The human genome contains an astonishingly large fraction of noncoding DNA, which is pervasively transcribed into thousands of long noncoding RNAs (lncRNAs) – long transcripts with no discernible protein-coding potential. However, little is known about lncRNAs' biological functions, and their genome annotations show evident signs of inadequacy: existing gene models are sketchy, and many lncRNAs remain uncatalogued. This annotation incompleteness hampers lncRNA functional characterization, notably by failing to accurately describe gene boundaries.

To address this issue, this work aims to advance towards a complete and accurate annotation of lncRNA genes in the human genome. Using a high-throughput, targeted long-read transcriptome sequencing methodology, this study uncovers thousands of novel lncRNAs, approximately doubling the annotated transcript complexity within targeted loci. The method presented vastly outperforms competing techniques in accuracy, and precisely maps many previously unknown, strongly supported lncRNA transcript boundaries. This augmented catalog provides the most definitive view of the genomic properties of lncRNAs to date, while contributing a robust foundation for future lncRNA functional characterization.

Julien Lagarde

Genomic Characterization of Human Long Noncoding RNAs

Genomic Characterization of Human Long Noncoding RNAs

Julien Lagarde