Demonstration cfdnakit

Pitithat

5/13/2021

This rmarkdown to demonstrate how cfdnakit work. First, load cfdnakit package to environment

```
library(cfdnakit)
```

Let cfdnakit read sequence alignment file (.bam) with function read_bamfile. This function will split sequence reads into equal-size non-overlapping windows. Available size of bin are 100, 500, and 1000 KB.

```
sample_bins =
  read_bamfile(bamfile_path = "/icgc/dkfzlsdf/analysis/OE0290_projects/pediatric_tumor/whole_genome_seq
               binsize = 1000)
```

```
## [1] "Reading bamfile"
## [1] "Filtering-out read on blacklist"
```

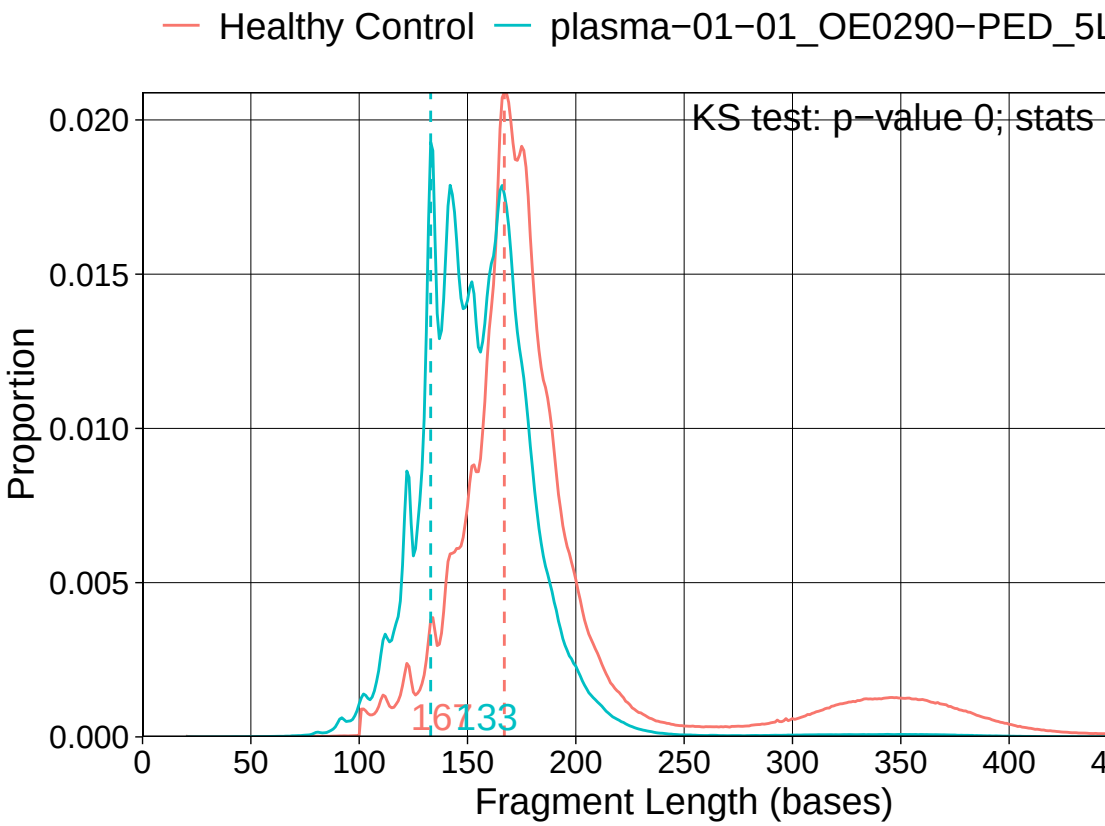
```
paste0("Number of bin : ",length(sample_bins))
```

```
## [1] "Number of bin : 2888"
```

Extract cfDNA fragment length information.

```
sample_profile =
  get_fragment_profile(sample_bins,
                       sample_id = "plasma-01-01_OE0290-PED_5LB-017")
plot_isize_distribution(sample_profile)
```

```
## Warning: Removed 200 row(s) containing missing values (geom_path).
```

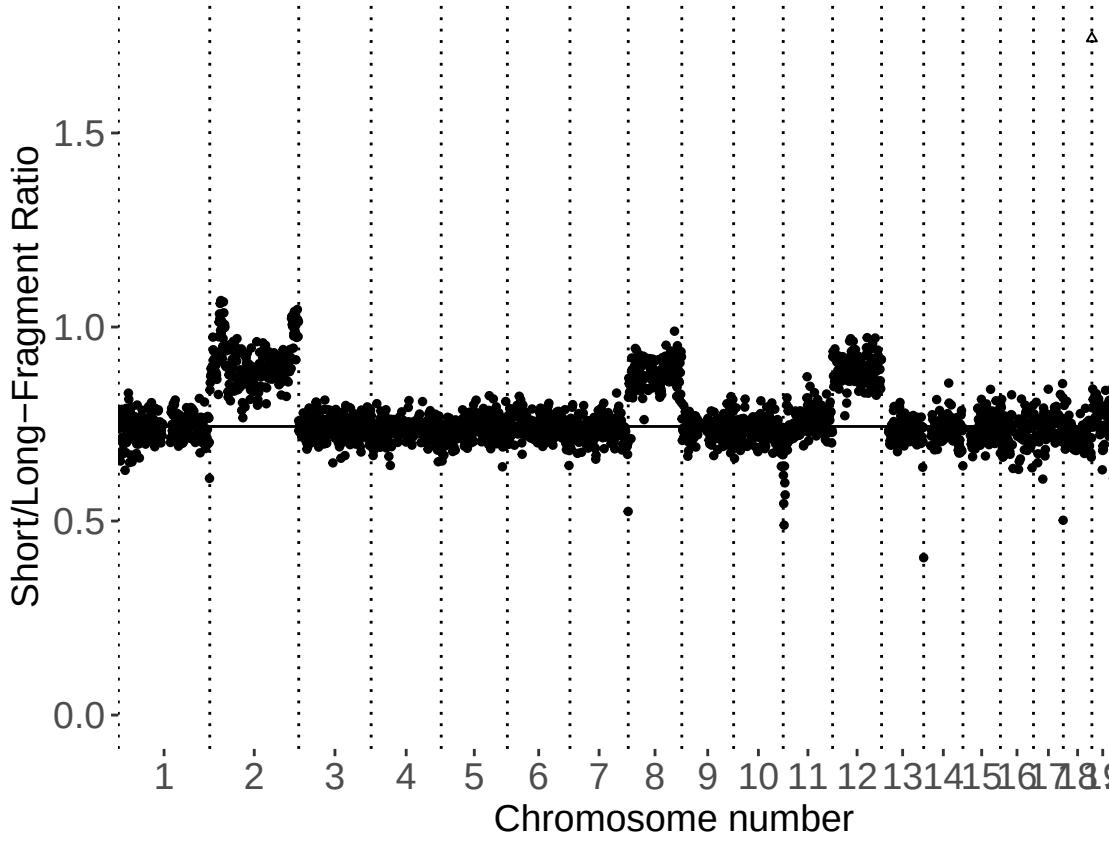


fragment length profile-1.pdf

```
plot_sl_ratio(sample_profile)
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```



fragment length profile-2.pdf

Save fragment profile as RDS file for later use or for creating Panel-of-Normal

```
save_fragment_profile(sample_profile,
                      output_dir = "/icgc/dkfzlsdf/analysis/OE0290_projects/pediatric_tumor/whole_genome
```

```
## [1] "Saving RDS : Done"
```

Making a Panel-of-Normal is necessary for downstream analysis as we want to compare fragment profile between a cfDNA from patient with pooled of healthy individuals. First, we create a text file where each line is a full path to rds file created by aforementioned function.

```
path_to_PoN_txt = "/icgc/dkfzlsdf/analysis/OE0290_projects/pediatric_tumor/whole_genome_sequencing/proc

create_PoN(path_to_PoN_txt,output_dir = "/icgc/dkfzlsdf/analysis/OE0290_projects/pediatric_tumor/whole_
```

```
## [1] "Done"
```

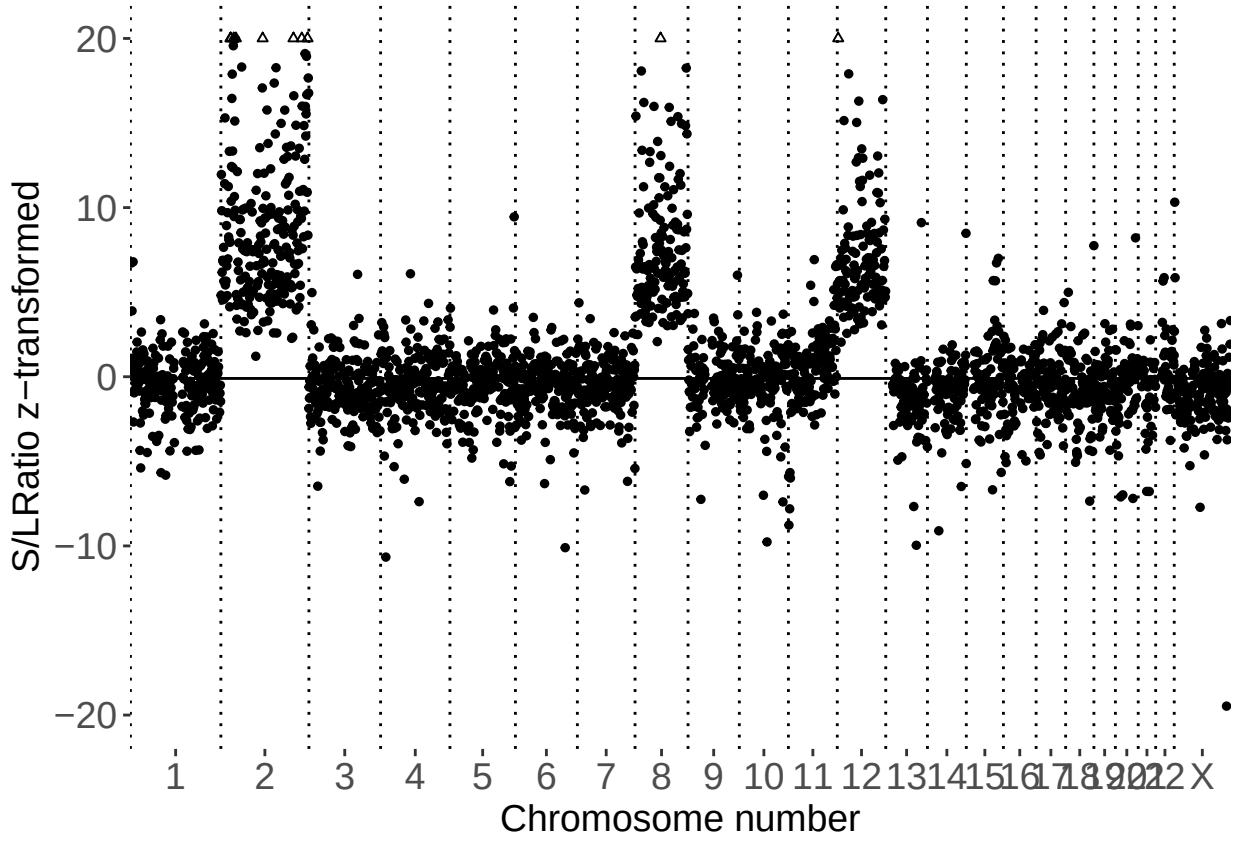
```
PoN_rdsfile = "/icgc/dkfzlsdf/analysis/OE0290_projects/pediatric_tumor/whole_genome_sequencing/processi
sample_zscore = get_zscore_profile(sample_profile,PoN_rdsfile)
```

```
## Reading PoN profile /icgc/dkfzlsdf/analysis/OE0290_projects/pediatric_tumor/whole_genome_sequencing/
```

```
plot_transformed_sl(sample_zscore)
```

```
## Warning: Removed 13 rows containing missing values (geom_point).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```



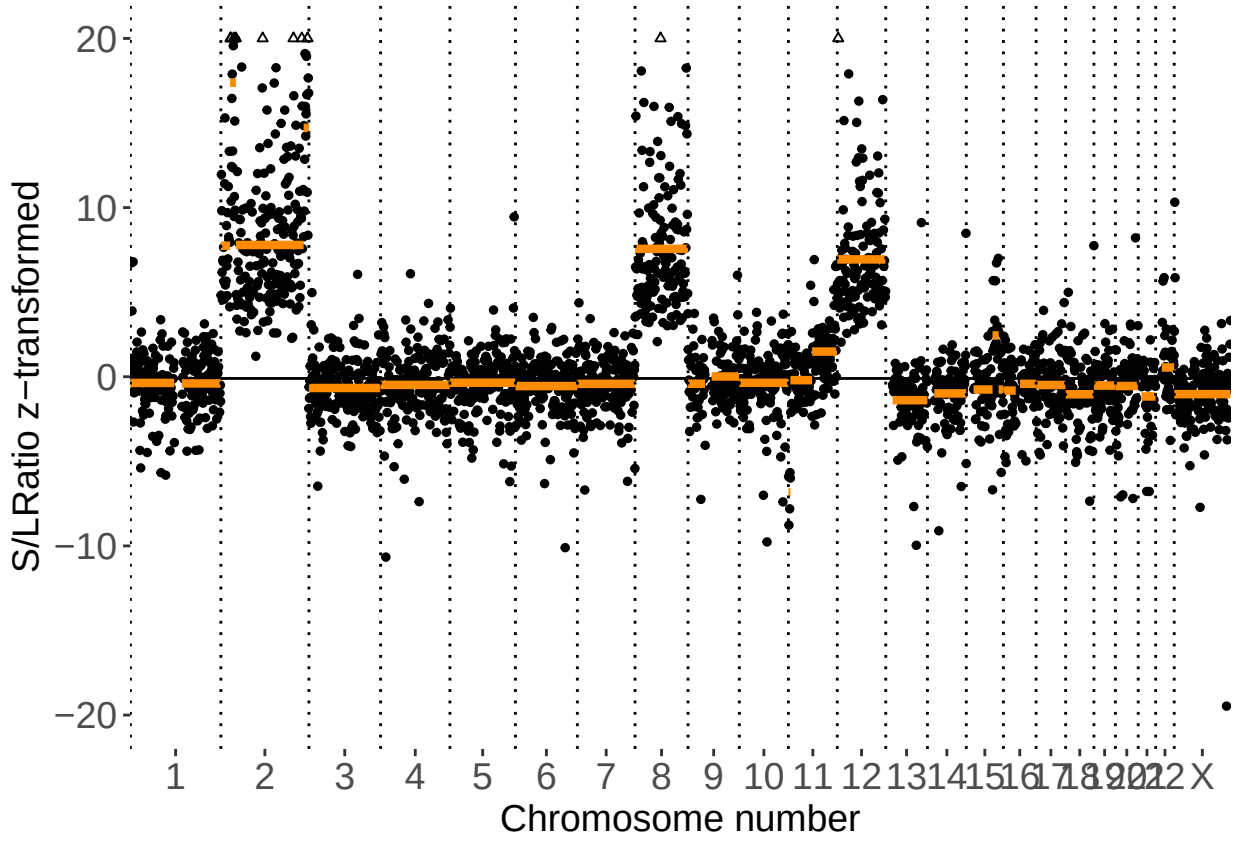

```
sample_zscore_segment = segmentByPSCB(sample_zscore)
```

```
## Warning in storage.mode(x) <- asMode: NAs introduced by coercion
```

```
plot_transformed_sl(sample_zscore,sample_zscore_segment)
```

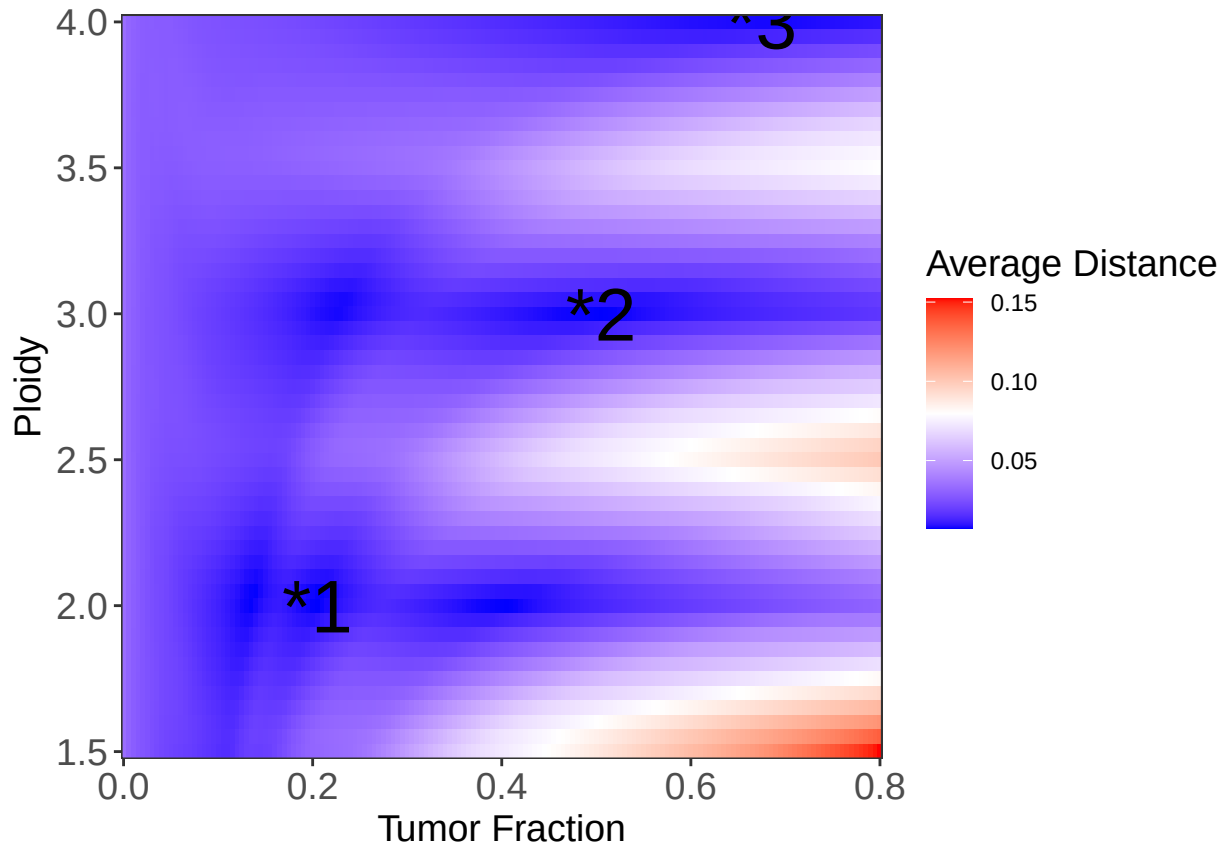
```
## Warning: Removed 13 rows containing missing values (geom_point).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```




```
sample_cnv = call_cnv(sample_zscore_segment,sample_zscore)

plot_distance_matrix(sample_cnv)
```

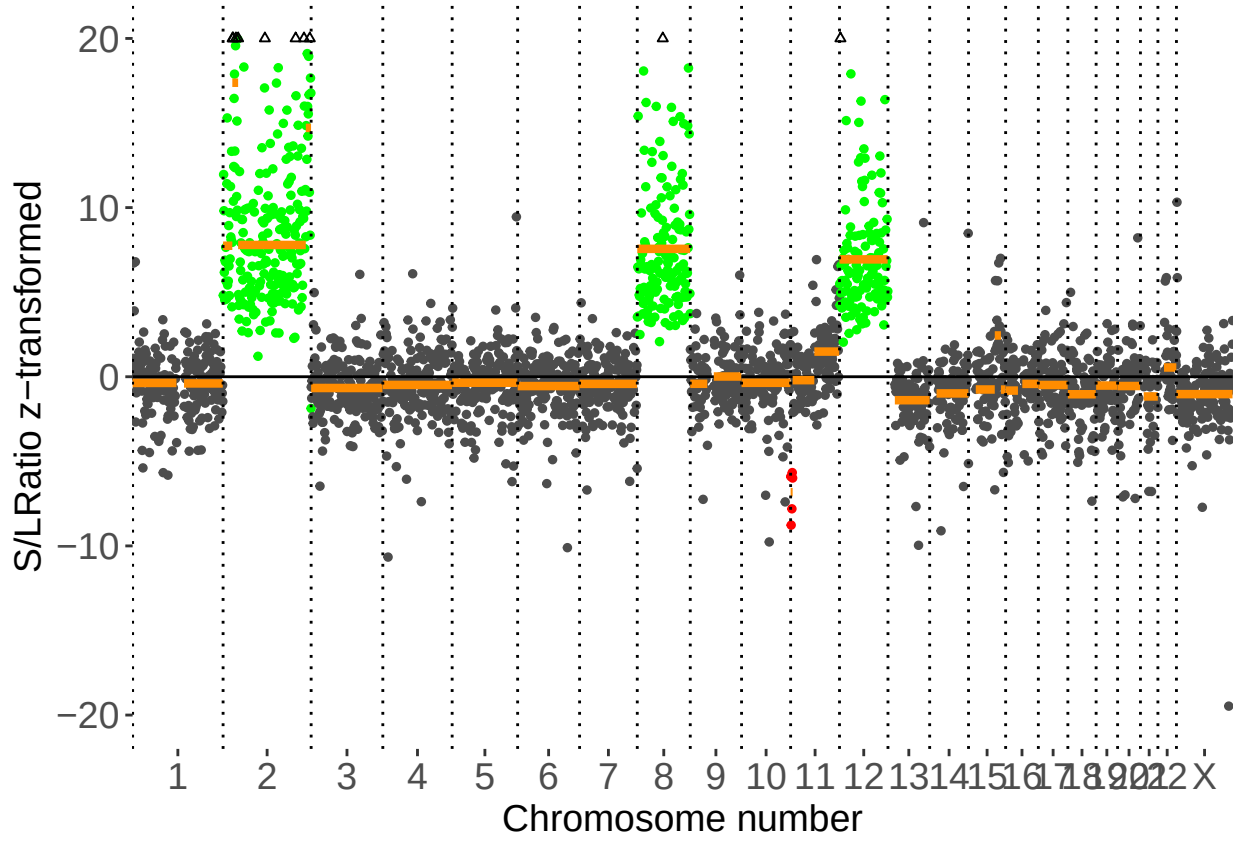


calling-1.pdf

```
plot_cnv_solution(sample_cnv,selected_solution = 1)
```

```
## Warning: Removed 13 rows containing missing values (geom_point).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```



calling-2.pdf

```
solution_table = get_solution_table(sample_cnv)
```

```
calculate_PCA_score(sample_zscore_segment)
```

```
## [1] 169.017
```