|  |  |  |
| --- | --- | --- |
|  | Plan | Status |
| Tuesday | 1, SRA download script and usage documents  2, trim\_galore script and usage documents  3, bismak alignment automatically and documents  4, BAM to haploinfo, MF, entropy, epi-polymorphism and documents | Green: completed  Red: on the way or waiting for predetermination |
| Wednesday | 1, Alice SRA file download (462 fastq)  2, Alignment (on the way)  3, BAM to hapinfo  4, hapinfo to MHB |  |
| Thursday | 1, Xliu2014’s dataset (Download)  2, Holger2016’s dataset((Download))  3, Alignment  4, BAM to MHL (on the way) |  |
| Friday | 1, recollect Figure 4 dataset  2, more gradient and build the standard curve.  3, estimate the average ctDNA fraction  4, check the minimum concentration of our approach  1, MHL and 5mC for enhance/super enhancer regions.  2, DMR and Classification. |  |
| Monday | 1, Prepare format for Nature Genetics  2, Prepare format for Nature method  3, Prepare to update the patents format.  1, submit or not? |  |

**8/16/2016-8/23/16**

|  |  |  |
| --- | --- | --- |
|  | Plan | Status |
| Tuesday | 1, SRA download script and usage documents  2, trim\_galore script and usage documents  3, bismak alignment automatically and documents  4, BAM to haploinfo, MF, entropy, epi-polymorphism and documents | Green: completed  Red: on the way or waiting for predetermination |
| Wednesday | 1, Alice SRA file download (462 fastq)  2, Alignment (on the way)  3, BAM to hapinfo  4, hapinfo to MHB |  |
| Thursday | 1, Xliu2014’s dataset (Download)  2, Holger2016’s dataset((Download))  3, Alignment  4, BAM to MHL (on the way) |  |
| Friday | 1, recollect Figure 4 dataset  2, more gradient and build the standard curve.  3, estimate the average ctDNA fraction  4, check the minimum concentration of our approach  1, MHL and 5mC for enhance/super enhancer regions.  2, DMR and Classification. |  |
| Monday | 1, Prepare format for Nature Genetics  2, Prepare format for Nature method  3, Prepare to update the patents format.  1, su |  |

**Prepare for better version:**

**(1)   Perform additional quality filtering of our data, remove some low-quality samples, and repeat the analysis. Make the whole pipeline a full fluent platform and make a package. (MHB, AML, AMF)**

Same as Alice mouse dataset and now almost integrate all the scripts. Alice’s 462 WGBS data (ES, 2i, adult tissue), by the way, including large number of single-cell-methylome data.

1, Download

1, Compress twins’ fastq

2, trim\_galore

3, Alignment of twin’s fastq files (running in the TSCC, estimated 2 days)

4, sortbam back to bam (sort by read name, since bismark\_methylation\_extractor need name sorted bam)

**(2)   Determine how low (percentage of tumor DNA in blood) can we go in tissue-of-origin mapping, and ctDNA detection, by computational mixing of tumor data with WB data at various ratios, and perform the analysis similar to Figure 4.**

1, Re-builded the matrix and fix the model

2, the coefficients have been extract.

3, resampling and estimate.

**(3)   Look out for additional WGBS data in the public database.**

Xliu2014’s dataset (**WGBS: 5 lung cancer vs 5 normal lung**)

1, Compress twins’ fastq

2, trim\_galore

3, Alignment of twin’s fastq files

4, sortbam back to bam (sort by read name, since bismark\_methylation\_extractor need name sorted bam)

5, bam to haploinfo and MHL

Holger2016’s dataset (**11 primary tumors, 2 metastases and 9 normal tissues**)

1, download

1, Compress twins’ fastq

2, trim\_galore

3, fast\_trim to bam by bismark

4, sortbam back to bam (sort by read name, since bismark\_methylation\_extractor need name sorted bam)

5, bam to haploinfo and MHL

**(4)   Look further into the potential differences of the spreading pattern between adding and removing methyl groups.**

(5) Try reference-free-deconvolution to detect tumor-DNA from plasma

(6) Any biological function for the interesting tissue-specific MHBs in the prediction model.

(7) The overlap between our prediction features and other studies? And check whether these biomarkers are reasonable through the comparison of the Figure.

(8) With our prediction model to predict some other public dataset and test the model stability.

(9) Cancer-specific methylation haplotype (here, we can include all the genomic region, not only in MHB) and then estimate the cancer fragment contribution. (DMR, supper enhancer)

10. (Alice) Try the mouse data and find some interesting MHB changing pattern between stem cell and adult cells during the differentiation.

Collected large number mouse WGBS data.