GeneMind

Quality Scores for GeneMind SURFSeq 5000

Quality Score Introduction

A next-generation sequencing experiment consists of a series of discrete steps that uniquely contribute to the overall quality of a data set. Sequencing quality metrics can provide important information about the accuracy of each step in this process, including library preparation, base calling, read alignment, and variant calling. Base calling accuracy, measured by the Phred quality score (Q score), is the most common metric used to assess the accuracy of a sequencing platform. It indicates the probability that a given base is called incorrectly by the sequencer. For example, when the base caller is 99% confident that a base call is correct, there is a 1% probability of an error, which translates to a Phred (Q) score of 20. Mathematically, Phred quality scores are defined as Q = -10 log10 P, where P is the probability of base calling error. As the Phred score increases, the confidence in the base call increases in a nonlinear fashion, as shown in Table below.

Q Score	Error Probability
10	10%
20	1.0%
30	0.1%
40	0.01%
50	0.001%

SURFSeq Data Quality

SURFSeq 5000 is the first module of ultra high-throughput gene sequencer of GeneMind. After it has been launched in Septembr 2023, it has been highly recognized by users in clinical diagnosis, molecular breeding, basic scientific research and other fields with its product advantages of "multiple application scenarios, fast sequencing speed, good data quality, and low operating cost and stable data performance".



GeneMind

"Good data quality" is the top priority, SURFSeq 5000 V1.0 sequencing kits (FCM, FCH) now could support two different mode: balanced mode(Default) and Enhanced mode(Option) after upgradtion, and with the Q40⩾90% high-quality data output.

✧ Dual-mode Q40 High Quality Data Output

SURFSeq 5000 provides users with two Q40 sequencing modes: Balanced Mode(BM) and Enhanced Mode(EM).

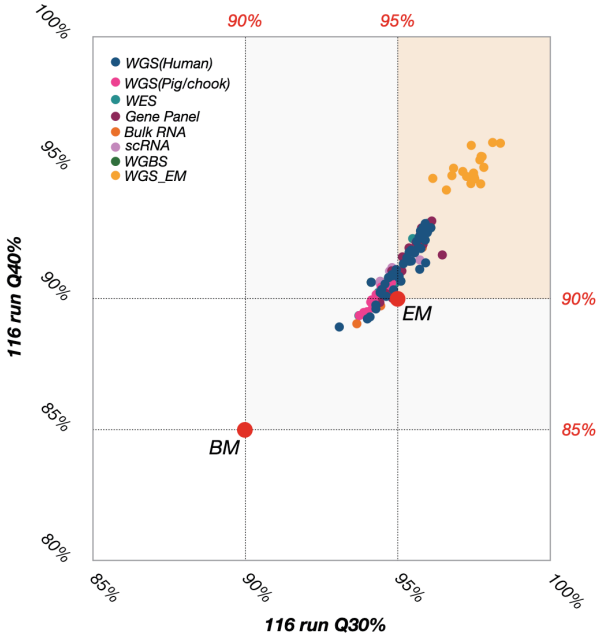
- ☑ In Balanced Mode, the SURFSeq 5000 achieves ⩾90% bases⩾Q30 and ⩾85% bases⩾Q40.

- ☑ In Enhanced Mode, the SURFSeq 5000 can achieve ⩾95% bases⩾Q30 and ⩾90% bases⩾Q40.

Mode	Q30*	Q40*
Balanced Mode	⩾90%	⩾85%
Enhanced Mode	⩾95%	⩾90%

The new Enhanced Mode adds multi-dimensional information such as sequence feature model constructed based on machine learning. The new "3D Basecall" can identify the "problem sequence regions" in the sequencing process more comprehensively, and perform base recognition and Q value scoring more accurately. Statistics of 16 runs of WGS PE150 sequencing data in Enhanced Mode showed that the mean values of Q30 and Q40 reached 97.1% and 95.2%, respectively, and the mean Output reads reached 2350M reads.

Application	Run number
WGS(Human)	*43*
WGS((Pig/chook)	*14*
WES	*10*
Gene Panel	*18*
Bulk RNA-Seq	*5*
scRNA-Seq	*7*
WGBS	*3*
BM Total	***100***
WGS_EM	*16*
EM Total	***16***



Dual-mode Lower Error Rate Data Performance

The SURFSeq 5000 is dual flow cells platform and supports asynchronous rolling loading and lane-split sample loading. As shown in the table below, we used two flow cells to perform both Enhanced Mode and Balanced Mode tests on the same machine. Lane1 and Lane2 were loaded with E.coli whole genome library samples (PCR-plus), and Lane3 and Lane4 were loaded with GIAB HG001 human whole genome library samples (PCR-plus).



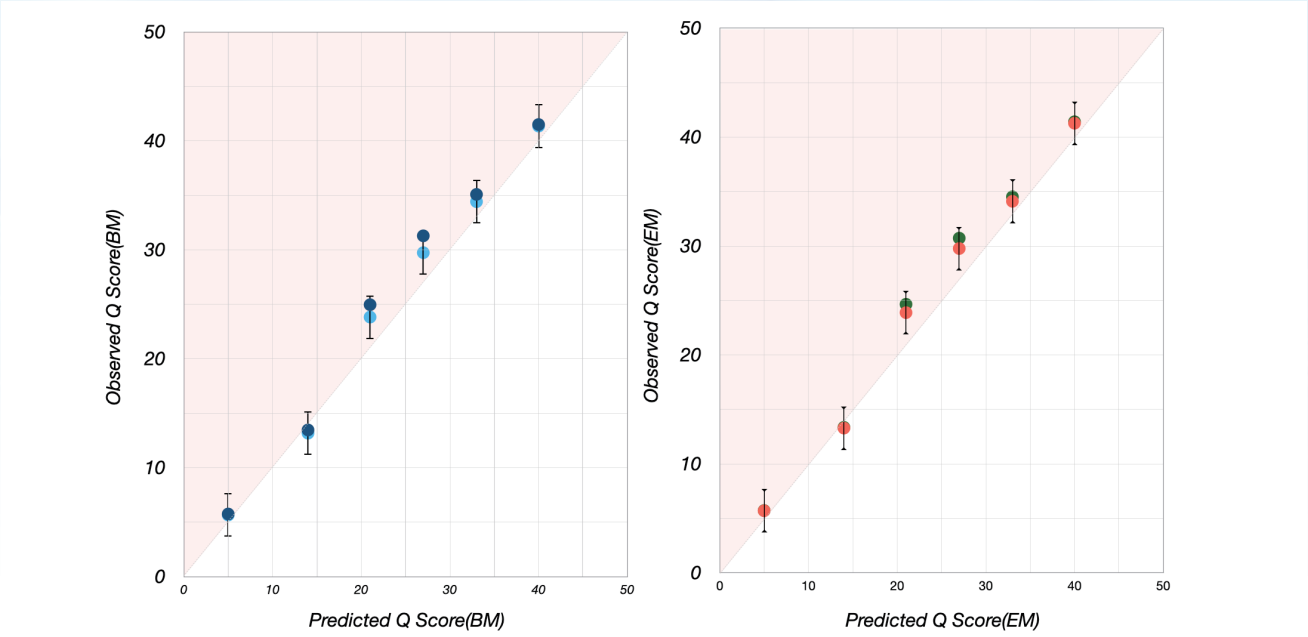
FCH	A side(BM_error rate)	B side(EM_error rate)
Lane 1	E.coli(0.124%)	E.coli(0.124%)
Lane 2	E.coli(0.135%)	E.coli(0.050%)
Lane 3	HG001(0.143%)	HG001(0.073%)
Lane 4	HG001(0.143%)	HG001(0.074%)



As shown in the figure above, the overall error rate of PE150 sequencing data in Enhanced Mode is reduced by 50-60% compared with Balanced Mode, down to 0.05%. The average value of Q30 /Q40 is 97.6%/96.0%.

✧ Accuracy Evaluation of Q Scoring

The accuracy of SURFSeq 5000 Q value scoring system is evaluated using Q/Q graph, and the Predicted Q Score(x) and Observed Q Score(y) in both Balanced Mode and Enhanced Mode show excellent consistency.



In Enhanced Mode, the SURFSeq 5000's overall sequencing error rate is reduced by more than 60% to as low as 0.05%.

Mode	Error Rate
SURFSeq 5000 BM	0.15±0.05%
SURFSeq 5000 EM	0.05+0.02%

GeneMind Biosciences Co., Ltd.

Technical support hotline: **+86-400-822-3660**
Website: en.genemind.com Email: info@genemind.com
Address: Room 502A/502B/602, Luohu Investment Holding Building, 112 Qingshuihe 1st Road, Luohu District, Shenzhen, Guangdong, China