Characterisation of mRNA isoform diversity in a transgenic model of tau pathology using targeted long-read sequencing

Szi Kay Leung[1], Aaron Jeffries[2], Eilis Hannon[1], Isabel Castanho[1], Karen Moore[2], Emma Dempster[1] , Tracey K.Murray[3], Zeshan Ahmed[3], David A Collier[3], Jonathan Mill[1]

[1]University of Exeter Medical School, University of Exeter, Exeter, UK;
[2]Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter, UK
[3]Beth Israel Deaconess Medical Center, Harvard University, Boston, US
Contact email: s.k.leung@exeter.ac.uk, j.mill@exeter.ac.uk

Abstract

This study uses ONT nanopore sequencing to assess splicing variation and isoform diversity in a transgenic mouse model of Alzheimer's Disease (AD) pathology. We focused on twenty genes robustly associated with AD from genetic studies, using selective gene enrichment to identify novel transcripts in entorhinal cortex tissue from the Tg4510 model of progressive tau pathology and wild-type controls. We detected many novel isoforms, revealing complex usage of alternative start sites and splicing events, not previously annotated in existing genomic datasets. These results further support a role for splicing dysregulation in development of AD pathology. Further work will be undertaken to characterize other AD-associated genes, and to extend these analyses to human post-mortem brain samples.

Introduction

Role of Alternative splicing in AD

An increasing number of studies implicate a role for alternative splicing in development and neuropathology of Alzheimer's disease (AD)[1]. Particularly prevalent in CNS development and function, alternative splicing of the same single gene can generate transcripts with very different and even antagonistic functions[2].

Advantages of Long-read isoforms sequencing

Oxford Nanopore Technologies (ONT) cDNA sequencing method generates long reads that span the full-length transcript (Figure 1). This allows complete and unambiguous inference of alternatively spliced exons, transcriptional start sites, and polyadenylation sites.

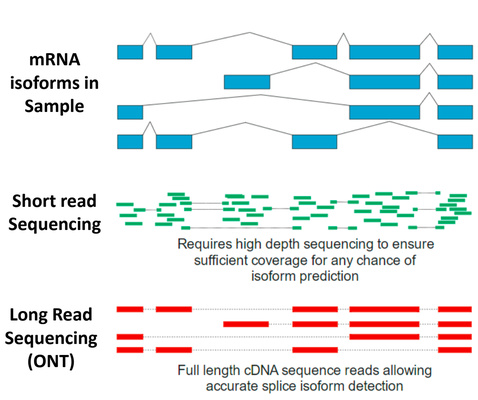
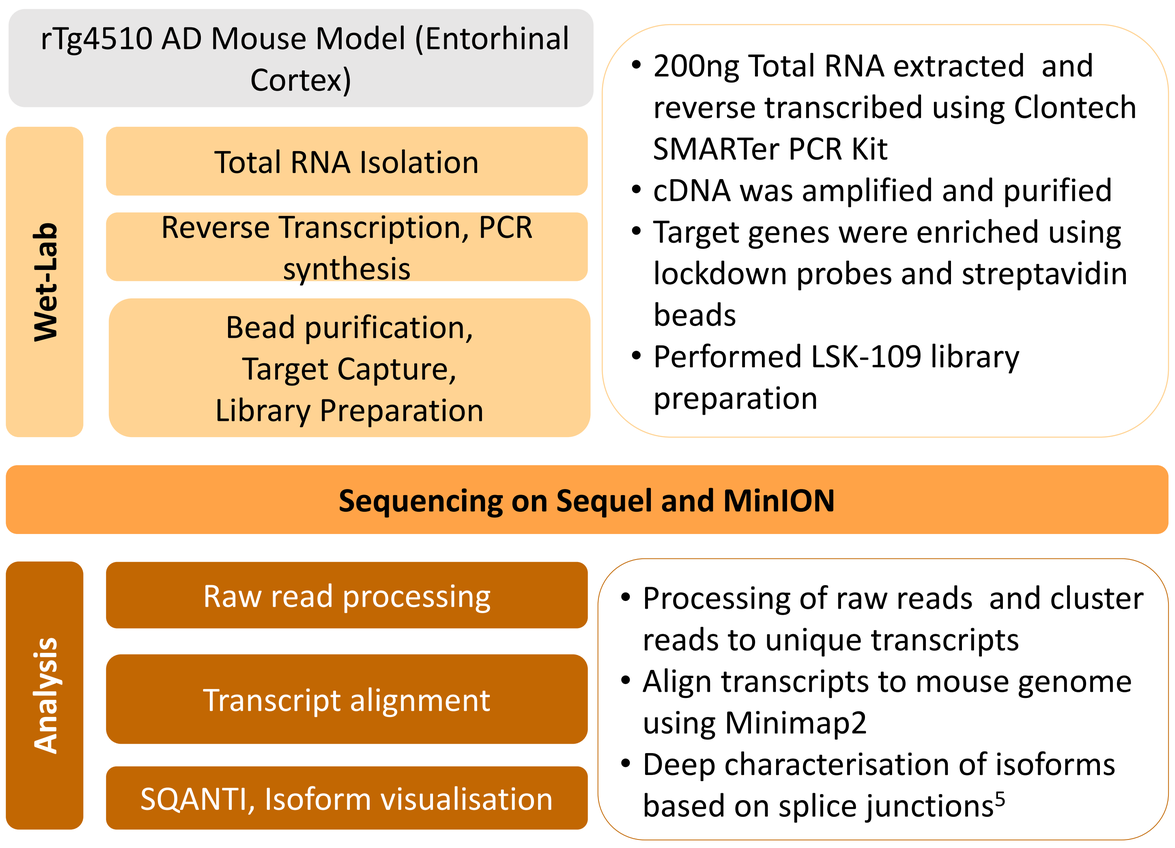


Figure 1: Schematic diagram of alternative splicing investigation using long read vs short read sequencing *The ONT method provides reads that span entire transcript isoforms, eliminating the need for computational reconstruction, and enabling functional characterisation of full-length transcript isoforms*

Methods



Results

Identified many novel isoforms with exon skipping and alternative 5' and 3' sites

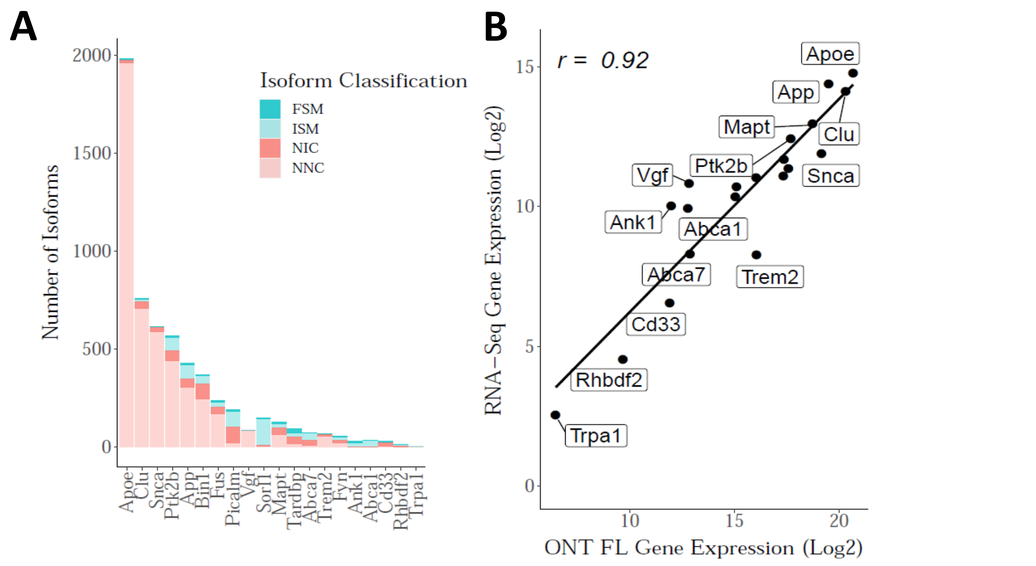


Figure 2: Deep sequencing coverage achieved from targeted enrichment reveal many novel isoforms annotated to AD-associated genes. A) Number of annotated (blue) vs novel isoforms (red) identified across target genes. A novel isoforms is defined as not having a fully matching exonic structure and agreement of internal junctions as reference. **B)** Strong correlation observed between ONT FL gene expression and RNA-Seq gene expression from matched samples

Figure 3: Alternative splicing events contributed to widespread isoform diversity. Exon skipping, and usage of alternative 5' and 3' sites was the most prevalent across the target genes

Widespread isoform diversity of AD-associated genes

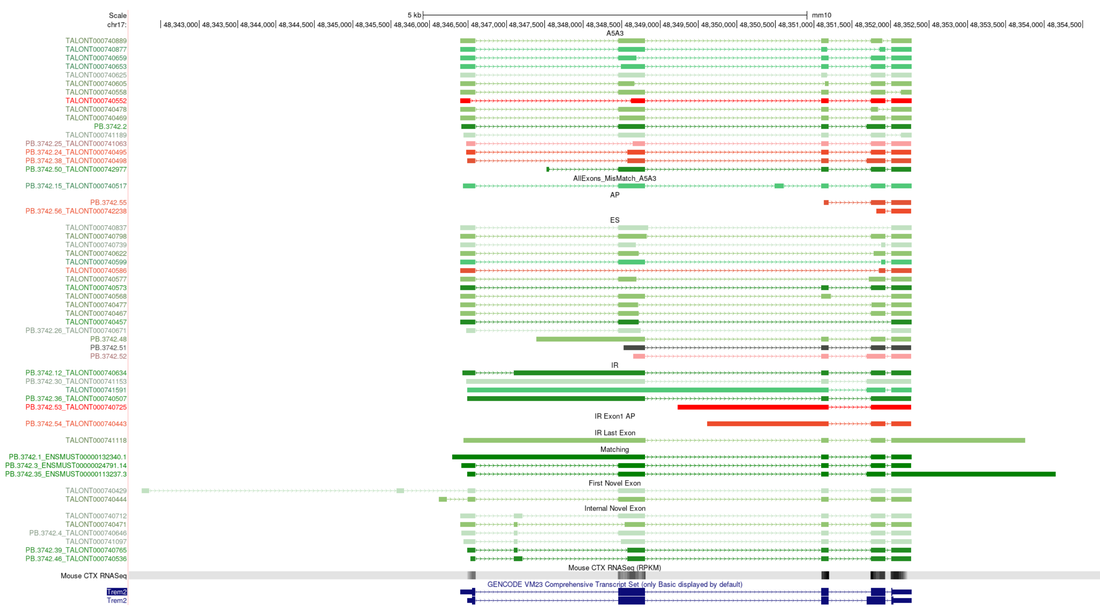


Figure 4: I*soform Visualisation of long reads (black) generated using PacBio Iso-Seq aligned to A) Trem2 on UCSC genome browser (blue - existing GENCODE transcript annotations)*. Isoforms are classified by splicing events, coloured based on coding status (green for protein coding, red for non protein coding) and shaded by abundance. ES – Exon skipping, IR – Intron Retention

References: 1) Raj et al. (2018) Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility; Nature Genetetics;50(11):1584-1592; 2) Wang et al. (2008) Alternative isoform regulation in human tissue transcriptomes; Nature. 27;456(7221):470-6, 3) Steijger et al. (2013) Assessment of transcript reconstruction methods for RNA-seq; Nature Methods;10(12):1177-84. 4) Tseng E. (2017) https://github.com/PacificBiosciences/IsoSeq_SA3nUP/; 5) Tardaguila Met al. (2018) SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification