Structural Variants

Structural variations

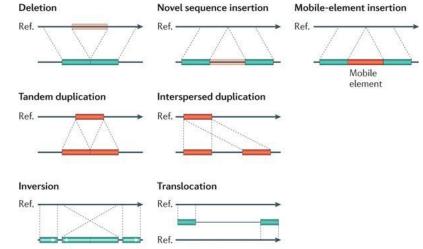
Structural genomic events > 50 bp

CNVs, but also structural rearrangements

Common in human genomes in normal population

Major cause of phenotypic variation

Common in some diseases, cancer +++

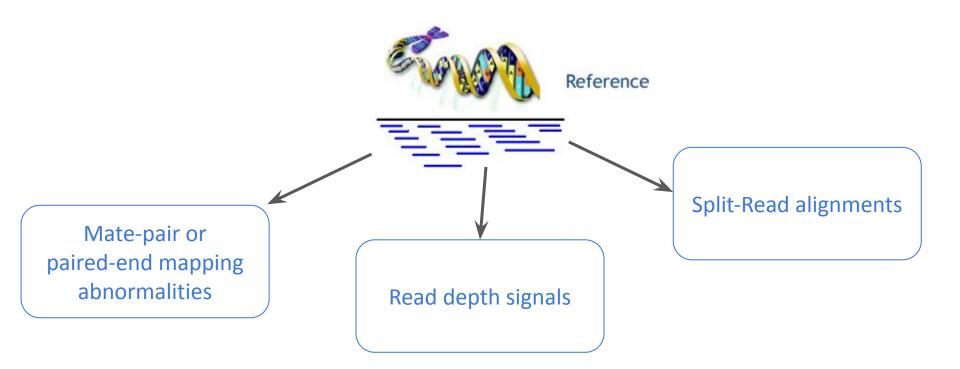


Nature Reviews | Genetics

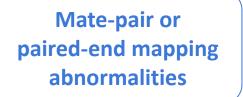
Growing effects on rare disease, autism, schizophrenia

How to detect Structural variations ?

Detection of genomic rearrangements



Adapted from http://www.cs.cmu.edu/~sssykim/teaching/s13/slides/LectureSVII.pdf



Read depth signals

Split-Read alignments

inferred insert size that is larger than expected (possible evidence of a deletion) Reference genome 1 · · · · X Inferred insert size Subject chr3 163,993,839-363,995,817 chv1 164.108.154-164.110.132 genome Expected insert size NA12878 SLX (CBJ cleaph fielding genes A12878-Ministers 109 chr3:164,109,473 303M of 505M

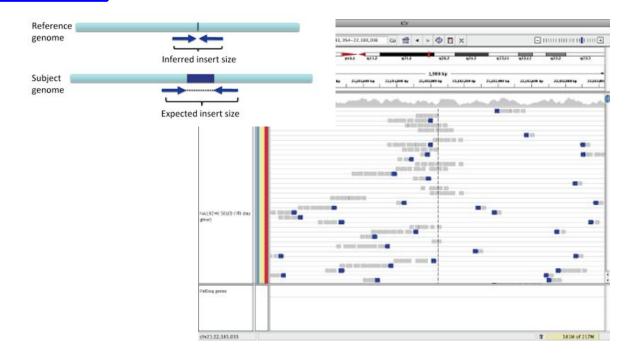
http://software.broadinstitute.org/software/igv/interpreting insert size

Mate-pair or paired-end mapping abnormalities

Read depth signals

Split-Read alignments

inferred insert size that is smaller than expected (possible evidence of an insertion)



Mate-pair or paired-end mapping abnormalities

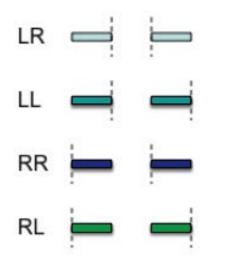
Read depth signals

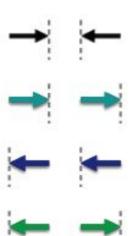
Split-Read alignments

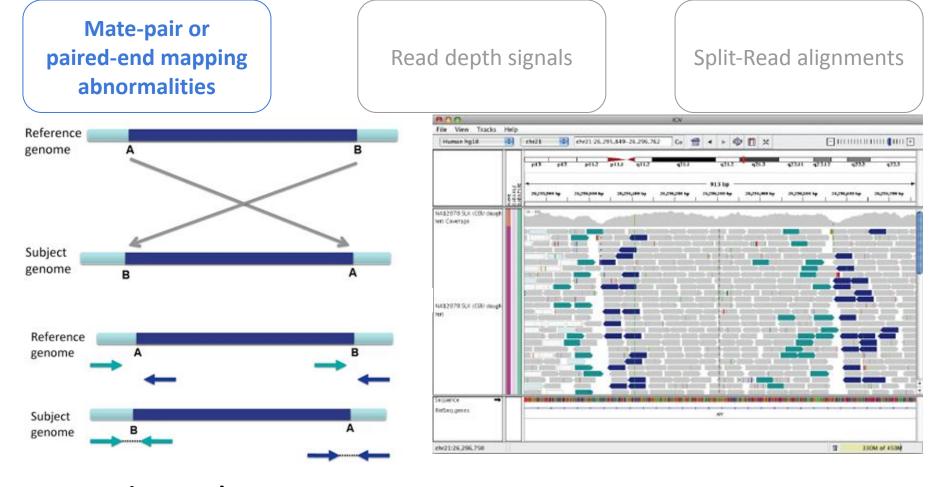
IGV Interpreting Color by Pair Orientation

Category

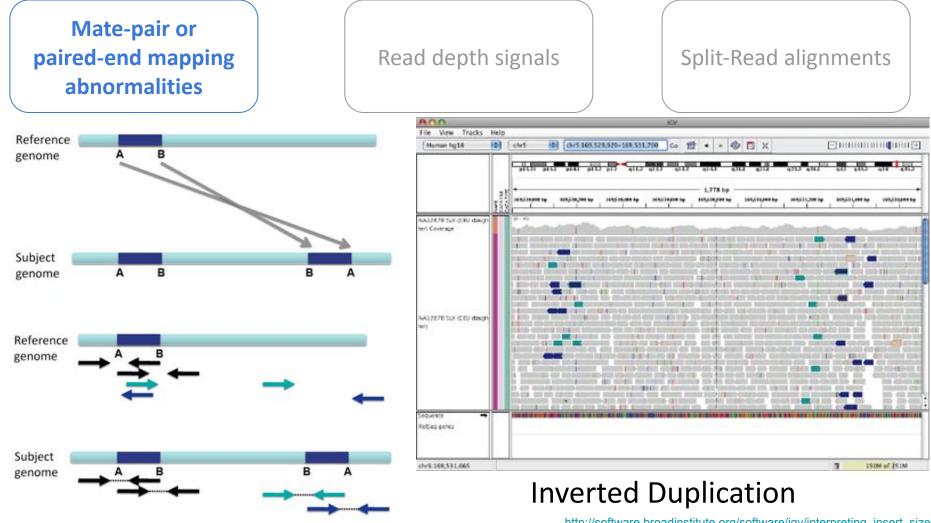
Illumina



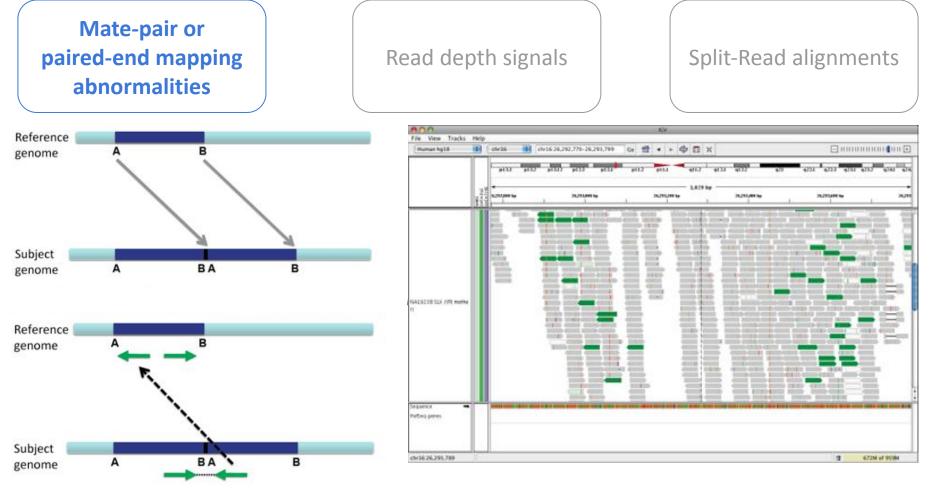




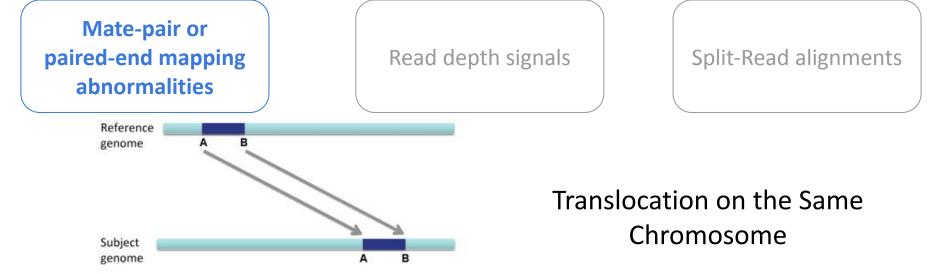
Inversions



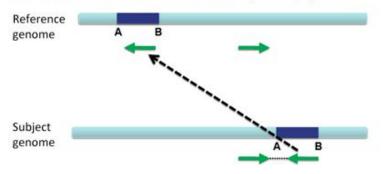
http://software.broadinstitute.org/software/igv/interpreting insert size

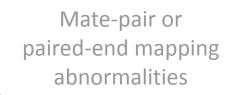


Tandem Duplication



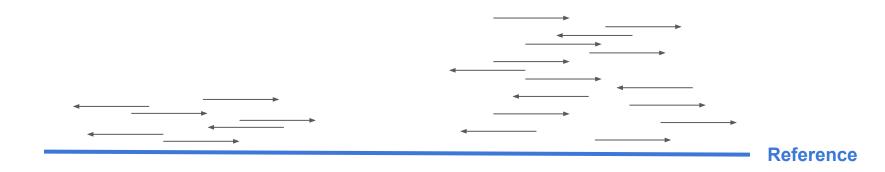
Translocations on the same chromosome can be detected by color-codin between two chromosomes can be detected by coloring by insert size.



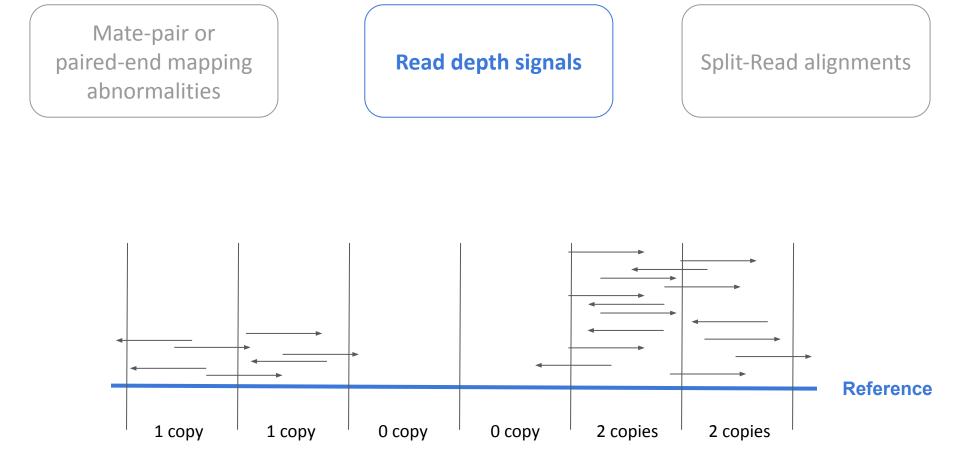


Read depth signals

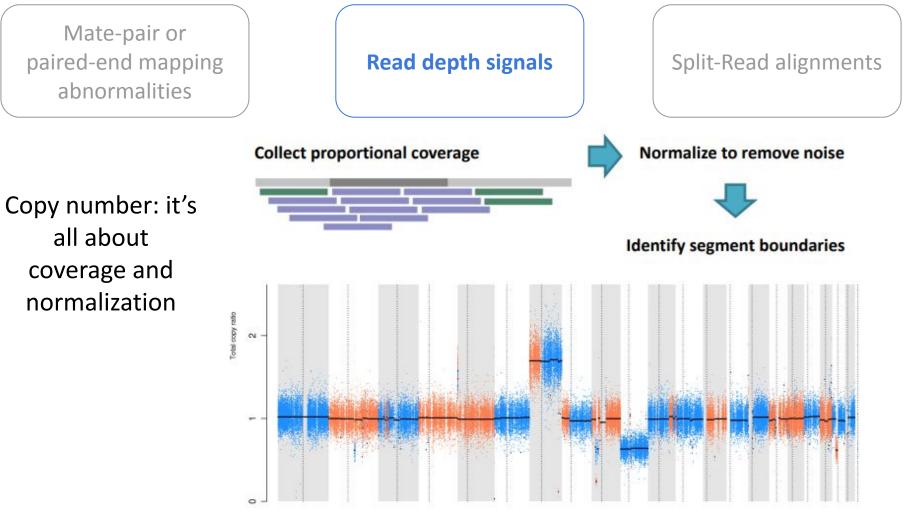
Split-Read alignments



Adapted from http://www.cs.cmu.edu/~sssykim/teaching/s13/slides/LectureSVII.pdf



Adapted from http://www.cs.cmu.edu/~sssykim/teaching/s13/slides/LectureSVII.pdf



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

Mate-pair or paired-end mapping abnormalities

Read depth signals

Split-Read alignments

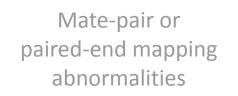
WES bait-capture and library amplification add to variability.



Coverage is variable across WES targets and kits



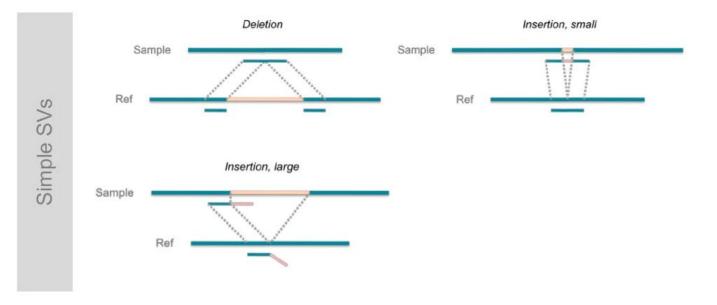
from Somatic Copy Number Alterations, GATK Best Practices for Variant Discovery



Read depth signals

Split-Read alignments

Reads spanning the exact breakpoint of a structural variation

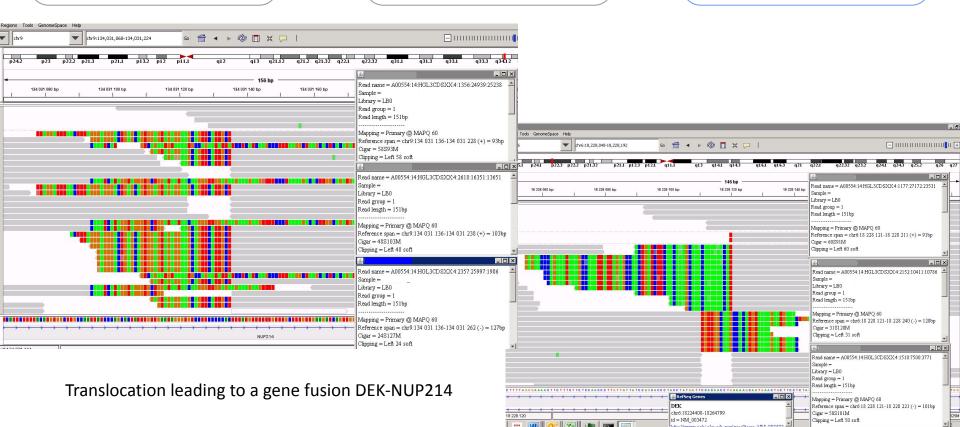


Zhang et al. Identification of genomic indels and structural variations using split reads. BMC Genomics 12, 375 (2011)

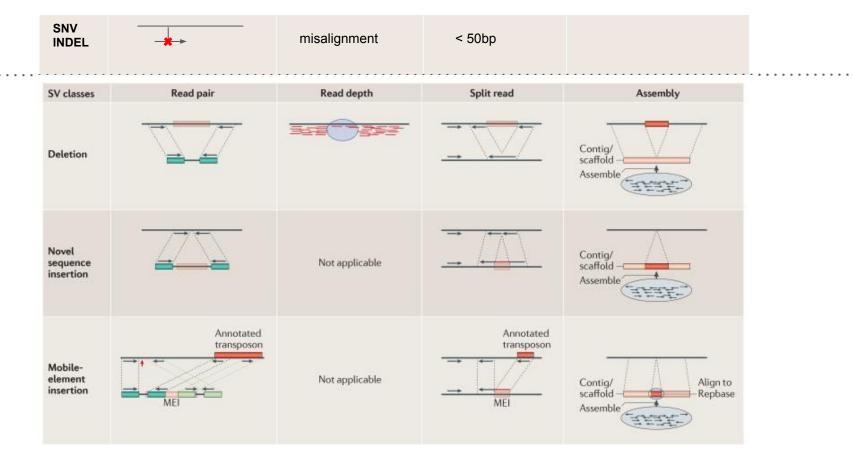
Mate-pair or paired-end mapping abnormalities

Read depth signals

Split-Read alignments

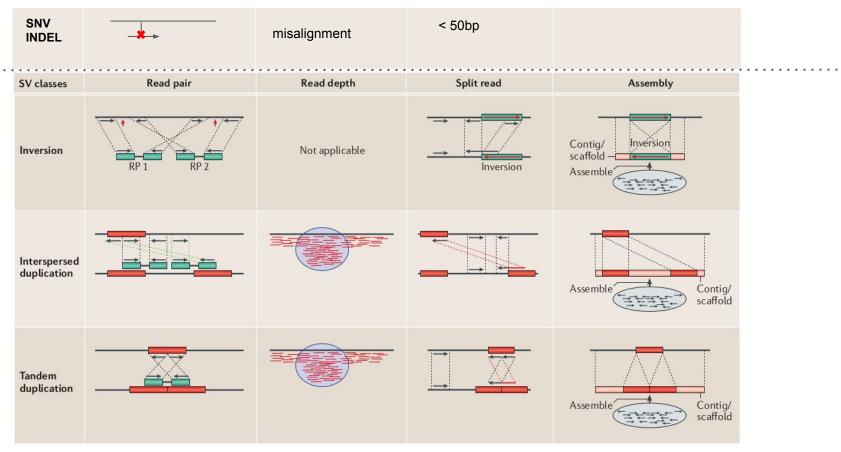


Structural Variation Calling



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?

****	leformat=V	Pred 2							
	ileDate=2010								
			ilot-NCBI36						
				000genomes/ft	n/rel	0000/0W	/breakpoint_assemblies.fasta		
					• • • • • • • • • • • • • • • • • • •		mbled 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">		
			=1,Type=String,De						
			n="Deletion">	scription- iy	pe or	Struct	Irai variant >		
			cription="Deletion"	n of ATH olon	ant IIs				
			ription="Deletion						
			n="Duplication">	of Li elemen					
			cription="Tandem	Duplication">					
			n="Insertion of n	and the second					
	and the second		cription="Inserti	· · · · · · · · · · · · · · · · · · ·		~			
			ription="Insertio			/			
		12 C	n="Inversion">	n of Li elemen					
			n="Copy number va	wishis wanter					
			Type=String,Desc			>			
			Type=Integer,Des	and the second					
							otype for imprecise events">		
							type quality for imprecise events">		
		ID	REF	ALT COPY		FILTER		FORMAT	NA00001
					QUAL				
			T		6				
			C		-				
			100	Contraction of the second s				Charles and the second second second	A CONTRACTOR OF A CONTRACT
		100						Contraction of the second s	
1 2 3 3 4	2827694 321682 14477084 9425916 12665100 18665128	:	CGTGGATGCGGGGAC T C C A T	C <del:me:alu> <ins:me:l1> <dup> <dup:tandem></dup:tandem></dup></ins:me:l1></del:me:alu>	23 14	PASS PASS PASS PASS PASS PASS	SVTYPE=DEL; END=2827708; HOMLEN=1; HOMSEQ=G; SVLEN=-14 SVTYPE=DEL; END=321887; SVLEN=-205; CIPOS=-56, 20; CIEND=-10, 62 SVTYPE=DEL; END=14477381; SVLEN=-297; CIPOS=-22, 18; CIEND=-12, 32 SVTYPE=INS; END=9425916; SVLEN=6027; CIPOS=-16, 22 SVTYPE=DUP; END=12686200; SVLEN=21100; CIPOS=-500, 500; CIEND=-500, 500 SVTYPE=DUP; END=18665204; SVLEN=76; CIPOS=-10, 10; CIEND=-10, 10	GT:GQ GT:GQ GT:GQ GT:GQ:CN:CNQ GT:GQ:CN:CNQ	

##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 of the assembled alternate allele in the assembly file"> ##INFO=<ID=CIEND,Number=2,Type=Integer,Description="Confidence interval around END for imprecise variants"> ##INFO=<ID=CIPOS,Number=2,Type=Integer,Description="Confidence interval around POS for imprecise variants"> ##INFO=<ID=END,Number=1,Type=Integer,Description="End position of the variant described in this record"> ##INFO=<ID=HOMLEN,Number=.,Type=Integer,Description="Length of base pair identical micro-homology at event breakpoints"> ##INFO=<ID=HOMSEQ,Number=.,Type=String,Description="Sequence of base pair identical micro-homology at event breakpoints"> ##INFO=<ID=SVLEN,Number=.,Type=Integer,Description="Difference in length between REF and ALT alleles"> ##INFO=<ID=SVTYPE.Number=1.Type=String.Description="Type of structural variant"> ##ALT=<ID=DEL.Description="Deletion"> ##ALT=<ID=DEL:ME:ALU.Description="Deletion of ALU element"> ##ALT=<ID=DEL:ME:L1,Description="Deletion of L1 element"> ##ALT=<ID=DUP,Description="Duplication"> ##ALT=<ID=DUP:TANDEM,Description="Tandem Duplication"> ##ALT=<ID=INS.Description="Insertion of novel sequence"> ##ALT=<ID=INS:ME:ALU.Description="Insertion of ALU element"> ##ALT=<ID=INS:ME:L1,Description="Insertion of L1 element"> ##ALT=<ID=INV.Description="Inversion"> ##ALT=<ID=CNV,Description="Copy number variable region"> precise deletion with known breakpoint ##FORMAT=<ID=GT.Number=1.Type=String.Description="Genotype"> ##FORMAT=<ID=GQ.Number=1.Type=Integer.Description="Genotype quality"> ##FORMAT=<ID=CN,Number=1,Type=Integer,Description="Copy number genotype for imprecise events"> ##FORMAT=<ID=CNQ,Number=1,Type=Float,Description="Copy number genotype quality for imprecise events"> OUAL ETLTER INFO #CHROM DOS TD DFF ATT FORMAT 2827694 rs2376870 CGTGGATGCGGGGGAC C SVTYPE=DEL; END=2827708; HOMLEN=1; HOMSEQ=G; SVLEN=-14 1 PASS GT:GQ SVTYPE=DEL; END=321887; SVLEN=-205; CIPOS=-56, 20; CIEND=-10, 62 2 321682 . т 6 PASS GT:GQ

2 14477084 . C <DEL:ME:ALU> 12 SVTYPE=DEL; END=14477381; SVLEN=-297; CIPOS=-22, 18; CIEND=-12, 32 PASS GT:GQ 0/1:12 3 9425916 C <INS: ME: L1> 23 PASS SVTYPE=INS; END=9425916; SVLEN=6027; CIPOS=-16, 22 GT:GQ 1/1:15 SVTYPE=DUP; END=12686200; SVLEN=21100; CIPOS=-500, 500; CIEND=-500, 500 3 12665100 . A <DUP> 14 PASS GT:GQ:CN:CNQ ./.:0:3:16.2 4 18665128 . T <DUP: TANDEM> 11 PASS SVTYPE=DUP; END=18665204; SVLEN=76; CIPOS=-10, 10; CIEND=-10, 10 GT:GQ:CN:CNQ ./.:0:5:8.3

NA00001

1/1:14

0/1:12

	fileformat=VCFv4	2						
	fileDate=2010050	170						
		nomesPilot-NCBI36						
			000genomes /fts	Irale	000/01	/breakpoint_assemblies.fasta		
						mbled 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">		
		Number=1, Type=String, De						
		ription="Deletion">	beription ij		001400			
		LU.Description="Deletion	of All alam	nt #>				
		1.Description="Deletion						
		ription="Duplication">	or br cremen.	5.55				
		EM, Description="Tandem	Duplication">					
		ription="Insertion of a	and the second second second	>				
	the second se	LU,Description="Insert:			0			
		1,Description="Insertion						
		ription="Inversion">						
		ription="Copy number va	riable region	>				
##.				vpe">		imprecise deletion of a	nproximate	elv 205 bn
##	FORMAT= <id=gt, nu<="" td=""><td>mber=1, Type=String, Des</td><td>ription="Genot</td><td></td><td></td><td>imprecise deletion of a</td><td>pproximate</td><td>ely 205 bp</td></id=gt,>	mber=1, Type=String, Des	ription="Genot			imprecise deletion of a	pproximate	ely 205 bp
##. ##.	FORMAT= <id=gt, nur<br="">FORMAT=<id=gq, nur<="" td=""><td>mber=1,Type=String,Desember=1,Type=Integer,Des</td><td>cription="Genot scription="Genot</td><td>type</td><td>qualit</td><td>y"></td><td>pproximate</td><td>ely 205 bp</td></id=gq,></id=gt,>	mber=1,Type=String,Desember=1,Type=Integer,Des	cription="Genot scription="Genot	type	qualit	y">	pproximate	ely 205 bp
##. ##! ##!	FORMAT= <id=gt, nur<br="">FORMAT=<id=gq, nur<br="">FORMAT=<id=cn, nur<="" td=""><td>mber=1,Type=String,Des mber=1,Type=Integer,Des mber=1,Type=Integer,Des</td><td>cription="Genot scription="Geno scription="Copy</td><td>type numb</td><td>qualit er gen</td><td>y"> otype for imprecise events"></td><td>pproximate</td><td>ely 205 bp</td></id=cn,></id=gq,></id=gt,>	mber=1,Type=String,Des mber=1,Type=Integer,Des mber=1,Type=Integer,Des	cription="Genot scription="Geno scription="Copy	type numb	qualit er gen	y"> otype for imprecise events">	pproximate	ely 205 bp
** ** ** **	FORMAT= <id=gt, num<br="">FORMAT=<id=gq, num<br="">FORMAT=<id=cn, num<br="">FORMAT=<id=cnq, num<="" td=""><td>mber=1,Type=String,Des mber=1,Type=Integer,Des mber=1,Type=Integer,Des</td><td>cription="Genot scription="Geno scription="Copy cription="Copy</td><td>number</td><td>qualit er gen er geno</td><td>y"> otype for imprecise events"> type quality for imprecise events"></td><td></td><td></td></id=cnq,></id=cn,></id=gq,></id=gt,>	mber=1,Type=String,Des mber=1,Type=Integer,Des mber=1,Type=Integer,Des	cription="Genot scription="Geno scription="Copy cription="Copy	number	qualit er gen er geno	y"> otype for imprecise events"> type quality for imprecise events">		
** ** ** **	FORMAT= <id=gt,num FORMAT=<id=gq,num FORMAT=<id=cn,num FORMAT=<id=cnq,num HROM POS ID</id=cnq,num </id=cn,num </id=gq,num </id=gt,num 	mber=1,Type=String,Des mber=1,Type=Integer,Des mber=1,Type=Integer,Des umber=1,Type=Float,Des	cription="Genot scription="Genot scription="Copy cription="Copy ALT	number	qualit er gen	y"> otype for imprecise events"> type quality for imprecise events">	FORMAT GT:GO	NA00001 1/1:14
** ** ** **	FORMAT= <id=gt,num FORMAT=<id=gq,num FORMAT=<id=cn,num FORMAT=<id=cnq,num HROM POS ID</id=cnq,num </id=cn,num </id=gq,num </id=gt,num 	mber=1, Type=String, Desc mber=1, Type=Integer, Desc mber=1, Type=Integer, Desc umber=1, Type=Float, Desc REF	cription="Genot scription="Genot scription="Copy cription="Copy ALT	number	qualit per gen r geno FILTER	y"> otype for imprecise events"> type quality for imprecise events"> INFO	FORMAT GT : GQ	NA00001
##, ##] ##] ##] #C] 1	FORMAT= <id=gt,nur FORMAT=<id=gq,nur FORMAT=<id=cn,nur FORMAT=<id=cnq,nu HROM POS ID 2827694 rs2</id=cnq,nu </id=cn,nur </id=gq,nur </id=gt,nur 	mber=1, Type=String, Desc mber=1, Type=Integer, Desc mber=1, Type=Integer, Desc umber=1, Type=Float, Desc REF 376870 CGTGGATGCGGGGGAC	cription="Genot scription="Genot scription="Copy cription="Copy ALT C	numbe QUAL	qualit per gen r geno FILTER PASS	y"> otype for imprecise events"> type quality for imprecise events"> INFO SVTYPE=DEL:END=2827708:HOMLEN=1:HOMSEO=G:SVLEN=-14	FORMAT GT:GQ GT:GQ	NA00001 1/1:14
##. ##! ##! #C! 1	FORMAT= <id=gt,nur FORMAT=<id=gq,nur FORMAT=<id=cn,nur FORMAT=<id=cnq,nu HROM POS ID 2827694 rs2: 321682 .</id=cnq,nu </id=cn,nur </id=gq,nur </id=gt,nur 	mber=1, Type=String, Desc mber=1, Type=Integer, Desc mber=1, Type=Integer, Desc umber=1, Type=Float, Desc REF 376870 CGTGGATGCGGGGGAC T	cription="Genot scription="Genot scription="Copy ALT C 	numbe QUAL	qualit per geno FILTER PASS PASS	y"> otype for imprecise events"> type quality for imprecise events"> INFO <u>SVTYPE=DEL:END=2827708:HOMLEN=1:HOMSEO=G:SVLEN=-14</u> SVTYPE=DEL;END=321887;SVLEN=-205;CIPOS=-56,20;CIEND=-10,62	FORMAT GT:GQ GT:GQ	NA000001 1/1:14 0/1:12
##, ##) ##) #C) 1 2 2	FORMAT= <id=gt,nur FORMAT=<id=gq,nur FORMAT=<id=cn,nur FORMAT=<id=cnq,nu FORMAT=<id=cnq,nu HROM POS ID 2827694 rs2: 321682 . 14477084 .</id=cnq,nu </id=cnq,nu </id=cn,nur </id=gq,nur </id=gt,nur 	mber=1,Type=String,Desc mber=1,Type=Integer,Desc mber=1,Type=Integer,Desc mber=1,Type=Float,Desc REF 376870 CGTGGATGCGGGGGAC T C	cription="Genot scription="Genot scription="Copy ALT C 	otype numbe QUAL 6 12 23	qualit per geno FILTER PASS PASS PASS	y"> otype for imprecise events"> type quality for imprecise events"> INFO SVTYPE=DEL:END=2827708:HOMLEN=1:HOMSEO=G:SVLEN=-14 SVTYPE=DEL;END=321887;SVLEN=-205;CIPOS=-56,20;CIEND=-10,62 SVTYPE=DEL;END=14477381;SVLEN=-297;CIPOS=-22,18;CIEND=-12,32	FORMAT GT:GQ GT:GQ GT:GQ	NA00001 1/1:14 0/1:12 0/1:12 1/1:15

****	ileformat=V	CEnt 2							
	ileDate=2010								
			ilot-NCBI36						
				000genomes/ft	/rel	ease/sv	/breakpoint_assemblies.fasta		
							mbled 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">		
			=1, Type=String, De						
			n="Deletion">						
			cription="Deletio	n of ALU eleme	ent">				
			ription="Deletion						
			n="Duplication">						
			cription="Tandem	Duplication">					
##A	LT= <id=ins,< td=""><td>Descriptio</td><td>n="Insertion of n</td><td>ovel sequence</td><td>'></td><td></td><td></td><td></td><td></td></id=ins,<>	Descriptio	n="Insertion of n	ovel sequence	'>				
##A	LT= <id=ins:< td=""><td>ME: ALU, Des</td><td>cription="Inserti</td><td>on of ALU eler</td><td>nent"</td><td>></td><td></td><td></td><td></td></id=ins:<>	ME: ALU, Des	cription="Inserti	on of ALU eler	nent"	>			
##A	LT= <id=ins:< td=""><td>ME:L1,Desc</td><td>ription="Insertio</td><td>n of L1 elemen</td><td>it"></td><td></td><td></td><td></td><td></td></id=ins:<>	ME:L1,Desc	ription="Insertio	n of L1 elemen	it">				
##A	LT= <id=inv,< th=""><th>Descriptio</th><th>n="Inversion"></th><th></th><th></th><th></th><th></th><th></th><th></th></id=inv,<>	Descriptio	n="Inversion">						
##A	LT= <id=cnv,< th=""><th>Descriptio</th><th>n="Copy number va</th><th>riable region</th><th>'></th><th></th><th></th><th></th><th></th></id=cnv,<>	Descriptio	n="Copy number va	riable region	'>				
##F	ORMAT= <id=g< td=""><td>r,Number=1</td><td>,Type=String,Desc</td><td>ription="Genot</td><td>ype"</td><td>></td><td>imprecise deletion of a</td><td>n ALU elem</td><td>ent</td></id=g<>	r,Number=1	,Type=String,Desc	ription="Genot	ype"	>	imprecise deletion of a	n ALU elem	ent
##F	ORMAT= <id=g< td=""><td>Q,Number=1</td><td>,Type=Integer,Des</td><td>cription="Gene</td><td>type</td><td>quality</td><td>y"></td><td></td><td></td></id=g<>	Q,Number=1	,Type=Integer,Des	cription="Gene	type	quality	y">		
##F	ORMAT= <id=c< td=""><td>N, Number=1</td><td>,Type=Integer,Des</td><td>cription="Copy</td><td>num</td><td>ber gen</td><td>otype for imprecise events"></td><td></td><td></td></id=c<>	N, Number=1	,Type=Integer,Des	cription="Copy	num	ber gen	otype for imprecise events">		
##F	ORMAT= <id=c< td=""><td>NQ,Number=</td><td>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=c<>	NQ,Number=	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	rs2376870	CGTGGATGCGGGGAC	C		PASS	SVTYPE=DEL; END=2827708; HOMLEN=1; HOMSEQ=G; SVLEN=-14	GT: GQ	1/1:14
2	321682		T		6	PASS	SVTYPE=DEL; END=321887; SVLEN=-205; CIPOS=-56, 20; CIEND=-10, 62	GT: GQ	0/1:12
2	14477084		C	<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		C	<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	1	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=\								
	fileDate=201								
	reference=10								
							/breakpoint_assemblies.fasta		
		1. C. C. M. C. C. M. C. C. M.					mbled 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">		
			=1, Type=String, Dea	scription="Typ	be of	struct	ural variant">		
		· · · · · · · · · · · · · · · · · · ·	n="Deletion">						
			cription="Deletion						
			ription="Deletion	of L1 element	:">				
			n="Duplication">						
			cription="Tandem						
			n="Insertion of n						
			cription="Insertion			>			
			ription="Insertion	n of L1 elemen	it">				
	Caller and the second		n="Inversion">						
			n="Copy number va:	· · · · · · · · · · · · · · · · · · ·					
#	FORMAT= <id=0< td=""><td>T,Number=1</td><td>Type=String, Desc</td><td>ription="Genot</td><td>ype"</td><td>></td><td>imprecise insertion of a</td><td>in L1 eleme</td><td>ent</td></id=0<>	T,Number=1	Type=String, Desc	ription="Genot	ype"	>	imprecise insertion of a	in L1 eleme	ent
			Type=Integer, Des						
							otype for imprecise events">		
#	FORMAT= <id=0< td=""><td>NQ,Number=</td><td>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=0<>	NQ,Number=	1, Type=Float, Desc	ription="Copy	numb	er geno	type quality for imprecise events">		
#	CHROM POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	NA00001
1	2827694	rs2376870	CGTGGATGCGGGGGAC	C		PASS	SVTYPE=DEL; END=2827708; HOMLEN=1; HOMSEQ=G; SVLEN=-14	GT: GQ	1/1:14
2	321682	1.	T		6	PASS	SVTYPE=DEL; END=321887; SVLEN=-205; CIPOS=-56, 20; CIEND= 10, 62	GT: GQ	0/1:12
2	14477084		C	<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		C	<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		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

##file	format=VCFv4.3							
##file	Date=20100501							
##refe	rence=1000Genomes	ilot-NCBI36						
##asse	mblv=ftp://ftp-tra	ce.ncbi.nih.gov/1	000genomes/ft	/rel	ease/sv	/breakpoint_assemblies.fasta		
						mbled 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">		
##INFO	= <id=svtype, number<="" td=""><td>=1, Type=String, De</td><td>scription="Typ</td><td>e of</td><td>struct</td><td>ural variant"></td><td></td><td></td></id=svtype,>	=1, Type=String, De	scription="Typ	e of	struct	ural variant">		
	<id=del,descriptio< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td></id=del,descriptio<>							
##ALT=	<id=del:me:alu,des< td=""><td>cription="Deletio</td><td>n of ALU eleme</td><td>ent"></td><td></td><td></td><td></td><td></td></id=del:me:alu,des<>	cription="Deletio	n of ALU eleme	ent">				
##ALT=	<id=del:me:l1,desc< td=""><td>ription="Deletion</td><td>of L1 element</td><td><"></td><td></td><td></td><td></td><td></td></id=del:me:l1,desc<>	ription="Deletion	of L1 element	<">				
##ALT=	<id=dup,descriptio< td=""><td>on="Duplication"></td><td></td><td></td><td></td><td></td><td></td><td></td></id=dup,descriptio<>	on="Duplication">						
##ALT=	<id=dup:tandem, des<="" td=""><td>scription="Tandem</td><td>Duplication"></td><td></td><td></td><td></td><td></td><td></td></id=dup:tandem,>	scription="Tandem	Duplication">					
##ALT=	<id=ins,descriptio< td=""><td>n="Insertion of n</td><td>ovel sequence</td><td>'></td><td></td><td></td><td></td><td></td></id=ins,descriptio<>	n="Insertion of n	ovel sequence	'>				
##ALT=	<id=ins:me:alu, des<="" td=""><td>cription="Inserti</td><td>on of ALU eler</td><td>nent"</td><td>></td><td></td><td></td><td></td></id=ins:me:alu,>	cription="Inserti	on of ALU eler	nent"	>			
##ALT=	<id=ins:me:l1,desc< td=""><td>ription="Insertio</td><td>n of L1 element</td><td>it"></td><td></td><td></td><td></td><td></td></id=ins:me:l1,desc<>	ription="Insertio	n of L1 element	it">				
##ALT=	<id=inv, descriptio<="" td=""><td>n="Inversion"></td><td></td><td></td><td></td><td></td><td></td><td></td></id=inv,>	n="Inversion">						
##ALT=	<id=cnv,descriptio< td=""><td>n="Copy number va</td><td>riable region</td><td>'></td><td></td><td></td><td></td><td></td></id=cnv,descriptio<>	n="Copy number va	riable region	'>				
##FORM	AT= <id=gt, number="1</td"><td>,Type=String,Desc</td><td>ription="Genot</td><td>type"</td><td>></td><td>imprecise duplication of</td><td>of approxim</td><td>ately 21Kb</td></id=gt,>	,Type=String,Desc	ription="Genot	type"	>	imprecise duplication of	of approxim	ately 21Kb
	AT= <id=gq,number=1< td=""><td>· · · ·</td><td>A</td><td></td><td></td><td></td><td></td><td>-</td></id=gq,number=1<>	· · · ·	A					-
##FORM	AT= <id=cn, number="1</td"><td>,Type=Integer,Des</td><td>cription="Copy</td><td>num</td><td>ber gen</td><td>otype for imprecise events"></td><td></td><td></td></id=cn,>	,Type=Integer,Des	cription="Copy	num	ber gen	otype for imprecise events">		
##FORM	AT= <id=cnq,number=< td=""><td>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,number=<>	1, Type=Float, Desc	ription="Copy	numb	er geno	type quality for imprecise events">		
#CHROM		REF	ALT	QUAL	FILTER		FORMAT	NA00001
1	2827694 rs2376870	CGTGGATGCGGGGGAC	C		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 .	C	<del:me:alu></del:me:alu>		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>	10.0 × 10.	PASS	SVTYPE=TNS; END=9425916; SVLEN=6027; CTPOS=-16, 22	GT:GQ	1/1:15
	12665100 .	A	<dup></dup>	14	PASS	SVTYPE=DUP; END=12686200; SVLEN=21100; CIPOS=-500, 500; CIEND=-500, 500	GT: GQ CN CNQ	
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

	ileformat=V0								
	eference=100								
							/breakpoint_assemblies.fasta		
							mbled 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">		
##]	NFO= <id=svl< td=""><td>EN,Number=</td><td>., Type=Integer, De</td><td>scription="Di</td><td>fere</td><td>nce in</td><td>length between REF and ALT alleles"></td><td></td><td></td></id=svl<>	EN,Number=	., Type=Integer, De	scription="Di	fere	nce in	length between REF and ALT alleles">		
##]	NFO= <id=svt< td=""><td>PE, Number</td><td>=1, Type=String, De</td><td>scription="Typ</td><td>e of</td><td>struct</td><td>ural variant"></td><td></td><td></td></id=svt<>	PE, Number	=1, Type=String, De	scription="Typ	e of	struct	ural variant">		
##/	LT= <id=del,i< td=""><td>escriptio</td><td>n="Deletion"></td><td></td><td></td><td></td><td></td><td></td><td></td></id=del,i<>	escriptio	n="Deletion">						
##/	LT= <id=del:< td=""><td>E:ALU,Des</td><td>cription="Deletio</td><td>n of ALU eleme</td><td>ent"></td><td></td><td></td><td></td><td></td></id=del:<>	E:ALU,Des	cription="Deletio	n of ALU eleme	ent">				
##/	LT= <id=del:< td=""><td>E:L1,Desc</td><td>ription="Deletion</td><td>of L1 element</td><td>:"></td><td></td><td></td><td></td><td></td></id=del:<>	E:L1,Desc	ription="Deletion	of L1 element	:">				
##/	LT= <id=dup,i< td=""><td>escriptio</td><td>n="Duplication"></td><td></td><td></td><td></td><td></td><td></td><td></td></id=dup,i<>	escriptio	n="Duplication">						
##/	LT= <id=dup:< td=""><td>ANDEM, Des</td><td>cription="Tandem</td><td>Duplication"></td><td></td><td></td><td></td><td></td><td></td></id=dup:<>	ANDEM, Des	cription="Tandem	Duplication">					
##/	LT= <id=ins,i< td=""><td>escriptio</td><td>n="Insertion of n</td><td>ovel sequence</td><td>></td><td></td><td></td><td></td><td></td></id=ins,i<>	escriptio	n="Insertion of n	ovel sequence	>				
##/	LT= <id=ins:< td=""><td>E:ALU,Des</td><td>cription="Inserti</td><td>on of ALU eler</td><td>nent"</td><td>></td><td></td><td></td><td></td></id=ins:<>	E:ALU,Des	cription="Inserti	on of ALU eler	nent"	>			
		and the second	ription="Insertio						
		12 C	n="Inversion">						
			n="Copy number va	riable region	>				
			,Type=String,Desc			>	imprecise tandem dup	lication of 7	'6bp
			.Type=Integer,Des	and the second	*** C				
							otype for imprecise events">		
							type quality for imprecise events">		
	ROM POS	ID	REF	ALT		FILTER		FORMAT	NA00001
1	and a second		CGTGGATGCGGGGGAC	C	quint	PASS	SVTYPE=DEL; END=2827708; HOMLEN=1; HOMSEQ=G; SVLEN=-14		1/1:14
2	321682		T		6	PASS	SVTYPE=DEL; END=321887; SVLEN=-205; CIPOS=-56, 20; CIEND=-10, 62		0/1:12
2	14477084		c	<del: alu="" me:=""></del:>		PASS	SVTYPE=DEL; END=14477381; SVLEN=-297; CIPOS=-22, 18; CIEND=-12, 32		0/1:12
3	9425916	-	c	<ins:me:l1></ins:me:l1>		PASS	SVTYPE=INS: END=9425916: SVLEN=6027: CIPOS=-16.22	GT:GQ	1/1:15
3	12665100				14	PASS	SVTYPE=DUP:END=12686200:SVLEN=21100:CIP0S=-500.500:CIEND=-500.500	GT : GQ : CN : CNQ	Contraction of the second second second
4	18665128		T	<dup: tandem=""></dup:>	-	PASS	SVTPE=DUP;END=126565204;SVLEN=21100;CIPUS=-300;S00;CIPAD=-300;S00 SVTYPE=DUP;END=18665204;SVLEN=76;CIPUS=-10,10;CIEND=-10,10	GT: GQ CN CNQ	
	10000120			SPOL T KUDEUS		TRUD	011112-D01,280-10000201,0128-10,01100-10,10,01280-10,10	ar ad on ond	.,

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

INFO
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

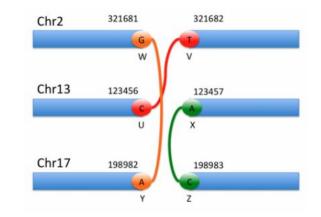
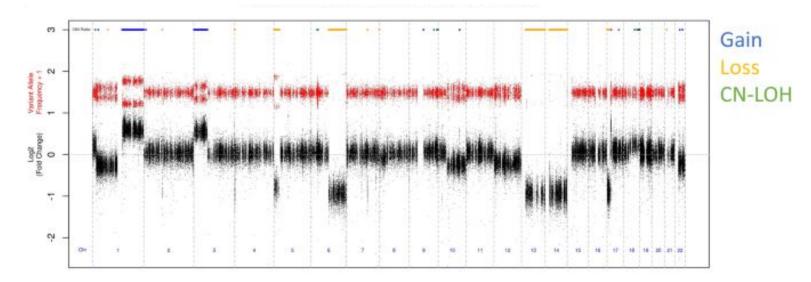


Figure 1: All possible orientations of breakends

How to visualize Structural Variations ?

Visualization of structural variants

When alteration is quantitative



Position

Log2 ratios distribution along chromosomes # numbers of copy

Visualization of structural variants

1 ŝ 2 Sol ≤ 1 16 15 A 6 10 8 9

When it gets more complex



https://training.galaxyproject.org/training-material/topics/visualisation/tutorials/circos/tutorial.html