Oxford Genomics CENTRE

Oxford Genomics Centre
The Wellcome Trust Centre for Human Genetics

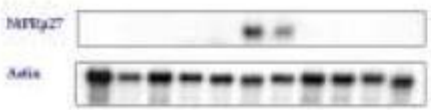
Next Generation Sequencing
Expression analysis

Dr Christine Blancher, Project & Service Manager

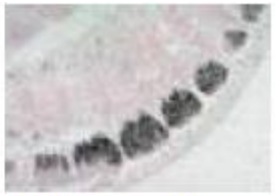
wellcome trust

UNIVERSITY OF OXFORD

Techniques available to study RNA expression

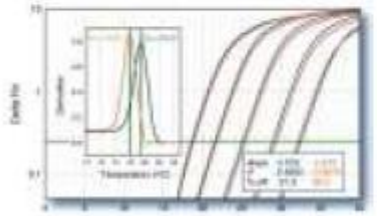


Gel based methods, e.g. Northern blots, RT-PCR

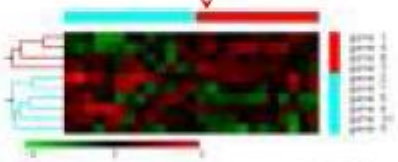


RNA in situ: Quantigene View

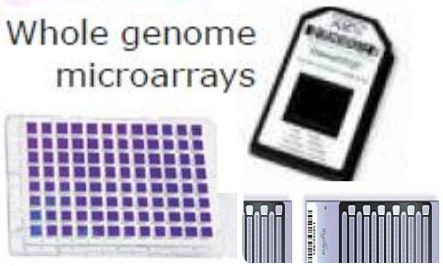
RNA expression



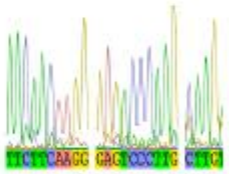
qPCR
Fluidigm Biomark HD



Whole genome microarrays



Affymetrix & Illumina Array



RNAseq

Illumina HiSeq NGS

wellcome trust

UNIVERSITY OF OXFORD

Oxford Genomics CENTRE

Planning your Gene expression project

Project set up: Question, Experimental conditions, Biological replicates, Sample type, Expected RNA yield, RNA extraction kit suited for downstream application.

Discussing with a bioinformatician to address question of sample number and how many biological replicates needed to get significant data, sample randomisation, best platform needed RNA seq or array?

Price, amount of data needed, bioinformatics analysis support available.

Expression Array Key limitations

- **Background hybridisation issues**

  Affects detection of low-level expression

  Correction algorithms improve accuracy but worsen precision (see Irizarry et al., *Bioinformatics* **22,** 789–794 (2006))

Low expression

- **Limited dynamic range**

  3-4 logs typically (as reported on array data sheets)

Small changes

- **Under-estimation of logFC**

  see Shi, L. *et al. BMC Bioinformatics* **6 Suppl 2,** S11 (2005) for review

- **Unable to uncover novel features**

  Requires prior knowledge of targets

Biased

RNA Extraction Methods

Different kits available depending on origin of samples (Blood, cells, tissue, plant…), availability of material (more than 100ng or less) and RNA quality (Cells, Biopsy, FFPE samples…).

Think ahead!

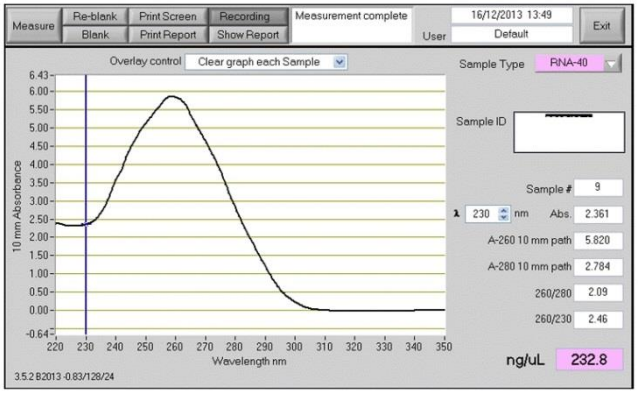
- If you wish to analyse **miRNA** on the same samples later on use a kit that isolate Total RNA including small RNAs.

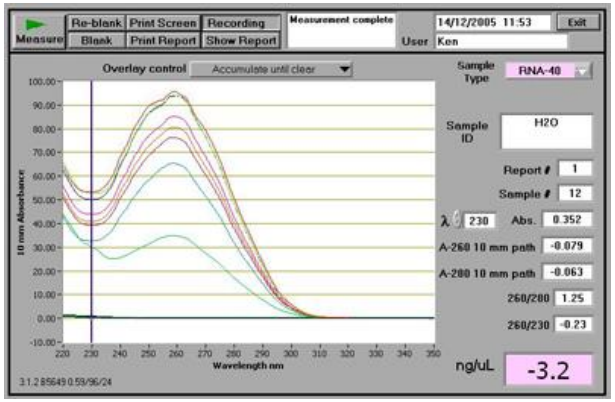
- 70% of the total mRNA in blood is globin mRNA and up to 95% can be removed with **globin depletion kits**.

- **Ribosomal depletion** and **mRNA selection** are important for RNA Sequencing but it is not necessary for microarray as the step of Reverse transcription into cDNA will eliminate rRNA.

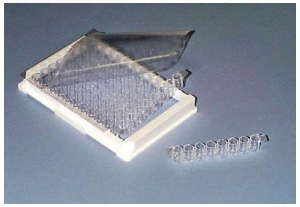
RNA Quantification & QC

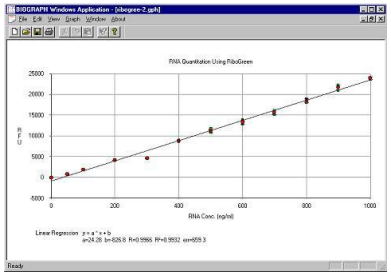
Nanodrop spectro



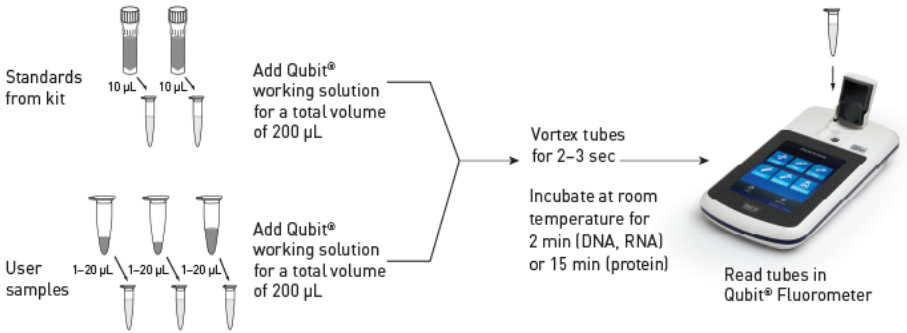


Quant iT Ribogreen assay kit





Qubit Assay



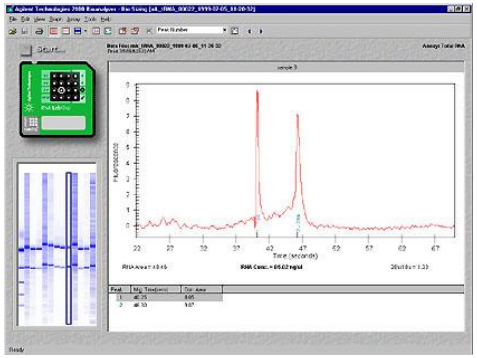
wellcome trust

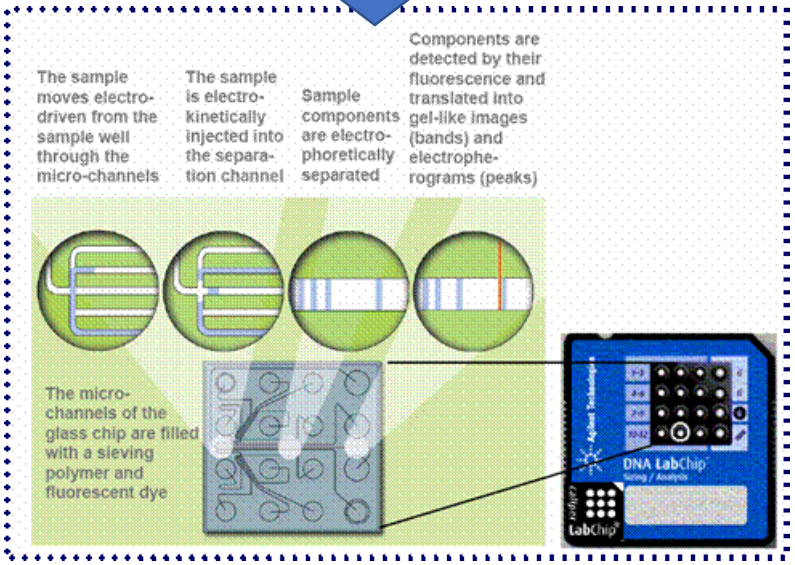
UNIVERSITY OF OXFORD

Oxford Genomics CENTRE

RNA Quantitation & QC

Agilent Bioanalyser





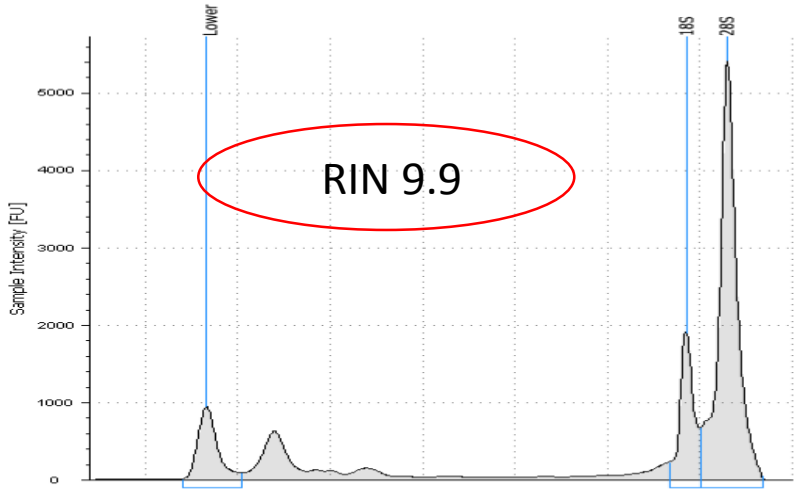
wellcome trust

UNIVERSITY OF OXFORD

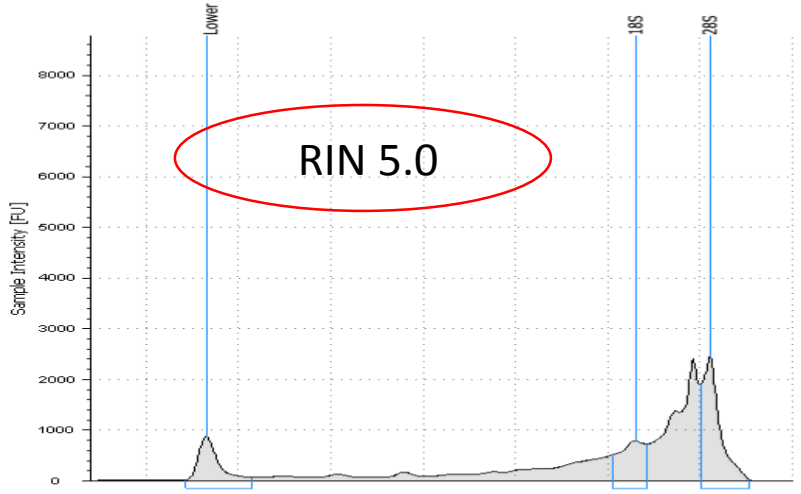
RNA quantification & QC

The **RNA integrity number** (**RIN**) is an algorithm for assigning integrity **values** to RNA measurements.

RNA



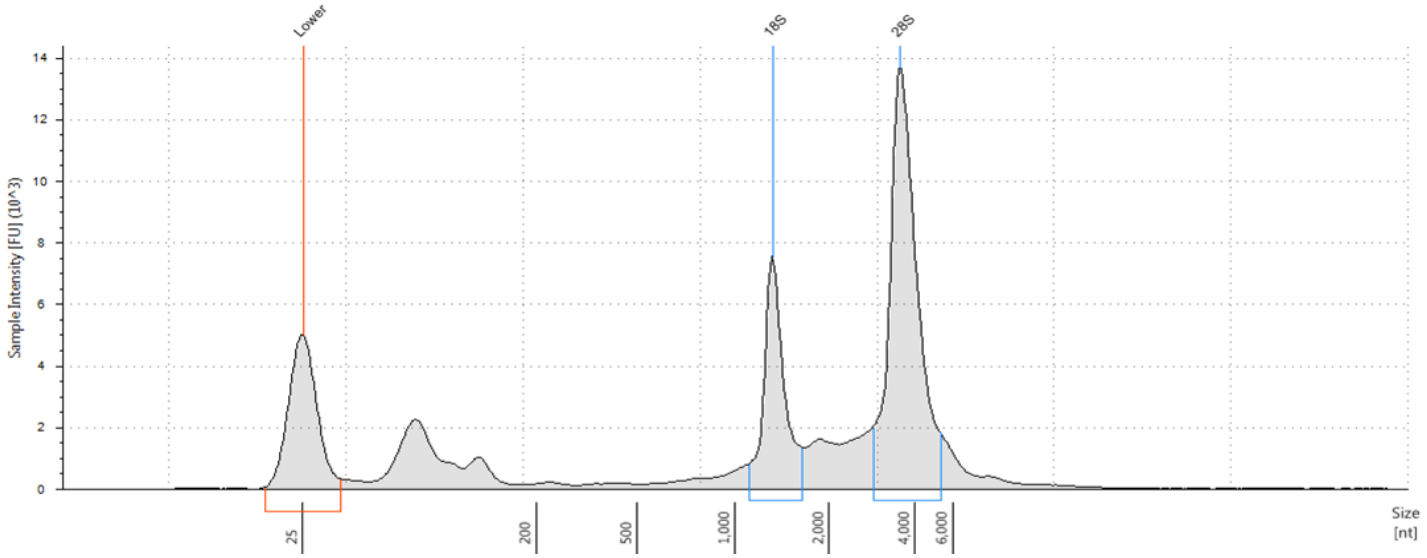
✓



✗

RNA Specification

- 260/280 ~ 1.8 – 2.0
- 260/230 ~ 1.8
- RIN ≥ 7.0
- 28S/18S ≥ 1.6



Common RNA library types

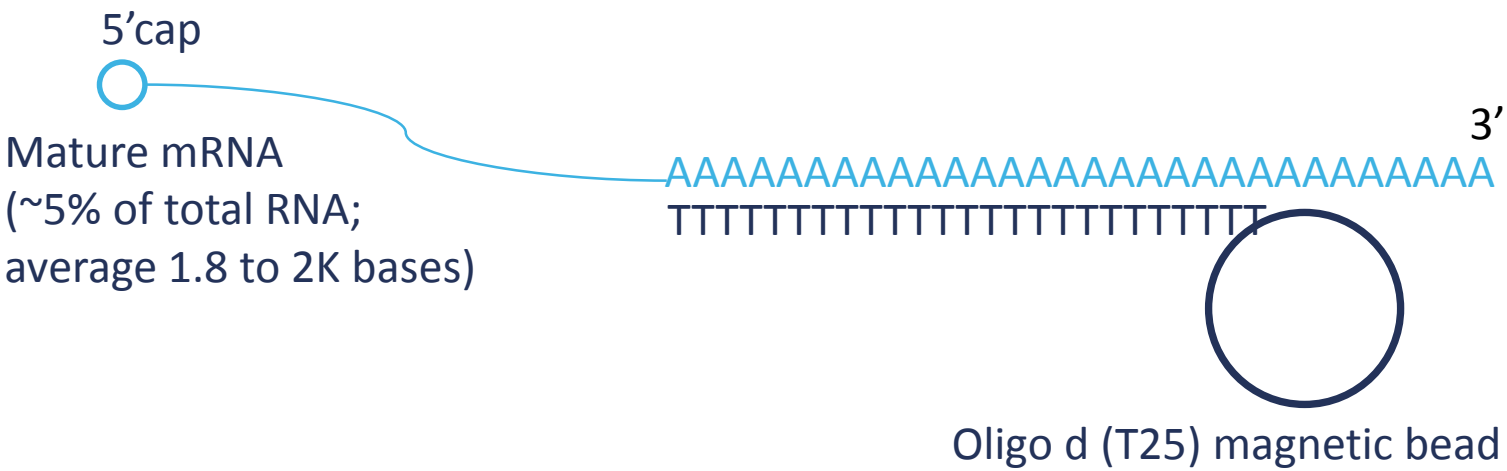
- Enriched for polyadenylated transcripts (poly-A)
- Depleted for ribosomal RNA transcripts (ribo-depleted)
- Amplified from low input material (SMARTer)
- Small RNA species
- 3’ mRNA

Poly-A

Polyadenylation is the post-transcriptional modification added to the 3’ end of mRNA

Polyadenylation is required for nuclear export and stability of mature transcripts and for efficient translation of mRNAs

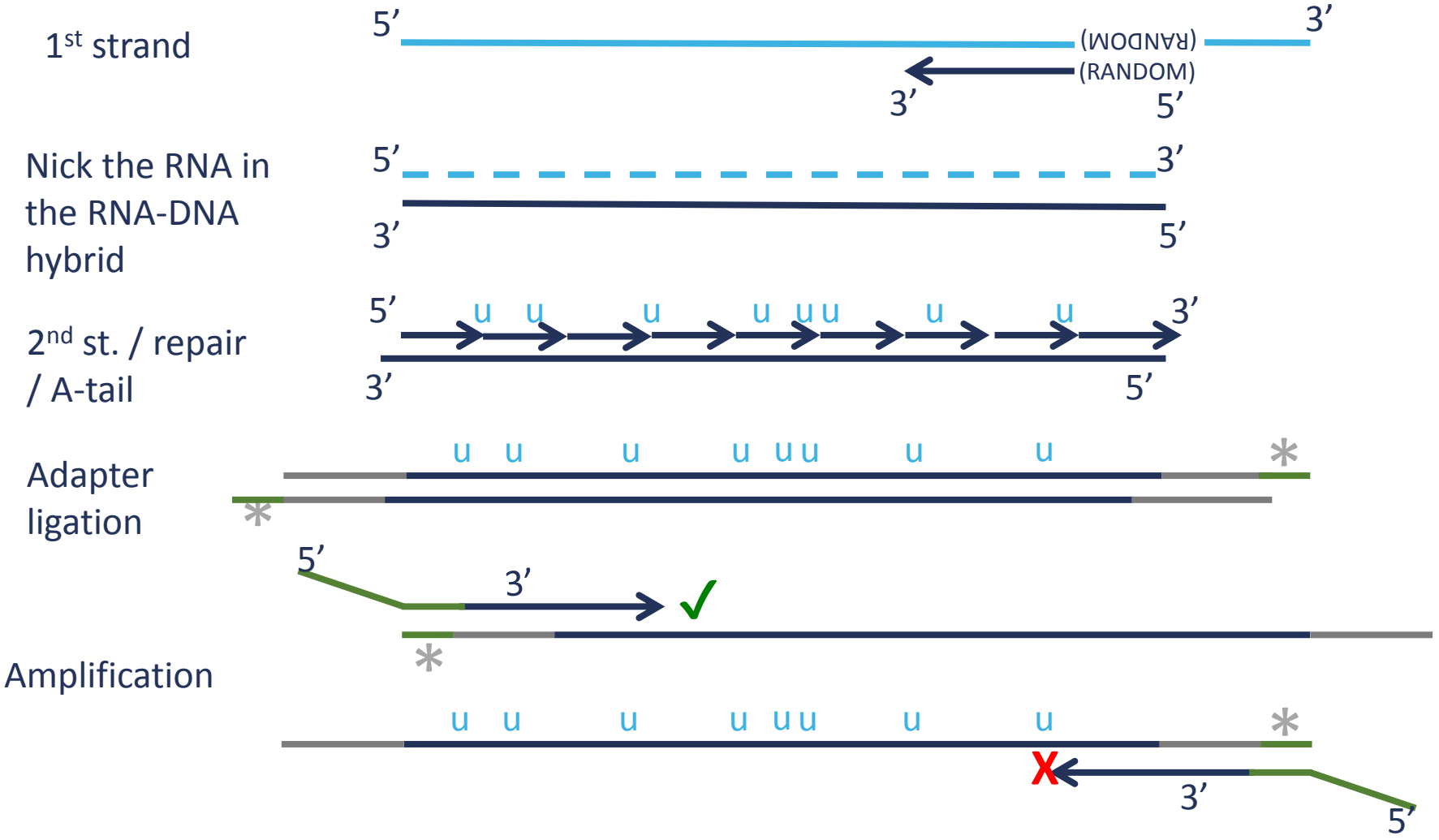
PolyA Library Preparation (TruSeq Illumina)



- Fragmentation
- 1st strand cDNA synthesis
- 2nd strand cDNA synthesis*
- End repair and A-tail
- Adapter ligation
- Index-incorporating PCR amplification

*directionality ensure mapping to coding strand – good for discovery and quantification of antisense transcripts and overlapping genes

RNA Library Preparation with directionality



wellcome trust

UNIVERSITY OF OXFORD

PolyA Libraries

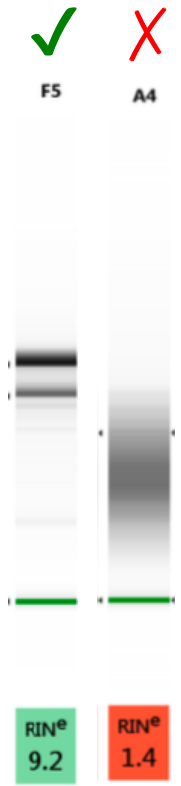
+ve

-focus on mature protein coding transcripts

-good use of lane capacity

-ve

-seq data quality influenced by sample quality



Standard project

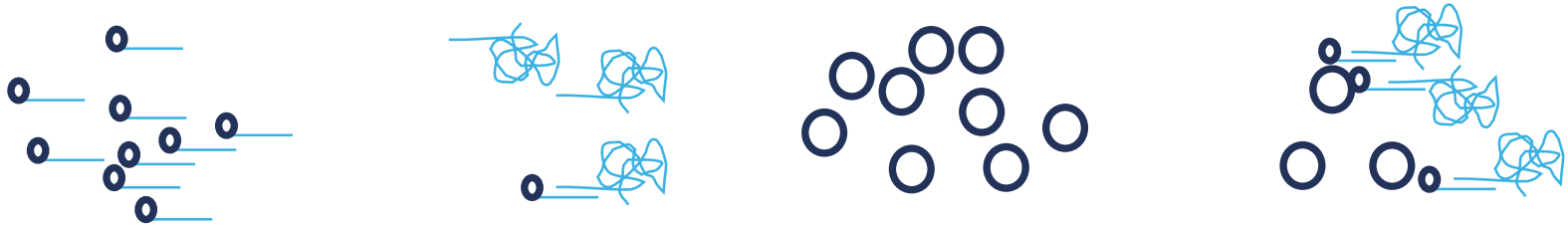
Input: purified total RNA 100 ng to 1 ug

Multiplex: equivalent of 8-10 libraries per lane (HiSeq 4000)

Read length: 75 or 150 bp PE

Ribodepleted Library Preparation (TruSeq Illumina)

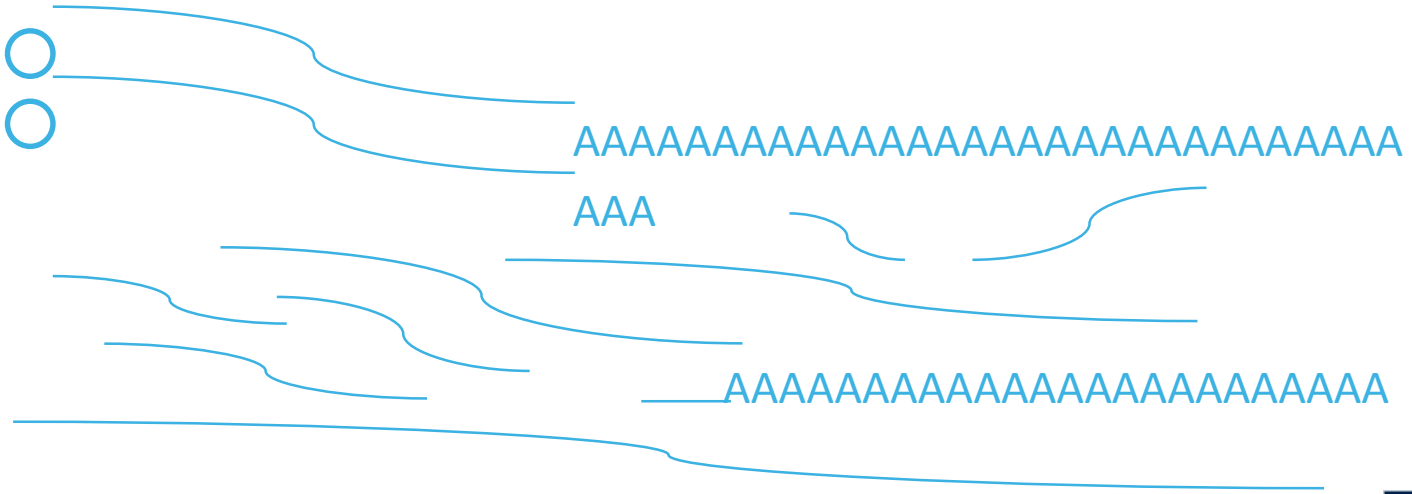
Ribosomal RNA is the RNA component of the ribosome. The ribosome is a complex of RNA and protein responsible for protein synthesis (translation). rRNA sequences can make up ~80% of total RNA.



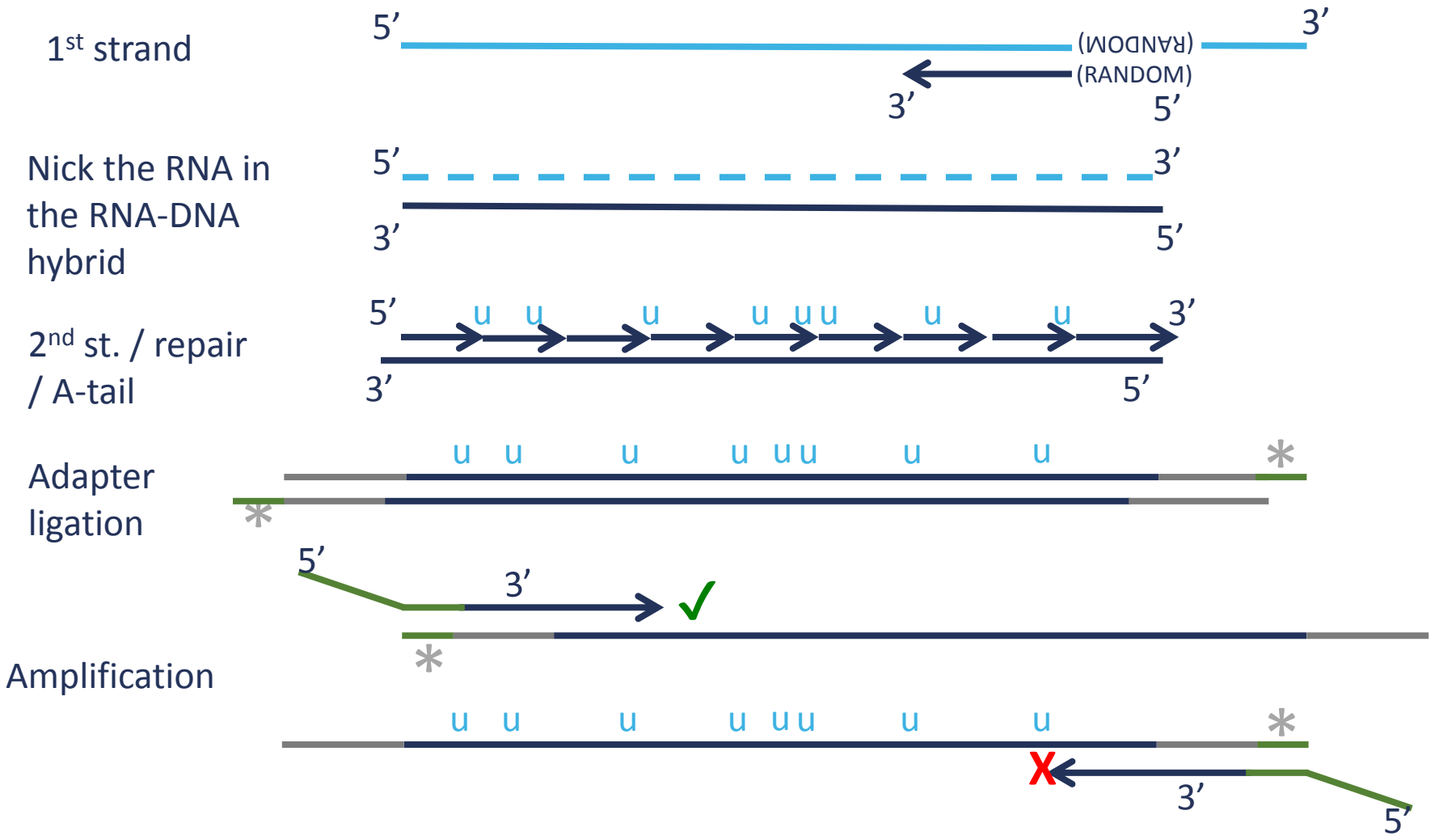
Biotinylated RNA probes

Streptavidin magnetic beads

Discard



RNA library preparation with directionality



wellcome trust

UNIVERSITY OF OXFORD

Ribodepleted Libraries

+ve

-good for poorer quality RNA (e.g. FFPE)

-Sequencing data included non-polyadenylated transcripts

(nascent pre-mRNA (unspliced), and a number of other functionally relevant long non-coding transcripts)

-probes can be combined for other highly expressed transcripts

-ve

-additional sequencing depth required

- some issues with probe efficiencies

Standard project

Input: purified total RNA 100 ng to 1 ug

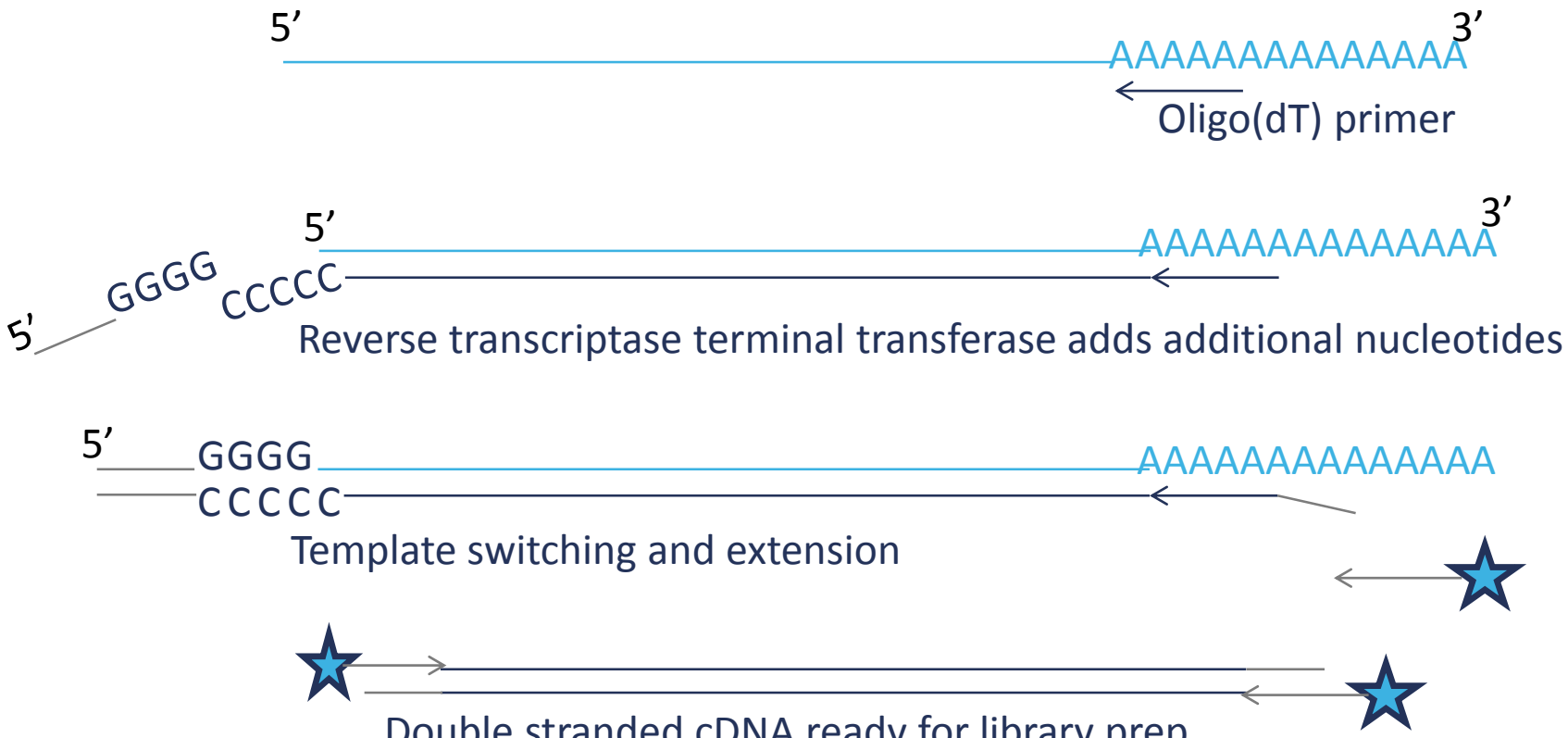
Multiplex: equivalent of 4-6 libraries per lane (HiSeq 4000)

Read length: 75 or 150 bp PE

Low RNA input library prep SMARter (Clontech)

Limiting RNA can be amplified to sufficient quantities of double stranded cDNA in order to generate libraries.

SMARTer technology using total RNA or single cells



wellcome trust

UNIVERSITY OF OXFORD

SMARTer Libraries

+ve

-low input (<10 ng) down to single cell (~10 pg)

-ve

-loss of directionality in the data

-3’ bias

-variable adapter sequence in sequencing read 2

Standard project

Input: purified total RNA <10 ng

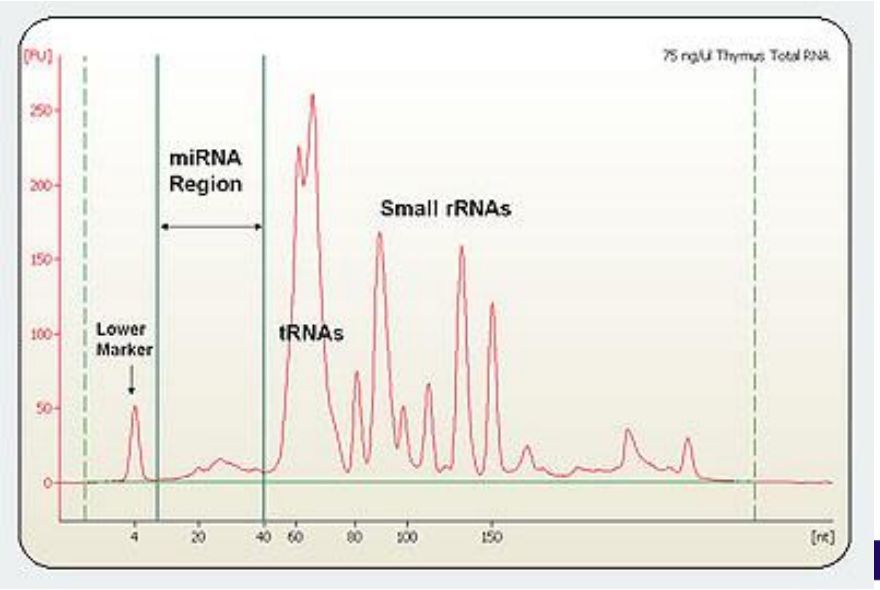
Multiplex: equivalent of 8-10 libraries per lane (HiSeq4000)

Read length: 75 or 50 bp, PE or SR

Small RNAs

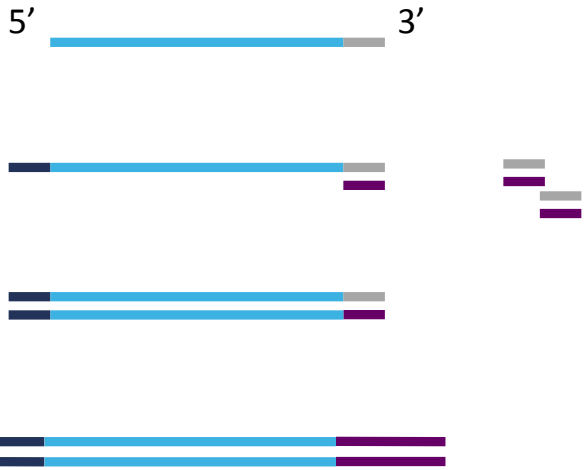
<200 nt non-coding regulatory RNA

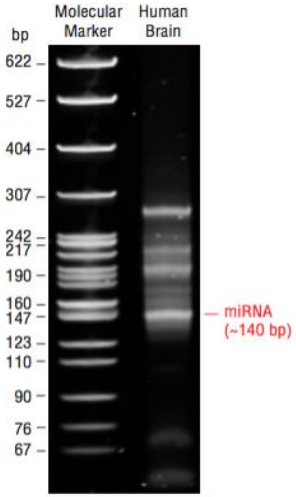
Main class of interest are 21-25 nt miRNA/siRNA involved in gene regulation through interaction with cognate mRNA transcripts or through directing epigenetic alteration such as DNA methylation and histone modification

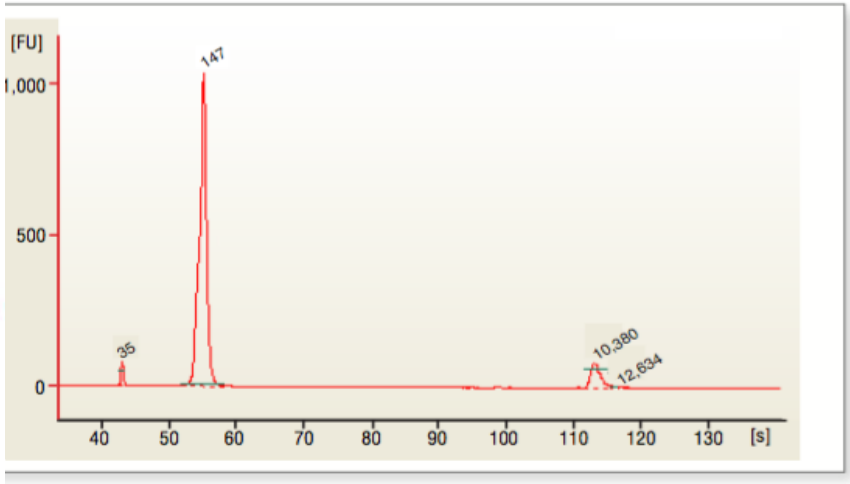


Small RNA Library Preparation (NEB)

- 3’ adapter ligation
- 5’ adapter ligation
- 1st strand cDNA synthesis
- PCR enrichment
- Size selection







wellcome trust

UNIVERSITY OF OXFORD

Small RNA Libraries

+ve

-highly selective for size range
-directional
-low input possible (eg. circulating miRNA)

-ve

-lower efficiencies with degraded material

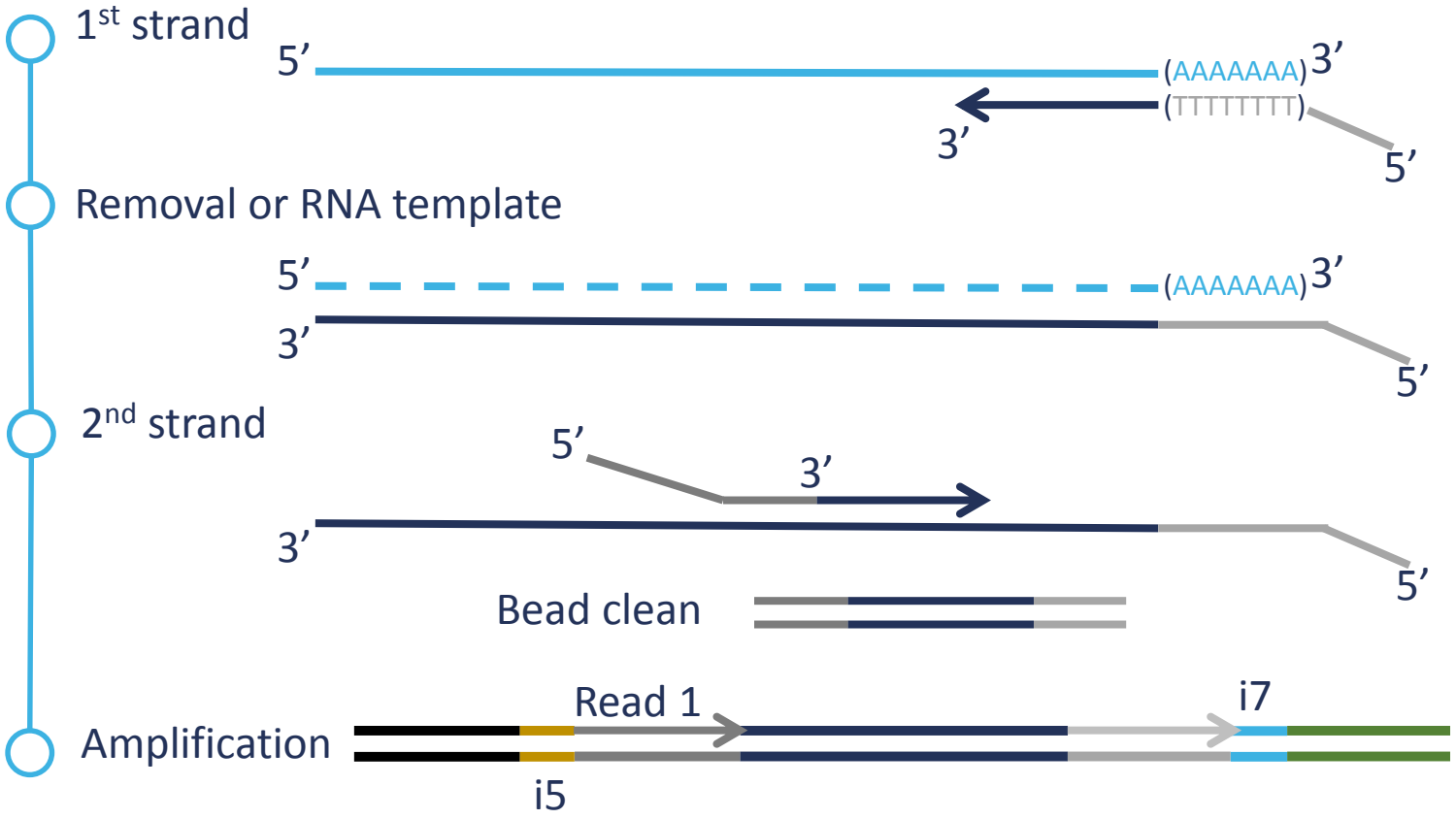
Standard project

Input: purified total RNA 1 ug
Multiplex: 20+ libraries per flow cell (HiSeq2500)
Read length: 50 bp SR

3’ mRNA Library Preparation (Lexogen)

- Relative gene expression analysis is based on ‘counting’ sequencing read alignments to annotated features of a reference genome, while compensating for the different transcript sizes.
- Good quality sequencing reads can be mapped to the full length of the reference feature and thus can also provide information for variant splicing events from the same gene.
- Another approach for expression analysis is to limit the sequenced read to the 3’ end of the transcript. The ‘counting’ is then based on the same region of each gene.

3’ mRNA Library Preparation (Lexogen)



Library amplification and bead cleaning leads to libraries with insert sizes between 200 and 300 bp (shift to 360bp possible)

3’ mRNA Libraries

+ve

-lower depth of sequencing required

-tolerance to degradation (eg FFPE)

-speed of analysis (no transcript length normalisation)

-ve

-no splice variant information

Standard project

Input: purified total RNA (0.5 ng to 2 ug)

Multiplex: equivalent of 48 samples per lane (HiSeq4000)

Read length: 50 or 75bp, PE (read 1 only used in analysis)

Oxford Genomics Centre