

NextGENe™

Next Generation Sequencing Software

Quick Start Guide

...

Select Application

1. Open NextGENe Run Wizard
2. Select Instrument Type used to generate data
3. Select Application Type

Steps not necessary for specific Instrument/Application combinations will be grayed out.

Enabled Steps with checkmark will be performed by NextGENe.

4. Click Next

The screenshot shows the 'Project Wizard - Application Type' dialog box. It features a vertical sidebar on the left with buttons for 'Application', 'Load Data', 'Condensation', 'Assembly', 'Alignment', and 'Post Processing'. The main area is divided into three sections: 'Instrument type' with radio buttons for Roche/454, Illumina (selected), SOLiD, and Ion Torrent; 'Application type' with radio buttons for de novo Assembly, SNP/Indel discovery (selected), Transcriptome, ChIP-Seq, SAGE, STR analysis, CNV-Seq, HLA, and Other, plus checkboxes for Alternative splicing and Mitochondrial amplicon; and 'Steps' with checkboxes for Sequence condensation, Sequence assembly, and Sequence alignment (checked). A 'Performance settings' section at the bottom has a text box for 'Number of cores to be used' with the value '3' and '(1~4)' next to it. At the bottom of the dialog are buttons for 'Save Settings', 'Load Settings', 'Next >>', 'Cancel', and 'Finish'. A 'Show Project Log>>' button is in the top right corner.

Load Data

1. Load FASTA or BAM format files into the Sample Files field

If sample is not in FASTA or BAM format, or if low quality calls need to be removed, open Format Conversion Tool (described on next page).

2. Load Reference Files
 1. Load (fasta or GBK format)
 2. Preloaded (large indexed)
3. Select Save Location in Output field
4. Click Next

Project Wizard - Load Data

Show Project Log >>

Step

Application

Load Data

Condensation

Assembly

Alignment

Post Processing

Load data

Previous run result: Load Previous Run Result

To convert to fasta: Format Conversion

Sample files:

F:\Data\Demo Data\RainDance Cancer Panel Data\ffpe_lung_9_Output\Pre
F:\Data\Demo Data\RainDance Cancer Panel Data\ffpe_lung_9_Output\Pre

Load

Remove

RemoveAll

Reference files:

Human_v37p10_dbsnp135

Load

Preloaded

Remove

RemoveAll

Output:

F:\Data\Demo Data\RainDance Cancer Panel Data\ffpe_lung_9_Output\Pre

Set

Available disk space: 2603 GB free of 4000 GB.

Set Amplicon BED file Set ROI regions from GBK files

Set

<< Back Next >> Cancel Finish

Load Data – Format Conversion

- Instruments produce several output formats.
- NextGENe input is FASTA.
- Use Format Conversion Tool to convert instrument output to FASTA.

1. Select the Instrument

2. Add sample files to be converted to the Input field

3. Select Output Format and set Output location

4. Select Settings to Clean-up sample file

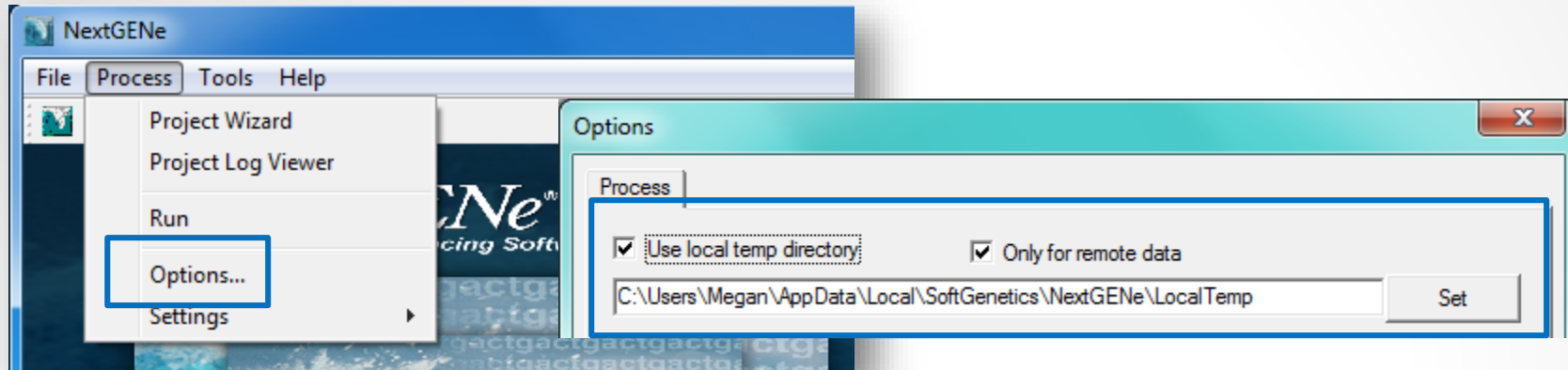
5. Click OK

The screenshot shows the 'Format Conversion' window with the following sections highlighted:

- Instrument:** A red box highlights the 'Instrument Type' dropdown menu, which is currently set to 'Illumina'.
- Input:** A yellow box highlights the 'Sample files' list area, which is currently empty, and the 'Input format type' dropdown menu, which is set to 'FASTQ'. Buttons for 'Add', 'Remove', and 'Remove All' are visible to the right of the list.
- Output:** A green box highlights the 'Output format type' dropdown menu, which is set to 'FASTA', and the 'Output' text field.
- Settings:** A blue box highlights the 'Settings' section, which includes several checked options: 'Median score threshold >= 20', 'Max # of uncalled bases <= 3', 'Called base number of each read >= 25', and 'Trim or reject read when >= 3 base(s) with score <= 16'. Other options like 'Paired reads data', 'Remove 5' 2 base(s) and 3' 4 base(s)', 'Keep only bases 0 to 0', 'Trim by sequences', 'Trim by sequences in file', and 'Custom linker' are unchecked.

At the bottom of the window, there are buttons for 'Save', 'Load', 'Add Job', 'Remove Job', 'OK', and 'Cancel'.

Load Data – Analyze Locally to increase speed



When any of the input or output fields contain files that are or will be on a network drive, performance would be improved by processing the project locally.

Within the NextGENe Process Options, choose to use a local temp directory for remote data and specify a local directory with adequate resources.

Alignment

1. Set Alignment Specificity – Default value work well for most datasets. Click on the Inspect input files to have NextGENe adjust automatically, or manually adjust.
2. Set Mutation Detection Sensitivity – mutation percentage and total depth of coverage for each base position.
3. For sample files with only single end reads, keep the Load paired reads deselected. For paired sample files, select this option and adjust range accordingly.
4. Click Next.

Project Wizard - Alignment Show Project Log >>

Alignment

Step: Application, Load Data, Condensation, Assembly, Alignment, Post Processing

Reads: Allowable mismatched bases (0-2)
 Allowable ambiguous alignments
 Seeds: bases, move step bases
 Allowable alignments (1-1000)
 Overall: Matching base percentage >= Detect large indels

Sample trim
 Select sequence range
 from bases to bases
 Hide unmatched ends

Mutation filter Use original Except for homozygous

	SNPs	Indels	HomopolymerIndels
Mutation percentage <=	<input type="text" value="20"/>	<input type="text" value="20"/>	<input type="text" value="20"/>
SNP allele count <=	<input type="text" value="3"/>	<input type="text" value="3"/>	<input type="text" value="3"/>
Total coverage count <=	<input type="text" value="5"/>	<input type="text" value="5"/>	<input type="text" value="5"/>

Perform in-read phasing
 Max gap between two variants (0-3)
 Phaseable reads percentage >=

File type
 Load assembled result files
 Load paired reads
 Library size range : from bases to bases
 454 Sequence:

Save matched reads Highlight anchor sequence Ambiguous gain/loss
 Detect structural variations Mismatch: length and bases

Post-Processing Reports

- Set the reports that you want to be automatically generated in each project.

1. Choose the report.

2. Choose a configuration file that you previously saved from the NextGENe Viewer report's Settings dialog.

3. Click Finish to Run NextGENe.

Project Wizard - Post Processing

Show Project Log >>

Step

Application

Load Data

Condensation

Assembly

Alignment

Post Processing

Post processing Report

Report	Settings	Set	Remove
Mutation	MR_General.ini	Set	Remove
Expression	ExpressionReport.ini	Set	Remove
Coverage Curve	CC_100x.ini	Set	Remove

Save summary report

Add Remove All

Export Settings

Export BAM

Output to Geneticist Assistant Details

Run Name Sample Name

Add Remove All

Save Settings Load Settings << Back Cancel Finish

Project Analysis – Run Log

1. Statistics are calculated

2. Log denotes project analysis completion

The screenshot displays the SOFTGENETICS® Software PowerTools for Genetics Analysis interface. The top header features the company logo and name. The main content area is divided into two sections. The upper section, highlighted with a red box, contains alignment statistics: Matched Reads Count: 248806, Unmatched Reads Count: 138368, Short Reads Count: 0, Number of Matched Bases: 19856355, Number of Unmatched Bases That are Recorded as Mutations: Mismatches: 4701, Deletions: 9417, Insertions: 35, and Number of Unmatched Bases That are NOT Recorded as Mutations: Mismatches: 30402, Deletions: 35812, Insertions: 31241. The lower section, highlighted with a yellow box, shows the completion message: [Monday, August 20, 2012, 08:54:23] Processing Complete.

SOFTGENETICS®
Software PowerTools for Genetics Analysis

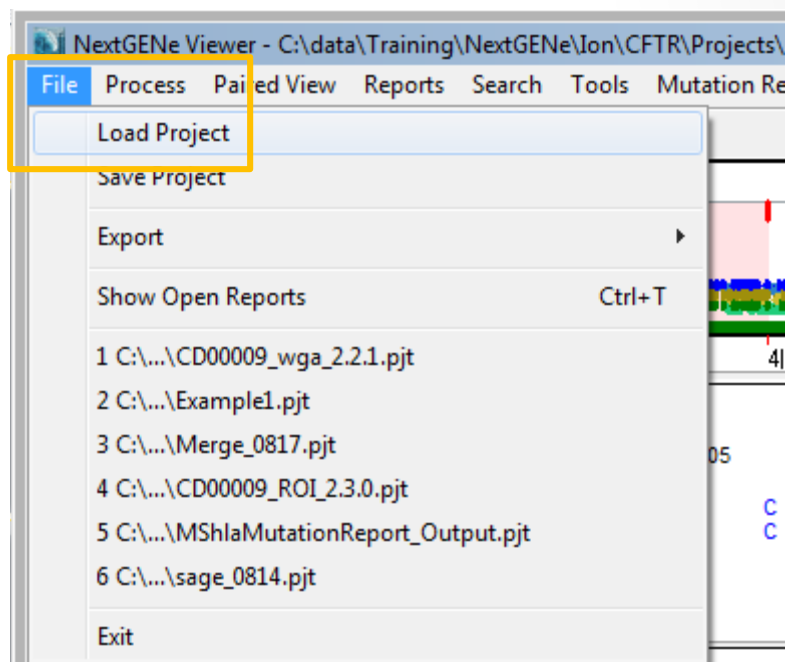
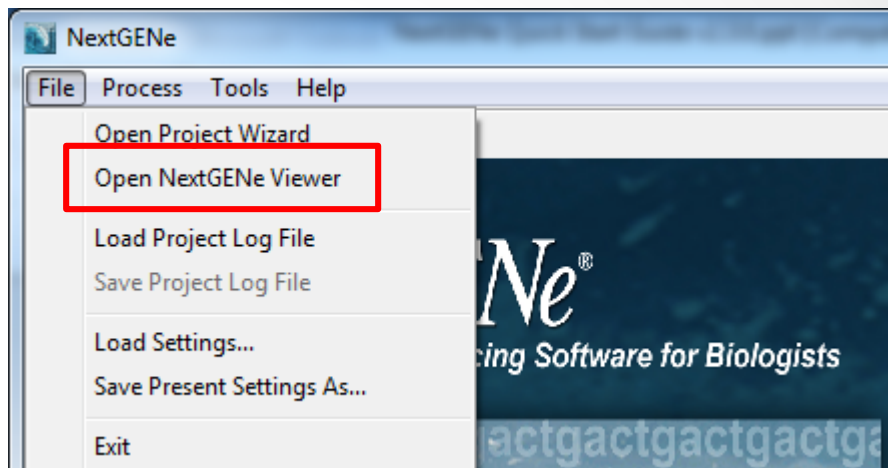
[Alignment Statistics Information]
Matched Reads Count: 248806
Unmatched Reads Count: 138368
Short Reads Count: 0
Number of Matched Bases: 19856355
Number of Unmatched Bases That are Recorded as Mutations:
Mismatches: 4701
Deletions: 9417
Insertions: 35
Number of Unmatched Bases That are NOT Recorded as Mutations:
Mismatches: 30402
Deletions: 35812
Insertions: 31241
Average Read Length: 80
Average Coverage: 856
Reference Length: 188703
Number of Covered Bases: 23273

[Monday, August 20, 2012, 08:54:23] Processing Complete.

View Projects in NextGENe Viewer

Projects can be reviewed in the NextGENe Viewer Application.

1. Click on Open NextGENe Viewer from NextGENe toolbar.
2. Click on File menu from the NextGENe Viewer toolbar and select Load Project.



NextGENe Viewer

NextGENe Viewer/Browser

GLOBAL VIEW

LOCAL VIEW

Report hyperlinked to alignment view and dbSNP database

Index	Reference Position	Chromosome	Gene	CDS	Chr	Reference Nucleotide	Coverage	A (%)	C (%)	G (%)	T (%)	Ins (%)	Del (%)	SNP db_vref	Genotype	Mutation Call	AminoAcid Change
15	98706860	1980874	PRK CZ	4	1	A	650	71.54	0.00	28.46	0.00	0.00	0.00	rs1878745	AG	c.318A>AG	106P>FP
15	99162928	2436942	PANK4	10	1	T	13	30.77	0.00	69.23	0.00	0.00	0.00		GA	c.1293G>AG	431Y>YY
16	99168415	2442429	PANK4	3	1	G	33	0.00	90.91	0.00	9.09	0.00	0.00	rs2985862	CC	c.393T>C	131K>K

NextGENe Viewer – Mutation Report

Mutation Report Settings

Display Filter Summary Report Output

Annotation Statistics Phasing

Basic

Chrom Pos Contig Contig Pos Ref Pos

Ref Alt Genotype Zygosity

Gene

Gene Trans. Type Trans. Accession Strand

Exon CDS Gene Ref Gene Alt

Protein

Protein Accession

Amino Acid Change

Function

Mutation Call

Genomic

Relative to CDS

Relative to mRNA

HGVS Genomic

HGVS Coding

HGVS Protein

Forensic (SWGDAM mtDNA)

Reference Database ID

dbSNP

Miscellaneous

Show all transcripts

Comments

Settings OK Cancel

Mutation Report Settings

Display Filter Summary Report Output

Annotation Bias Score ROI

Function

Coding Variants

Synonymous substitutions

Missense substitutions

Nonsense substitutions

No-stop substitutions

Indels

Noncoding Variants

Splicing Sites

Intron Exon

All other substitutions

All other Indels

Zygosity

Homozygous

Heterozygous

Limit Regions

Gene Before After

Exon Before After

CDS Before After

Source

Called by software

Added manually

Confirmed

Deleted

Negative

Reference Database

dbSNP

Reported

Settings

Hyperlink to NCBI dbSNP database

Index	Chrom	Pos	Coverage	Ref	Alt	Alt%	Mutation Call: Relative To CDS	Function	Gene	Trans Accession	CDS	dbSNP	Amino Acid Change
14	13	32915005	1564	G	C	97.89	c.6513G>CG	Synonymous	BRCA2	NM_000059.3	10	rs206076	p.V2171W
15	13	32929007	2296	G	C	1.96	c.7017G>CG	Missense	BRCA2	NM_000059.3	13	rs45574331	p.K2339NK
16	13	32929232	819	A	G	14.29	c.7242A>AG	Synonymous	BRCA2	NM_000059.3	13	rs1799955	p.S2414SS
17	13	32929331	1410	T	A	23.76	c.7341T>AT	Missense	BRCA2	NM_000059.3	13	rs4986858	p.N2447KN
18	13	32929387	1962	T	C	97.66	c.7397T>CT	Missense	BRCA2	NM_000059.3	13	rs169547	p.V2466AV
19	13	32971071	1551	C	T	2.13	c.9538C>CT	Missense	BRCA2	NM_000059.3	25		p.L3180LF
20	13	32972626	1524	A	T	1.64	c.9976A>AT	Nonsense	BRCA2	NM_000059.3	26	rs11571833	p.K3326KX
21	13	32972884	896	A	G	2.90	c.10234A>AG	Missense	BRCA2	NM_000059.3	26	rs1801426	p.I3412IV

NextGENe Viewer – Variant Comparison Tool

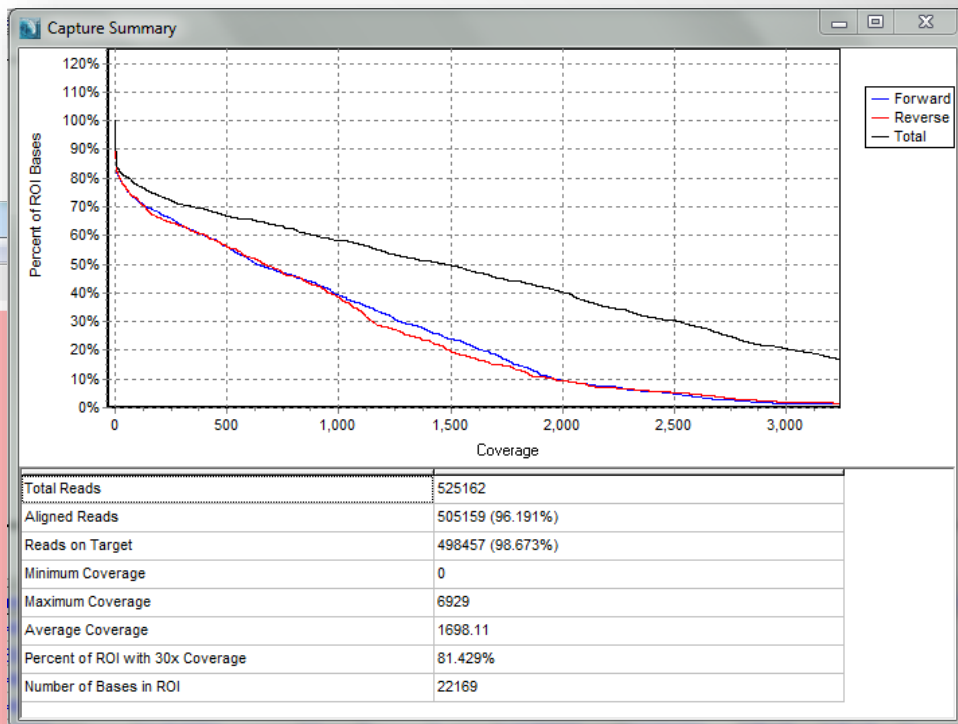
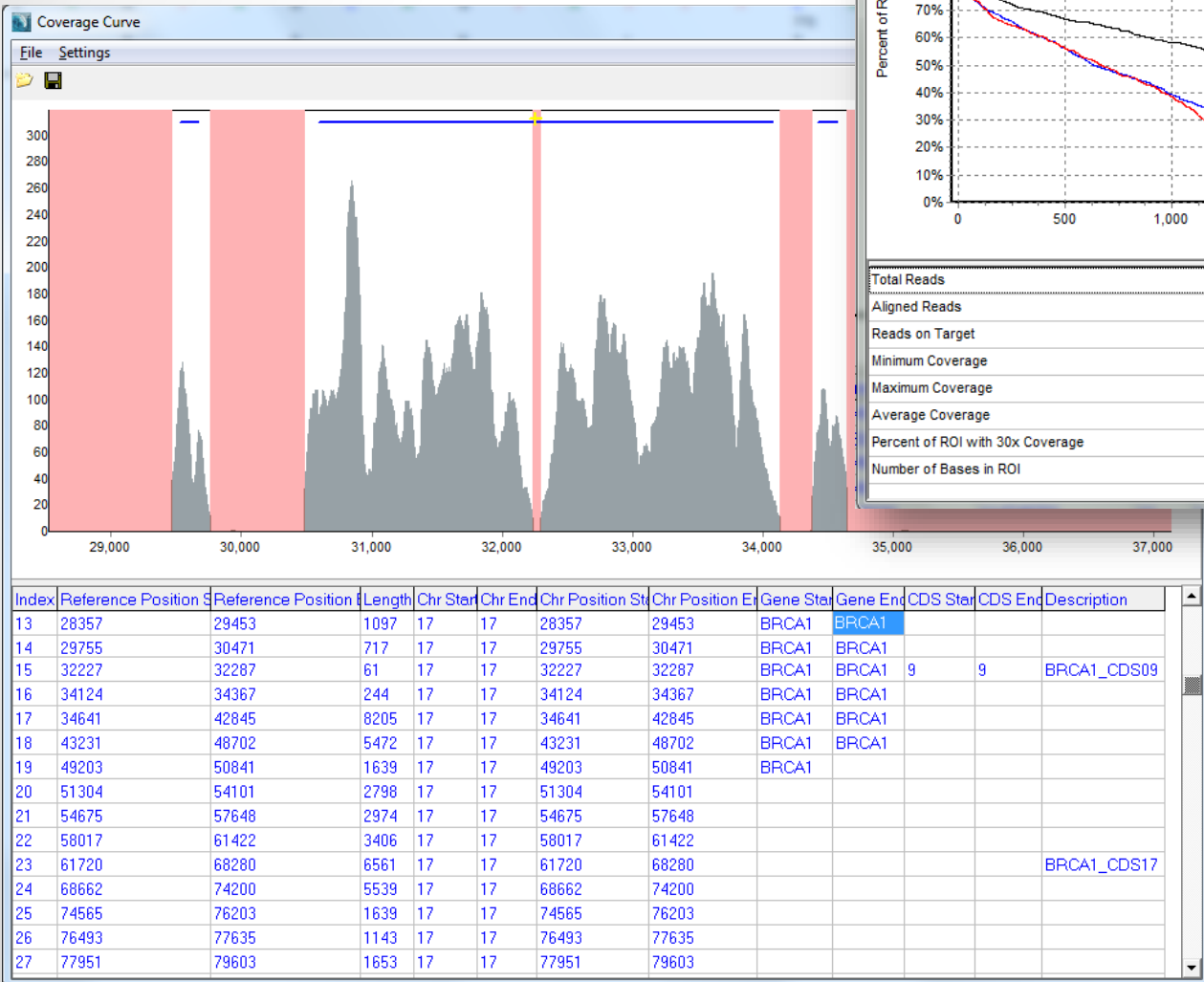
Compare mutations detected in two or more samples aligned to the same reference.

Use the Rare Disease tool to filter by relationship and phenotype.



NextGENe Viewer – Coverage Curve

Identify regions below a user-defined threshold of coverage



NextGENe Viewer – Expression Report

Expression Report Settings

General | Display | Filters | Summary Report

Regions

Use segments as defined in reference files

Contig ROI Amplicon

Gene mRNA CDS

Continuous mRNA Continuous CDS

Set incremental segment length bases

Contig Chromosome

Input region of interest (*.bed)

Limits

Limit to first bp

Limit to last bp

Expression Report Settings

General | Display | Filters | Summary Report

Include columns

Index Description Min Coverage Max Coverage

Chr Contig Average Coverage

Name Locus Tag Minimum Forward Read Coverage

Number Start Minimum Reverse Read Coverage

Chr Position Start End Read Counts

Chr Position End Reference Length Forward Read Counts

Chr Length Fragment Counts

Gene RPKM RPK

CDS FPKM

RNA Accession Original Max Coverage

Protein Accession Original Average Coverage

Original Read Counts

Index	Chr	Chr Position Start	Chr Position End	Gene	CDS	RNA Accession	Min Coverage	Minimum Forward R	Average Coverage	Read Counts	RPKM
1	chr13	32890598	32890664	BRCA2; +	1	NM_000059.3	403	13	805.10	1073	38469.2122
2	chr13	32893214	32893462	BRCA2; +	2	NM_000059.3	609	230	1824.30	6590	63573.2539
3	chr13	32899213	32899321	BRCA2; +	3	NM_000059.3	417	28	546.15	903	19899.8419
4	chr13	32900238	32900287	BRCA2; +	4	NM_000059.3	514	456	605.80	532	25558.1845
5	chr13	32900379	32900419	BRCA2; +	5	NM_000059.3	342	52	453.39	256	14998.3845
6	chr13	32900636	32900750	BRCA2; +	6	NM_000059.3	975	582	1715.73	3355	70078.2192
7	chr13	32903580	32903629	BRCA2; +	7	NM_000059.3	705	479	1147.82	1144	54959.7050
8	chr13	32905056	32905167	BRCA2; +	8	NM_000059.3	340	17	827.46	1368	29339.7526