

Isoform-Resolved RNA-Sequencing with RNA BaseCode

Gert-Jan Hendriks[1], Anton Larsson[1,2], Paloma Ruiz de Castroviejo Teba[1], Rickard Sandberg[1,2]

[1]Basic Genomics AB, Nobels Väg 16, 171 65 Solna, Sweden, [2]Department of Cell and Molecular Biology, Karolinska Institute, Solnavägen 9, 171 65 Solna, Sweden.

RNA sequencing has transformed molecular biology, enabling deep insights into gene expression across tissues and individual cells. Most mammalian genes are however expressed in multiple isoforms and their reliable detection and quantification remain a key limitation. Current short- and long-read sequencing methods struggle to distinguish closely related transcript variants, hindering our understanding of disease-related processes where specific isoforms, or splicing more generally can play important roles.

We have pioneered synthetic long-read RNA sequencing with the RNA BaseCode chemistry and software platform. Unique error patterns are introduced during reverse transcription and are used to uniquely identify all short reads originating from the same RNA molecule. Using this approach, RNA BaseCode combines the power of short-read sequencing to produce high-quality and highly quantitative data, with a proprietary chemistry to reconstruct full-length RNA sequences, enabling scalable and reliable detection and quantification of RNA isoforms applicable to transcriptomic studies across fields from cancer biology to developmental genomics.

RNA BaseCode can be performed on populations of cells or individual cells, with dramatic improvements in quantifications of full-length RNA isoforms, single-nucleotide mutations or polymorphism, fusion transcripts. Compared to long-read sequencing, RNA BaseCode provides dramatically improved base accuracy and full-length RNA sequencing reconstructions.

RNA BaseCode offers a scalable strategy to survey RNA isoform expression, important for exploring RNA isoform regulation across cells and tissues and their contribution to health and disease.