

第10回 シーケンス講習会

RNA-seq library調製法の特徴と選び方

理化学研究所 (RIKEN)
ライフサイエンス技術基盤研究センター (CLST)
機能性ゲノム解析部門 (DGT)
ゲノムネットワーク解析支援施設 (GeNAS)

野間 将平

- シーケンスをする目的は？
- よいシーケンスライブラリーとは？
 - RNA-seq ライブラリーのムリ・ムダ・ムラ
- いろいろなRNA-seqライブラリーの特徴
 - 性能比較実験の結果から

RNA シーケンスをする目的

- RNAを研究することで分かること
 - RNA-seq は細胞・組織の”今”の状態を知る
- ライブラリー作製・シーケンスは解明のための通過点(一手段)
 - RNA抽出が完了した時点で結果は出ている
 - そこから いかにかに情報を失わず、シーケンスに持ち込むか

減らしたいRNA シーケンスのムリ・ムダ・ムラ

• ムリ

- RNAが本来持っていた情報が正確に反映されていない状態
 - 結果にバイアスがかかった状態
 - RNAの品質、サンプル量

• ムダ

- RNAの持つ情報は極力失いたくない
 - とはいえ、Total RNA をそのままシーケンスするとほとんど(9割) がrRNA
 - 現行のシーケンサー能力やコストを考慮するとrRNA を効率的に除去する必要がある
 - Oligo dT Beads
 - RiboZero (Epicentre/Illumina)
 - GeneRead rRNA Depletion Kit (QIAGEN) etc...
- PCR条件の最適化
 - PCR duplicate の 割合を減らす

• ムラ

- 結果が再現しない
 - サンプルング、RNAの品質、実験環境、手技

Stranded RNA-seq 手法比較

- 評価対象
 - TruSeq Stranded RNA (Illumina)
 - ScritSeq v2 (Epicentre)
 - RNA ligase base method (GeNAS)
- Human Brain total RNA 1ug を使用して各手法 n=3 でlibrary 調製
- rRNA 除去にはRiboZero Goldを使用
 - Non-coding RNAも含まれる
- 3mix ibrary / lane でHiSeq2500 100PE でシーケンス
- Genomic Work Bench (CLC Bio) で解析

TruSeq Stranded RNA Sample prep

cDNA synthesis /Ligate adapter / PCR amplification



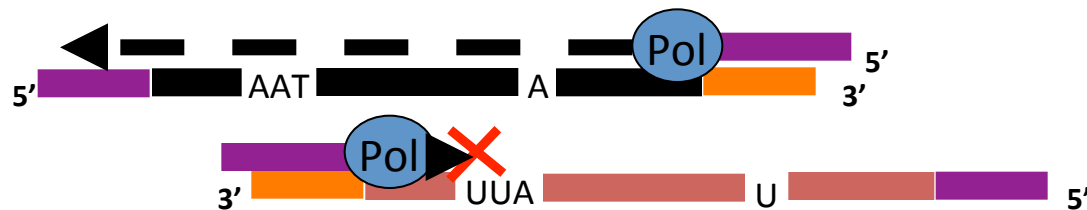
1st strand cDNA synthesis with random hexamer



2nd cDNA synthesis with random hexamer
Incorporates dUTP instead of dTTP



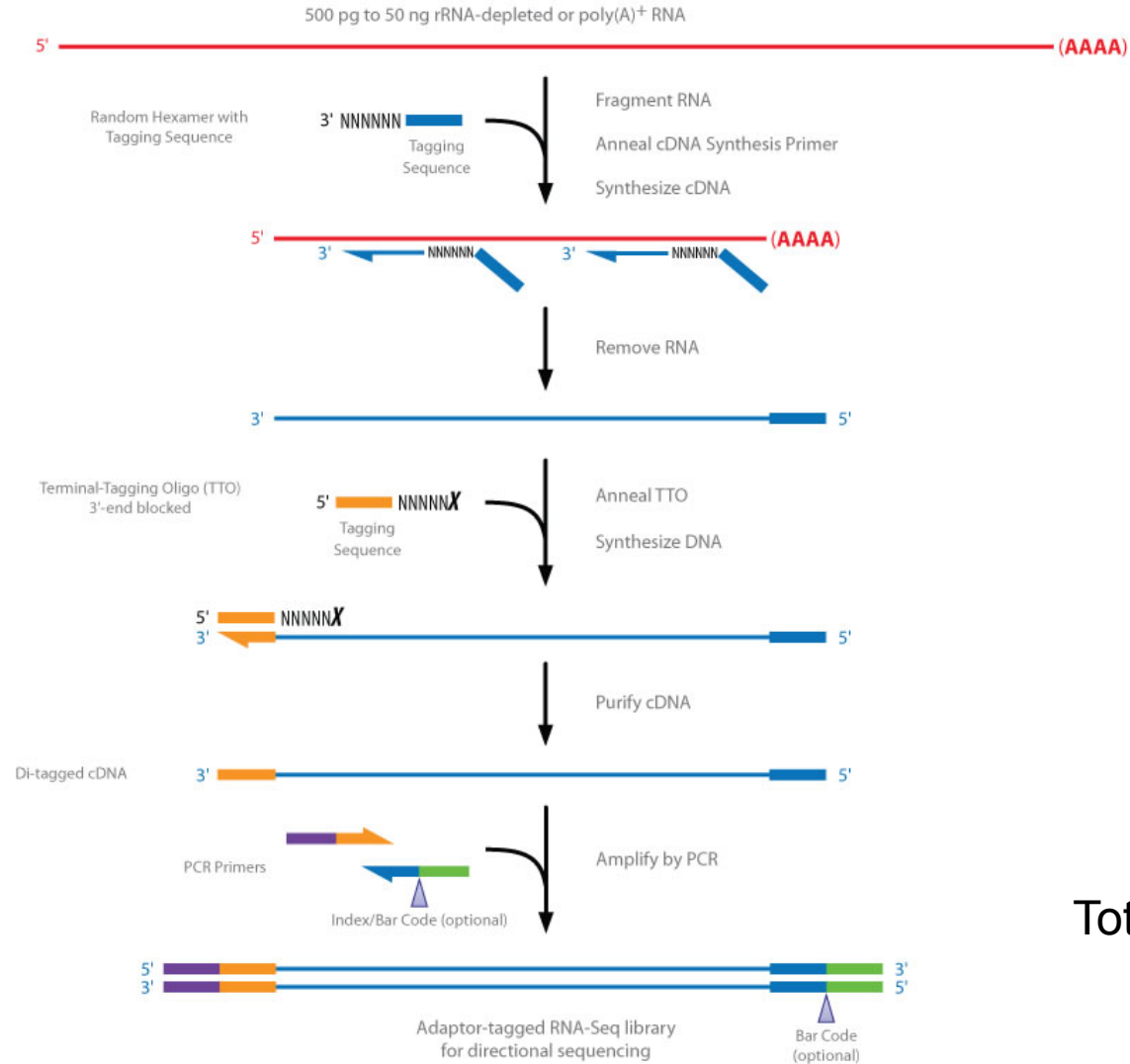
End repair, adenilation
Adapter ligation



Amplify with High-fidelity Taq
Strand selective amplification

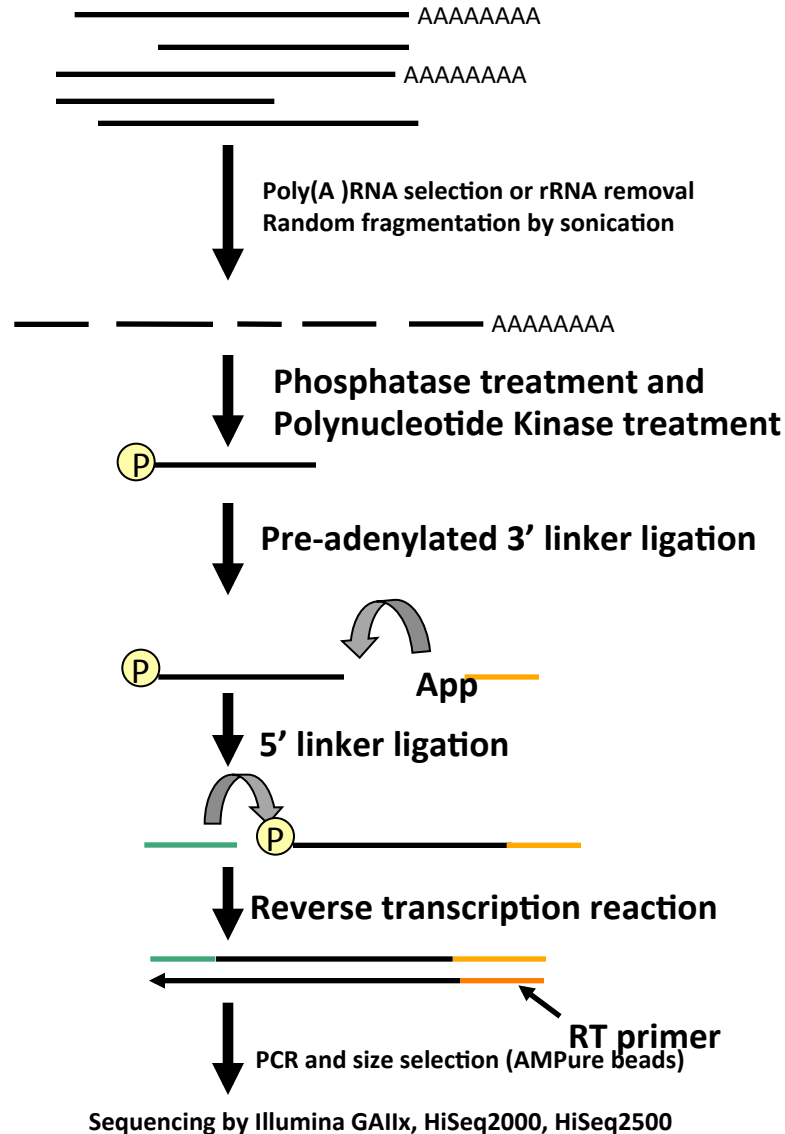
Total 2day

ScriptSeq v2



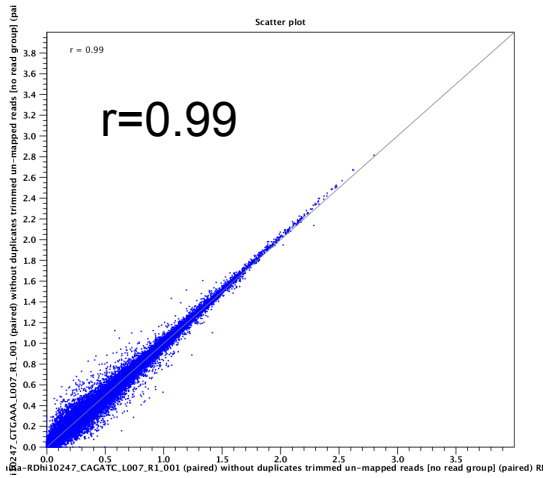
Total 1day

RNA ligase base method

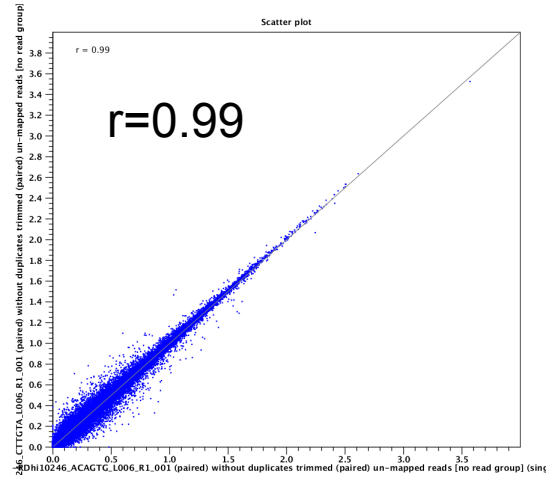


Total 3day

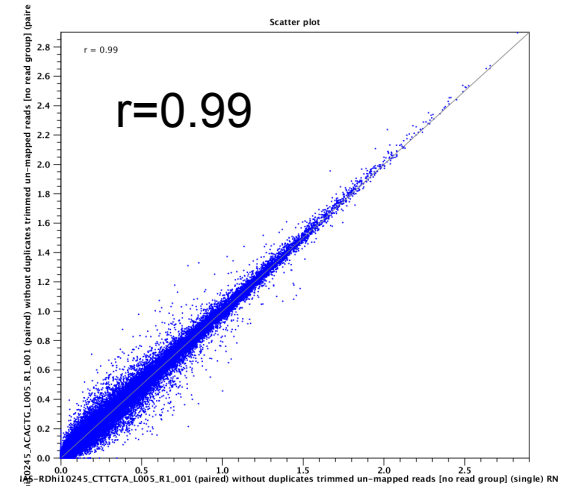
再現性 Technical replica



TruSeq stranded



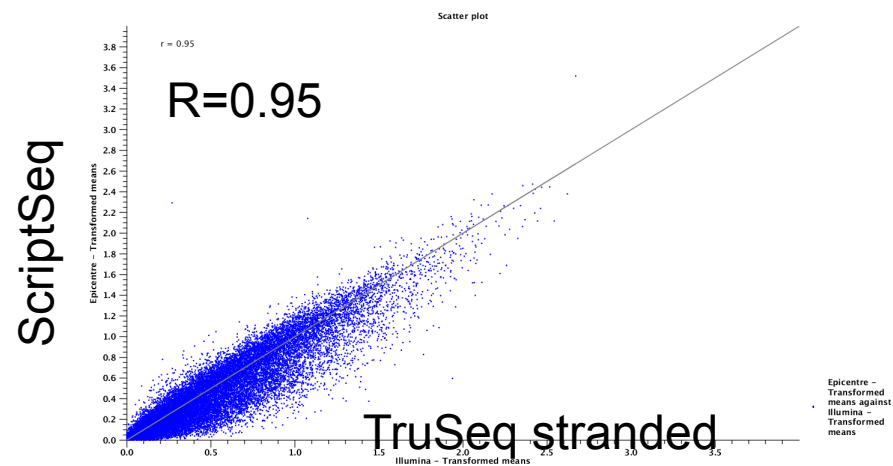
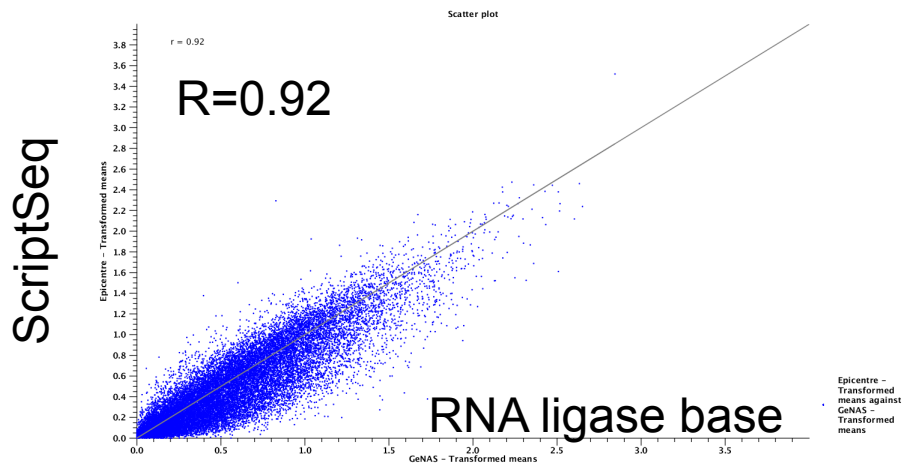
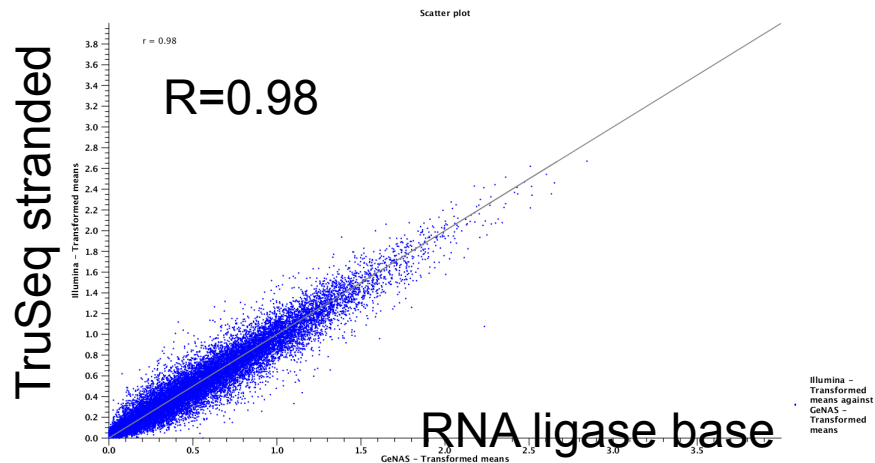
Epicentre



GeNAS
RNA ligase base

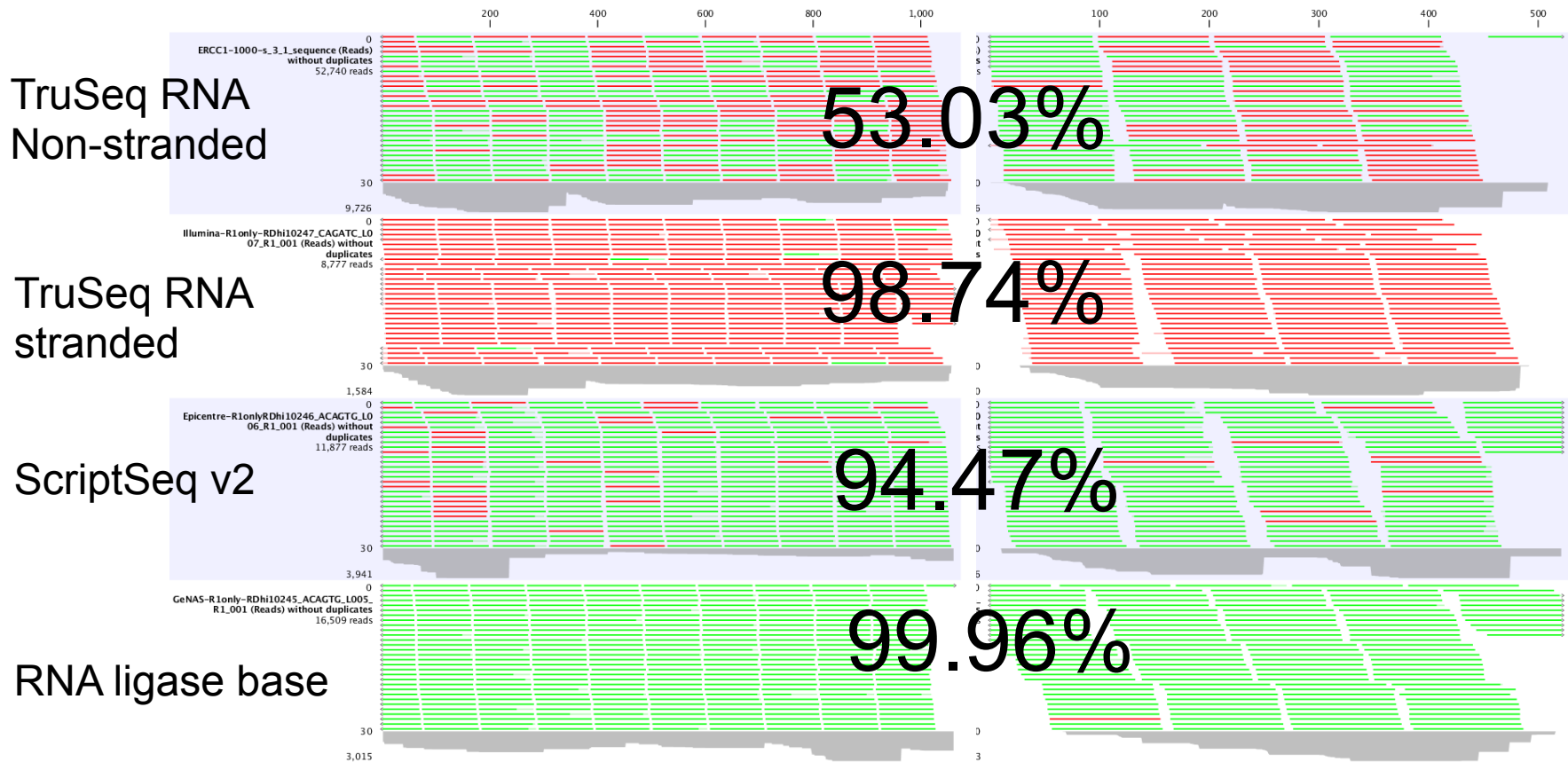
hg19 refseqにmappingしてRPKM算出
Log10(1+RPKM) でプロット
各手法とも再現性は高レベル

手法間での発現量相関



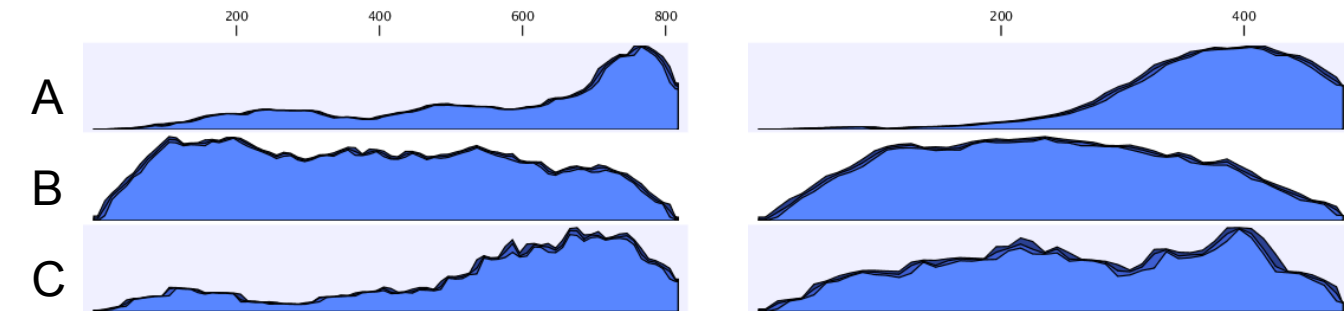
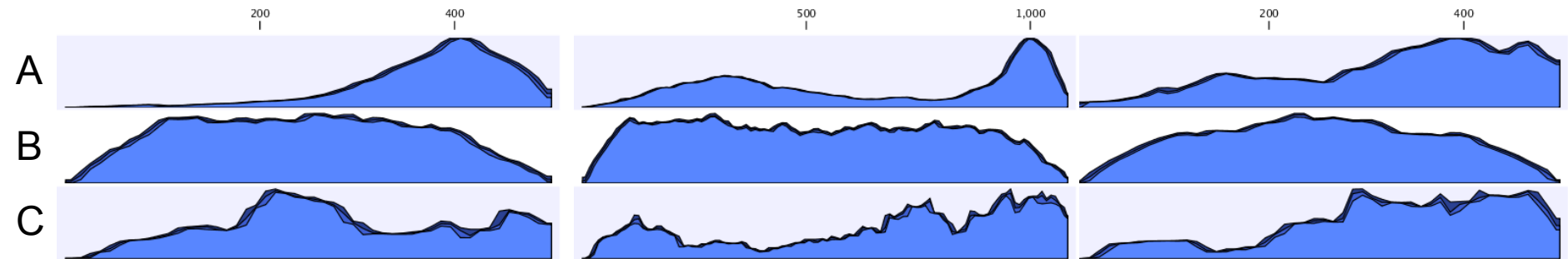
異なる調製法のライブラリーを単純比較するのは危険

Strand Specificity



- Read1 のFastq をCLC Genomics Workbench 6.0.2にimport
- hg19, ERCC ref seq にmapping
- PCR duplicateを除去
- ERCCに対するreadの方向を比較

Coverage evenness



A: RNA ligase base
 B: TruSeq Stranded
 C: ScriptSeq

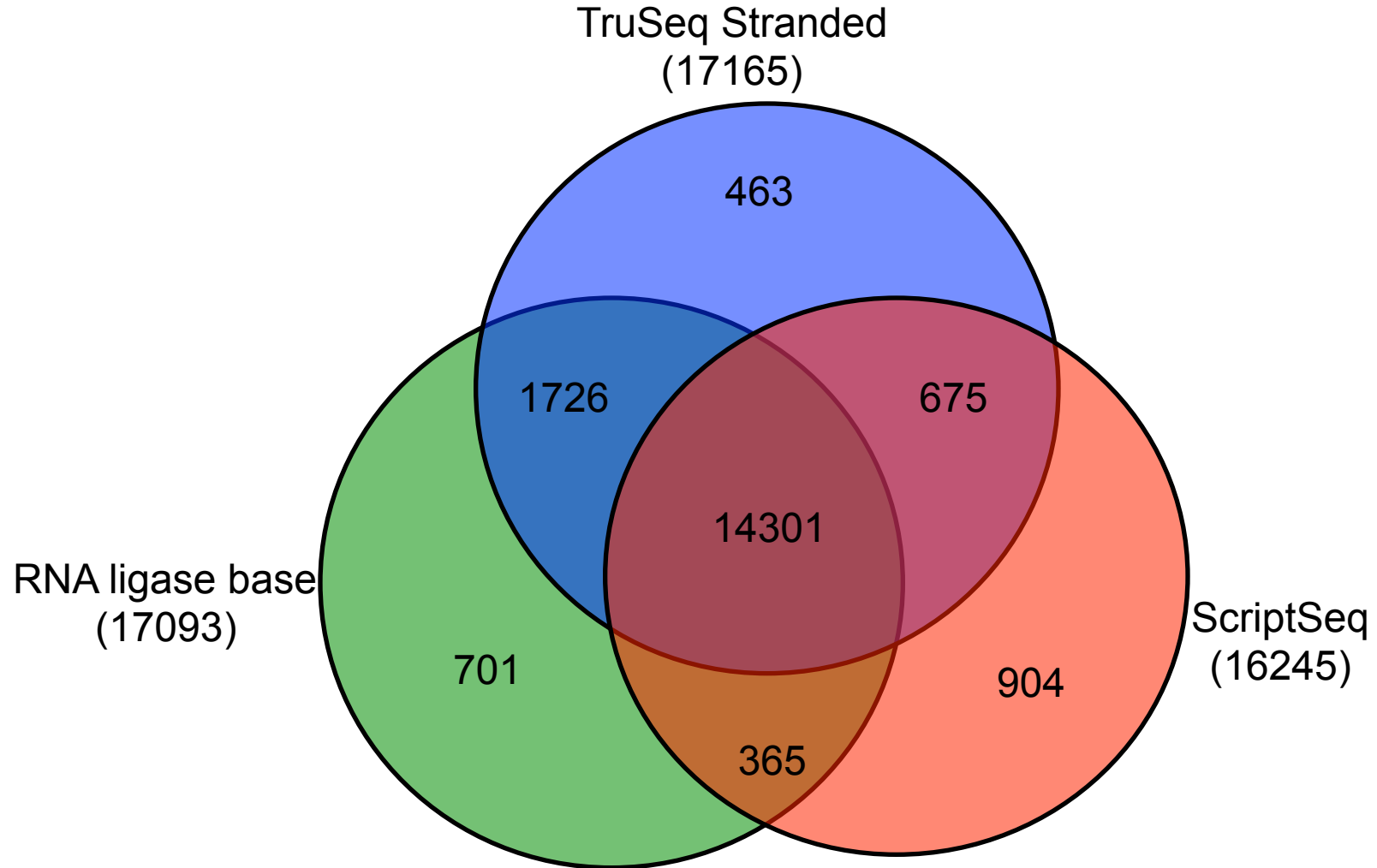
TruSeq Strandedが安定して全体をカバーできている

他はcoverageに偏りあり

Replicate間で偏り傾向は一致していたので手法に依存したバイアスがある推測される

Coverageスケールはそれぞれ異なる

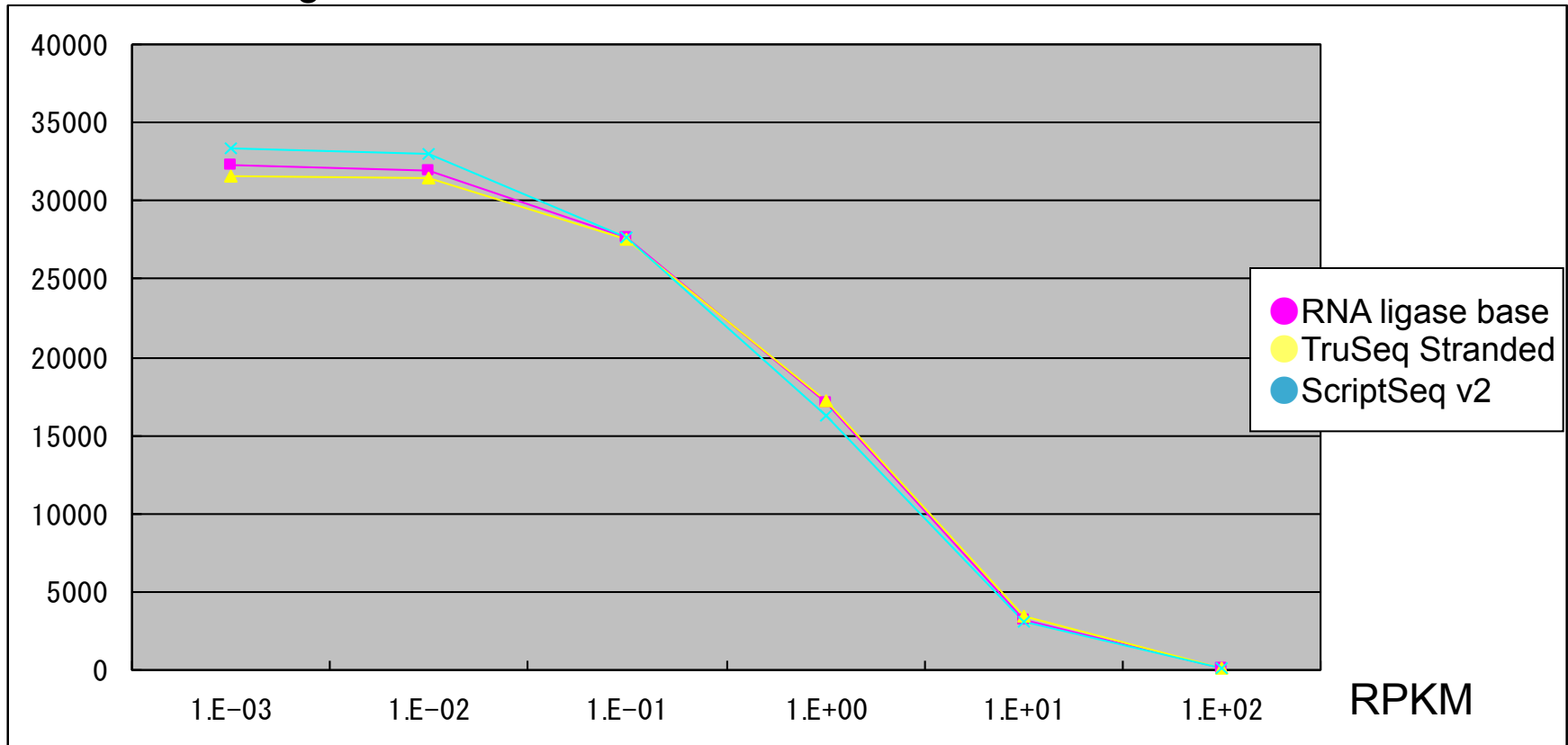
検出遺伝子数



TriplicateでRPKM 1< 示したものをカウント

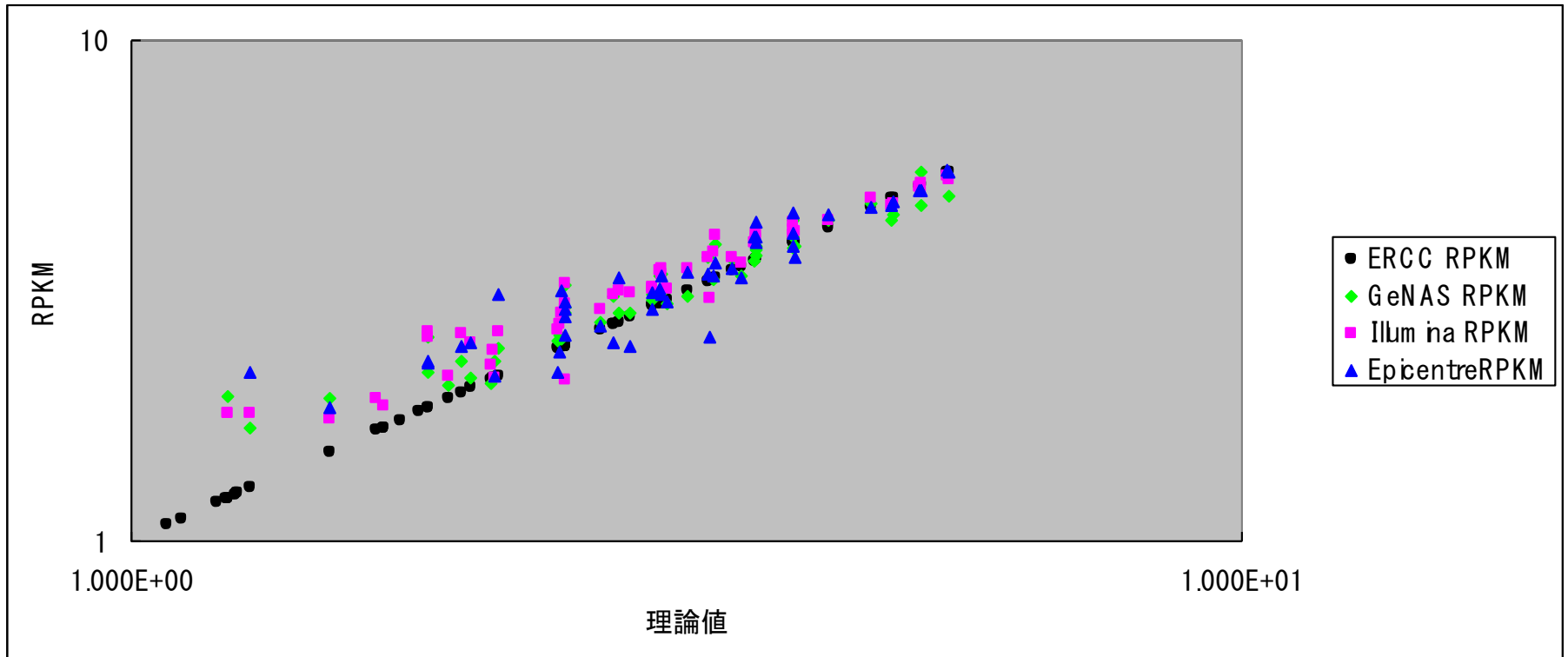
検出遺伝子数

Number of gene detected



一定のRPKM以上で検出された遺伝指数

ERCCによる定量性の検証



Read count 10未満を外れ値として除外してプロット

RPKMで取ったスピアマン相関係数

RNA ligase base : 0.90

TruSeq Stranded : 0.94

ScriptSeq v2 : 0.91

手法間の相対比較

- 手法毎に得手・不得手はある
 - 現時点でTruSeq Strandedが相対的に優れている点が多い

手法名	Illumina (TruSeq Stranded)	Epicentre (ScriptSeq v2)	GeNAS (RNA ligase base)
必要サンプル量	○ 0.1-1.0ug	○ 0.1-1ug	△ 1ug≦
コスト	△ 48反応/kit	△ 6反応/kit	△ 試薬を個別に用意 する必要あり
操作性	○ 2day	◎ 1day	△ 3day
再現性	◎ R=0.99	◎ R=0.99	◎ R=0.99
Strand Specificity	○	△	◎
Coverage evenness	○	△	△
定量性	◎	○	○

◎極めて優れる ○ 優れる △ 普通

Agilent, NuGEN などからもStrand specificな RNA-seq library prep kit は販売されている