Global analysis of CNV

Ricardo H. Ramirez-Gonzalez

We want to explore how the details of the called CNVs

There are `43,412,060` CNV events in the datasatet

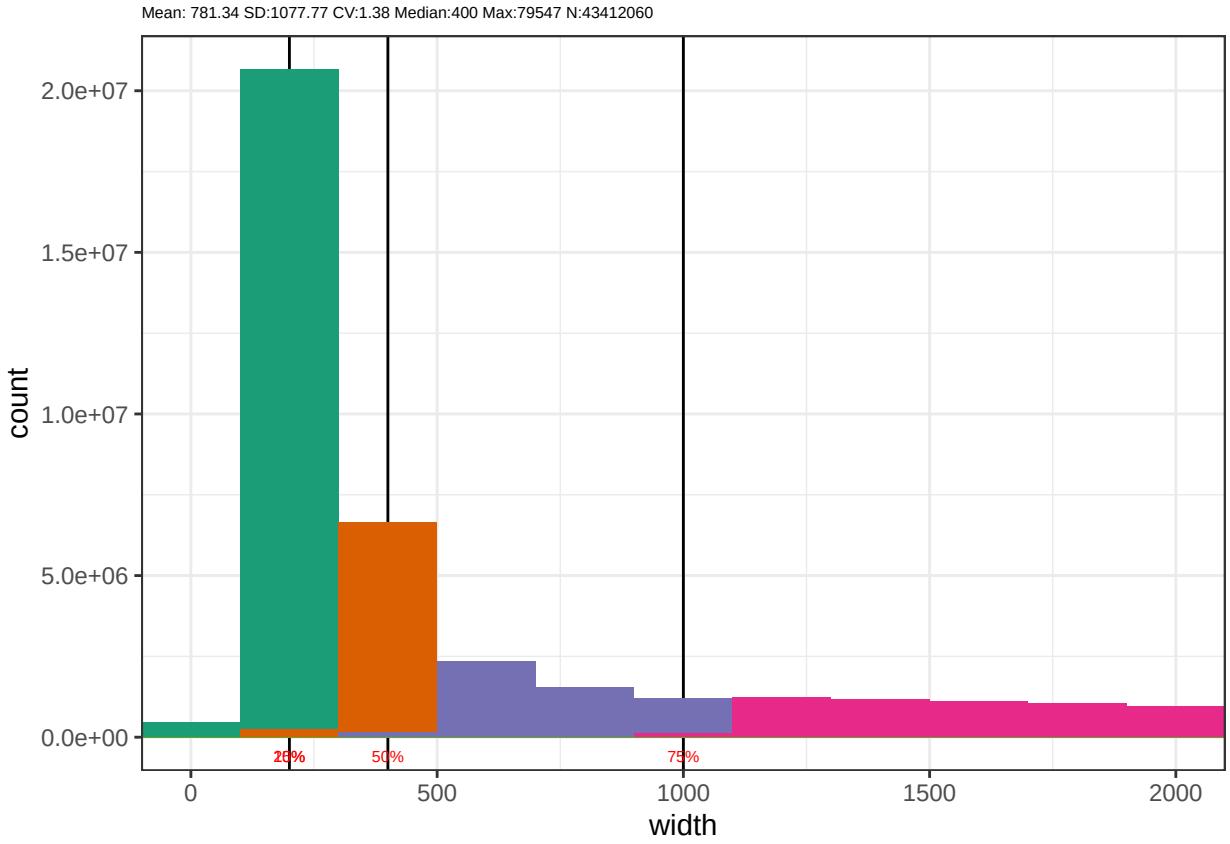
First we want to get an idea of the distribution of the deletion sizes.

```
kable(head(cnvs))
```

seqnames	start	end	width	strand	norm_cov	cnv_level	line	max_gap
chr1A_part1	1610157	1610356	200	*	0.0000000	0	B1190023.1	10
chr1A_part1	2393686	2394485	800	*	0.0000000	0	B1190023.1	10
chr1A_part1	2436043	2437842	1800	*	0.0266237	0	B1190023.1	10
chr1A_part1	2472568	2475367	2800	*	0.0170682	0	B1190023.1	10
chr1A_part1	2520813	2521012	200	*	0.0000000	0	B1190023.1	10
chr1A_part1	2557847	2558046	200	*	0.0000000	0	B1190023.1	10

```
plotHistogram(cnvs_df, column="width", binwidth=200)
```

```
## Warning: Use of `quantiles$value` is discouraged. Use `value` instead.
```



Most of them are under 1,500, however there are `7,630,091` larger than 1,500 (`17.58` %).

We want to look how one of the lines look randomly

```
plotCnvsInLine(cnvs)
```

```
## GRanges object with 6 ranges and 4 metadata columns:
##          seqnames          ranges strand |  norm_cov cnv_level        line
##             <Rle>       <IRanges>  <Rle> | <numeric> <integer> <character>
##   [1] chr1A_part1 1168522-1168721      * |  2.524868         3   WATDE0821
##   [2] chr1A_part1 1192644-1192863      * |  2.018469         2   WATDE0821
##   [3] chr1A_part1 1610157-1610356      * |  0.000000         0   WATDE0821
##   [4] chr1A_part1 1640797-1641196      * |  7.452139         7   WATDE0821
##   [5] chr1A_part1 1645597-1647325      * |  3.390315         3   WATDE0821
##   [6] chr1A_part1 1655115-1656514      * |  0.853136         1   WATDE0821
##         max_gap
##       <integer>
##   [1]        10
##   [2]        10
##   [3]        10
##   [4]        10
##   [5]        10
##   [6]        10
##   -------
##   seqinfo: 43 sequences from an unspecified genome
```