

Smart Deep Basecaller, AI-based Basecaller for Sanger Sequencing

Yong Chu (yong.chu@thermofisher.com), Thermo Fisher Scientific, **Yong Chu**, Thermo Fisher Scientific, **Stephanie Schneider**, Thermo Fisher Scientific, **Ming Jiang**, Thermo Fisher Scientific, **Harrison Leong**, Thermo Fisher Scientific, **Youngmin Kim**, Thermo Fisher Scientific, **Megan McCluskey**, Thermo Fisher Scientific, **Madalin Radu**, Thermo Fisher Scientific

Deep learning has delivered outstanding performance across domains. Smart Deep Basecaller (SDB) developed by Thermo Fisher Scientific is the first AI-based basecaller for Sanger Sequencing enabled by deep learning. A set of algorithms including neural networks has been applied to improve basecalling accuracy in Sanger Sequencing traces. Large, in-house, annotated Sanger Sequencing datasets were used to train and test SDB networks.

SDB enables significantly improved basecalling performance when compared to KB Basecaller. SDB has increased read lengths with more high quality basecalls at 5' and 3' ends. It enables more accurate pure and mixed basecalls. SDB's advanced algorithm provides new functionality to support basecalling through heterozygous insertion deletion variants. SDB supports greater robustness for basecalling through common artifacts, such as dye blobs, as well as difficult sequences, such as GC-rich templates. The enhanced view enables greater resolution and clarity at the 3' end of plasmid sequencing data. Manual review time is significantly reduced because SDB provides fewer low quality basecalls and fewer false positive calls, which eases data review.

SDB, an innovative new basecaller, allows users to obtain increased throughput, greater confidence, and greater robustness for Sanger sequencing with reduced manual review time. SDB will benefit low/mid-throughput sequencing, next-generation sequencing confirmation, and applied re-sequencing applications.