# Validation setup for cost-efficient RNA-sequencing of pooled samples

**Per Qvist**[1,2], **Anto P. Rajkumar**[1,2], Jane H. Christensen[1,2], Hayan Song[3], Qin Wang[3] and Anders D. Børglum[1,2]
1) Department of Biomedicine, Aarhus University, Aarhus, Denmark 2) The Lundbeck Foundation Initiative for Integrative Psychiatric Research, *i*PSYCH 3) BGI Europe

## Introduction

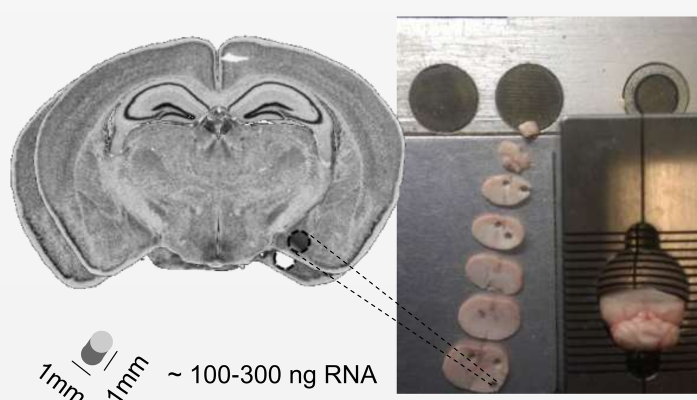
Next generation sequencing (NGS) technology provides fast generation of large amounts of data and has revolutionized genetic research. Despite continuous reductions in NGS costs, the methods are still expensive and not routinely applied in low budget projects. Sequencing pools of individual RNA samples can reduce the cost of RNA sequencing but the validity of such pooling strategy to detect differentially regulated genes remains uncertain. Hence, we aimed to validate a RNA sequencing strategy involving pooling of individual RNA samples, derived from brains of genetically modified mice and of their wild genotype littermate controls.

## Materials & Methods

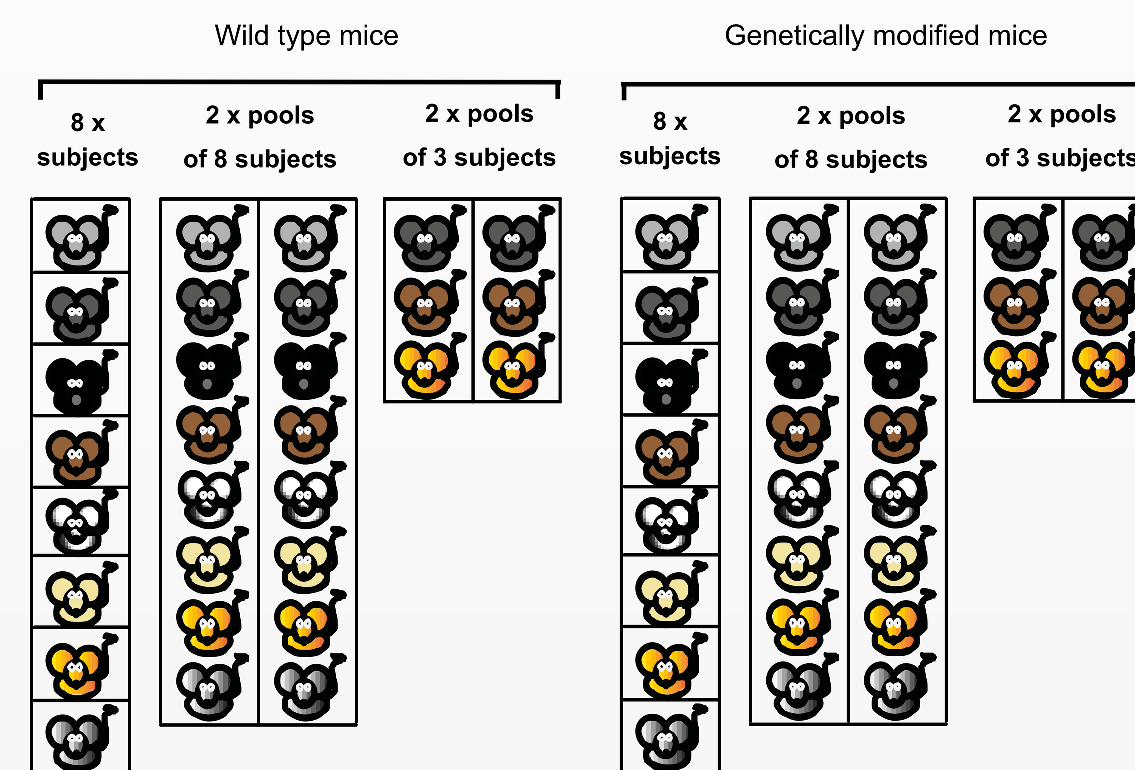
### Tissue Collection & RNA extraction

In the present study, a genetically modifed line of C57Bl/6 mice were investigated. Brains were obtained from 8 wild type and 8 transgenic mice and sectioned manually in a 1 mm coronal mouse brain matrix. Micro-punches (1 mm in diameter) containing amygdala were collected from each section.

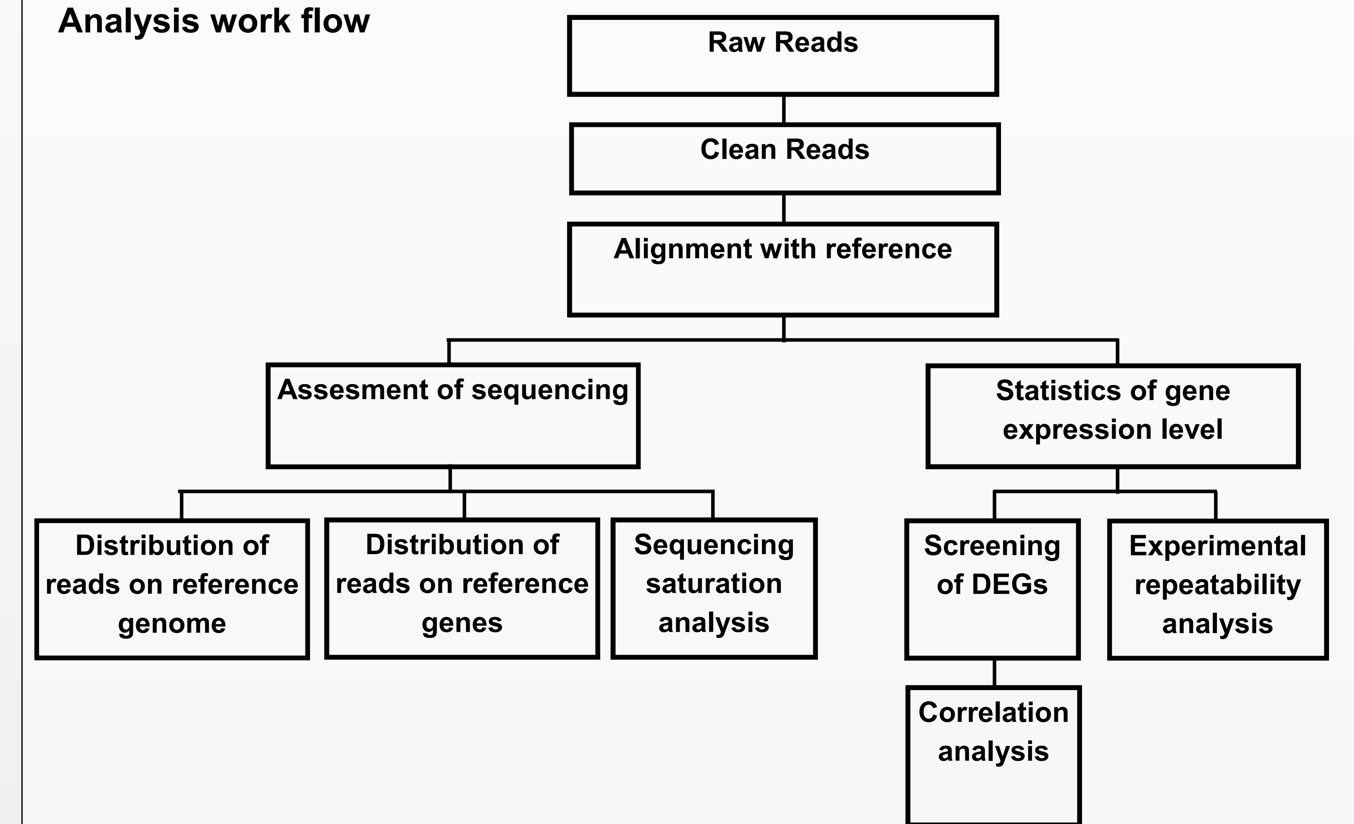
RNA was extracted from tissue using the Maxwell automated system (Promega) and quality assessed on a Agilent 2100 system.



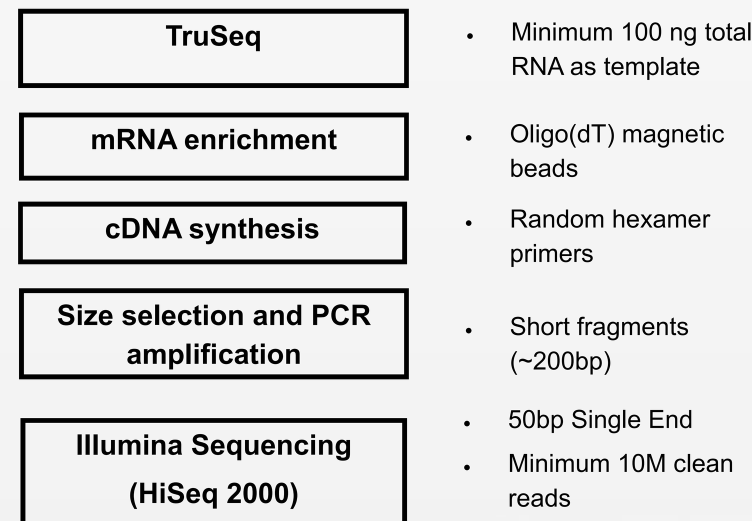

### Pooling strategy




### Analysis work flow




### Library construction & sequencing setup




- Minimum 100 ng total RNA as template
- Oligo(dT) magnetic beads
- Random hexamer primers
- Short fragments (~200bp)
- 50bp Single End
- Minimum 10M clean reads

### Correlation analysis

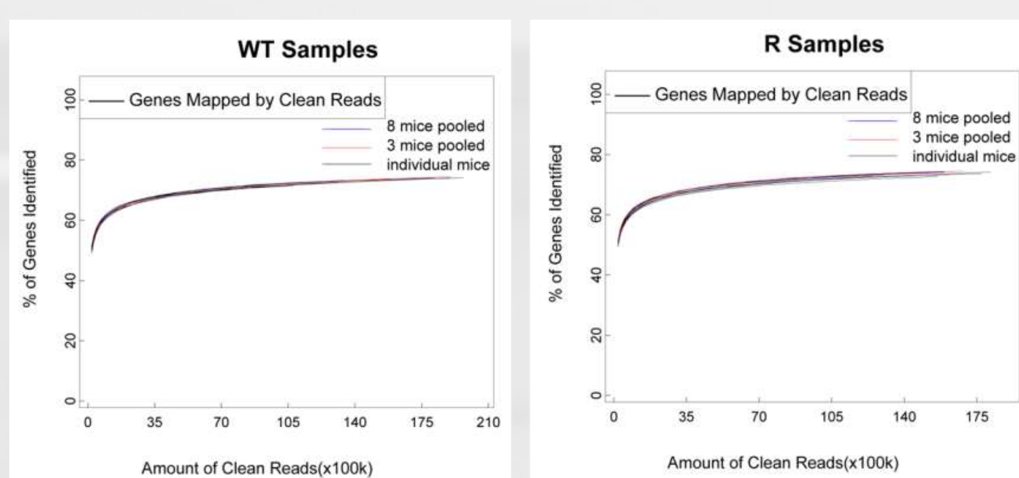
We assessed the correlation between the estimated $Log_2$Ratios of the individual and pooled samples by Spearman rank order correlation. We employed Receiver Operating Characteristic (ROC) curve analysis to demonstrate the ability of estimated $Log_2$Ratios of the pooled samples to predict the Differentially Expressed Genes, identified by individual Samples.

## Results

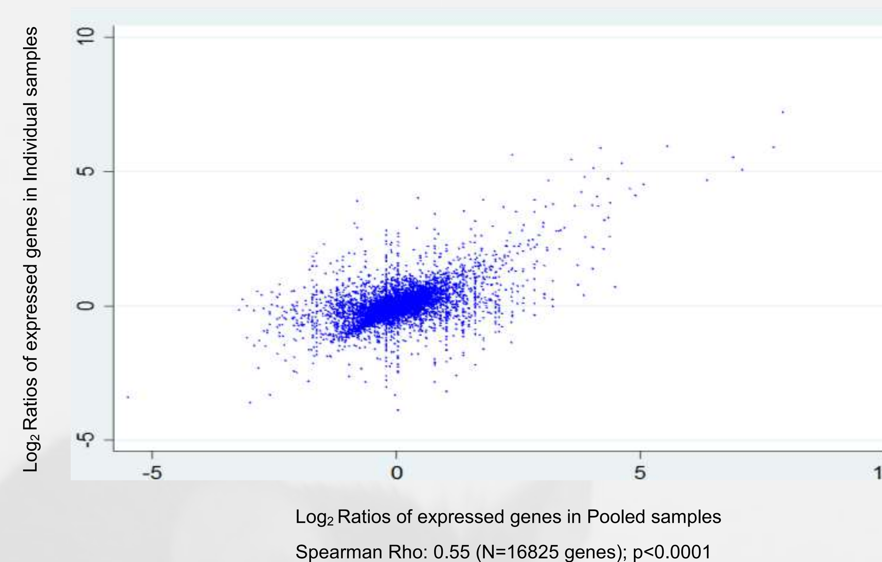
### Read and map statistics

| | | Raw reads | Total mapped ratio | Covered gene ratio |
|---|---|---|---|---|
| Wild type | Individual (average) | 14561877 | 69.88% | 82.71% |
| | 8 pool (average) | 14624534 | 69.59% | 78.20% |
| Genetically modified | Individual (average) | 15361114 | 68.82% | 82.89% |
| | 8 pool (average) | 15917878 | 68.38% | 78.82% |

### Saturation analysis

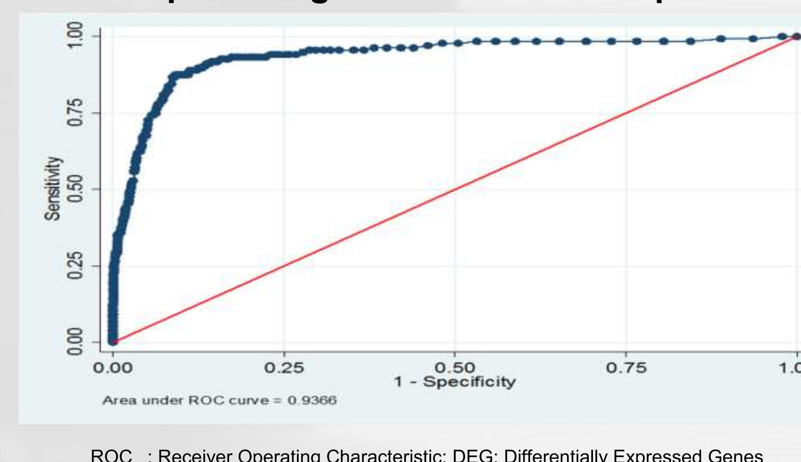



### Correlation between two strategies



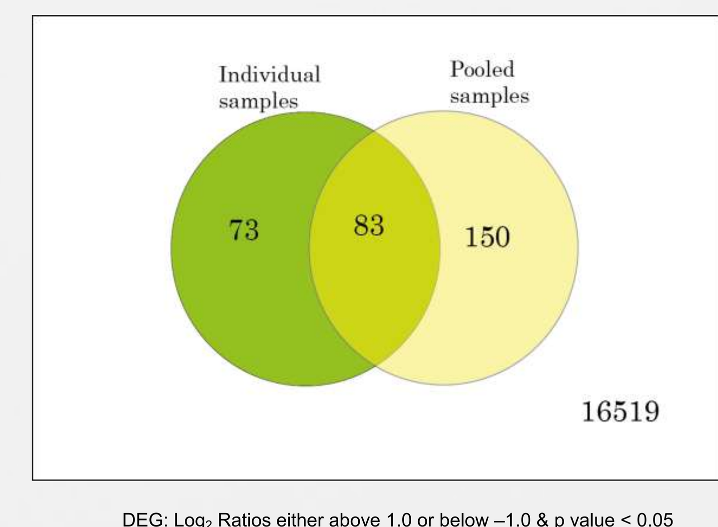

Spearman Rho: 0.55 (N=16825 genes); p<0.0001

### ROC curve to predict the DEG using the $Log_2$ Ratios of expressed genes in Pooled samples




ROC : Receiver Operating Characteristic; DEG: Differentially Expressed Genes

### Reliability of Pooling strategy to identify Differentially Expressed Genes (DEG)




DEG: $Log_2$ Ratios either above 1.0 or below –1.0 & p value < 0.05

| | |
|---|---|
| Sensitivity | 53.21 % |
| Specificity | 98.62 % |
| Positive predictive value | 35.62 % |
| Negative predictive value | 99.56 % |
| Correctly classified genes | 98.67 % |

## Discussion

Quality of pooled samples were comparable to the individual samples and they yielded similar mappable clean reads as well as covered gene ratios. Correlation between the fold changes of expression of genes, estimated by pooled and individual samples was statistically significant. However, the sensitivity of the pooling strategy to identify differentially expressed genes varied depending on the definition of threshold for the fold changes of gene expression.

Sequencing pools of RNA samples is a reliable strategy to measure the fold changes of gene expression and it substantially reduces the cost of RNA-seq. Pooling strategy is desirable, while working with low amounts of RNA samples. However, the validity of this approach to detect differentially expressed genes demands further evaluation.

Investigations to validate the identified differentially expressed genes by qPCR and by Ion Proton RNA sequencing are currently ongoing.

AARHUS UNIVERSITY
DEPARTMENT OF BIOMEDICINE
*i*PSYCH
The Lundbeck Foundation Initiative for Integrative Psychiatric Research