

Pause report

Sathish Thiyagarajan
sathish.thiyagarajan@earlham.ac.uk

March 11, 2025

Contents

1	Input json configuration file	2
2	Source files/directories	3
3	Output data files	4
4	Count forks and pauses (at ROI and elsewhere)	5
5	Statistics of ROI intervals	7
6	List of folders with other related files/documents	7

1 Input json configuration file

```
{
  "mod_bam": "/ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/dnascent_sam_minimap2_fastq_20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f.detect.mod.sorted.bam",
  "forksense_dir": "/ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/forkSenseOverallBedgraphs/3kb_30kb_noChrM_files_copy_renamed",
  "pause_file": "/ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/forkLen10AlignLen30/modelFitResults/sensitivity_10b_scale/pause_sensitivity",
  "bed_file": "/ei/.project-scratch/e/e52ea1f0-1ba9-4074-9055-1b2aab7a1606/input_json_files/11mar25/etc2.bed",
  "op_dir": "/ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/forkLen10AlignLen30/modelFitResults/analysis_11mar25/etc2",
  "dataset": "20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f",
  "feature": "etc2",
  "division": "0",
  "analysis_label": "11mar25",
  "mod_bam_left": "/ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/dnascent_sam_minimap2_fastq_20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f.forkSense.mod.left.sorted.bam",
  "mod_bam_right": "/ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/dnascent_sam_minimap2_fastq_20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f.forkSense.mod.right.sorted.bam",
  "alignment_file": "/ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/sam_minimap2_fastq_20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f.sorted.bam",
  "fasta": "/ei/projects/e/e52ea1f0-1ba9-4074-9055-1b2aab7a1606/data/references/GCA_002163515.1_ASM216351v1_genomic_edited_st_29oct21.fna",
  "n": 30,
  "prefix": "collection",
  "prefix_plot_option": false,
  "relative_direction_option": "all",
  "delete_new_mod_bam_made_on_the_fly": false,
  "genome_size_bp_optional": 12297572,
  "AT_ratio_optional": 0.62,
  "pause_sensitivity_bedgraphs_prefix_optional": "/ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/forkLen10AlignLen30/modelFitResults/sensitivity_10b_scale/pause_sensitivity.pause_sensitivity"
}
```

2 Source files/directories

Mod bam file: /ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/dnascent_sam_minimap2_fastq_20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f.detect.mod.sorted.bam

Forksense directory: /ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/forkSenseOverallBedgraphs/3kb_30kb_noChrM_files_copy_renamed

Pause file: /ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/forkLen10AlignLen30/modelFitResults/sensitivity_10b_scale/pause_sensitivity

Mod bam left fork probability file: /ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/dnascent_sam_minimap2_fastq_20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f.forkSense.mod.left.sorted.bam

Mod bam right fork probability file: /ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/dnascent_sam_minimap2_fastq_20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f.forkSense.mod.right.sorted.bam

Fasta file: /ei/projects/e/e52ea1f0-1ba9-4074-9055-1b2aab7a1606/data/references/GCA_002163515.1_ASM216351v1_genomic_edited_st_29oct21.fna

Regions of interest (ROI): from /ei/.project-scratch/e/e52ea1f0-1ba9-4074-9055-1b2aab7a1606/input_json_files/11mar25/etc2.bed

ROI bed file contents:

done by ST on Tuesday 11/03/2025 17:39 +00:00

NOTE that the strand below is not necessarily the strand of the ETC2 site.

any strand orientation is fine.

chrXV 52000 66000 neighbourhood_of_etc2 1000 +

3 Output data files

Relative direction option: all . Restrict fork direction option: Important note: Since a read can contain a fork in either direction, we can restrict reads in the mod bam file using the criteria that they overlap with the ROI and ignoring strands. But, we have no such problem with the pause file(s). So, the pause files below include a relative-direction between fork and region of interest criterion as requested by the user.

Mod bam file of reads overlapping with ROI: /ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/forkLen10AlignLen30/modelFitResults/analysis_11mar25/etc2/collection.mod.bam

Data from forks overlapping with ROI: /ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/forkLen10AlignLen30/modelFitResults/analysis_11mar25/etc2/collection_all pauses

Data from forks overlapping with ROI but no pauses in them: /ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/forkLen10AlignLen30/modelFitResults/analysis_11mar25/etc2/collection_no pauses

Data from forks overlapping with ROI but pauses not in ROI: /ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/forkLen10AlignLen30/modelFitResults/analysis_11mar25/etc2/collection_pauses_elsesewhere

Data from forks overlapping with ROI and pauses in ROI: /ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/forkLen10AlignLen30/modelFitResults/analysis_11mar25/etc2/collection_pauses_in_region

Report of statistics in this pdf in json format: /ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/forkLen10AlignLen30/modelFitResults/analysis_11mar25/etc2/collection_pauses_report.json

Plots of pause statistics: /ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_ARY01730uMpromethion_e50860f/forkLen10AlignLen30/modelFitResults/analysis_11mar25/etc2/collection_pause_statistics_plots

4 Count forks and pauses (at ROI and elsewhere)

Relative direction option: all

- [collection_n_reads] Total number of reads in mod bam file
2704457
- [collection_n_forks] Total number of forks in pause file
332096
- [collection_n_pauses] Total number of pauses in pause file $N_p^{tot} =$
17033
- [collection_n_reads_roi] Total number of reads in mod bam file that overlap the ROI
6258
- [collection_n_forks_roi] Total number of forks in pause file that overlap the ROI
815
- [collection_n_forks_roi_no_pause] Total number of forks in pause file that overlap the ROI but have no pauses
762
- [collection_n_pauses_roi_elsewhere] Total number of pauses in pause file occurring on forks that overlap the ROI but not in the ROI
28
- [collection_n_pauses_roi] Total number of pauses in pause file occurring on forks that overlap the ROI and in the ROI $N_p^{reg} =$ NOTE: In this counting, pauses are not counted multiple times. For example, let's say a pause occurs in a region that is specified twice in the ROI bed file. In spite of this, the pause is only counted once in this counting.
25
- [collection_n_pauses_roi_sens] Total number of pauses in ROI expected from sensitivity analysis $N_p^{sens} =$
18.973175114
- [collection_total_roi_length] Total length of ROI
14000
- [collection_total_n_roi] Total number of regions in ROI
1
- [collection_total_merged_roi_length] Total length of ROI after merging overlapping regions. $L_{ROI,merge} =$ NOTE: this could be slightly different depending on whether the relative direction option was set to all or otherwise.
14000
- [collection_total_genome_length] Total length of genome $L_{genome} =$
12297572
- [collection_total_merged_roi_length_fraction_genome] Fraction of genome covered by (merged) ROI $f_{ROI} = L_{ROI,merge}/L_{genome}$
0.00113843
- [collection_total_merged_roi_length_fraction_genome_halved_if_need_be] Fraction of genome covered by (merged) ROI halved (assuming equal co-directional and head-on forks) $f_{ROI,2} =$. If relative direction option is 'all', then same as above
0.00113843
- [collection_total_merged_roi_length_fraction_genome_halved_if_need_be_sd] Expected error due to counting in the fraction above, $\sqrt{\frac{f_{ROI,2}}{N_p^{tot}}} =$
0.0002585280037115902
- [collection_total_genome_fraction_AT] Fraction of AT bases in the genome (user-supplied) $b_{AT,genome} =$
0.62
- [collection_total_T_same_strand] Total number of T bases in (merged) ROI
4414
- [collection_total_T_opposite_strand] Total number of T bases in complementary strand of (merged) ROI
4205
- [collection_total_T_both_strands_all_genome] Total number of T bases in both strands of genome
7624494.64

- [collection_null_hypothesis_for_uniform_T] Expected pause fraction using null pause hypothesis of pauses at every available T $f_{null,T}$ = NOTE: number of available Ts depends on the relative direction option and the fork direction restriction. For e.g.: if you choose head-on and leading strand synthesis, then a + element in the ROI will only see T bases on the - strand. For details, please see the calculations in the script.
0.00113043
- [collection_sd_null_hypothesis_for_uniform_T] s.d. of above null hypothesis $\sqrt{\frac{f_{null,T}}{N_p^{tot}}} =$
0.0002576180354571473
- [collection_total_fork_overlap_length] Total length of sections of forks that overlap with the ROI $N_f^{reg} =$. NOTE: In this counting, forks are not counted multiple times. For example, let's say a fork overlaps with a region that is specified twice in the ROI bed file. In spite of this, the length of the overlapping part of the fork is measured only once.
4792466
- [collection_total_fork_length] Total length of forks in the pause file $N_f^{tot} =$
4063957362
- [collection_ratio_fork_overlap_length_to_fork_length] Ratio of total length of sections of forks that overlap the ROI to total length of forks $N_f^{reg}/N_f^{tot} =$ NOTE: this is the expected ratio of pauses in ROI to total number of pauses if region is typical.
0.0011792608959955914
- [collection_ratio_fork_overlap_length_to_fork_length_th_sd] (theoretical) s.d. in expected ratio using counting statistics $\sqrt{\frac{N_f^{reg}}{N_f^{tot} N_p^{tot}}} =$
0.00026312334141248285
- [collection_ratio_sens_region_to_total_pauses] Ratio of number of pauses expected from sensitivity in ROI to total number of pauses $N_p^{sens}/N_p^{tot} =$
0.0011139
- [collection_ratio_sens_region_to_total_pauses_sd] s.d. in number of pauses expected from sensitivity in ROI to total number of pauses $\frac{\sqrt{N_p^{sens}}}{N_p^{tot}} =$
0.00025572
- [collection_ratio_pause_region_to_total_pauses] Ratio of number of pauses in ROI to total number of pauses $N_p^{reg}/N_p^{tot} =$
0.0014677390946985264
- [collection_mean_duration] Mean duration of valid pauses in region (in kb)
8058.8134704799995
- [collection_sd_duration] S.d. of duration of valid pauses in region (in kb)
2476.3571327933646
- [collection_mean_duration_all_pauses] Mean duration of valid pauses in the entire pause file (kb)
8843.939691922356
- [collection_sd_duration_all_pauses] S.d. of duration of valid pauses in the entire pause file (kb)
3251.6578329244994

5 Statistics of ROI intervals

NOTE: some of these statistics may be repeats of statistics from the section above.

```
{
  "bed_file": "/ei/projects/e/e52ea1f0-1ba9-4074-9055-1b2aab7a1606/scratch/tmp/tmp
.CddgYed190",
  "number_of_regions": 1,
  "number_of_+_regions": 1,
  "number_of_-_regions": 0,
  "number_of_..regions": 0,
  "total_length_bp": 14000,
  "number_of_intersections_ignoring_strand_and_ignoring_self": 0,
  "closest_distance_statistics_per_region_same_strand_bp": [
  ],
  "closest_distance_statistics_per_region_ignoring_strand_bp": [
  ],
  "comment": "NOTE: closest distance counts may be lower than overall region count
because there may be no neighbours for some intervals",
  "number_of_regions_merge_same_strand": 1,
  "number_of_+_regions_merge_same_strand": 1,
  "number_of_-_regions_merge_same_strand": 0,
  "total_A_merge_same_strand": 4205,
  "total_C_merge_same_strand": 2607,
  "total_G_merge_same_strand": 2764,
  "total_T_merge_same_strand": 4414,
  "ratio_of_AT_to_total_merge_same_strand": 0.616083,
  "total_length_bp_merge_same_strand": 14000,
  "total_length_bp_merge_ignore_strand": 14000,
  "column_value_statistics": {
    "5_mean": 1000,
    "5_median": 1000,
    "5_stddev": "",
    "5_min": 1000,
    "5_max": 1000,
    "length_mean": 14000,
    "length_median": 14000,
    "length_stddev": "",
    "length_min": 14000,
    "length_max": 14000
  },
  "comment_stats": "In the calculation above, we've assumed columns 5,7,8,9,10 ex
ist and are numeric. If this assumption is not correct for any field, then igno
re the corresponding output above"
}
```

6 List of folders with other related files/documents

Reads with at least one fork with no pause

```
/ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_A
RY01730uMpromethion_e50860f/forkLen10AlignLen30/modelFitResults/analysis_11mar25
/etc2/no_pause/no_pause_read_ids_subsampled
```

Reads with at least one fork with pause but not in ROI

```
/ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_A
RY01730uMpromethion_e50860f/forkLen10AlignLen30/modelFitResults/analysis_11mar25
/etc2/elsewhere_pause/elsewhere_pause_read_ids_subsampled
```

Reads with at least one fork with pause and at ROI

```
/ei/.project-scratch/8/8e1c31fd-918e-4bd7-8fb8-2de97ff7675f/20231109_AR_ONT_SC_A
RY01730uMpromethion_e50860f/forkLen10AlignLen30/modelFitResults/analysis_11mar25
/etc2/region_pause/region_pause_read_ids_subsampled
```