Table S1 Quality assessment and read processing

Name	OS	Input	Output	Supported platforms	Report	Tag (1) removal	Filtering	Trimming
ContEST [1]	Lin, Mac, Win	BAM, VCF, FASTA (ref)	TXT	Illumina, ABI SOLiD, 454	no	no	no	no
FastQC [2]	Lin, Mac, Win	(CS) FASTQ, SAM, BAM	HTML	Illumina, ABI SOLiD	yes	no	no	no
FASTX-Toolkit [3]	Lin, Mac, web interface	FASTA, FASTQ	FASTA, FASTQ	Illumina	yes	yes	yes	yes
Galaxy [4]	Lin, Mac, web interface, Cloud instance	FASTQ	FASTQ	Illumina	yes	yes	yes	yes
htSeqTools [5]	Lin, Mac, Win	FASTQ	Graphs	Illumina	yes	no	no	no
NGSQC [6]	Lin	FASTA (ref), FASTQ, CSFASTA, QUAL FASTA	HTML	Illumina, ABI SOLiD	yes	no	no	no
PIQA [7]	Lin, Mac, Win	FASTQ, bustard, output, SCARF	HTML, TXT	Illumina	yes	no	no	no
PRINSEQ [8]	Lin, Mac, Win, web interface	FASTA, FASTQ, QUAL FASTA	FASTA, FASTQ, QUAL FASTA, HTML	Illumina, 454	yes	no	yes	yes
SolexaQA [9]	Lin, Mac	FASTQ	FASTQ, PNG	Illumina, 454	yes	no	no	yes
TagCleaner [10]	Lin, Mac, web interface	FASTA, FASTQ	FASTA	454	no	yes	no	no
TileQC [11]	Lin, Mac	Eland output	Graphs	Illumina	yes	no	no	no

(1) Artifacts such as multiplex identifiers, adapters, primers, and linkers
OS..... Operating system: Lin ... Linux, Mac ... Mac OS X, Win ... Windows
ref..... reference genome

BAM ..... Binary SAM
Bustard ..... Illumina's base calling software
CSFASTQ..... color space (SOLiD encoded) FASTQ file
CSFASTA ..... color space (SOLiD encoded) FASTA
Eland..... Illumina's alignment algorithm

FASTA ..... text-based format for representing nucleotide sequences
FASTQ..... text-based format for repesenting nucleotide sequences and their corresponding quality scores
QUAL FASTA .. text-based format for representing quality scores
SAM ..... Sequence Alignment/Map
SCARF..... output of Illumina's Gerald software (single colon separated file with one record per line containing read name, sequence, and quality)
VCF..... Variant Call Format

Table S2 Alignment software

Name	OS	Input	Output	Supported platforms	Indexing method	Gapped alignment
BarraCUDA [12]	Lin	FASTQ	SAM	Illumina	FM index (BWT)	yes
BFAST [13]	Lin	FASTQ	SAM	Illumina, ABI SOLiD, 454	Multiple (hash, tree, …)	yes
Bowtie [14]	Lin, Mac, Win	FASTQ, FASTA	SAM	Illumina, ABI SOLiD	FM index (BWT)	no
Bowtie2 [15]	Lin, Mac, Win	FASTQ, FASTA, QSEQ	SAM	Illumina, 454	FM index (BWT)	yes
BWA [16]	Lin	(CS)FASTQ, FASTA	SAM	Illumina, ABI SOLiD(1)	FM index (BWT)	yes
BWA-SW [17]	Lin	FASTQ, FASTA	SAM	454	FM index (BWT)	yes
ELAND [18]	Lin	FASTQ, FASTA	SAM	Illumina	-	no
MAQ [19]	Lin	FASTQ, FASTA	Maq	Illumina	Hash based	yes
Mosaik [20]	Lin, Mac, Win	FASTQ, FASTA	SAM, BED, several others	Illumina, ABI SOLiD, 454	-	yes
mrFAST [21]	Lin	FASTQ, FASTA	SAM, DIVET	Illumina	Hash based	yes
mrsFAST [22]	Lin	FASTQ, FASTA	SAM, DIVET	Illumina	Hash based	no
Novoalign [23]	Lin, Mac	FASTQ, (CS)FASTA	SAM, TXT	Illumina, ABI SOLiD	-	yes
SOAP2 [24]	Lin	FASTQ, FASTA	SOAP (2)	Illumina	FM index (BWT)	yes
SOAP3 [25]	Lin	FASTQ, FASTA	SAM	Illumina	FM index (BWT)	no
SSAHA2 [26]	Lin, Mac	FASTA	SAM, GFF	Illumina, ABI SOLiD, 454	Tree index	yes
Stampy [27]	Lin, Mac (3)	FASTQ, FASTA	SAM	Illumina, 454	FM index (BWT)	-
YOABS [28]	Lin	-	-	Illumina	FM & Tree index	yes

(1) SOLiD support dropped with introduction of version 1.6.0
(2) Provides script for conversion to SAM
(3) Experimental

OS..... Operating system: Lin ... Linux, Mac ... Mac OS X, Win ... Windows

BED..... Browser Extensible Data, a text-based file format
CSFASTQ..... color space (SOLiD encoded) FASTQ file
CSFASTA ..... color space (SOLiD encoded) FASTA
DIVET ..... VariationHunter's comma separated input file format
FASTA ..... text-based format for representing nucleotide sequences

FASTQ ..... text-based format for repesenting nucleotide sequences and their corresponding quality scores
GFF ..... General Feature Format
Maq ..... MAQ proprietary alignment format
QSEQ ..... Illumina's base calling result file format
SAM ..... Sequence Alignment/Map
SOAP ..... SOAP proprietary alignment format

Variant identification

Table S3 Germline callers

Name	BAM/SAM input	Other inputs	VCF output	Other outputs	Illumina	Solid	SNP	INDEL	Last update	Notes
Atlas 2 [29]	yes	FASTA (ref. Genome)	yes		yes	yes	yes	yes	2011-08-29	Suite for variant analysis in WES data, which has been integrated into the Genboree Workbench; collects coverage information;
Bambino [30]	yes		no	CSV	yes	yes			2012-03-09	
Beagle [31]	no	Beagle input format	no	Beagle format	yes	y/y	yes	yes	2011-10-31	Software for imputation, phasing and association;
CoNAn-SNV [32]	no	pileup, model parameters, segmentation boundaries	no	CSV	yes	yes	yes	no	2010-06-23	
CORTEX [33]	no	fastq, FASTA	yes	FASTA-like	yes	yes	yes	yes	2011-11-03	Tool performs variant discovery by de novo assembly - no reference genome required; can also detect inversions, complex variants, and small haplotypes;
CRISP [34]	yes		yes		yes	yes	yes	yes	2012-04-24	
Dindel [35]	yes	FASTA (ref. genome)	yes		yes	yes	no	yes	2010-10-25	Tool can also test if detected variants are real INDELs or sequencing or mapping errors;
FreeBayes [36]	yes		yes		yes	yes	yes	yes	2012-05-30	Software also finds MNPs (multi-nucleotide polymorphisms), and complex events (composite insertion and substitution events) smaller than the length of a short-read sequencing alignment;
GATK (UnifiedGenotyper) [37]	yes		yes		yes	yes	yes	yes	2012-05-01	Package includes SNP and genotype caller, SNP filtering, and SNP quality recalibration;
GSNP [38]	no	SOAP alignment result	no	CSV	yes	no	yes	no	2011-11-11	GPGPU (cuda) implementation of SoapSNP;
IMPUTE2 [39]	no	genotype-file-format, recombination-map	no	impute format	yes	yes	yes	yes	2012-01-27	Software for imputation and phasing, including a mode for genotype calling; good documentation is provided but a steep learning curve;
Indelocator [40]	yes		no	CSV	yes	yes	no	yes	2012-02-22	Calling short indels; uses inputs

Name	BAM/SAM input	Other inputs	VCF output	Other outputs	Illumina	Solid	SNP	INDEL	Last update	Notes
										from normal and tumor samples; also runs with just one sample
Ion Variant Hunter [41]	yes		yes		-	-	yes	yes	2012-04-04	
MaCH [42]	no	glf	no	MaCH format	yes	yes	yes	no	-	
moDIL [43]	no	own CSV	no	CSV	yes	yes	no	yes	2012-01-20	No user guide could be found;
PolyScan [44]	no	Consed ace file	no	polyScan format	-	-	yes	yes	2007-07-24	May require some familiarization;
Qcall [45]	yes		-	-		-	yes	no	2010-09-22	No homepage, documentation, or binaries could be found
realSFS [46]	no	fastq	-	-	yes	no	yes	no	-	
SAMtools [47]	yes	FASTA	yes		yes	yes	yes	yes	2011-09-02	Suite performs computation of genotype likelihoods and SNP and genotype calling;
SliderII [48]	no	prb files	no	CSV	yes	no	yes	no	2009-06-09	
Sniper [49]	no	map file, reference genome, fastq	no	CSV	yes	no	yes	no	2011-10-04	
SNVer [50]	yes		yes		yes	yes	yes	yes	2012-04-21	Statistical framework to find rare and common variants in individual and pooled sequencing data;
SNVMix [51]	no	pileup/mpileup	no	CSV	yes	yes	yes	no	2012-03-21	
SOAPindel [52]	yes	SOAP, FASTA	no	CSV	yes	no	no	yes	2012-03-29	
SOAPsnp [24]	no	SOAP alignment result	no	CSV, GLFv2, GPFv2	yes	no	yes	y/n	2009-05-25	
Syzygy [53]	yes	pool info file + target info file	no	CSV	yes	yes	yes	no	2012-07-05	Provided is a good user guide; Needs in addition to BAM/SAM pool info and target info file;
VarScan 2	no	pileup/mpileup	yes	CSV	yes	yes	yes	yes	2012-05-01	
VARiD [54]	yes	SAM, FASTA	no	CSV	yes	yes	yes	yes	2010-10-05	
VipR [55]	no	pileup (from samtools)	yes		yes	yes	yes	yes	2012-04-12	Tool that uses data from multiple DNA pools;

Table S4 Somatic callers

Name	Input Format	Output Format	Illumina	Solid	SNP	INDEL	Last update	Notes
GATK (SomaticIndelDetector) [37]	BAM	VCF	yes	yes	no	yes	2012-05-01	No SNP caller;
MutationSeq [56]	BAM	CSV	yes	yes	yes	yes	2012-05-24	Software requires matlab;
MuTect [57]	BAM	CSV			yes	no	2012-04-10	Identifies somatic point mutations and is suggested by the Broad Institute; currently only for registered beta-tester;
SAMtools [47]	BAM	VCF	yes	yes	yes	yes	2011-09-02	Requires merging pairs into one BAM file; no separate parameters for tumor and normal BAM files;
SomaticCall [58]	BAM	CSV	yes	yes	yes	yes	2009-11-06	The tool is no longer maintained;
SomaticSniper [59]	BAM	VCF / somatic sniper output	yes	yes	yes	no	2012-06-12	
SPLINTER [60]	SCARF file	CSV	yes	no	yes	yes	2010-07-01	Registration is required; no dedicated download page is provided;
VarScan 2 [61]	pileup/mpileup	VCF / varscan CSV	yes	yes	yes	yes	2012-05-01	Can also predict cnv and be used as a germline variant caller;

Table S5 CNV identification

Name	SAM/BAM	Other input	Output	Illumina	Solid	Last Update	Notes
CNAseg [62]	yes	-	CSV	yes	yes	2010-09-14	Tool is using R; for tumor/normal pairs of cancer data;
CNVer [63]	yes	CSV	cnv files	yes	yes	2011-07-11	For Illumina data; authors recommend using the Bowtie with specific parameters; no paired-end reads supported;
cnvHMM [64]	no	cnsfile	CSV	yes	no	2009-06-04	
CNVnator [65]	yes	FASTA	CSV	yes	yes	2012-02-07	Tool for identifying, genotyping and characterizing CNVs; not very easy to install and authors don't offer a comprehensive tutorial;
CNV-seq [66]	yes	psl	cnv summary file	yes	yes	2011-07-15	Currently supports BLAT psl file and SOLiD matching pipeline output as input;
CONTRA [67]	yes	2 BAM files, BED	VCF, CSV	yes	yes	2012-07-24	For targeted resequencing data such as those from whole exome capture data; needs test and control BAM files;
CopySeq [68]	yes	-	CSV	yes	yes	2011-04-06	
ExomeCNV [69]	yes	pileup + bed + FASTA		yes	yes	2012-06-01	Most suitable when paired samples (e.g. tumor-normal pair) are available; requires R;
RDXplorer [70]	yes	FASTA	CSV	yes	yes	2012-01-13	Tool for copy number variants (CNV) detection in whole human genome sequence data using read depth (RD) coverage.
readDepth [71]	no	BED, R	segmented CNVs, CSV	yes	yes	2011-04-15	Tool is reported to perform good on low and high coverage; R package;
Segseq [72]	no	CSV	CSV	yes	no	2009-01-28	Tool requires Matlab;

Table S6 SV identification

Name	BAM/SAM	Input Format	Output Format	Illumina	Solid	Last Update	Notes
APOLLOH [73]	no	CSV (infile, cnfile)	CSV	yes	yes	2011-12-01	Tool requires Matlab; predicts somatic loss of heterozygosity and allelic imbalance in whole tumour genome sequencing data;
BreakDancer [74]	yes	BAM + config file	ctx, BED	yes	yes	2011-02-21	Tool can detect deletions, insertions, inversions, intra, and inter chromosomal translocations; computes the copy number for each BAM file;
Breakpointer [75]	yes	BAM	GFF	yes	yes	2012-01-20	The tool can also call INDELs;
BreakSeq [76]	no	GFF	GFF, CSV, various ouput files			2010-07-26	
Breakway [77]	yes	BAM	CSV	yes	yes	2011-04-01	Can also annotate the identified breakpoints;
CLEVER [78]	yes	BAM, FASTA	own CSV	yes	-	2012-03-27	Authors provide structured documentation;
ClipCrop [79]	yes	SAM, FASA	BED	yes	yes	2012-01-27	This is a tool for detecting structural variations using soft-clipping information.
CREST [80]	no	DIVET		yes	no	2011-11-08	The user can use his/her own programs in place of BLAT and CAP3, but needs to implement them by himself/herself;
FusionMap [81]	no	fastq	SAM, own format	yes	yes	2012-04-17	Aligns reads spanning fusion junctions directly to the genome without prior knowledge of potential fusion regions;
GASVPro [82]	yes	SAM/BAM	clusters file	yes	yes	2012-06-14	Software to detect SVs from paired-end mapping data;
Hydra-sv [83]	no	Hydra CSV file	BEDPE	-	-	2010-08-20	
PEMer [84]	no	Several input files	multiple output files	yes	yes	2009-02-02	Megablast and Smith Waterman programs are required; tool requires several configuration steps;
Pindel [85]	yes	BAM + FASTA	CSV	yes	??	2011-08-29	Tool that detects breakpoints of large deletions, medium sized insertions, inversions, tandem duplications and other SVs;
SPLITREAD [86]	no	FASTA, mrFast SAM	BED	yes	no	2011-10-13	Identifies INDELs and SVs;
SVDetect [87]	yes	BAM/SAM or ELAND or Bioscope output	Txt, BED, Circos link	yes	yes	2011-07-12	Detects large deletions and insertions, inversions, intra- and inter-chromosomal rearrangements;
SVMerge [88]	yes	BAM + FASTA(ref)	bed	yes	-	2012-02-16	SVMerge integrates calls from several existing SV callers: BreakDancerMax, Pindel, RDXplorer, cnD and SECluster;
Tigra [89]	yes	BAM	CSV	yes	yes	-	Tool is not available anymore;
VariationHunter [90]	no	own CSV input file	txt for insertionas, deletions and inversions	yes	yes	2010-01-12	

Table S7 Varaint annotation

Name	Input Format	Output Format	SNP	INDEL	CNV	GUI	CLI	Web	Notes
ABSOLUTE [91]	HAPSEG output, sample level variance, precomputed models of cancer types, sigma values	Plot showing the Purity/Ploidy, R data file	yes	no	yes	no	yes	no	Comes bundles with HAPSEG;
Align-GVGD [92]	FASTA, substitutions list	Web report	yes	no	no	no	no	yes	Estimates SNP risk;
ANNOVAR [93]	VCF4, Complete Genomics, GFF3-SOLiD, CSV in Annovar format;	Gene-based annotation; Region-based annotations; Filter-based annotation. For all categories	yes	yes	yes	no	yes	no	Integrated tool providing gene annotation, db ids and various scores;
AnnTools [94]	VCF, pileup, CSV	VCF	yes	yes	yes	no	yes	no	Provides a set of helper tools for custom annotation;
Auto-mute [95]	PDB ID, Chain, Mutation	Web report	yes	no	no	no	no	yes	The tool performs stability and disease potential predictions.
CandiSNPer [96]	dbSNP ID, population	Web report	yes	no	no	no	no	yes	
CHASM and SNVBox [97]	Passenger mutation rates, AA changes	CSV including CHASM score, p-value, and FDR	yes	no	no	no	yes	no	Predicts the functional significance of somatic missense mutations observed in the genomes of cancer cells and features prioritization of mutations;
CUPSAT [98]	PDB ID; PDB file format	Web report	yes	no	no	no	no	yes	Performs protein stability prediction;
dbNSFP [99]			yes	no	no	no	yes	no	Integrated SNP database; provides a simple JAVA CLI tool for searching;
VEP (Ensembl - Variant Effect Predictor) [100]	CSV, VCF, Pileup, HGVS, Variant Identifiers	Web report	yes	no	no	no	yes	yes	
ESEfinder [101]	FASTA	Web report, CSV	-	-	-	no	no	yes	Analyzes sequences for the presence of ESE motifs;
ESRSearch [102]	plain sequence; FASTA	Web report	-	-	-	no	no	yes	Finds ESR sequences;
FANS [103]	FASTA format; or variation information via web interface	Web report, CSV	yes	no	no	no	no	yes	Prioritized variations based on risk levels; divided into: Genome View, Gene View, Transcript View, Variation View;
FastSNP [104]	Gene Symbol, dbSNP ID	Web report	yes	no	no	no	no	yes	Outputs prioritized list of SNPs with risk assessment;
FESD [105]	Gene name	Web report	yes	no	no	no	no	yes	Output includes regions: promoter, CpG, islands, translation start, splice site, translation stop, poly(A) signal, transcript;

Name	Input Format	Output Format	SNP	INDEL	CNV	GUI	CLI	Web	Notes
FOLD-X [106]			yes	no	no	no	yes	yes	It performs protein stability analysis.
F-SNP [107]	SNP ID; disease; gene; chromosomal region		yes	no	no				The software integrates information obtained from 16 bioinformatics tools and databases about the functional effects of SNPs.
GERP++ [108]		Web report	yes	no	no	no	yes	yes	It produces evolutionary conservation scores.
GSITIC [109]	Segmentation File, Markers File, FASTA, (Array List File, CNV File)	Lesions, Amplification Genes, Deletion Genes, Gistic Scores, Plots	no	no	yes	no	yes	no	Identifies regions of the genome that are significantly amplified or deleted across a set of samples;
HOPE [110]	FASTA, accession code for protein	Web report on structural differences between wild type and mutations	yes	no	no	no	no	yes	The web-based tool offers a simple web interface for entering protein sequence and amino acid mutation.
Human Splicing Finder (HSF) [111]	Ensembl / RefSeq ID, plain text sequences		yes	no	no	no	no	yes	
I-Mutant2.0 [112]	One letter residue code, sequence residue number	Web report	yes	no	no	no	yes	yes	The tool is based on support vector machines.
LS-SNP [113]	SwissProt ID, dbSNP ID, Kegg Pathway ID, HUGO Gene ID	Web report	yes	no	no	no	no	yes	The tool offers prediction of disease association and confidence of prediction and is based on support vector machines (SVM).
MAPP [114]	FASTA	CSV in MAPP format	yes	no	no	no	yes	no	
MuD [115]		Web report	yes	no	no	no	no	yes	
MutaGeneSys [116]		Web report / CSV	yes	no	no	no	yes	yes	The query Interface is not working.
MutationAssessor [117]	CSV in MutationAssessor format, Uniprot ID, Refseq ID	CSV in MutationAssessor format	yes	no	no	no	no	yes	
MutationTaster [118]	ORF, cDNA sequence, genomic sequence, alteration	Web report	yes	yes	no	no	yes	yes	
MutPred [119]	FASTA sequence, CSV file of mutations	Web report	yes	no	no	no	no	yes	Calculates the impact of mutation on different protein properties; is based on SIFT and offers precomputed dbSNP results;
MutSig [120]	List of mutations, regions to investigate	CSV	yes	yes	no	no	yes	no	Still in beta testing – available upon request
NGS-SNP [121]	VCF, pileup, CSV	VCF	yes	no	no	no	yes	no	
nsSNPAnalyzer [122]	FASTA, substitutions list	Web report	yes	no	no	no	no	yes	The tool outputs various SNP features and predicts the phenotypic class.
Oncotator [123]	Oncotator format	CSV	yes	yes	no	no	no	yes	Annotations with data relevant to cancer researcher; collects Genomic Annotations, Protein Annotations, Cancer Annotations

Name	Input Format	Output Format	SNP	INDEL	CNV	GUI	CLI	Web	Notes
PANTHER [124]	Protein sequence and Substitution	subPSEC score	yes	no	no	no	yes	yes	Uses subPSEC score;
Parepro [125]	Protein sequence and Substitution	-	yes	no	no	no	yes	no	It is based on support vector machines (SVM).
PESX [126]	plain sequence; FASTA	Web report	-	-	-	no	no	yes	Finds ESE sequences;
pfSNP [127]	SNP ID; chromosome region; Gene ID;	Web report	yes	no	no	no	no	yes	
PHAST [128]	FASTA, PHYLIP, MPM, MAF, SS	Conservation score	-	-	-	no	yes	no	Phylogenetic analysis toolbox, including phastCons and phyloP;
PhD-SNP [129]	One letter residue code, Swiss-Prot protein code, Sequence file	Effect preditction	yes	no	no	no	yes	no	
PMUT [130]	FASTA sequence/file SWISSProt code	Web report	yes	no	no	no	no	yes	Offers different prediction modes and is able to output detailed mutation analysis reports;
PolyDoms [131]	Gene/protein symbol(s), RefSeqID dbSNP ID	Web report	yes	no	no	no	no	yes	
PolyMAPr [132]	-	-	-	-	-	no	yes	no	No longer available;
PolyPhen-2 [133]	UniProt ID, FASTA, dbSNP ID	CSV in PolyPhen format	yes	no	no	no	yes	yes	
PupaSNP Finder [134]	dbSNP ID, Gene/Transcript ID; PED format	Web report	yes	no	no	no	no	yes	
QuickSNP [135]	genomic position; HUGO gene symbol	Web report	yes	no	no	no	no	yes	
RescueESE [136]	plain text; multi-FASTA	predicts sequences with ESE activity	-	-	-	no	no	yes	
SAPRED [137]	FASTA and mutation file		yes	no	no				The website is offline.
SCAN [138]		Web report	yes	no	yes	no	no	yes	
SCONE [139]	MAF	Conservation score	-	-	-	no	yes	no	
SeattleSeq Annotation [140]	Maq, GFFm CASAVA, VCF, GATK bed	VCF, own format	yes	yes	no	no	no	yes	
SeqAnt [141]	FASTA sequence file	Web report	yes	yes	no	no	no	yes	
SeqProfCod [142]	-	-	yes	no	no	-	-	-	Not available online;
SVA (Sequence Variant Analyser) [143]	VCF of variants, project file (for command line version)	--potential biological function -- dbSNP/Kegg/GO/1000Genomes/DGV annotation --identifies protein-truncating variants -- filtering by function	yes	yes	no	yes	yes	no	

Name	Input Format	Output Format	SNP	INDEL	CNV	GUI	CLI	Web	Notes
SIFT [144]	Multiple proteins, dbSNP ID, NCBI GI number, protein sequence, protein sequence alignment, Pileup, VCF4, maq, soap, gff3, casava, cg	XXX in SIFT format	yes	no	no	no	yes	yes	
SIFT Indel [145]			no	yes	no	no	no	yes	
SiPhy [146]	FASTA, MAF, PHYLIP		-	-	-	no	yes	no	
SNAP [147]	AA in FASTA, substitutions format	Web report	yes	no	no	no	no	yes	This tool offers a user friendly web interface.
SNP Function Portal [148]	RefSNP Ids, OMIM Ids	Web report	yes	no	no	no	no	yes	
SNP@Domain [149]			yes	no	no	-	-	-	Not available anymore;
SNPdbe [150]	Gene/protein symbol, FASTA	Web report	yes	no	no	no	no	yes	The protein function is predicted using SNAP and SIFT and entries are augmented with experimental information from public databases.
SNPeffect 4.0 [151]	FASTA, PDB file, PDB ID, UniProt ID	Web report	yes	no	no	no	no	yes	This tool mainly uses protein structure information.
SNPHunter [152]	Gene symbol; dbSNP ID;	Web report	yes	no	no	yes	no	no	
SNPnexus [153]	CSV in SNPnexus input format	CSV in SNPnexus output format	yes	yes	yes	no	no	yes	Outputs CNV, INDELs, inversions;
SNPper [154]	dbSNP ID, TSC ID, position	Web report	yes	no	no	no	no	yes	
SNPs&GO [155]	One letter residue code; Swiss-Prot protein code; Sequence file; GO terms; CSV	Web report	yes	no	no	no	no	yes	Predicts neutral/deleterious; calculates reliability index and disease probability;
SNPs3D [156]	Gene symbol, SNP ID	Web report	yes	no	no	no	no	yes	
SNPseek [157]	-	-	-	-	-	-	-	-	Tool that performs neural network based protein stability prediction which is not available anymore;
SNPselector [158]	-	-	-	-	-	no	no	yes	No longer available;
SnpSIFT + snpEff [159]	VCF, SNPs, insertions, deletions, and MNPs	CSV	yes	yes	no	no	yes	no	A collection of tools to manipulate VCF files;
SPOT [160]	SNPs and p-values,	Web report	yes	no	no	no	no	yes	Outputs various DB ids and scores;
StSNP [161]	protein sequence; protein name; dbSNP ID; gene symbol	Web report	yes	no	no	no	no	yes	
TAMAL [162]	-	-	-	-	-	-	-	-	No longer available;
TopoSNP [163]	Protein ID, protein sequence	Web report	yes	no	no	no	no	yes	Predicts whether substitution is on surface of the protein structure;

Name	Input Format	Output Format	SNP	INDEL	CNV	GUI	CLI	Web	Notes
									conservation score based on Pfam protein alignments;
VARIANT [164]	VCF	Web report, text	yes	no	no	no	yes	yes	

CSV = comma separated value

Variant visualization

Table S8 Genome Browsers

Name	BAM/SAM	VCF	Other formats	GUI	Web	Annotation	Notes
ABrowse [165]	yes	no	GFF, WIG	no	yes (local)	yes	Shows tracks as large images similar to google maps;
AnnoJ [166]	no	no	own format	no	yes (local)	yes	
Apollo [167]	no	no	DAS, GFF, GFF3, WIG	yes	no	yes	
Argo / Combo [168]	no	no	FASTA, Genbank, GFF, BLAST, BED, Wiggle (WIG), Genscan files	yes	no	yes	Argo is a standalone genome viewer which integrates combo as a comparative genome browser.
Artemis [169]	yes	yes	BCF, FASTA	yes	no	yes	Standalone tool where BAMView has been integrated;
Bambino [30]	yes	no	FASTA, UCSC, 2bit, nib	yes(JWS)	no	yes	
BamView [170]	yes	no	-	yes	no	no	This tool has been integrated into Artemis and is capable of calculating read counts and RPKM values.
Consed [171]	no	no	Newbler, Cross_match, Phrap, MIRA, Velvet and PCAP	yes	no	no	The standalone tool has been designed to display genome assemblies.
DiProGB [172]	no	no	GenBank, FASTA, GFF PTT	yes	no	yes	Is able to display sequence graphs and a feature graphs;
EagleView [173]	no	no	ACE, READS, EGL, MAP	yes	no	no	A genome assembler viewer;
Ensembl [174]	yes	yes	BED, BedGraph, GFF, GTF, PSL, WIG, BigWig	no	yes	yes	Web-based tool with a variety of reference genome and integrated annotations;
Gaggle [175]	no	no	SQL, GFF	yes	no	yes	For systems biology;
Gap5 [176]	yes	no	ACE, BAF	yes	no	no	This standalone tool has been developed to facilitate the process of finishing assemblies.
GBrowse [177]			GFF	no	yes	no	This web-based tool is the precursor of JBrowse.
G-compass [178]	no	no	-	no	yes	yes	Web-based tool that shows comparison of different homolog genomes;
Genome Environment Browser [179]	no	no	-	yes	no		This tool has its strength for viewing repeat elements and other non-genic sequence features.
GenomeView [180]	yes	no	FASTA, GFF; BED, WIG, TDF, MAF, MAQ		no	yes	
GenoViewer [181]	yes	no	FASTA, GFF	yes	no	yes	It is a standalone genome viewer that is not developed or supported anymore.
Hawkeye [182]	no	no	fastq, fastq	yes	no	no	A genome assembler viewer;
Integrated Genome Browser (IGB) [183]	yes	no	DAS, wig	yes (JWS)	no	yes	Standalone, Java tool with export feature into PDF, EPS, PNG, …;

Name	BAM/SAM	VCF	Other formats	GUI	Web	Annotation	Notes
Integrative Genomics Viewer (IGV) [184]	yes	yes	(> 30 formats) TDF, CN, SNP, GCT, RES, GFF, GFF3, BED, GISTIC, LOH, MUT, GCT, SEG, CBS, IGV, TAB, WIG	yes	no	yes	Can be started locally or from websites; offers lots of customization features;
JalView [185]	no	no	DAS	yes	no	no	This tools is capable of performing multiple sequence alignment.
JBrowse [186]	no	no	FASTA, BED, GFF, GFF3, WIG	no	yes (local)	yes	It is a web based tool where tracks are rendered on the client side. Tracks need to be prepared by the user in advance.
LookSeq [187]	yes	no	MAQ, CIGAR	no	yes	no	Web based alignment viewer;
MagicViewer [188]	no	no	ACE	yes	no	yes	This tool is aimed at users who work with DNA methylation data.
MapView [189]	no	no	MVF	yes	no	no	
NGSView [190]	no	no	XML, BED, BLAST, Eland, mapview processed MAQ, Corona, GFF	yes	no	yes	Sequence alignment editor;
SAMSCOPE [191]	yes	no	BIP (specific file format)	yes	no	no	
samtools tview [47]	yes	no	-	CLI	no	no	
Savant [192]	yes	yes	FASTA, BED, GFF, WIG, any tab-delimited	yes	no	yes	Standalone, Java based genome viewer which allows users to create their own plug-ins;
SeqWord [193]				no	yes	no	A web-based tool to visualize the natural compositional polymorphism of DNA sequences.
SNUGB [194]				no	yes	no	The tool has been developed for comparative genomics.
Tablet [195]	yes	no	ACE, AFG, MAQ, SOAP2, FASTA, FASTQ, GFF3	yes	no	yes	
UCSC cancer genomics browser [196]	no	no	-	no	yes	yes	This tool displays cancer related datasets, but does not allow the upload of local data.
UCSC Genome Browser [197]	yes	yes	BED, bigBed, bedGraph, GFF, GTF, WIG, bigWig, MAF, BED, SNP, PSL	no	yes	yes	Web-based tool with a variety of public databases; It offers many customization features and allows the user to upload new tracks.
UTGB toolkit [198]	yes	no	FASTA, BED, WIG, DAS	no	yes (local)	yes	The tool is web-based and uses a dedicated database and web-server. It offers flexible customization possibilities and tracks can hold private or public data.
VEGA [199]	yes	yes	BED, bedGraph, BigBed, BigWig, GBrowse, GFF, GTF, PSL, WIG.	no	yes	yes	This application contains manually annotated genomes from different species. Large parts of the human genome are annotated.
Vista [200]	no	no	-	no	yes	no	The web-based viewer can be used to perform comparative genomics.

Table S9 CNV & SV visualization

Name	Input	Output	Types	GUI	Web	Notes
Circos [201]	2D tracks, CSV	scatter, line, and histogram plots, heat maps, tiles, connectors, and text	CNV, INDEL, TRANS, INV	yes	no	Also runs on Windows;
Gremlin [202]	?	?	?	no	yes	Currently Gremlin cannot be downloaded.

Table S10 Pipelines

Name	Input Format	Output Format	Illumina	Solid	Requirements	GUI	CLI	Cloud	Algn	Var	Anno	Notes
Bcbio-nextgen [203]	fastq	fatsq, sam, BAM, bed, VCF, pdf	yes	?	Linux, MaC OS X, Windows	yes	yes	no	yes	yes	no	Fully automated pipeline which includes alignment, SNP calling, summary collection, and integration into Galaxy;
Crossbow [204]	fastq, sra	SoapSNP output file	yes	yes		yes	yes	yes	yes	yes	no	Software pipeline which combines Bowtie and SoapSNP;
Games [205]	BAM+FASTA(ref)	summary and annotation text html files, output files that can in SIFT and PolyPhen	yes	yes	Linux	no	yes	no	no	yes	yes	Tool for identifying and predicting mutations; does not include alignment;
GATK [37]	BAM, FASTA(ref), dbSNP(rod files), refSeq table	CSV, VCF, txt	yes	yes	Linux	no	yes	no	no	yes	yes	Performs no alignment;
HugeSeq [206]	fastq, FASTA	VCF, gff, Annovar output	yes	-	Linux	no	yes	no	yes	yes	yes	Combines tools for alignment, variant calling, and annotation. Also identifies CNVs and SVs;
inGap [207]	fastq, FASTA	CSV	yes	no	Linux, MaC OS X, Windows	yes	no	no	yes	yes	no	Application with a graphical user interface integrating alignment and variant detection; it can be used for comparing genomes and simulating reads.

Name	Input Format	Output Format	Illumina	Solid	Requirements	GUI	CLI	Cloud	Algn	Var	Anno	Notes
MutationTaster NGS pipeline [208]	FASTA, fastq, csFASTA (reads) + FASTA (ref. Genome) + Annotation file in tab separated format with all Ensembl transcripts + bed (target region - optional)	MutationTaster snippets, read coverage, SNP positions, variation counts	yes	yes	Linux	no	yes	no	yes	yes	no	Performs mapping, variant calling, and variant annotation using MutationTaster and offers filtering options for the SNPs; requires basic knowledge of Linux.
Ngs-backbone [209]	fastq, FASTA	VCF, gff,text files	yes	no	Linux	no	yes	no	yes	yes	no (not functional)	Pipeline which includes read cleaning, mapping, transcriptome assembly, annotation and SNV calling;
RTG [210]	fastq, FASTA, Complete Genomics format	SAM, BAM, BED, VCF, snp files	yes	no	Linux, MAC OS X, Windows	no	yes	no	yes	yes	no	This tool performs alignment, variant detection, and calculation of various summary statistics.
SeqGene [211]	SAM	snpa, wig, svg	yes	yes	Linux, Windows, Mac OS X	no	yes	no	no	yes	yes	This pipeline supports SNP/INDEL detection, SNP filtering, and performing SNP-expression association tests but has no alignment capability. It offers pre-built annotation packages for latest Ensembl human, mouse, and rat genomes.
SHORE [212]	fastq, CSFASTA	various text files (for SNPs, SVs, CNVs)	yes	yes	Linux or MacOS; Dual Core; 2GB RAM; 500GB storage	no	yes	no	yes	yes	no	The pipeline covers alignment and performs its own statistical analysis to detect variants.
Simplex [213]	fastq	sam, BAM, gtf, amt, xls	yes	yes	Linux	no	yes	yes	yes	yes	yes	

Name	Input Format	Output Format	Illumina	Solid	Requirements	GUI	CLI	Cloud	Algn	Var	Anno	Notes
Treat [214]	fastq, BAM or variant file	summary files	yes	yes (annotation module)	Linux; 4-cores - 16 GB of RAM, ~175 GB storage space	no	yes	yes	yes	yes	yes	A tool for analyzing and interpreting NGS data covering alignment, variant calling, and variant annotation; It provides four different categories of variant annotations and links variants to a genome viewer. Currently, it provides no hg19 reference.

Table S11 Workflow systems

Name	Illumina	Solid	Requirements	GUI	CLI	Online	Cloud	Notes
Ergatis [215]	yes	yes	Linux, MAC OS X, Windows	yes	no	yes	Yes	Web-based workflow management system for configuring and monitoring pipelines; offers components for BWA, bowtie without executables and is due to its complexity aimed at bioinformaticians;
Galaxy [216]	yes	yes	Linux, MAC OS X	yes	no	yes	yes	Web-based platform for performing, reproducing, and sharing complete analyses; The system offers graphical workflow editing and includes many built-in NGS tools. Users can add new tools and share them with the community.
Genboree Workbench [217]	yes	yes	Linux, MAC OS X, Windows	yes	no	yes	Yes	It is a platform for deploying genomic tools as a service and offers a web-based drag and drop interface. Tasks are executed on a compute cluster and data can be uploaded as tracks. The system cannot be installed locally.
GenePattern [218]	yes	yes	Linux, MAC OS X, Windows	yes	no	yes	No	Scientific workflow system that provides access to more than 150 genomic analysis tools;
GeneProf [219]	yes	yes	Linux (it is not tested on Others yet)	yes	no	yes	No	A web-based, graphical software suite for the analysis of NGS data; To complete the installation IT experience is required.
Kepler (bioKepler) [220]	yes	yes	Linux, MAC OS X, Windows; > 1 GB RAM, 2 GHz CPU	yes	no	no	No	Free software with a graphic user interface system for managing scientific workflows; supports hierarchy in workflows to create modular components.;
KNIME [221]	yes	-	Linux, MAC OS X, Windows	yes	yes	no	Yes	Open source platform for graphically building and editing workflows and data analysis pipelines; includes some NGS analysis programs, which are mostly for filtering and manipulating VCF files;
LONI Pipeline [222]	yes	yes	Linux, MAC OS X, Windows	yes	yes	no	No	Workflow processing application where executables can be used by creating dedicated wrappers; provides a few NGS analysis pipelines and users can apply for an account to download and use the software;
Moa [223]	yes	yes	Linux	yes	yes	no	No	The command line based management system allows users to write and executes workflows; offers some NGS support, mainly for aligning reads to a reference genome;
Tavaxy [224]	yes	yes	Linux	yes	no	yes	Yes	The tool combines the web interface of Galaxy, with the complexity of Taverna.
Taverna [225]	yes	yes	Linux, MAC OS X, Windows	yes	yes	no	yes	Open source, web-service based workflow management application with a large library of existing tools that operate on genomic sequences; does not ship with any prepackaged sequence analysis tools and integrating the tools requires some programming experience;
Yabi [226]	-	-	Linux	yes	yes	yes	yes	Currently, only a trial version is available and users need to request for an account. The public available version does not contain predefined workflows for NGS analysis.

Supplementary Information I1 Additional information about evaluated software tools

Crisp

The tool requires an existing Python installation and provides a tool to convert BAM files into their pileup format. The pileup format created by SAMtools is not supported.

GATK

It requires Ant and Java to build and execute the tools. GATK offers an extensive documentation and wiki system as well as a very active community.

SAMtools

The included software suite "BCFtools" has the ability to call SNPs and short INDELs from a single alignment file in the pileup format, which in turn can easily be generated from multiple BAM or SAM files, using SAMtools' mpileup. SAMtools is self-contained and has no further dependencies. It needs to be compiled from source code, and the steps required to do so are listed in its detailed documentation.

SNVer

The tool is written in Java and hence, it is operating system independent and requires no dependencies besides a Java runtime environment. SNVer provides a straightforward command-line interface, which is well documented. Additionally, the tool includes sample data for quick testing.

VarScan 2

It takes input files in pileup format, thus requiring a preparation step using SAMtools to convert BAM input files. Installation is only dependent on an existing Java environment and extensive usage documentation is provided.

SomaticSniper

The required input is a tumor BAM file, a normal BAM file, and the corresponding reference genome in FASTA format. SomaticSniper outputs results either in VCF format or in a format similar to the SAMtools consensus format. The program is available

as a Debian APT package or can be manually installed. Both the installation and user manual are detailed and standard usage is straightforward.

CNVnator

It uses read-depth analysis based on mean-shift to detect CNVs. In order to compile the tool, the ROOT data analysis framework is required, which has to be compiled and configured as well. Though it provides a script for setting the necessary environment variables, the script did not work on our test system and consequently the variables had to be set manually. CNVnator provides a command-line interface for analysis. However, multiple separate commands are necessary in order obtain CNV calls.

CONTRA

The tool requires Python and R and is dependent on BEDtools (which is included in the installation package) and SAMtools. Installation of the tool is straightforward and well documented and test data to check correct configuration is provided. It is important that BEDtools is correctly configured as CONTRA will not run without it and does not check for its correct configuration on startup. The application takes as input SAM/BAM files for test and control samples and outputs results in VCF and CSV format. The control sample can additionally be provided as a baseline file in BED format.

ExomeCNV

It takes as input BAM files, which need to be converted into a proprietary pileup format. The conversion can be done using a provided script or GATK's *DepthOfCoverage* function, where the latter is preferred due to superior runtime performance. Moreover, BAM files need to be sorted canonically and require appropriate read group information. Further required input is an exome definition file in *chr#:start-end* format and a conservative approximation of sample admixture rate (for tumor samples only). ExomeCNV is dependent on R including the package DNAcopy.

RDXplorer

The tool is distributed as a stand-alone program but can also be used on a high performance computer cluster. It accepts as input sorted BAM files where duplicates need to be either marked or removed, e.g.: using Picard's *MarkDuplicates* function (http://picard.sourceforge.net). Installation of the program is dependent on multiple programs (blas, lapack, nose, SAMtools), requires Python including multiple packages, and an R installation with special configuration parameters. The installation is challenging for non-experienced users. To start the analysis, a configuration file and the run file need to be adapted.

BreakDancer

The tool depends on Perl and the installation of five Perl modules. As BreakDancer has only been compiled for MacOSX, manual compilation for Linux systems is required, which depends on the installation of SAMtools and manual adaptation of the make file.

Breakpointer

It requires aligned BAM files as input and outputs predicted regions containing breakpoints in GFF format. As the tool does not investigate SV content, additional methods for the classification of SVs are required. Breakpointer depends on the BamTools API (https://github.com/pezmaster31/bamtools), which needs to be installed before compilation. Installation and usage is documented in a readme file. No test data is provided and no further information how to interpret the output is given.

CLEVER

The command-line tool provides easy to use compilation and installation scripts. CLEVER offers an intuitive script with default parameters called "clever-all-in-one", but parameters can also be fine-tuned. It takes BAM files as input and delivers as output detected variations in a proprietary text file format. BAM files need to be sorted by read name which can be achieved using the SortSam command of the Picard suite with *SORT_ORDER=queryname*.

GASVPro

It requires the installation of the GASV software, which is dependent on Perl, Java, and Ant. The tool provides an extensive user manual, example data, and the scripts *GASVPro.sh* and *GASVPro-HQ.sh*, which facilitate the streamlined use of the pipeline. *GASVPro-HQ* works with datasets containing high-quality unique read mappings in BAM format whereas *GASVPro* requires a combination of a high-quality unique read BAM and a lower-quality possibly multiple mapping BAM file.

SVMerge

SVMerge requires a multitude of dependencies that need to be installed before usage, including Perl, LSF queuing system, and all integrated tools. The software does not offer ready-to-use virtual box or cloud implementations, which would enhance the usability of the installation.

ANNOVAR

The tool includes a separate script called *auto_annovar.pl*, which implements an iterative filter workflow that has been used to identify two causal mutations for Miller syndrome. If the user wants to use all features of ANNOVAR, numerous manual commands need to be executed to download all databases. The tool requires its own text-based input format, though there are easy to use scripts provided to convert most popular genotype calling formats.

AnnTools

Although the authors provide an easy to use installation script, the program requires a MySQL database, which has to be installed and configured manually. Once the database is configured, the tool provides a straightforward command-line interface.

NGS-SNP

The scripts have numerous dependencies which all have to be installed manually. To facilitate the usage of the tool, the authors provide a pre-configured Linux virtual image, which requires only one simple configuration script. While testing, the regular annotation script lost connection to the Ensembl database several times, which aborted the annotation process. However, the authors provide a separate script, which should fix this issue.

SeattleSeq

The output can be provided either in VCF format or SeattleSeq's own text output format. The latter can then be displayed in a simple web interface providing overview metrics, as well as some basic sorting and filtering options.

SVA

Variation lists can be exported in CSV format. Required input parameters are list of variants in VCF format, the reference genome, and binary coverage files. Furthermore, a project configuration file has to be written, defining the input files, the reference genome, and various other parameters or data sources. The required database files have to be downloaded manually and are denoted in the script file.

snpEff

The program can be downloaded as a jar file and besides downloading the desired reference databases, no additional set up steps are required. The results are provided either in VCF or text format, and additionally, in an easy to read html summary file, including various diagrams, is generated.

VARIANT

The tool requires that VCF files use GRCh references and even though it recognizes inputs with a UCSC hg version as valid VCF files it does not annotate mutations based on a UCSC hg reference genome.

VEP

The command-line version includes a simple installation script that installs missing modules and downloads a minimal set of the Ensembl API. For the annotation process the online database is used per default, but database files can be cached locally. The Perl API addresses developers who want to integrate the functionality of VEP into their own programs.

UCSC Genome Browser

The tested BAM and VCF files needed to be sorted and indexed. However, given the correct format, the tool offers great usability and supports many different data formats.

Vega

BAM and VCF files can be easily uploaded into Vega, and the usage is similar to the Ensemble Genome Browser.

Artemis

It is important that users load the appropriate reference genome first, otherwise uploading BAM and VCF files is not possible. The tool offers many different features and customization options.

IGV

Furthermore, MATLAB users can load files directly into IGV and jump to specific loci. Upon startup, IGV automatically loads a specified reference genome. All test files could be loaded into the genome browser and the interface is intuitive and responsive.

Savant

Savant features a plug-in mechanism where users are encouraged to develop new features. All tested files could be loaded into Savant. The interface is responsive, and additional information can be directly loaded from the UCSC web server.

Circos

The tool requires an existing Perl installation and is dependent on several Perl modules. A test script is included in order to determine which modules are still missing or are not configured properly. The visual appearance of the output is defined using configuration files.

References

1. Cibulskis K, McKenna A, Fennell T, et al. ContEst: estimating cross-contamination of human samples in next-generation sequencing data. *Bioinformatics* 2011;**27**:2601–2602.
2. FastQC. http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc.
3. FASTX-Toolkit. http://hannonlab.cshl.edu/fastx_toolkit.
4. Blankenberg D, Gordon A, Von Kuster G, et al. Manipulation of FASTQ data with Galaxy. *Bioinformatics* 2010;**26**:1783–1785.
5. Planet E, Attolini CS-O, Reina O, et al. htSeqTools: high-throughput sequencing quality control, processing and visualization in R. *Bioinformatics* 2012;**28**:589–590.
6. Dai M, Thompson RC, Maher C, et al. NGSQC: cross-platform quality analysis pipeline for deep sequencing data. *BMC Genomics* 2010;**11 Suppl 4**:S7.
7. Martínez-Alcántara A, Ballesteros E, Feng C, et al. PIQA: pipeline for Illumina G1 genome analyzer data quality assessment. *Bioinformatics* 2009;**25**:2438–2439.
8. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 2011;**27**:863–864.
9. Cox MP, Peterson DA, Biggs PJ. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* 2010;**11**:485.
10. Schmieder R, Lim YW, Rohwer F, et al. TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. *BMC Bioinformatics* 2010;**11**:341.
11. Dolan PC, Denver DR. TileQC: a system for tile-based quality control of Solexa data. *BMC Bioinformatics* 2008;**9**:250.
12. Klus P, Lam S, Lyberg D, et al. BarraCUDA - a fast short read sequence aligner using graphics processing units. *BMC Res Notes* 2012;**5**:27.
13. Homer N, Merriman B, Nelson SF. BFAST: an alignment tool for large scale genome resequencing. *PLoS ONE* 2009;**4**:e7767.
14. Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;**10**:R25.
15. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 2012;**9**:357–359.
16. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;**25**:1754–1760.
17. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;**26**:589–595.
18. Cox AJ. ELAND: Efficient large-scale alignment of nucleotide databases. Illumina. 2007;
19. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 2008;**18**:1851–1858.
20. Mosaik. http://code.google.com/p/mosaik-aligner/.
21. Alkan C, Kidd JM, Marques-Bonet T, et al. Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat. Genet.* 2009;**41**:1061–1067.
22. Hach F, Hormozdiari F, Alkan C, et al. mrsFAST: a cache-oblivious algorithm for short-read mapping. *Nat. Methods* 2010;**7**:576–577.

Formatted: French (Belgi

Formatted: French (Belgi

23. Novoalign. http://novocraft.com.
24. Li R, Yu C, Li Y, et al. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 2009;**25**:1966–1967.
25. Liu C-M, Wong T, Wu E, et al. SOAP3: ultra-fast GPU-based parallel alignment tool for short reads. *Bioinformatics* 2012;**28**:878–879.
26. Ning Z, Cox AJ, Mullikin JC. SSAHA: a fast search method for large DNA databases. *Genome Res.* 2001;**11**:1725–1729.
27. Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* 2011;**21**:936–939.
28. Galinsky VL. YOABS: yet other aligner of biological sequences--an efficient linearly scaling nucleotide aligner. *Bioinformatics* 2012;**28**:1070–1077.
29. Challis D, Yu J, Evani US, et al. An integrative variant analysis suite for whole exome next-generation sequencing data. *BMC Bioinformatics* 2012;**13**:8.
30. Edmonson MN, Zhang J, Yan C, et al. Bambino: a variant detector and alignment viewer for next-generation sequencing data in the SAM/BAM format. *Bioinformatics* 2011;**27**:865–866.
31. Browning BL, Yu Z. Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. *Am. J. Hum. Genet.* 2009;**85**:847–861.
32. CoNAn-SNV. http://compbio.bccrc.ca/software/conan-snv/.
33. Iqbal Z, Caccamo M, Turner I, et al. De novo assembly and genotyping of variants using colored de Bruijn graphs. *Nat. Genet.* 2012;**44**:226–232.
34. Bansal V. A statistical method for the detection of variants from next-generation resequencing of DNA pools. *Bioinformatics* 2010;**26**:i318–324.
35. Albers CA, Lunter G, MacArthur DG, et al. Dindel: accurate indel calls from short-read data. *Genome Res.* 2011;**21**:961–973.
36. FreeBayes. http://bioinformatics.bc.edu/marthlab/FreeBayes.
37. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 2011;**43**:491–498.
38. GSNP. http://jil.genomics.org.cn/index.php/en/software/gsnp.html.
39. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009;**5**:e1000529.
40. Indelocator. https://confluence.broadinstitute.org/display/CGATools/Indelocator.
41. Ion Variant Hunter. https://github.com/iontorrent/Ion-Variant-Hunter.
42. Li Y, Willer CJ, Ding J, et al. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* 2010;**34**:816–834.
43. Lee S, Hormozdiari F, Alkan C, et al. MoDIL: detecting small indels from clone-end sequencing with mixtures of distributions. *Nat. Methods* 2009;**6**:473–474.
44. Chen K, McLellan MD, Ding L, et al. PolyScan: an automatic indel and SNP detection approach to the analysis of human resequencing data. *Genome Res.* 2007;**17**:659–666.
45. Le SQ, Durbin R. SNP detection and genotyping from low-coverage sequencing data on multiple diploid samples. *Genome Res.* 2011;**21**:952–960.
46. realSFS. http://jil.genomics.org.cn/index.php/en/software/realsfs.html.
47. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;**25**:2078–2079.
48. Malhis N, Jones SJM. High quality SNP calling using Illumina data at shallow coverage. *Bioinformatics* 2010;**26**:1029–1035.
49. Simola DF, Kim J. Sniper: improved SNP discovery by multiply mapping deep sequenced reads. *Genome Biol.* 2011;**12**:R55.
50. Wei Z, Wang W, Hu P, et al. SNVer: a statistical tool for variant calling in analysis of pooled or individual next-generation sequencing data. *Nucleic Acids Res.* 2011;**39**:e132.
51. Goya R, Sun MGF, Morin RD, et al. SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. *Bioinformatics* 2010;**26**:730–736.
52. SOAPindel. http://soap.genomics.org.cn/soapindel.html.
53. Rivas MA, Beaudoin M, Gardet A, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat. Genet.* 2011;**43**:1066–1073.
54. Dalca AV, Rumble SM, Levy S, et al. VARiD: a variation detection framework for color-space and letter-space platforms. *Bioinformatics* 2010;**26**:i343–349.
55. Altmann A, Weber P, Quast C, et al. vipR: variant identification in pooled DNA using R. *Bioinformatics* 2011;**27**:i77–84.
56. Ding J, Bashashati A, Roth A, et al. Feature-based classifiers for somatic mutation detection in tumour-normal paired sequencing data. *Bioinformatics* 2012;**28**:167–175.
57. MuTect. https://confluence.broadinstitute.org/display/CGATools/MuTect.
58. SomaticCall. http://www.broadinstitute.org/science/programs/genome-biology/computational-rd/somaticcall-manual.
59. Larson DE, Harris CC, Chen K, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics* 2012;**28**:311–317.
60. Vallania FLM, Druley TE, Ramos E, et al. High-throughput discovery of rare insertions and deletions in large cohorts. *Genome Res.* 2010;**20**:1711–1718.
61. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 2012;**22**:568–576.
62. Ivakhno S, Royce T, Cox AJ, et al. CNAseg--a novel framework for identification of copy number changes in cancer from second-generation sequencing data. *Bioinformatics* 2010;**26**:3051–3058.
63. Medvedev P, Fiume M, Dzamba M, et al. Detecting copy number variation with mated short reads. *Genome Res.* 2010;**20**:1613–1622.
64. cnvHMM. http://genome.wustl.edu/software/cnvhmm.
65. Abyzov A, Urban AE, Snyder M, et al. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* 2011;**21**:974–984.
66. Xie C, Tammi MT. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics* 2009;**10**:80.

Formatted: French (Belgi

Formatted: French (Belgi

67. Li J, Lupat R, Amarasinghe KC, et al. CONTRA: copy number analysis for targeted resequencing. *Bioinformatics* 2012;**28**:1307–1313.
68. Waszak SM, Hasin Y, Zichner T, et al. Systematic inference of copy-number genotypes from personal genome sequencing data reveals extensive olfactory receptor gene content diversity. *PLoS Comput. Biol.* 2010;**6**:e1000988.
69. Sathirapongsasuti JF, Lee H, Horst BAJ, et al. Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics* 2011;**27**:2648–2654.
70. Yoon S, Xuan Z, Makarov V, et al. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res.* 2009;**19**:1586–1592.
71. Miller CA, Hampton O, Coarfa C, et al. ReadDepth: a parallel R package for detecting copy number alterations from short sequencing reads. *PLoS ONE* 2011;**6**:e16327.
72. Chiang DY, Getz G, Jaffe DB, et al. High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nat. Methods* 2009;**6**:99–103.
73. Ha G, Roth A, Lai D, et al. Integrative analysis of genome-wide loss of heterozygosity and mono-allelic expression at nucleotide resolution reveals disrupted pathways in triple negative breast cancer. *Genome research* 2012;
74. Chen K, Wallis JW, McLellan MD, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat. Methods* 2009;**6**:677–681.
75. Sun R, Love MI, Zemojtel T, et al. Breakpointer: using local mapping artifacts to support sequence breakpoint discovery from single-end reads. *Bioinformatics* 2012;**28**:1024–1025.
76. Lam HYK, Mu XJ, Stütz AM, et al. Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. *Nat. Biotechnol.* 2010;**28**:47–55.
77. Clark MJ, Homer N, O'Connor BD, et al. U87MG decoded: the genomic sequence of a cytogenetically aberrant human cancer cell line. *PLoS Genet.* 2010;**6**:e1000832.
78. Marschall T, Costa I, Canzar S, et al. CLEVER: Clique-Enumerating Variant Finder. *arXiv:1203.0937* 2012;
79. Suzuki S, Yasuda T, Shiraishi Y, et al. ClipCrop: a tool for detecting structural variations with single-base resolution using soft-clipping information. *BMC Bioinformatics* 2011;**12 Suppl 14**:S7.
80. Wang J, Mullighan CG, Easton J, et al. CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat. Methods* 2011;**8**:652–654.
81. Ge H, Liu K, Juan T, et al. FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. *Bioinformatics* 2011;**27**:1922–1928.
82. Sindi SS, Onal S, Peng L, et al. An integrative probabilistic model for identification of structural variation in sequencing data. *Genome biology* 2012;**13**:R22.
83. Quinlan AR, Clark RA, Sokolova S, et al. Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. *Genome Res.* 2010;**20**:623–635.
84. Korbel JO, Abyzov A, Mu XJ, et al. PEMer: a computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data. *Genome Biol.* 2009;**10**:R23.
85. Ye K, Schulz MH, Long Q, et al. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 2009;**25**:2865–2871.
86. Karakoc E, Alkan C, O'Roak BJ, et al. Detection of structural variants and indels within exome data. *Nat. Methods* 2012;**9**:176–178.
87. Zeitouni B, Boeva V, Janoueix-Lerosey I, et al. SVDetect: a tool to identify genomic structural variations from paired-end and mate-pair sequencing data. *Bioinformatics* 2010;**26**:1895–1896.
88. Wong K, Keane TM, Stalker J, et al. Enhanced structural variant and breakpoint detection using SVMerge by integration of multiple detection methods and local assembly. *Genome Biol.* 2010;**11**:R128.
89. Mills RE, Walter K, Stewart C, et al. Mapping copy number variation by population-scale genome sequencing. *Nature* 2011;**470**:59–65.
90. Hormozdiari F, Hajirasouliha I, Dao P, et al. Next-generation VariationHunter: combinatorial algorithms for transposon insertion discovery. *Bioinformatics* 2010;**26**:i350–357.
91. Carter SL, Cibulskis K, Helman E, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* 2012;**30**:413–421.
92. Mathe E, Olivier M, Kato S, et al. Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids Res.* 2006;**34**:1317–1325.
93. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;**38**:e164.
94. Makarov V, O'Grady T, Cai G, et al. AnnTools: a comprehensive and versatile annotation toolkit for genomic variants. *Bioinformatics* 2012;**28**:724–725.
95. Masso M, Vaisman II. AUTO-MUTE: web-based tools for predicting stability changes in proteins due to single amino acid replacements. *Protein Eng. Des. Sel.* 2010;**23**:683–687.
96. Schmitt AO, Assmus J, Bortfeldt RH, et al. CandiSNPer: a web tool for the identification of candidate SNPs for causal variants. *Bioinformatics* 2010;**26**:969–970.
97. Wong WC, Kim D, Carter H, et al. CHASM and SNVBox: toolkit for detecting biologically important single nucleotide mutations in cancer. *Bioinformatics* 2011;**27**:2147–2148.
98. Parthiban V, Gromiha MM, Schomburg D. CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Res.* 2006;**34**:W239–242.
99. Liu X, Jian X, Boerwinkle E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum. Mutat.* 2011;**32**:894–899.
100. McLaren W, Pritchard B, Rios D, et al. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 2010;**26**:2069–2070.
101. Cartegni L, Wang J, Zhu Z, et al. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* 2003;**31**:3568–3571.
102. Fairbrother WG, Yeh R-F, Sharp PA, et al. Predictive identification of exonic splicing enhancers in human genes. *Science* 2002;**297**:1007–1013.
103. Liu C-K, Chen Y-H, Tang C-Y, et al. Functional analysis of novel SNPs and mutations in human and mouse genomes. *BMC Bioinformatics* 2008;**9 Suppl 12**:S10.
104. Yuan H-Y, Chiou J-J, Tseng W-H, et al. FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Res.* 2006;**34**:W635–641.
105. Kang HJ, Choi KO, Kim B-D, et al. FESD: a Functional Element SNPs Database in human. *Nucleic Acids Res.* 2005;**33**:D518–522.
106. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* 2002;**320**:369–387.

Formatted: French (Belgi

Formatted: French (Belgi

107. Lee PH, Shatkay H. F-SNP: computationally predicted functional SNPs for disease association studies. *Nucleic Acids Res.* 2008;**36**:D820–824.
108. Davydov EV, Goode DL, Sirota M, et al. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* 2010;**6**:e1001025.
109. Mermel CH, Schumacher SE, Hill B, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 2011;**12**:R41.
110. Venselaar H, Te Beek TAH, Kuipers RKP, et al. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* 2010;**11**:548.
111. Desmet F-O, Hamroun D, Lalande M, et al. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* 2009;**37**:e67.
112. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* 2005;**33**:W306–310.
113. Karchin R, Diekhans M, Kelly L, et al. LS-SNP: large-scale annotation of coding non-synonymous SNPs based on multiple information sources. *Bioinformatics* 2005;**21**:2814–2820.
114. Stone EA, Sidow A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res.* 2005;**15**:978–986.
115. Wainreb G, Ashkenazy H, Bromberg Y, et al. MuD: an interactive web server for the prediction of non-neutral substitutions using protein structural data. *Nucleic Acids Res.* 2010;**38**:W523–528.
116. Stoyanovich J, Pe'er I. MutaGeneSys: estimating individual disease susceptibility based on genome-wide SNP array data. *Bioinformatics* 2008;**24**:440–442.
117. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* 2011;**39**:e118.
118. Schwarz JM, Rödelsperger C, Schuelke M, et al. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* 2010;**7**:575–576.
119. Li B, Krishnan VG, Mort ME, et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* 2009;**25**:2744–2750.
120. Banerji S, Cibulskis K, Rangel-Escareno C, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* 2012;**486**:405–409.
121. Grant JR, Arantes AS, Liao X, et al. In-depth annotation of SNPs arising from resequencing projects using NGS-SNP. *Bioinformatics* 2011;**27**:2300–2301.
122. Bao L, Zhou M, Cui Y. nsSNPAnalyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Res.* 2005;**33**:W480–482.
123. Oncotator. http://www.broadinstitute.org/oncotator/.
124. Thomas PD, Campbell MJ, Kejariwal A, et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 2003;**13**:2129–2141.
125. Tian J, Wu N, Guo X, et al. Predicting the phenotypic effects of non-synonymous single nucleotide polymorphisms based on support vector machines. *BMC Bioinformatics* 2007;**8**:450.
126. Zhang XH-F, Kangsamaksin T, Chao MSP, et al. Exon inclusion is dependent on predictable exonic splicing enhancers. *Mol. Cell. Biol.* 2005;**25**:7323–7332.
127. Wang J, Ronaghi M, Chong SS, et al. pfSNP: An integrated potentially functional SNP resource that facilitates hypotheses generation through knowledge syntheses. *Hum. Mutat.* 2011;**32**:19–24.
128. Hubisz MJ, Pollard KS, Siepel A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinformatics* 2011;**12**:41–51.
129. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics* 2006;**22**:2729–2734.
130. Ferrer-Costa C, Gelpí JL, Zamakola L, et al. PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics* 2005;**21**:3176–3178.
131. Jegga AG, Gowrisankar S, Chen J, et al. PolyDoms: a whole genome database for the identification of non-synonymous coding SNPs with the potential to impact disease. *Nucleic Acids Res.* 2007;**35**:D700–706.
132. Freimuth RR, Stormo GD, McLeod HL. PolyMAPr: programs for polymorphism database mining, annotation, and functional analysis. *Hum. Mutat.* 2005;**25**:110–117.
133. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat. Methods* 2010;**7**:248–249.
134. Conde L, Vaquerizas JM, Santoyo J, et al. PupaSNP Finder: a web tool for finding SNPs with putative effect at transcriptional level. *Nucleic Acids Res.* 2004;**32**:W242–248.
135. Grover D, Woodfield AS, Verma R, et al. QuickSNP: an automated web server for selection of tagSNPs. *Nucleic Acids Res.* 2007;**35**:W115–120.
136. Yeo G, Hoon S, Venkatesh B, et al. Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc. Natl. Acad. Sci. U.S.A.* 2004;**101**:15700–15705.
137. Ye Z-Q, Zhao S-Q, Gao G, et al. Finding new structural and sequence attributes to predict possible disease association of single amino acid polymorphism (SAP). *Bioinformatics* 2007;**23**:1444–1450.
138. Gamazon ER, Zhang W, Konkashbaev A, et al. SCAN: SNP and copy number annotation. *Bioinformatics* 2010;**26**:259–262.
139. Asthana S, Roytberg M, Stamatoyannopoulos J, et al. Analysis of sequence conservation at nucleotide resolution. *PLoS Comput. Biol.* 2007;**3**:e254.
140. Seattle Seq Annotation. http://snp.gs.washington.edu/SeattleSeqAnnotation/.
141. Shetty AC, Athri P, Mondal K, et al. SeqAnt: a web service to rapidly identify and annotate DNA sequence variations. *BMC Bioinformatics* 2010;**11**:471.
142. Capriotti E, Arbiza L, Casadio R, et al. Use of estimated evolutionary strength at the codon level improves the prediction of disease-related protein mutations in humans. *Hum. Mutat.* 2008;**29**:198–204.
143. Ge D, Ruzzo EK, Shianna KV, et al. SVA: software for annotating and visualizing sequenced human genomes. *Bioinformatics* 2011;**27**:1998–2000.
144. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;**4**:1073–1081.
145. Hu J, Ng PC. Predicting the effects of frameshifting indels. *Genome biology* 2012;**13**:R9.
146. Garber M, Guttman M, Clamp M, et al. Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* 2009;**25**:i54–62.

Formatted: French (Belgi
Formatted: French (Belgi
Formatted: French (Belgi

147. Bromberg Y, Rost B. SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res.* 2007;**35**:3823–3835.
148. Wang P, Dai M, Xuan W, et al. SNP Function Portal: a web database for exploring the function implication of SNP alleles. *Bioinformatics* 2006;**22**:e523–529.
149. Han A, Kang HJ, Cho Y, et al. SNP@Domain: a web resource of single nucleotide polymorphisms (SNPs) within protein domain structures and sequences. *Nucleic Acids Res.* 2006;**34**:W642–644.
150. Schaefer C, Meier A, Rost B, et al. SNPdbe: constructing an nsSNP functional impacts database. *Bioinformatics* 2012;**28**:601–602.
151. De Baets G, Van Durme J, Reumers J, et al. SNPeffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants. *Nucleic Acids Res.* 2012;**40**:D935–939.
152. Wang L, Liu S, Niu T, et al. SNPHunter: a bioinformatic software for single nucleotide polymorphism data acquisition and management. *BMC Bioinformatics* 2005;**6**:60.
153. Chelala C, Khan A, Lemoine NR. SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. *Bioinformatics* 2009;**25**:655–661.
154. Riva A, Kohane IS. A SNP-centric database for the investigation of the human genome. *BMC Bioinformatics* 2004;**5**:33.
155. Calabrese R, Capriotti E, Fariselli P, et al. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum. Mutat.* 2009;**30**:1237–1244.
156. Yue P, Melamud E, Moult J. SNPs3D: candidate gene and SNP selection for association studies. *BMC Bioinformatics* 2006;**7**:166.
157. SNPseek. http://snp.wustl.edu/cgi-bin/SNPseek/index.cgi.
158. Xu H, Gregory SG, Hauser ER, et al. SNPselector: a web tool for selecting SNPs for genetic association studies. *Bioinformatics* 2005;**21**:4181–4186.
159. Cingolani P, Patel VM, Coon M, et al. Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front Genet* 2012;**3**:35.
160. Saccone SF, Bolze R, Thomas P, et al. SPOT: a web-based tool for using biological databases to prioritize SNPs after a genome-wide association study. *Nucleic Acids Res.* 2010;**38**:W201–209.
161. Uzun A, Leslin CM, Abyzov A, et al. Structure SNP (StSNP): a web server for mapping and modeling nsSNPs on protein structures with linkage to metabolic pathways. *Nucleic Acids Res.* 2007;**35**:W384–392.
162. Hemminger BM, Saelim B, Sullivan PF. TAMAL: an integrated approach to choosing SNPs for genetic studies of human complex traits. *Bioinformatics* 2006;**22**:626–627.
163. Stitziel NO, Binkowski TA, Tseng YY, et al. topoSNP: a topographic database of non-synonymous single nucleotide polymorphisms with and without known disease association. *Nucleic Acids Res.* 2004;**32**:D520–522.
164. Medina I, De Maria A, Bleda M, et al. VARIANT: Command Line, Web service and Web interface for fast and accurate functional characterization of variants found by Next-Generation Sequencing. *Nucleic Acids Res.* 2012;**40**:W54–58.
165. Kong L, Wang J, Zhao S, et al. ABrowse--a customizable next-generation genome browser framework. *BMC Bioinformatics* 2012;**13**:2.
166. Lister R, O'Malley RC, Tonti-Filippini J, et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 2008;**133**:523–536.
167. Lee E, Harris N, Gibson M, et al. Apollo: a community resource for genome annotation editing. *Bioinformatics* 2009;**25**:1836–1837.
168. Engels R, Yu T, Burge C, et al. Combo: a whole genome comparative browser. *Bioinformatics* 2006;**22**:1782–1783.
169. Carver T, Harris SR, Berriman M, et al. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 2012;**28**:464–469.
170. Carver T, Harris SR, Otto TD, et al. BamView: visualizing and interpretation of next-generation sequencing read alignments. *Briefings in bioinformatics* 2012;
171. Gordon D. Viewing and editing assembled sequences using Consed. *Curr Protoc Bioinformatics* 2003;**Chapter 11**:Unit11.2.
172. Friedel M, Nikolajewa S, Sühnel J, et al. DiProGB: the dinucleotide properties genome browser. *Bioinformatics* 2009;**25**:2603–2604.
173. Huang W, Marth G. EagleView: a genome assembly viewer for next-generation sequencing technologies. *Genome Res.* 2008;**18**:1538–1543.
174. Spudich GM, Fernández-Suárez XM. Touring Ensembl: a practical guide to genome browsing. *BMC Genomics* 2010;**11**:295.
175. Bare JC, Koide T, Reiss DJ, et al. Integration and visualization of systems biology data in context of the genome. *BMC Bioinformatics* 2010;**11**:382.
176. Bonfield JK, Whitwham A. Gap5--editing the billion fragment sequence assembly. *Bioinformatics* 2010;**26**:1699–1703.
177. Donlin MJ. Using the Generic Genome Browser (GBrowse). *Curr Protoc Bioinformatics* 2009;**Chapter 9**:Unit 9.9.
178. Kawahara Y, Sakate R, Matsuya A, et al. G-compass: a web-based comparative genome browser between human and other vertebrate genomes. *Bioinformatics* 2009;**25**:3321–3322.
179. Huntley D, Tang YA, Nesterova TB, et al. Genome Environment Browser (GEB): a dynamic browser for visualising high-throughput experimental data in the context of genome features. *BMC Bioinformatics* 2008;**9**:501.
180. Abeel T, Van Parys T, Saeys Y, et al. GenomeView: a next-generation genome browser. *Nucleic Acids Res.* 2012;**40**:e12.
181. Laczik M, Tukacs E, Uzonyi B, et al. Geno viewer, a SAM/BAM viewer tool. *Bioinformation* 2012;**8**:107–109.
182. Schatz MC, Phillippy AM, Sommer DD, et al. Hawkeye and AMOS: visualizing and assessing the quality of genome assemblies. *Briefings in bioinformatics* 2011;
183. Nicol JW, Helt GA, Blanchard SG Jr, et al. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics* 2009;**25**:2730–2731.
184. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* 2012;
185. Waterhouse AM, Procter JB, Martin DMA, et al. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009;**25**:1189–1191.
186. Westesson O, Skinner M, Holmes I. Visualizing next-generation sequencing data with JBrowse. *Briefings in bioinformatics* 2012;
187. Manske HM, Kwiatkowski DP. LookSeq: a browser-based viewer for deep sequencing data. *Genome Res.* 2009;**19**:2125–2132.

188. Hou H, Zhao F, Zhou L, et al. MagicViewer: integrated solution for next-generation sequencing data visualization and genetic variation detection and annotation. *Nucleic Acids Res.* 2010;**38**:W732–736.
189. Bao H, Guo H, Wang J, et al. MapView: visualization of short reads alignment on a desktop computer. *Bioinformatics* 2009;**25**:1554–1555.
190. Arner E, Hayashizaki Y, Daub CO. NGSView: an extensible open source editor for next-generation sequencing data. *Bioinformatics* 2010;**26**:125–126.
191. Popendorf K, Sakakibara Y. SAMSCOPE: an OpenGL-based real-time interactive scale-free SAM viewer. *Bioinformatics* 2012;**28**:1276–1277.
192. Fiume M, Williams V, Brook A, et al. Savant: genome browser for high-throughput sequencing data. *Bioinformatics* 2010;**26**:1938–1944.
193. Ganesan H, Rakitianskaia AS, Davenport CF, et al. The SeqWord Genome Browser: an online tool for the identification and visualization of atypical regions of bacterial genomes through oligonucleotide usage. *BMC Bioinformatics* 2008;**9**:333.
194. Jung K, Park J, Choi J, et al. SNUGB: a versatile genome browser supporting comparative and functional fungal genomics. *BMC Genomics* 2008;**9**:586.
195. Milne I, Bayer M, Cardle L, et al. Tablet--next generation sequence assembly visualization. *Bioinformatics* 2010;**26**:401–402.
196. Sanborn JZ, Benz SC, Craft B, et al. The UCSC Cancer Genomics Browser: update 2011. *Nucleic Acids Res.* 2011;**39**:D951–959.
197. Dreszer TR, Karolchik D, Zweig AS, et al. The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res.* 2012;**40**:D918–923.
198. Saito TL, Yoshimura J, Sasaki S, et al. UTGB toolkit for personalized genome browsers. *Bioinformatics* 2009;**25**:1856–1861.
199. Loveland J. VEGA, the genome browser with a difference. *Brief. Bioinformatics* 2005;**6**:189–193.
200. Dubchak I. Comparative analysis and visualization of genomic sequences using VISTA browser and associated computational tools. *Methods Mol. Biol.* 2007;**395**:3–16.
201. Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009;**19**:1639–1645.
202. O'Brien TM, Ritz AM, Raphael BJ, et al. Gremlin: an interactive visualization model for analyzing genomic rearrangements. *IEEE Trans Vis Comput Graph* 2010;**16**:918–926.
203. Bcbio-nextgen. https://github.com/chapmanb/bcbb/blob/master/nextgen/README.md.
204. Langmead B, Schatz MC, Lin J, et al. Searching for SNPs with cloud computing. *Genome Biol.* 2009;**10**:R134.
205. Sana ME, Iascone M, Marchetti D, et al. GAMES identifies and annotates mutations in next-generation sequencing projects. *Bioinformatics* 2011;**27**:9–13.
206. Lam HYK, Pan C, Clark MJ, et al. Detecting and annotating genetic variations using the HugeSeq pipeline. *Nat. Biotechnol.* 2012;**30**:226–229.
207. Qi J, Zhao F, Buboltz A, et al. inGAP: an integrated next-generation genome analysis pipeline. *Bioinformatics* 2010;**26**:127–129.
208. MutationTaster. http://www.mutationtaster.org/NextGenerationSequencing.html.
209. Blanca JM, Pascual L, Ziarsolo P, et al. ngs_backbone: a pipeline for read cleaning, mapping and SNP calling using next generation sequence. *BMC Genomics* 2011;**12**:285.
210. RTG. http://www.realtimegenomics.com/.
211. Deng X. SeqGene: a comprehensive software solution for mining exome- and transcriptome- sequencing data. *BMC Bioinformatics* 2011;**12**:267.
212. Ossowski S, Schneeberger K, Clark RM, et al. Sequencing of natural strains of Arabidopsis thaliana with short reads. *Genome Res.* 2008;**18**:2024–2033.
213. Fischer M, Snajder R, Pabinger S, et al. SIMPLEX: Cloud-Enabled Pipeline for the Comprehensive Analysis of Exome Sequencing Data. *PLoS ONE* 2012;**7**:e41948.
214. Asmann YW, Middha S, Hossain A, et al. TREAT: a bioinformatics tool for variant annotations and visualizations in targeted and exome sequencing data. *Bioinformatics* 2012;**28**:277–278.
215. Orvis J, Crabtree J, Galens K, et al. Ergatis: a web interface and scalable software system for bioinformatics workflows. *Bioinformatics* 2010;**26**:1488–1492.
216. Goecks J, Nekrutenko A, Taylor J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 2010;**11**:R86.
217. Genboree. www.genboree.org/.
218. Reich M, Liefeld T, Gould J, et al. GenePattern 2.0. *Nat. Genet.* 2006;**38**:500–501.
219. Halbritter F, Vaidya HJ, Tomlinson SR. GeneProf: analysis of high-throughput sequencing experiments. *Nat. Methods* 2012;**9**:7–8.
220. Kepler. https://kepler-project.org/.
221. Jagla B, Wiswedel B, Coppée J-Y. Extending KNIME for next-generation sequencing data analysis. *Bioinformatics* 2011;**27**:2907–2909.
222. Rex DE, Ma JQ, Toga AW. The LONI Pipeline Processing Environment. *Neuroimage* 2003;**19**:1033–1048.
223. Moa. http://mfiers.github.com/Moa/.
224. Abouelhoda M, Issa SA, Ghanem M. Tavaxy: Integrating Taverna and Galaxy workflows with cloud computing support. *BMC bioinformatics* 2012;**13**:77.
225. Hull D, Wolstencroft K, Stevens R, et al. Taverna: a tool for building and running workflows of services. *Nucleic Acids Res.* 2006;**34**:W729–732.
226. Hunter AA, Macgregor AB, Szabo TO, et al. Yabi: An online research environment for grid, high performance and cloud computing. *Source Code Biol Med* 2012;**7**:1.