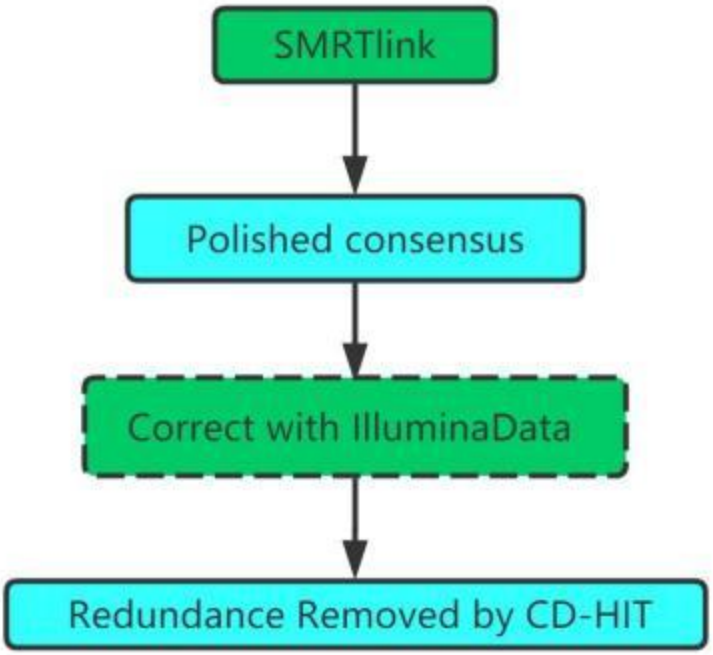
# Data flow combing of report without reference




| Statistics in final report | sample data | Data Sources | Data description |
|---|---|---|---|
| Polymerase Reads | 423875 | Table 3.1.1 Statistical results of Polymerase read | 1 |
| Polymerase Read Bases(G) | 23.99G | Table 3.1.1 Statistical results of Polymerase read | 2 |
| Subreads number | 13784046 | Table 3. 1.2.1 Subreads statistical results | 3 |
| Subreads base(G) | 22.95G | Table 3. 1.2.1 Subreads statistical results | 4 |
| CCS_ number | 374105 | Table 3.1.3.1 CCS statistical results | 5 |
| FLNC_ number | 315540 | Table 3.1.4.1 Statistical results of FLNC | 6 |
| Consensus_ number | 32540 | Table 3. 1.6.1 Polished consensus statistical results | 7 |
| Total_ number | 32540 | Table 3.2.1 Statistical table of transcript length distribution before and after correction | 8 |
| Number of transcripts | 32540 | Table 3.3.1 Statistical table of transcript length frequency distribution before and after deredundancy | 9 |
| Number of Genes | 19207 | | |

1 The number of high-quality sequencing reads produced by a single molecule during the sequencing process, and each read is a multi-copy sequence containing adapters obtained by cycle sequencing;

2 Polymerase Read data volume, obtained by multiplying Polymerase Reads by Polymerase Read Length (mean);

3 Remove the linker in Polymerase Read and the number of reads with a length less than 50bp to obtain the number of subreads. After removing the linker, the multi-copy sequence is interrupted, so the number of subreads is much greater than the number of Polymerase Reads;

4 Subreads data volume (same as Polymerase Read data volume);

5 The number of a consistent sequence obtained by self-calibration of multiple Subreads sequences in each ZMW (zero-mode waveguide hole) hole. In principle, each hole has and only one CCS;

6 Full-length non-chimeric sequences in CCS sequences (CCS also includes full-length chimeric sequences and non-full-length sequences);

7 FLNC sequence clustering to remove redundancy and correct the number of Consensus;

8 The second generation data is the number of reads before and after Consensus correction, since the bases in the reads are corrected,

Therefore, the number of reads remains unchanged before and after correction;

9 According to the 95% similarity between the sequences, perform clustering and de-redundancy on the corrected transcript sequences, and make statistics on the length-frequency distribution of transcripts before and after de-redundancy;