

BED Format

(modified from: <https://learn.gencore.bio.nyu.edu>)

The official documentation for BED format can be found [here](#).

BED format is a simple way to define basic sequence features to a sequence. It consists of one line per feature, each containing 3-12 columns of data, plus optional track definition lines. These are generally used for user defined sequence features as well as graphical representations of features.

Here are some links to the formal definitions of each field.

- [Required fields](#)
- [Optional fields](#)
- [Track lines](#)
- [BedGraph format](#)

Required fields

The first three fields in each feature line are required:

1. Chromosome Name
 - Name of the chromosome or scaffold. Any valid seq_region_name can be used, and chromosome names can be given with or without the 'chr' prefix.
2. Chromosome Start
 - Start position of the feature in standard chromosomal coordinates \ (i.e. first base is 0\).
3. Chromosome End
 - End position of the feature in standard chromosomal coordinates

```
chr1 213941196 213942363
chr1 213942363 213943530
chr1 213943530 213944697
chr2 158364697 158365864
chr2 158365864 158367031
chr3 127477031 127478198
chr3 127478198 127479365
chr3 127479365 127480532
chr3 127480532 127481699
```

Optional fields

Nine additional fields are optional. Note that columns cannot be empty – lower-numbered fields must always be populated if higher-numbered ones are used.

1. Name
 - Label to be displayed under the feature, if turned on in “Configure this page”.
2. Score

- A score between 0 and 1000. See [track lines](<http://asia.ensembl.org/info/website/upload/bed.html#tracklines>) , below, for ways to configure the display style of scored data.
- 3. Strand
 - defined as + \ (forward\) or – \ (reverse\).
- 4. thickStart
 - coordinate at which to start drawing the feature as a solid rectangle
- 5. thickEnd
 - coordinate at which to stop drawing the feature as a solid rectangle
- 6. itemRgb
 - an RGB colour value \ (e.g. 0,0,255\). Only used if there is a track line with the value of itemRgb set to “on” \ (case-insensitive\).
- 7. blockCount
 - the number of sub-elements \ (e.g. exons\) within the feature
- 8. blockSizes
 - the size of these sub-elements
- 9. blockStarts
 - the start coordinate of each sub-element

E.g.

```
chr7 127471196 127472363 Pos1 0 + 127471196 127472363 255,0,0
chr7 127472363 127473530 Pos2 0 + 127472363 127473530 255,0,0
chr7 127473530 127474697 Pos3 0 + 127473530 127474697 255,0,0
chr7 127474697 127475864 Pos4 0 + 127474697 127475864 255,0,0
chr7 127475864 127477031 Neg1 0 – 127475864 127477031 0,0,255
chr7 127477031 127478198 Neg2 0 – 127477031 127478198 0,0,255
chr7 127478198 127479365 Neg3 0 – 127478198 127479365 0,0,255
chr7 127479365 127480532 Pos5 0 + 127479365 127480532 255,0,0
chr7 127480532 127481699 Neg4 0 – 127480532 127481699 0,0,255
```

Track lines

Track definition lines can be used to configure the display further, e.g. by grouping features into separate tracks. Track lines should be placed at the beginning of the list of features they are to affect.

The track line consists of the word ‘track’ followed by space-separated key=value pairs – see the example below. Valid parameters used by Ensembl are:

- **name:** unique name to identify this track when parsing the file
- **description:** Label to be displayed under the track in Region in Detail
- **priority:** integer defining the order in which to display tracks, if multiple tracks are defined.
- **color:** as RGB, hex or X11 named color
- **useScore:** a value from 1 to 4, which determines how scored data will be displayed. Additional parameters may be needed, as described below.
- tiling array ([example file](#))
- colour gradient – defaults to Yellow-Green-Blue, with 20 colour grades. Optionally you can specify the colours for the gradient (cgColour1, cgColour2, cgColour3) as either RGB, hex or X11 colour names, and the number of colour grades (cgGrades). ([example file](#))
- histogram ([example file](#))
- wiggle plot ([example file](#))
- **itemRgb**
- if set to 'on' (case-insensitive), the individual RGB values defined in tracks will be used.

E.g.

```
track name="ItemRGBDemo" description="Item RGB demonstration" itemRgb="On"
chr7 127471196 127472363 Pos1 0 + 127471196 127472363 255,0,0
chr7 127472363 127473530 Pos2 0 + 127472363 127473530 255,0,0
chr7 127473530 127474697 Pos3 0 + 127473530 127474697 255,0,0
chr7 127474697 127475864 Pos4 0 + 127474697 127475864 255,0,0
chr7 127475864 127477031 Neg1 0 - 127475864 127477031 0,0,255
chr7 127477031 127478198 Neg2 0 - 127477031 127478198 0,0,255
chr7 127478198 127479365 Neg3 0 - 127478198 127479365 0,0,255
chr7 127479365 127480532 Pos5 0 + 127479365 127480532 255,0,0
chr7 127480532 127481699 Neg4 0 - 127480532 127481699 0,0,255
```

BedGraph format

BedGraph is a suitable format for moderate amounts of scored data. It is based on the BED format (see above) with the following differences:

1. The score is placed in column 4, not column 5
2. Track lines are compulsory, and must include type=bedGraph. Currently the only optional parameters supported by Ensembl are:
 - * **name**
 - * see above
 - * **description**
 - * see above
 - * **priority**
 - * see above
 - * **graphType**
 - * either 'bar' or 'points'.

E.g.

```
track type=bedGraph name="BedGraph Format" description="BedGraph format" priority=20
chr19 59302000 59302300 -1.0
chr19 59302300 59302600 -0.75
chr19 59302600 59302900 -0.50
chr19 59302900 59303200 -0.25
chr19 59303200 59303500 0.0
chr19 59303500 59303800 0.25
chr19 59303800 59304100 0.50
chr19 59304100 59304400 0.75
```

What software use bed files?

- Alignment viewers can use these data to graphically display certain features.
- [bedtools](<http://bedtools.readthedocs.io/en/latest/index.html>) uses this format to query for nearby features.
- Some annotation files are in this format.
- Feature detection packages use this as output.

How are these files generated?

- Feature detection algorithms.
- Lots of databases that hold certain genomic features report their data in this format.
- Sometimes manually curated from alignments (via bedtools, bamtools, etc.).

Type to enter text

Type to enter text

Type to enter text

Type to enter text

Type to enter text

Type to enter text

Type to enter text

Type to enter text

Type to enter text

Type to enter text

Type to enter text

Type to enter text