A brief history of splicing:

Direct RNA sequencing of mouse brain samples from the RIKEN aging project

Callum Parr[1], *Katrin Blondrath[1], Kayoko Yasuzawa[1,]* and Jay Shin[1,2]

[1]Laboratory for Advanced Genomics Circuit, Center for Integrative Medical Sciences, RIKEN, ***Japan***
[2]Laboratory of Regulatory Genomics, A Genomics Institute of Singapore,* ***Singapore***

Project funding

科研費 KAKENHI

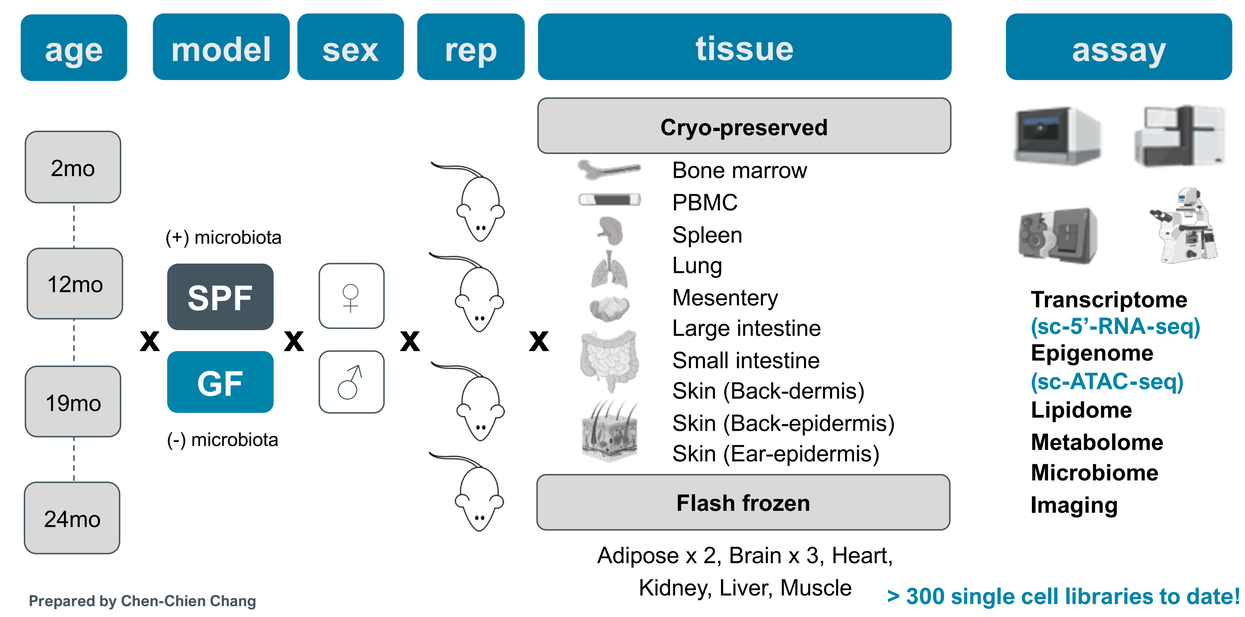
BACKGROUND

The **RIKEN Aging Project** is a multi-center and multi-disciplinary project that looks to provide a detailed account of genetic, transcriptomic, metabolic and phenotypic changes of the model mouse during aging. With a particular focus inflammaging, the study of how inflammatory insults during life can drive overall health of an individual during the aging process. on single-cell analysis of immune cells throughout multiple organs, and the influence microbiota by comparing standard specific-pathogen-free (SPF) housing versus specially reared germ-free (GF) mice.

Alternative splicing of RNA allows for a greater number of RNA isoforms and protein variants to be expressed from the same genome and involved in cell specification during development. However, little is known of how dynamic the expression at the isoform level is within a cell type during the ageing process of humans. The brain is resident to the oldest cells in our body, and therefore provides a history of ageing at the molecular level. Due to the complexity of splicing in the brain, we will leverage long-read direct RNA Nanopore sequencing to overcome current sequencing limitations. We will map the dynamic nature of RNA splicing of brain cell types and -states of the brain.

We envisage discovering linked isoforms and RNA modifications to the ageing process and age-associated diseases, such as Alzheimer's disease. Overall, the gene changes during aging differ between animals housed in standard specific-pathogen free housing and those in germ-free environment.

EXPERIMENTAL DESIGN

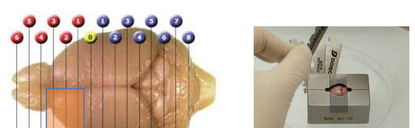


> **> 300 single cell libraries to date!**

OPTIMISATION

To maximize the read length of the RNA extracted from the brain tissue we tried various methods and compared by read length with low-coverage sequencing. As we did not see any appreciable difference in RNA integrity and read length between them, and TissueLyser provided the highest throughput we proceeded with this method for all samples going forward. Libraries where multiplexed and sequenced with the **direct cDNA kit (SQK-DCS108)**.

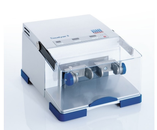
dissection



added to RNAlater O/N
Stored at -80 ℃

homogenization

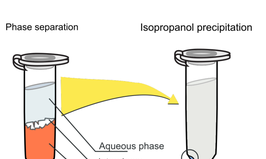
TissueLyser
High throughput
Thorough homogenisation



GentleMACS
Maybe severe?
Faster and 2 samples at once
No cross-contamination

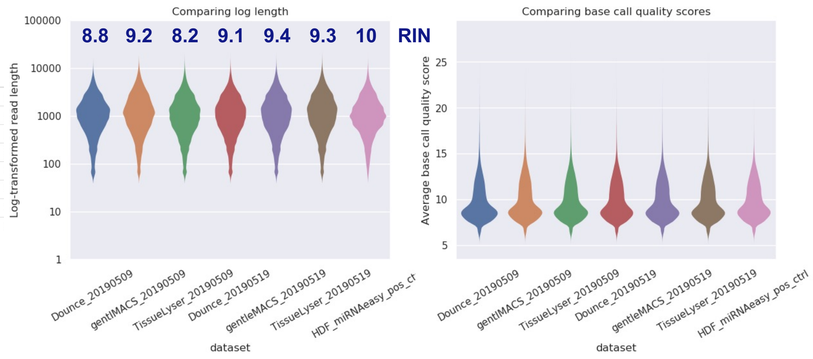
Dounce homogenizer
More gentle
slower and low throughput

extraction



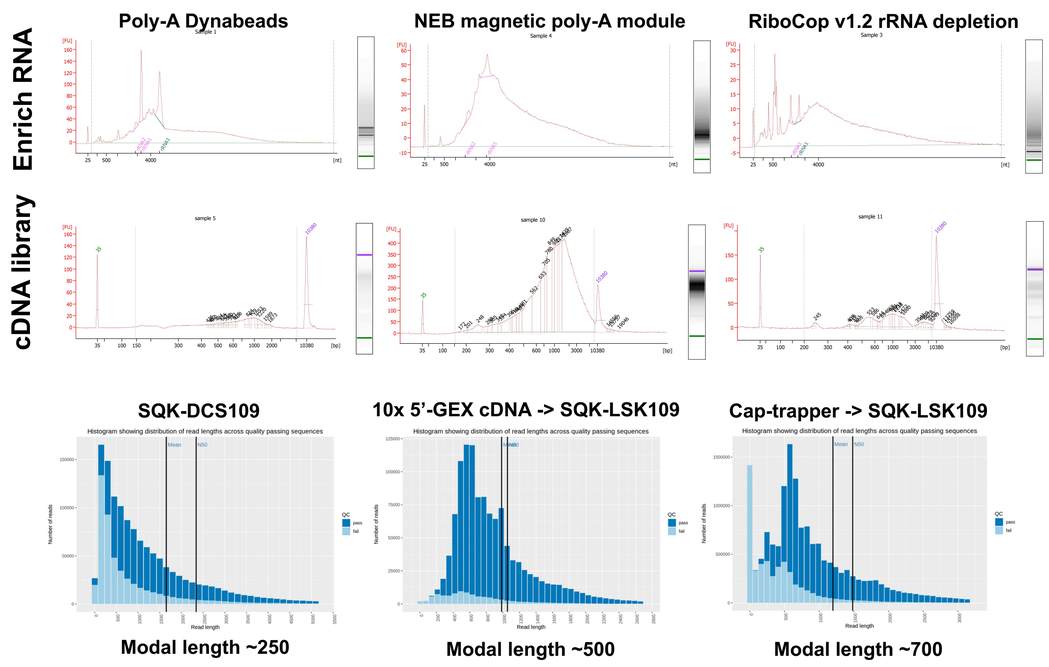
Direct cDNA kit (SQK-LSK108)

	barcode	freq	%	Mb	min	max	mean	N50
1	barcode01	53453	6.7	82	69	18406	1527	2082
2	barcode02	72124	9.1	129	69	18307	1784	2626
3	barcode03	124456	15.7	195	65	24759	1565	2212
4	barcode04	103608	13.1	163	73	20323	1575	2230
5	barcode05	132090	16.7	226	72	20806	1709	2417
6	barcode06	104733	13.2	187	77	23336	1786	2582
7	barcode07	129812	16.4	203	70	23116	1562	2104
12	unclassified	72774	9.2	125	1	24370	1721	2507



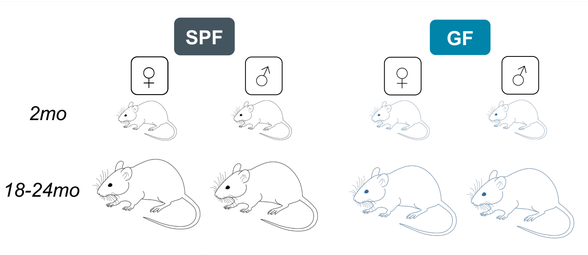
enrichment

We found that magnetic poly-A beads (NEB) worked the best for enriching mRNA over rRNA. While RiboCop kit effectively depleted rRNA the resulting bioanalyzer trace for the finished double-stranded, end-prepped cDNA library had biased appearance. For this reason, we stuck to using. During initial optimization steps we also found many short fragments were sequenced with the direct cDNA kit, that were not observed from other full-length cDNA generation methods. We therefore switched to **direct RNA kit to avoid RT- and PCR bias**.

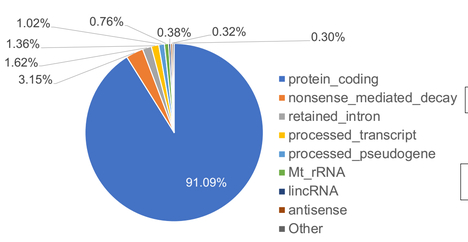


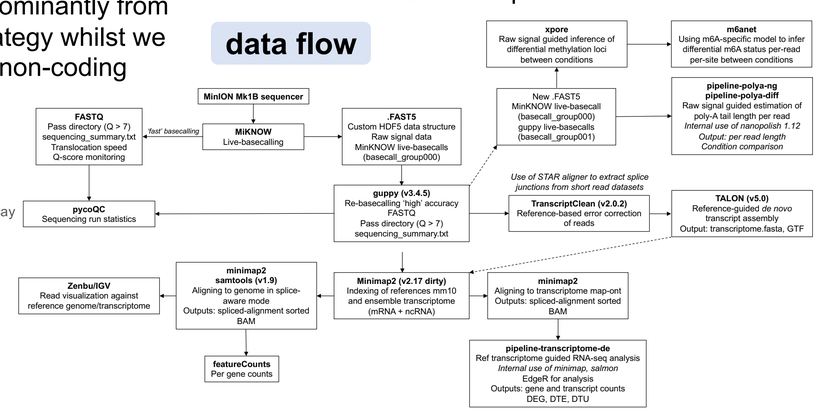
RESULTS

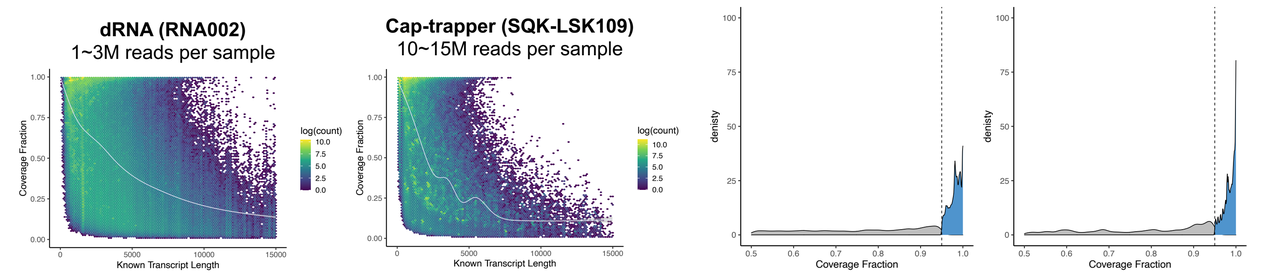
Using the direct RNA kit (RNA002) we proceeded to sequence all libraries from the each of the 8 conditions, **generating roughly 1~3M reads per sample**. We detected approximately 33,000 genes and 50,000 isoforms expressed across all the samples sequenced. **Each sample expressed approximately 13~15k genes and 20~30k isoforms** with the limited coverage offered by the dRNA kit. Reads detected were predominantly from mRNA given the poly-A targeting strategy whilst we could detect some highly expressed non-coding RNAs.



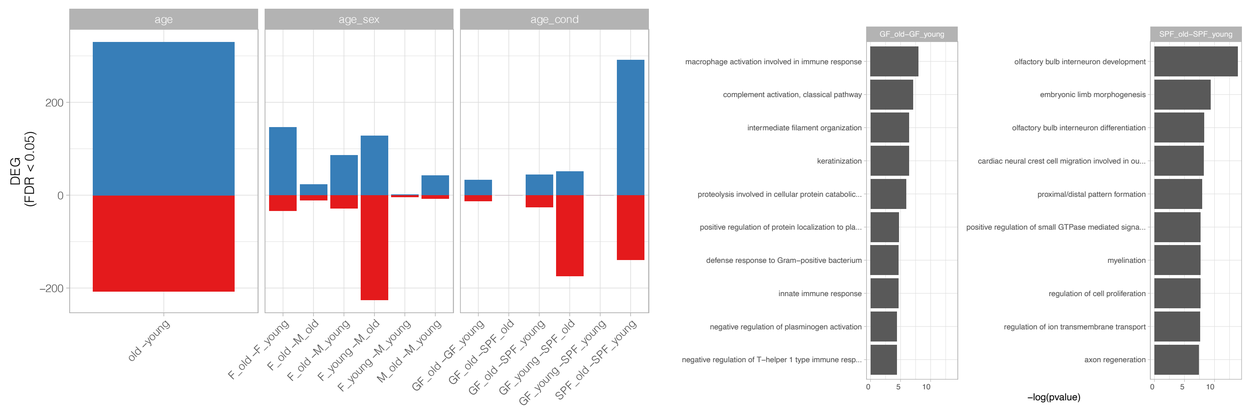
N=3-5 animal per condition





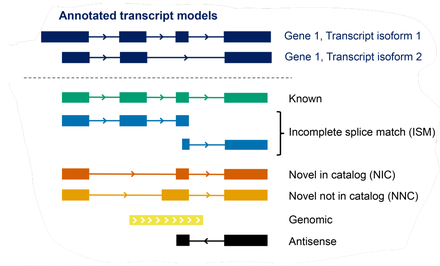


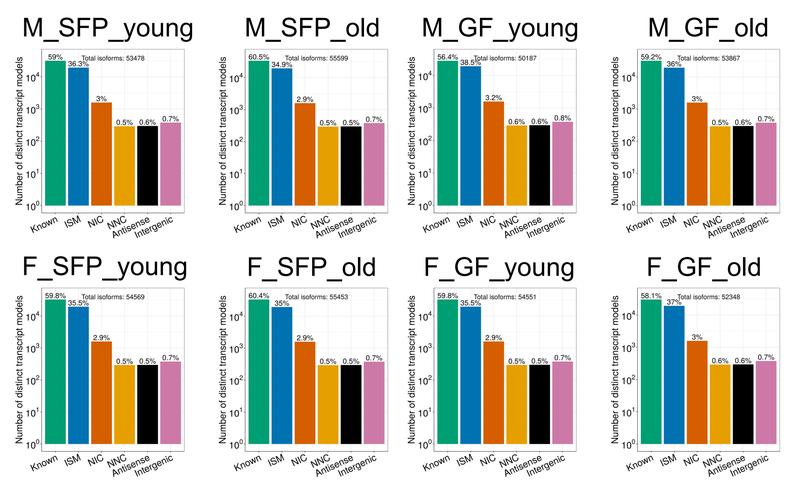
Direct RNA shows relatively better coverage of longer isoforms compared to cDNA-PCR method, Cap-trapper despite the lower sequencing depth, 1~3M reads compared to 10~15M reads respectively.



No. of DEGs across all comparisons. Left most shows no. of DEGs between old and young mice from any condition. The GO terms from DEGs between old and young mice in SPF and GF differ significantly. Indicative of differing molecular aging between the two models.

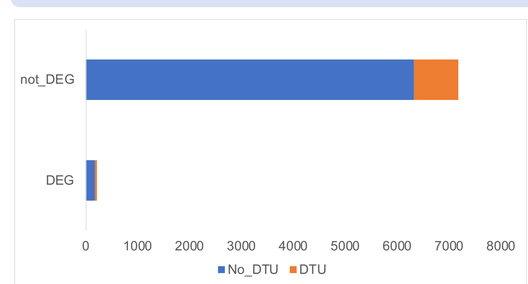
TALON: discovery of novel transcripts from Mouse frontal cortex





TALON discovers 1000s of new transcripts of various categories across the differing conditions from Mouse frontal cortex tissue. Many of which have 5'-end support from CAGE data from Mouse tissue.

Differential isoform usage between young and old mice



Many genes that are not significantly differentiated between young and old mice show at least one or more differentially expression transcripts. Now exploring these more.