=.,type=integer,description="length>		
##INFO= <id=homseq,number=.,type=string,description="sequence at="" base="" breakpoints"="" event="" identical="" micro-homology="" of="" pair=""></id=homseq,number=.,type=string,description="sequence>		
##INFO= <id=svlen.number=.,type=integer,description="difference alleles"="" alt="" and="" between="" in="" length="" ref=""></id=svlen.number=.,type=integer,description="difference>		
##INFO= <id=svtype,number=1,type=string,description="type of="" structural="" variant"=""></id=svtype,number=1,type=string,description="type>		
##ALT= <id=del.description="deletion"></id=del.description="deletion">		
##ALT= <id=del:me:alu_description="deletion alu="" element"="" of=""></id=del:me:alu_description="deletion>		
##ALT= <id=del:me:l1,description="deletion element"="" l1="" of=""></id=del:me:l1,description="deletion>		
##ALT= <id=dup_description="duplication"></id=dup_description="duplication">		
##ALT= <id=dup:tandem,description="tandem duplication"=""></id=dup:tandem,description="tandem>		
##ALT= <id=ins,description="insertion novel="" of="" sequence"=""></id=ins,description="insertion>		
##ALT= <id=ins:me:alu,description="insertion alu="" element"="" of=""></id=ins:me:alu,description="insertion>		
##ALT= <id=ins:me:l1,description="insertion element"="" l1="" of=""></id=ins:me:l1,description="insertion>		
##ALT= <id=inv,description="inversion"></id=inv,description="inversion">		
##ALT= <id=cnv,description="copy number="" region"="" variable=""></id=cnv,description="copy>		
##FORMAT= <id=gt,number=1,type=string,description="genotype"> imprecise duplication of</id=gt,number=1,type=string,description="genotype">	of approxim	ately 21Kb
##FORMAT= <id=gq,number=1,type=integer,description="genotype quality"=""></id=gq,number=1,type=integer,description="genotype>		
##FORMAT= <id=cn,number=1,type=integer,description="copy events"="" for="" genotype="" imprecise="" number=""></id=cn,number=1,type=integer,description="copy>		
##FORMAT= <id=cnq,number=1,type=float,description="copy events"="" for="" genotype="" imprecise="" number="" quality=""></id=cnq,number=1,type=float,description="copy>		
#CHROM POS ID REF ALT QUAL FILTER INFO	FORMAT	NA00001
1 2827694 rs2376870 CGTGGATGCGGGGAC C . PASS SVTYPE=DEL;END=2827708;HOMLEN=1;HOMSEQ=G;SVLEN=-14	GT:GQ	1/1:14
2 321682 . T <pre> </pre>		0/1:12
2 14477084 . C <pre><pre><pre><pre>C </pre><pre><pre><pre><pre><pre><pre><pre><</pre></pre></pre></pre></pre></pre></pre></pre></pre></pre>		0/1:12
3 9425916 . C <ins:me:l1> 23 PASS SVTYPE=INS:END=9425916;SVLEN=6027;CIP0S=-16;22</ins:me:l1>		1/1:15
	GT: GQ CN CNQ	
4 18665128 . T <pre><pre><pre><pre>4 18665128 . T </pre><pre><pre><pre><pre><pre><pre><pre><</pre></pre></pre></pre></pre></pre></pre></pre></pre></pre>	GT:GQ:CN:CNQ	./.:0:5:8.3

##fileformat=VCFv4.3	
##fileDate=20100501	
##reference=1000GenomesPilot-NCBI36	
##assembly=ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/release/sv/breakpoint_assemblies.fasta	
##INFO= <id=bkptid,number=.,type=string,description="id allele="" alternate="" assembled="" assembly="" file"="" in="" of="" the=""></id=bkptid,number=.,type=string,description="id>	
##INF0= <id=ciend,number=2,type=integer,description="confidence around="" end="" for="" imprecise="" interval="" variants"=""></id=ciend,number=2,type=integer,description="confidence>	
##INF0= <id=cip0s,number=2,type=integer,description="confidence around="" for="" imprecise="" interval="" pos="" variants"=""></id=cip0s,number=2,type=integer,description="confidence>	
##INF0= <id=end.number=1.type=integer.description="end described="" in="" of="" position="" record"="" the="" this="" variant=""></id=end.number=1.type=integer.description="end>	
##INFO= <id=homlen,number=.,type=integer,description="length at="" base="" breakpoints"="" event="" identical="" micro-homology="" of="" pair=""></id=homlen,number=.,type=integer,description="length>	
##INFO= <id=homseq.number=type=string.description="sequence at="" base="" breakpoints"="" event="" identical="" micro-homology="" of="" pair=""></id=homseq.number=type=string.description="sequence>	
##INFO- <id=nulen,number=.,type=string,description="difference alleles"="" alt="" and="" between="" in="" length="" ref=""></id=nulen,number=.,type=string,description="difference>	
##INFO- <id=svtype,number=1,type=string,description="type of="" structural="" variant"=""></id=svtype,number=1,type=string,description="type>	
##ARD-STD=DEL.Description="Deletion">	
##ALT= <id=del:me:alu.description="deletion alu="" element"="" of=""></id=del:me:alu.description="deletion>	
##ALT= <id=del:me:al0,description="deletion al0="" element"="" of=""> ##ALT=<id=del:me:l1,description="deletion element"="" l1="" of=""></id=del:me:l1,description="deletion></id=del:me:al0,description="deletion>	
##ALT= <id=dup.description="duplication"></id=dup.description="duplication">	
##ALT= <id=dup:tandem,description="tandem duplication"=""></id=dup:tandem,description="tandem>	
##ALT= <id=ins,description="insertion novel="" of="" sequence"=""></id=ins,description="insertion>	
##ALT= <id=ins:me:alu,description="insertion alu="" element"="" of=""></id=ins:me:alu,description="insertion>	
##ALT= <id=ins:me:l1,description="insertion element"="" l1="" of=""></id=ins:me:l1,description="insertion>	
##ALT= <id=inv,description="inversion"></id=inv,description="inversion">	
##ALT= <id=cnv,description="copy number="" region"="" variable=""></id=cnv,description="copy>	lighting of 70km
##FORMAT= <id=gt,number=1,type=string,description="genotype"> imprecise tandem dup</id=gt,number=1,type=string,description="genotype">	lication of 760p
##FORMAT= <id=gq,number=1,type=integer,description="genotype quality"=""></id=gq,number=1,type=integer,description="genotype>	
##FORMAT= <id=cn,number=1,type=integer,description="copy events"="" for="" genotype="" imprecise="" number=""></id=cn,number=1,type=integer,description="copy>	
##FORMAT= <id=cnq,number=1,type=float,description="copy events"="" for="" genotype="" imprecise="" number="" quality=""></id=cnq,number=1,type=float,description="copy>	
#CHROM POS ID REF ALT QUAL FILTER INFO	FORMAT NA00001
1 2827694 rs2376870 CGTGGATGCGGGGAC C . PASS SVTYPE=DEL;END=2827708;HOMLEN=1;HOMSEQ=G;SVLEN=-14	GT:GQ 1/1:14
2 321682 . T <pre> </pre>	GT:GQ 0/1:12
2 14477084 . C <pre><pre><pre><pre><pre>C </pre><pre><pre><pre><pre><pre><pre><pre><</pre></pre></pre></pre></pre></pre></pre></pre></pre></pre></pre>	GT:GQ 0/1:12
3 9425916 . C <ins:me:l1> 23 PASS SVTYPE=INS;END=9425916;SVLEN=6027;CIPOS=-16,22</ins:me:l1>	GT:GQ 1/1:15
<u>3 12665100</u> , <u>A</u> <u><dup> 14 PASS SVTYPE=DUP;END=12686200;SVLEN=21100;CIPOS=-500,500;CIEKD=-500,500</dup></u>	GT:GQ:CN:CNQ ./.:0:3:16.2
4 18665128 . T <pre><pre>CDUP:TANDEM> 11 PASS SVTYPE=DUP;END=18665204;SVLEN=76;CIPOS=-10,10;CIEND=-10,10</pre></pre>	GT:GQ CN CNQ ./.:0:5:8.3

Encoding complex rearrangements with breakends in VCF

Rearrangement breakpoint defined as 2 breakends \rightarrow novel adjacency

Breakend is encoded by SVTYPE=BND in the INFO field

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO
2	321681	bnd_W	G	G]17:198982]	6	PASS	SVTYPE=BND
2	321682	bnd_V	т]13:123456]T	6	PASS	SVTYPE=BND
13	123456	bnd_U	\mathbf{C}	C[2:321682[6	PASS	SVTYPE=BND
13	123457	bnd_X	Α	[17:198983[A	6	PASS	SVTYPE=BND
17	198982	bnd_Y	Α	A]2:321681]	6	PASS	SVTYPE=BND
17	198983	bnd_Z	\mathbf{C}	[13:123457[C	6	PASS	SVTYPE=BND

REF	ALT	Meaning
s	t[p[piece extending to the right of p is joined after t
s	t]p]	reverse comp piece extending left of p is joined after
s]p]t	piece extending to the left of p is joined before t
s	[p[t	reverse comp piece extending right of p is joined before t

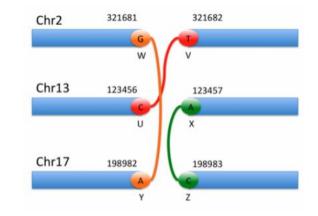
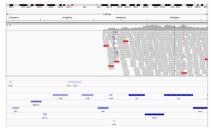


Figure 1: All possible orientations of breakends

How to visualize Structural Variations ?

A. Linear Genome Browser



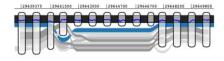
F. Linear Coordinate Plot

D. SV Table

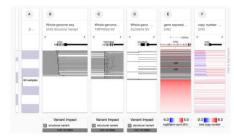
E. Circos Plot

		Sorts			Terpet			lers .
60	(from	Transport.	- 4/1	170105	brauport	1.46	procity	ntere
	Dise Al	1		Dom N	1			
•	· civită	47.028.102		+uk10	47 (164,582		26430	
•	• ov5	179.080.866		• evs	179.000.007		24902	= nev
•	 ms 	35.758.099		•m•	38,706,090		24433	
	e unité	76236.052		w unr16	PERMIT		10070	- 14/
•	• cm12	80.842.171		Cavity .	85391,791		19610	
	it chrif	1100,0007.0000		× 1007	158,287,836		19546	
•	* CHR	24/372,424		* cnik	24,980,945		18811	
•	en18	44,544,968		E che 18	44,345,398		17748	
•	• cv18	34,893,086		*m18	34,853,089		10645	
	• chr2	181065.349		• m2	195.001.890		14211	

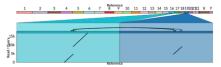
I. Graph View

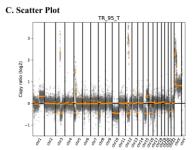


J. Population View



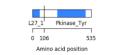
B. Dot Plot



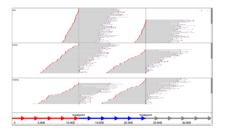


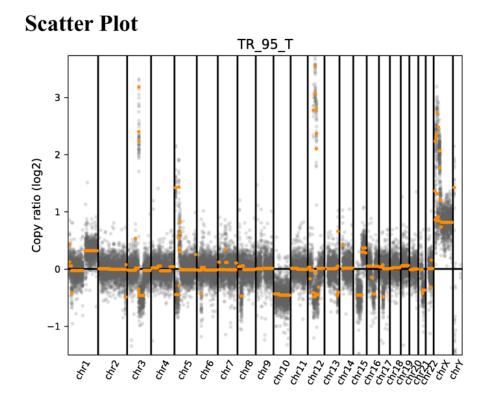
G. Two-way View

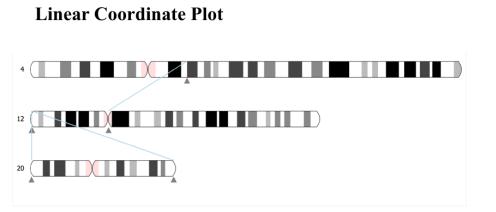
ENSMUST0000023454-ENSMUST0000002487

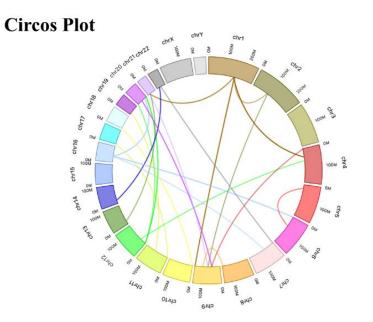


H. Multi-way View









Graph View

