User Guide for Segtor

February 28, 2011

Contents

Chapter 1

Introduction

Segtor is a software written in Perl to:

1. Determine the relative position of genomic coordinates and intervals with respect to genes
2. Annotate single nucleotide variations with respect to genes
3. Determine the closest transcription start site to various coordinates
4. Determine the relative position of insertion, deletions and translocations

Our software is designed for rapidly handling large datasets. We also designed a web interface for small queries for test purposes. See the quick start guide for installation instructions.

Chapter 2

Command line

2.1 Modes

There are 5 available modes:

1. **Coordinates:** To determine the relative position of genomic coordinates to genes (e.g. in an exon, upstream of a gene).
2. **Intervals:** To determine the relative position of intervals to genes (e.g. overlapping an intron, overlapping the 5' end of a gene).
3. **SNVs:** To determine the relative position of SNVs to genes and the impact of SNVs landing in exons. (e.g. within an intron, within a protein coding exon causing a non-synonymous mutation.
4. **TSS:** To determine the closest transcription start site of a gene for a genomic coordinate indepedently of the relative position.
5. **Insertions/deletions/translotions:** To determine the relative position of insertions/deletions/translocation to genes and give a putative protein sequence for each site landing within coding exons.

The following sections describe each mode.

2.1.1 Mode 1 :Coordinates

Input format

The input format is:

```
chromosome coordinate id
```

Example:

```
chr1 12393 coord1
```

The **chromosome** must be the same as the one used on the UCSC Genome Browser, the **coordinate** must be the same as the one used on the UCSC Genome Browser (1-based) and the **id** must be unique for each line.

Relevance of hits

The program will sort the results according to the relevance of the hits, here is the order used:

1. Within exons
2. Within introns
3. Upstream
4. Downstream

Output format

Results:

Field	Meaning
#INPUTID	The unique identifier provided for the input record
CHR	The chromosome for the input record
COORD	The coordinate for the input record
TRANSCRIPT	The transcripts that were found for each input
GENE	The gene for the transcript that was found
POSITION	The relative position of each input with respect to the genes/transcripts
INDEX	Index of the exons/introns when a coordinate falls within an exon or intron
DISTANCE5P	Distance to the 5' end
DISTANCE3p	Distance to the 3' end
PARTIAL	When the coordinates is within range of the 5' or 3' end and the transcript only aligns partially to the genome

Output files and meaning

File name	Information reported with respect to:	Which inputs will reported ?	Which genes will reported ?
[range].input.all.out	Inputs	all inputs	all the transcripts, all the genes
[range].input.single.out	Inputs	all inputs	Pick the most relevant transcript per gene but report all the genes
[range].input.best.out	Inputs	all inputs	Pick the most relevant transcript for all genes
[range].genes.all.out	Genes	all inputs	all the transcripts, all the genes
[range].genes.single.out	Genes	all inputs	Pick the most relevant transcript per gene but report all the genes
[range].genes.best.out	Genes	all inputs	Pick the most relevant transcript for all genes
[range].input.none.out	inputs	only those without a hit to a gene	no genes will be reported
[range].input.exon.out	inputs	only those in exon	pick one transcript per gene but report all the different genes
[range].input.intron.out	inputs	only those in introns	pick one transcript per gene but report all the different genes
[range].input.upstrm.out	inputs	only those upstream of a gene	pick one transcript per gene but report all the different genes
[range].input.dwnstrm.out	inputs	only those downstream of a gene	pick one transcript per gene but report all the different genes

2.1.2 Mode 2: Intervals

Input format

The input format is:

```
chromosome coordinate1 coordinate2 id
```

Example:

```
chr9 230894 231123 interval1
```

The **chromosome** must be the same as the one used on the UCSC Genome Browser, the **coordinate1** must be the same as the one used on the UCSC Genome Browser (1-based) and represent the start of the interval, **coordinate2** is the end coordinate which should be greater than **coordinate1** and the **id** must be unique for each line.

Relevance of hits

The program will sort the results according to the relevance of the hits, here is the order used:

1. Contained in transcript and totally contained in a single exon
2. Contained in transcript and overlaps exonic and intronic regions

3. Contained in transcript and totally contained in intronic region
4. Range completely contains the transcript
5. Overlaps 5' end and first exon only
6. Overlaps 5' end and exons, introns
7. Overlaps 3' end and last exon only
8. Overlaps 3' end and exons, introns
9. Upstream
10. Downstream

Output format

Results:

Field	Meaning
#INPUTID	The unique identifier provided for the input record
CHR	The chromosome for the input record
COORD	The coordinate for the input record
TRANSCRIPT	The transcripts that were found for each input
GENE	The gene for the transcript that was found
POSITION	The relative position of each input with respect to the genes/transcripts
INDEXexons	The index of the exons that are overlapped by the interval
INDEXintrons	The index of the introns that are overlapped by the interval
PARTIAL	When the interval is within range of the 5' or 3' end and the transcript only aligns partially to the genome

Output files and meaning

File name	Information reported with respect to:	Which inputs will reported ?	Which genes will reported ?
[range].input.all.out	Inputs	all inputs	all the transcripts, all the genes
[range].input.single.out	Inputs	all inputs	Pick the most relevant transcript per gene but report all the genes
[range].input.best.out	Inputs	all inputs	Pick the most relevant transcript for all genes
[range].genes.all.out	Genes	all inputs	all the transcripts, all the genes
[range].genes.single.out	Genes	all inputs	Pick the most relevant transcript per gene but report all the genes
[range].genes.best.out	Genes	all inputs	Pick the most relevant transcript for all genes
[range].input.none.out	inputs	only those without a hit to a gene	no genes will be reported
[range].input.exon.out	inputs	only those in exon	pick one transcript per gene but report all the different genes
[range].input.intron.out	inputs	only those in introns	pick one transcript per gene but report all the different genes
[range].input.upstrm.out	inputs	only those upstream of a gene	pick one transcript per gene but report all the different genes
[range].input.dwnstrm.out	inputs	only those downstream of a gene	pick one transcript per gene but report all the different genes

2.1.3 Mode 3: SNVs/SNPs

Input format

The input format is:

```
chromosome coordinate1 bpRef bpRead id
```

Example:

```
chrX 23094 A C snv1
```

The **chromosome** must be the same as the one used on the UCSC Genome Browser, the **coordinate** must be the same as the one used on the UCSC Genome Browser (1-based), **bpRef** is the reference base pair must either be A,C,G,T, **bpRead** is the read base pair must either be A,C,G,T and the **id** must be unique for each line.

Relevance of hits

The program will sort the results according to the relevance of the hits, here is the order used:

1. SNVs falling within the coding region of exons for which the reading frames can be ascertained causing a change in amino acid (non-synonymous)

2. SNVs falling within the coding region of exons for which the reading frames can be ascertained not causing a change in amino acid (synonymous)
3. SNVs falling within the coding region of exons for which the reading frames cannot be ascertained
4. SNVs falling within the non-coding region of exons
5. SNVs falling within introns
6. SNVs falling upstream
7. SNVs falling downstream

Output format

Results:

Field	Meaning
#INPUTID	The unique identifier provided for the input record
CHR	The chromosome for the input record
COORD	The coordinate for the input record
TRANSCRIPT	The transcripts that were found for each input
GENE	The gene for the transcript that was found
POSITION	The relative position of each input with respect to the genes/transcripts
INDEX	Index of the exons/introns when a SNV falls within an exon or intron
DISTANCE5p	Distance to the 5' end
DISTANCE3p	Distance to the 3' end
PARTIAL	When the SNV is within range of the 5' or 3' end and the transcript only aligns partially to the genome, it can also contain "potential splice disruption" for SNVs landing in the first or last 2 bp of an intron, or "untranslated 5'UTR", "untranslated 3'UTR" or "untranslated" for those SNVs landing in the untranslated parts of an exon
CODONREF	DNA codon with the reference base pair.
CODONREAD	DNA codon with the read base pair.
AAREF	Amino acid produced by the DNA codon with the reference base pair.
AAREAD	Amino acid produced by the DNA codon with the read base pair.

Output files and meaning

File name	Information reported with respect to:	Which inputs will be reported ?	Which genes will be reported ?
[range].input.all.out	Inputs	all inputs	all the transcripts, all the genes
[range].input.single.out	Inputs	all inputs	Pick the most relevant transcript per gene but report all the genes
[range].input.best.out	Inputs	all inputs	Pick the most relevant transcript for all genes
[range].genes.all.out	Genes	all inputs	all the transcripts, all the genes
[range].genes.single.out	Genes	all inputs	Pick the most relevant transcript per gene but report all the genes
[range].genes.best.out	Genes	all inputs	Pick the most relevant transcript for all genes
[range].input.none.out	inputs	only those without a hit to a gene	no genes will be reported
[range].input.cde.out	inputs	only those in coding exons	pick one transcript per gene but report all the different genes
[range].input.ncde.out	inputs	only those in non-coding exons	pick one transcript per gene but report all the different genes
[range].input.syn.out	inputs	only those causing synonymous mutations	pick one transcript per gene but report all the different genes
[range].input.nsyn.out	inputs	only those causing non-synonymous mutations	pick one transcript per gene but report all the different genes
[range].input.exon.out	inputs	only those in exon	pick one transcript per gene but report all the different genes
[range].input.intron.out	inputs	only those in introns	pick one transcript per gene but report all the different genes
[range].input.upstrm.out	inputs	only those upstream of a gene	pick one transcript per gene but report all the different genes
[range].input.dwnstrm.out	inputs	only those downstream of a gene	pick one transcript per gene but report all the different genes
[range].input.dbsnp.out	inputs	all inputs	Only information about dbSNP records matching input records (optional file)

2.1.4 Mode 4: TSS

Input format

The input format is:

```
chromosome coordinate id
```

Example:

```
chr19 903423 test1
```

The **chromosome** must be the same as the one used on the UCSC Genome Browser, the **coordinate** must be the same as the one used on the UCSC Genome Browser (1-based) and the **id** must be unique for each line.

Output format

Results:

Field	Meaning
#INPUTID	The unique identifier provided for the input record
CHR	The chromosome for the input record
COORD	The coordinate for the input record
TRANSCRIPT	The transcripts whose TSS was the closest to the input coordinate were found for each input
GENE	The gene for the transcript that was found
POSITION	The relative position of each input with respect to the genes/transcripts
DISTANCE5P	Distance to the 5' end
DISTANCE3p	Distance to the 3' end
PARTIAL	When the coordinates is within range of the 5' or 3' end and the transcript only aligns partially to the genome

Output files and meaning

File name	Information reported with respect to:	Which inputs will be reported ?	Which genes will be reported ?
input.closest.out	inputs	All inputs for which a transcription start site was found	The gene corresponding to the transcription start site
input.none.out	inputs	All inputs lying on chromosomes for which there are no genes to report	no genes will be reported

2.1.5 Mode 5: Insertions, deletions and translocations

Input format

The input format can either be:

```
INS chromosome coordinate DNAsequence id
DEL chromosome coordinate1 coordinate2 id
TRANS  chromosome1 coordinate1 strand1 chromosome2 coordinate2 strand2 id
```

Example:

```
INS chr7 902344 CAGT ins1
DEL chr10 230984 230996 del1
TRANS  chr9 324023 + chr4 23135 -trans1
```

The **chromosome** must be the same as the one used on the UCSC Genome Browser, **coordinate**, **coordinate1** and **coordinate2** must be the same as the one used on the UCSC Genome Browser (1-based) and the **id** must be unique for each line. **strand1** and **strand2** are the strands indicating the which strand was paired the translocation.

Output format

The 3 types of inputs are mixed together in the output. However, they do not always have the same number of columns.

Insertions:

Field	Meaning
1	The unique identifier provided for the input record
2	The chromosome for the input record
3	The coordinate for the input record
4	The transcripts whose TSS was the closest to the input coordinate were found for each input
5	The gene for the transcript that was found
6	The relative position of each input with respect to the genes/transcripts
7	Distance to the 5' end
8	Distance to the 3' end
9	When the coordinates is within range of the 5' or 3' end and the transcript only aligns partially to the genome
10	Coordinate of the insertion in the amino acid sequence
11	The sequence without the insertion
12	A putative sequence with the insertion (does not take splicing and other events into account)

Deletions:

Field	Meaning
1	The unique identifier provided for the input record
2	The chromosome for the input record
3	The start coordinate for the input record
4	The end coordinate for the input record
5	The transcripts whose TSS was the closest to the input coordinate were found for each input
6	The gene for the transcript that was found
7	The relative position of each input with respect to the genes/transcripts
8	The index of the exons that are overlapped by the interval
9	The index of the introns that are overlapped by the interval
10	When the interval is within range of the 5' or 3' end and the transcript only aligns partially to the genome
11	The sequence without the deletion
12	A putative sequence with the deletion (does not take splicing and other events into account)

Translocations

Field	Meaning
1	The unique identifier provided for the input record
2	The chromosome for the first input record
3	The coordinate for the first input record
4	The chromosome for the second input record
5	The coordinate for the second input record
6	The transcript found for the first input record (**DD** = different direction, **SD**=same direction)
7	The gene found for the first input record
8	The beginning of the protein sequence due to the first input record
9	The transcript found for the second input record (**DD** = different direction, **SD**=same direction)
10	The gene found for the second input record
11	The beginning of the protein sequence due to the second input record

Output files and meaning

File name	Information reported with respect to:	Which inputs will be reported ?	Which genes will be reported ?
[range].input.all.out	Inputs	all inputs	all the transcripts, all the genes
[range].genes.all.out	Genes	all inputs	all the transcripts, all the genes
[range].input.none.out	inputs	only those without a hit to a gene	no genes will be reported

Chapter 3

Web Interface

3.1 Modes

3.1.1 Coordinate (coord)

This mode is to annotate genomic coordinates

Input format

The input format is:

```
chromosome coordinate id
```

Example:

```
chr1 12393 coord1
```

The **chromosome** must be the same as the one used on the UCSC Genome Browser, the **coordinate** must be the same as the one used on the UCSC Genome Browser (1-based) and the **id** must be unique for each line.

Relevance of hits

The program will sort the results according to the relevance of the hits, here is the order used:

1. Within exons
2. Within introns
3. Upstream
4. Downstream

Output

Results:

Field	Meaning
Transcript	The genes/transcripts that were found for each input
Position	The relative position of each input with respect to the genes/transcripts
Index	Index of the exons/introns when a coordinate falls within an exon or intron
Distance5p	Distance to the 5' end
Distance3p	Distance to the 3' end
Partial	When the coordinates is within range of the 5' or 3' end and the transcript only aligns partially to the genome

Statistics:

Field	Meaning
Genome Used:	The genome/assembly that was used
Database Used:	The database that was used
Range Used:	The range that was used
Database file:	The database files that was in the construction of the segment tree and the data at which they were downloaded
Total number of sites:	Total number of coordinates in the input that were entered by the user
Sites upstream of genes:	Number of coordinates that are located upstream of their respective most relevant transcript
Sites downstream of genes:	Number of coordinates that are located downstream of their respective most relevant transcript
Sites within exons of genes:	Number of coordinates that are located within an exon of their respective most relevant transcript
Sites within introns of genes:	Number of coordinates that are located within an intron of their respective most relevant transcript
Exons:	The index of the exons for the "Sites within exons of genes"
Introns:	The index of the introns for the "Sites within introns of genes"
Pie Chart	A pie chart showing distribution of the coordinates

Options

Available option	Effect
Range	This will consider genes within the range specified by the user (default = 0 bp)

3.1.2 Intervals (interval)

This mode is to annotate genomic intervals

Input format

The input format is:

```
chromosome coordinate1 coordinate2 id
```

Example:

```
chr9 230894 231123 interval1
```

The **chromosome** must be the same as the one used on the UCSC Genome Browser, the **coordinate1** must be the same as the one used on the UCSC Genome Browser (1-based) and represent the start of the interval, **coordinate2** is the end coordinate which should be greater than **coordinate1** and the **id** must be unique for each line.

Relevance of hits

The program will sort the results according to the relevance of the hits, here is the order used:

1. Contained in transcript and totally contained in a single exon
2. Contained in transcript and overlaps exonic and intronic regions
3. Contained in transcript and totally contained in intronic region
4. Range completely contains the transcript
5. Overlaps 5' end and first exon only
6. Overlaps 5' end and exons, introns
7. Overlaps 3' end and last exon only
8. Overlaps 3' end and exons, introns
9. Upstream
10. Downstream

Output

Results:

Field	Meaning
Partial	When the coordinates is within range of the 5' or 3' end and the transcript only aligns partially to the genome
Exons#	The index of the exons that are overlapped by the interval
Introns#	The index of the introns that are overlapped by the interval
Position	The relative position of each input with respect to the genes/transcripts

Statistics:

Field	Meaning
Genome Used:	The genome/assembly that was used
Database Used:	The database that was used
Range Used:	The range that was used
Database file:	The database files that was in the construction of the segment tree and the data at which they were downloaded
Total number of sites:	Total number of coordinates in the input that were entered by the user
Intervals upstream :	Number of intervals that are located upstream of their respective most relevant transcript
Intervals downstream:	Number of intervals that are located downstream of their respective most relevant transcript
Intervals in single exon:	Number of intervals that are contained within a single exon
Intervals in exons/introns:	Number of intervals spanning exonic and intronic regions
Intervals in single intron:	Number of intervals that are contained within a single intron
Intervals spanning a gene:	Number of intervals that totally span a gene
Intervals 5' end, first exon	Number of intervals that span the 5' end and the first exon only
Intervals 5' end, exons/introns	Number of intervals that span the 5' end and exonic and intronic regions
Intervals 3' end, last exon	Number of intervals that span the 3' end and the last exon only
Intervals 3' end, exons/introns	Number of intervals that span the 3' end and exonic and intronic regions
Exons:	The index of the exons for the "Sites within exons of genes"
Introns:	The index of the introns for the "Sites within introns of genes"
Pie Chart	A pie chart showing distribution of the coordinates

Options

Available option	Effect
Range	This will consider genes within the range specified by the user (default = 0 bp)

3.1.3 Single nucleotide variation (or polymorphism) (SNP)

This mode is to annotate genomic SNVs (or SNPs).

Input format

The input format is:

```
chromosome coordinate1 bpRef bpRead id
```

Example:

```
chrX 23094 A C snv1
```

The **chromosome** must be the same as the one used on the UCSC Genome Browser, the **coordinate** must be the same as the one used on the UCSC Genome Browser (1-based), **bpRef** is the reference base pair must either be A,C,G,T, **bpRead** is the read base pair must either be A,C,G,T and the **id** must be unique for each line.

Relevance of hits

The program will sort the results according to the relevance of the hits, here is the order used:

1. SNVs falling within the coding region of exons for which the reading frames can be ascertained causing a change in amino acid (non-synonymous)
2. SNVs falling within the coding region of exons for which the reading frames can be ascertained not causing a change in amino acid (synonymous)
3. SNVs falling within the coding region of exons for which the reading frames cannot be ascertained
4. SNVs falling within the non-coding region of exons
5. SNVs falling within introns
6. SNVs falling upstream
7. SNVs falling downstream

Output

In the online interface, non-synonymous mutations are highlighted in red.
Results:

Field	Meaning
Transcript	The genes/transcripts that were found for each input
Position	The relative position of each input with respect to the genes/transcripts
Index	Index of the exons/introns when a SNP falls within an exon or intron
Distance5p	Distance to the 5' end
Distance3p	Distance to the 3' end
Partial	When the SNP is within range of the 5' or 3' end and the transcript only aligns partially to the genome
Codon Reference	DNA codon with the reference base pair.
Codon Read	DNA codon with the read base pair.
AA Reference	Amino acid produced by the DNA codon with the reference base pair.
AA Read	Amino acid produced by the DNA codon with the read base pair.
Comment	Comments regarding the annotation

Statistics:

Field	Meaning
Genome Used:	The genome/assembly that was used
Database Used:	The database that was used
Range Used:	The range that was used
Database file:	The database files that was in the construction of the segment tree and the data at which they were downloaded
Total number of sites:	Total number of SNPs in the input that were entered by the user
Sites upstream of genes:	Number of SNPs that are located upstream of their respective most relevant transcript
Sites downstream of genes:	Number of SNPs that are located downstream of their respective most relevant transcript
Sites within exons of genes:	Number of SNPs that are located within an exon of their respective most relevant transcript
Sites within introns of genes:	Number of SNPs that are located within an intron of their respective most relevant transcript
Exons:	The index of the exons for the "Sites within exons of genes"
Introns:	The index of the introns for the "Sites within introns of genes"
Site hitting reliable genes:	Number of SNPs hitting genes for which the reading frames can be established with certainty
Site hitting unreliable genes:	Number of SNPs hitting genes for which the reading frames cannot be established with certainty
Site hitting coding exons:	Number of SNPs located within protein coding regions of exons
Site hitting non-coding exons:	Number of SNPs located within exons but outside of a protein coding region
Site causing non-synonymous mutations:	Number of SNPs in coding exons causing a change in the produced amino acid sequence
Site causing non-synonymous mutations:	Number of SNPs in coding exons not causing a change in the produced amino acid sequence
Pie Chart	A pie chart showing distribution of the SNPs

Options

Available option	Effect
Range	This will consider genes within the range specified by the user (default = 0 bp)
Produce AAseq?	If checked, it will report the 2 amino-acid sequences, one using the reference base pair and the other, using read base pair
No reference base pair	Check this if you do not want to specify the reference base pair, USE AT YOUR OWN RISK

3.1.4 Transcription start site (tss)

This mode is to determine the closest transcription start site.

Input format

The input format is:

```
chromosome coordinate id
```

Example:

```
chr19 903423 test1
```

The **chromosome** must be the same as the one used on the UCSC Genome Browser, the **coordinate** must be the same as the one used on the UCSC Genome Browser (1-based) and the **id** must be unique for each line.

Output

Results:

Field	Meaning
Transcript	The genes/transcripts whose TSS was closest to each input
Position	The relative position of each input with respect to the genes/transcripts
Index	Index of the exons/introns when a coordinate falls within an exon or intron
Distance5p	Distance to the 5' end
Distance3p	Distance to the 3' end
Partial	When the coordinates is within range of the 5' or 3' end and the transcript only aligns partially to the genome

Statistics:

Field	Meaning
Genome Used:	The genome/assembly that was used
Database Used:	The database that was used
Database file:	The database files that was in the construction of the segment tree and the data at which they were downloaded
Chart:	This represents the number of inputs that can be found on each side on each side of the TSS, the number of bins and relative size of bins in the chart can be changed upon launching Segtor. The graph can be saved in jpg or png in various resolution.

Options

This mode produces a bar chart showing the distribution of the sites around transcription start sites, here are the options:

Available option	Effect
bins	Number of bars to create on either side of the TSS
Base pair per bin	Number of base pairs that a bin represents

3.1.5 Insertion, deletions and translocations (INS/DEL/TRANS)

This mode is to annotate insertion, deletions and translocations

Input format

The input format can either be:

```
INS chromosome coordinate DNAsequence id
DEL chromosome coordinate1 coordinate2 id
TRANS  chromosome1 coordinate1 strand1 chromosome2 coordinate2 strand2 id
```

Example:

```
INS chr7 902344 CAGT ins1
DEL chr10 230984 230996 del1
TRANS  chr9 324023 + chr4 23135 -trans1
```

The **chromosome** must be the same as the one used on the UCSC Genome Browser, **coordinate**, **coordinate1** and **coordinate2** must be the same as the one used on the UCSC Genome Browser (1-based) and the **id** must be unique for each line. **strand1** and **strand2** are the strands indicating the which strand was paired the translocation.

Output

Insertions:

Field	Meaning
Transcript	The genes/transcripts that were found for each input
Position	The relative position of each input with respect to the genes/transcripts
Index	Index of the exons/introns when an insertion falls within an exon or intron
Distance5p	Distance to the 5' end
Distance3p	Distance to the 3' end
Partial	When the insertion is within range of the 5' or 3' end and the transcript only aligns partially to the genome
CoordAA	Coordinate of the insertion in the amino acid sequence
Comment	Comments regarding the annotation
Sequence Reference:	The sequence without the insertion
Sequence Read:	A putative sequence with the insertion (does not take splicing and other events into account)

Deletions:

Field	Meaning
Transcript	The genes/transcripts that were found for each input
Position	The relative position of each insertion with respect to the genes/transcripts
Index	Index of the exons/introns when an insertion falls within an exon or intron
Distance5p	Distance to the 5' end
Distance3p	Distance to the 3' end
Partial	When the insertion is within range of the 5' or 3' end and the transcript only aligns partially to the genome
Sequence Reference:	The sequence without the deletion
Sequence Read:	A putative sequence with the deletion (does not take splicing and other events into account)

Translocations

Field	Meaning
Transcript	The genes/transcripts that were found for each input
Position	The relative position in terms of strand of the translocation with respect to the genes/transcripts, may have 2 lines for each input if the translocation fuses 2 genes
Sequence:	The sequence produced by either truncating or fusing genes due to a translocation

Options

Available option	Effect
Range	This will consider genes within the range specified by the user (default = 0 bp)

3.2 Available databases

To set the database, first select the species you want by selecting the clade (mammal, verterbrate, etc) the species (human, chicken, etc) and the assembly for the given species (Human Feb. 2009 (GRCh37/hg19) , Human Mar. 2006 (NCBI36/hg18)). Finally, select the database (NCBI RefSeq, Ensembl Genes, etc). Please see the references for details on each database.