Comprehensive isoform characterisation and splicing signatures of AD-risk genes from long-read sequencing of tau mouse model

Szi Kay Leung[1], Aaron Jeffries[1], Isabel Castanho[2], Rosemary Bamford[1], Karen Moore[1], Jonathan T. Brown[1], Emma Dempster[1] , Eilis Hannon[1] , Jonathan Mill[1]

[1] University of Exeter Medical School, Department of Clinical and Biomedical Sciences, University of Exeter, Exeter, UK,
[2] Beth Israel Deaconess Medical Center, Harvard University, Boston, US.
Contact email: s.k.leung@exeter.ac.uk, j.mill@exeter.ac.uk

Abstract

In this study, we use two complementary targeted long-read sequencing approaches, Pacific Biosciences (PacBio) isoform sequencing (Iso-Seq) and Oxford Nanopore Technologies (ONT) nanopore cDNA sequencing, to profile the cortex of a mouse model of tau pathology (rTg4510) with ultra-deep sequencing of a panel of 20 genes robustly implicated in AD. The depth of sequencing achieved using targeted gene enrichment enabled us to identify thousands of novel transcripts. We reveal unprecedented diversity of alternatively-spliced isoforms with widespread usage of alternative novel 5' and 3' splice sites, and exon skipping events. We further identify robust transcriptional and splicing differences in these AD-risk genes associated with the development of tau pathology. Among these changes, we observe global up-regulation of *Trem2*-associated isoforms, further supporting a role for the dysregulation of the immune response in AD pathology. Our analyses confirm the importance of alternative RNA splicing in AD and identify evidence of differential transcript usage, even in the absence of gene-level expression alterations. Further work will be undertaken to validate these novel isoforms, and to extend these analyses to human post-mortem AD brain samples.

Introduction

Role of alternative splicing in AD

An increasing number of studies implicate a role for alternative splicing in the progressive neuropathology associated with 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 isoform sequencing

While it has been historically challenging to characterise splicing events with traditional RNA-sequencing (RNA-Seq) approaches[3], the Pacific Bioscience's Isoform-Sequencing (Iso-Seq) and Oxford Nanopore Technologies (ONT) cDNA sequencing methods generate 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. We have recently used this approach to characterize widespread alternative splicing in both the mouse and human cortex [4].

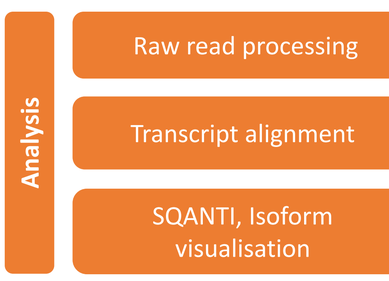

Figure 1: Schematic diagram of alternative splicing investigation using long read vs short read sequencing.
PacBio Iso-Seq and ONT nanopore approaches generate long reads that span the entire transcript length, eliminating the need for computational reconstruction, and enabling functional characterisation of full-length transcript isoforms.

Methods

Age (months): 2 | 4 | 6 | 8

Wet-Lab

- rTg4510 AD mouse model (Entorhinal cortex)
- Total RNA Isolation
- Reverse Transcription, PCR synthesis
- Bead purification, Target Capture Library Preparation

- 200ng Total RNA extracted using Qiagen Rneasy Mini Kit (RIN of 9)
- RNA was reverse transcribed using Clontech SMARTer PCR Kit and amplified for 15 cycles
- Target genes were enriched using lockdown probes and streptavidin beads (IDT)
- Performed Iso-Seq and ONT (LSK-109) library preparation

Sequencing on PacBio Sequel and ONT MinION

Analysis

- Raw read processing
- Transcript alignment
- SQANTI, Isoform visualisation

- Processing of raw reads to remove primers, poly(A) tails and cluster reads to unique transcripts (*Cupcake* for Iso-Seq reads, *TALON* for ONT reads)
- Align transcripts to mouse genome using *Minimap2*
- Merge Iso-Seq and ONT targeted dataset, and perform comprehensive characterisation of isoforms using *SQANTI* and custom tool (*FICLE*)

Results

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

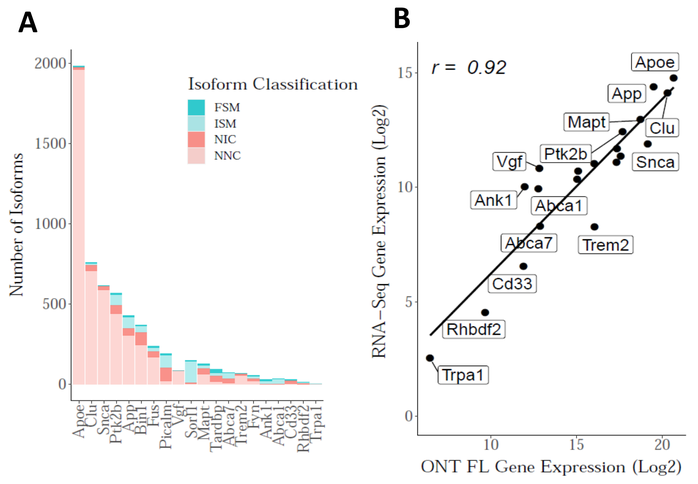

Figure 2: Deep sequencing coverage achieved from targeted enrichment reveals thousands of novel isoforms annotated to AD-associated genes in the mouse cortex.
(A) Number of annotated (blue) vs novel isoforms (red) identified across target genes. A novel isoform is defined as not existing in reference genome.
(B) Strong correlation observed between ONT full-length gene expression and RNA-Seq gene expression from matched samples.

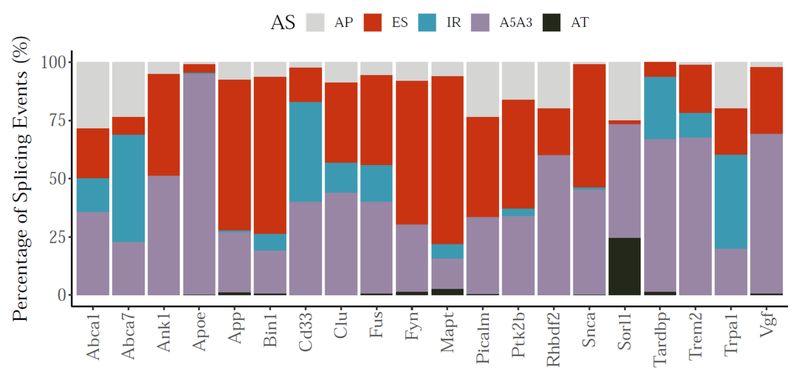

Figure 3: A range of different alternative splicing events contribute to widespread isoform diversity.
Exon skipping, and usage of alternative 5' and 3' sites were the most prevalent AS events across the target genes.

Extensive isoform diversity for selected AD-associated genes: *Trem2*, *Bin1*

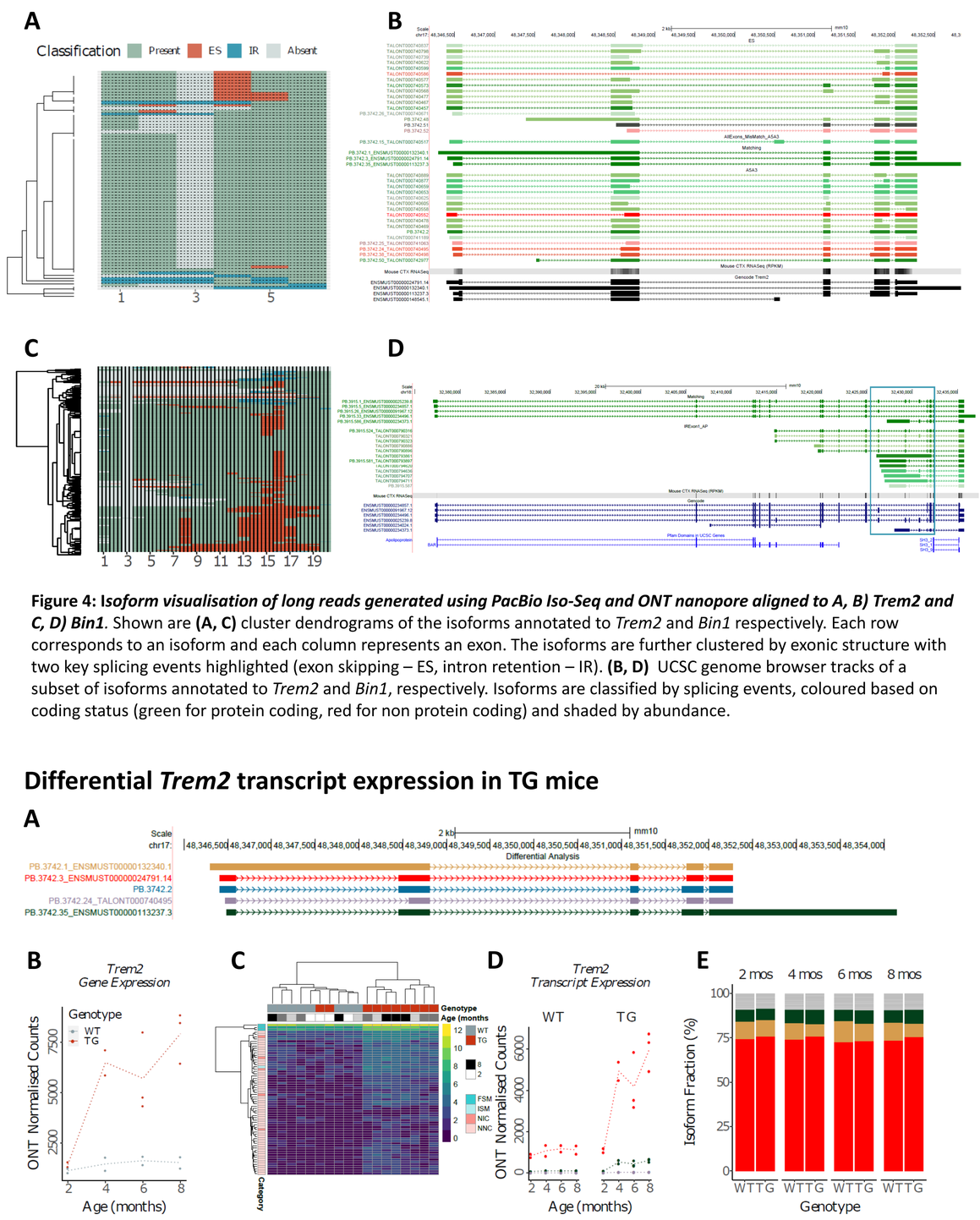

Figure 4: I*soform visualisation of long reads generated using PacBio Iso-Seq and ONT nanopore aligned to A, B) Trem2 and C, D) Bin1*. Shown are **(A, C)** cluster dendrograms of the isoforms annotated to *Trem2* and *Bin1* respectively. Each row corresponds to an isoform and each column represents an exon. The isoforms are further clustered by exonic structure with two key splicing events highlighted (exon skipping – ES, intron retention – IR). **(B, D)** UCSC genome browser tracks of a subset of isoforms annotated to *Trem2* and *Bin1*, respectively. Isoforms are classified by splicing events, coloured based on coding status (green for protein coding, red for non protein coding) and shaded by abundance.

Differential *Trem2* transcript expression in TG mice

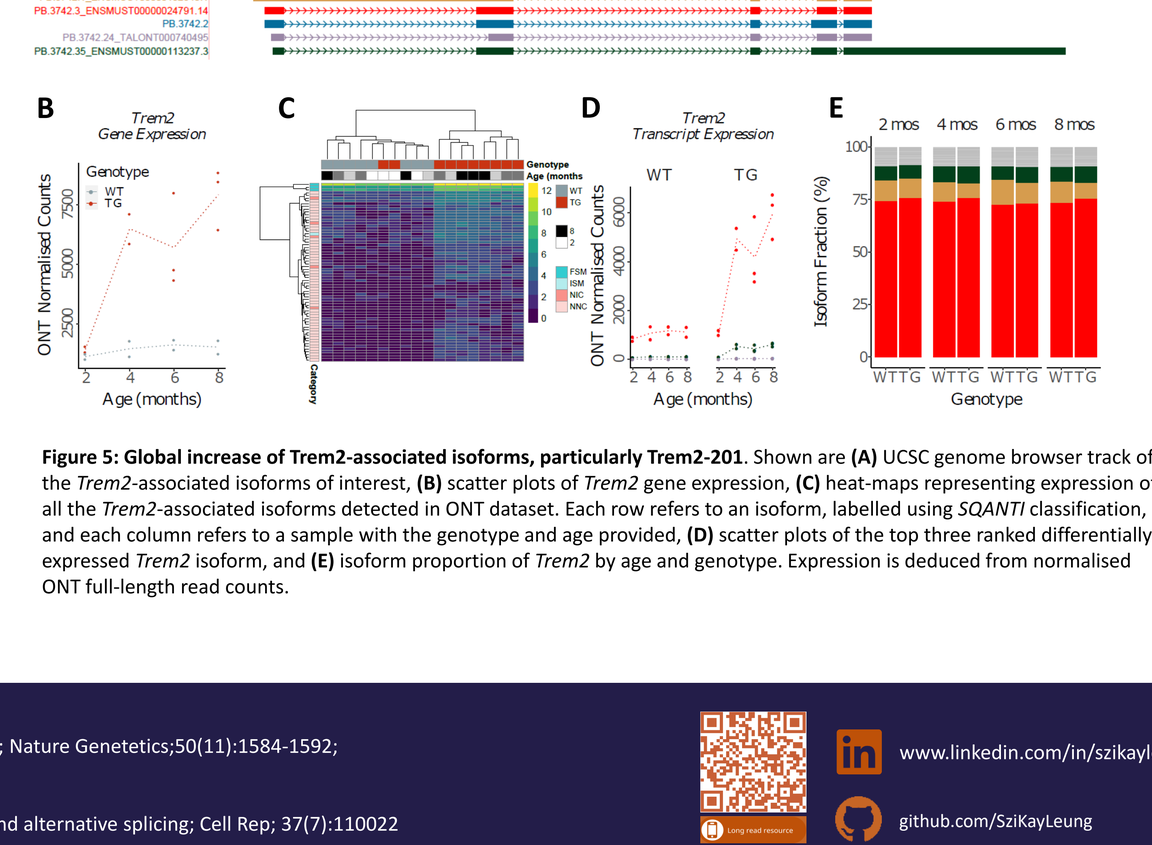

Figure 5: Global increase of Trem2-associated isoforms, particularly Trem2-201. Shown are **(A)** UCSC genome browser track of the *Trem2*-associated isoforms of interest, **(B)** scatter plots of *Trem2* gene expression, **(C)** heat-maps representing expression of all the *Trem2*-associated isoforms detected in ONT dataset. Each row refers to an isoform, labelled using *SQANTI* classification, and each column refers to a sample with the genotype and age provided, **(D)** scatter plots of the top three ranked differentially-expressed *Trem2* isoform, and **(E)** isoform proportion of *Trem2* by age and genotype. Expression is deduced from normalised ONT full-length read counts.

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) Leung et al. (2021) Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing; Cell Rep; 37(7):110022

www.linkedin.com/in/szikayleung
github.com/SziKayLeung