# An assessment of q-PCR accuracy based on RNA-seq

presented by V. Sherina

October 7, 2015

# Introduction

The **National Center for Biotechnology Information** (NCBI) is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health.

The NCBI is located in Maryland and was founded in 1988 through legislation sponsored by Senator Claude Pepper.

# Introduction

The **National Center for Biotechnology Information** (NCBI) is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health.

The NCBI is located in Maryland and was founded in 1988 through legislation sponsored by Senator Claude Pepper.

# Introduction

The NCBI houses

- a series of databases biotechnology and biomedicine;
- and an important resource for bioinformatics tools and services.

# Introduction

The NCBI houses

- a series of databases biotechnology and biomedicine;
- and an important resource for bioinformatics tools and services.

Major databases

- **GenBank** for DNA sequences,

# Introduction

The NCBI houses

- a series of databases biotechnology and biomedicine;
- and an important resource for bioinformatics tools and services.

Major databases

- **GenBank** for DNA sequences,
- **PubMed**, a bibliographic database for the biomedical literature,

# Introduction

The NCBI houses

- a series of databases biotechnology and biomedicine;
- and an important resource for bioinformatics tools and services.

Major databases

- **GenBank** for DNA sequences,
- **PubMed**, a bibliographic database for the biomedical literature,
- the **NCBI Epigenomics** database,

# Introduction

The NCBI houses

- a series of databases biotechnology and biomedicine;
- and an important resource for bioinformatics tools and services.

Major databases

- **GenBank** for DNA sequences,
- **PubMed**, a bibliographic database for the biomedical literature,
- the **NCBI Epigenomics** database,
- **Gene Expression Omnibus** (GEO).

# Introduction

The NCBI houses

- a series of databases biotechnology and biomedicine;
- and an important resource for bioinformatics tools and services.

Major databases

- **GenBank** for DNA sequences,
- **PubMed**, a bibliographic database for the biomedical literature,
- the **NCBI Epigenomics** database,
- **Gene Expression Omnibus** (GEO).

All these databases are available online through the **Entrez** search engine.

# Introduction

The NCBI houses

- a series of databases biotechnology and biomedicine;
- and an important resource for bioinformatics tools and services.

Major databases

- **GenBank** for DNA sequences,
- **PubMed**, a bibliographic database for the biomedical literature,
- the **NCBI Epigenomics** database,
- **Gene Expression Omnibus** (GEO).

All these databases are available online through the **Entrez** search engine.

# GEO

**GEO** is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted.
Tools are provided to help users query and download experiments and curated gene expression profiles.

# Motivation

- "A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the sequencing quality control consortium." Nature Biotechnology (2014), by SEQC/MAQC-III Consortium.
- Bioconductor package with data "RNA-seq data generated from SEQC (MAQC-III) study", by Yang Liao and Wei Shi with contributions from Steve Lianoglou.

# Motivation

- "A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the sequencing quality control consortium." Nature Biotechnology (2014), by SEQC/MAQC-III Consortium.
- Bioconductor package with data "RNA-seq data generated from SEQC (MAQC-III) study", by Yang Liao and Wei Shi with contributions from Steve Lianoglou.

# Motivation

- "A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the sequencing quality control consortium." Nature Biotechnology (2014), by SEQC/MAQC-III Consortium.
- Bioconductor package with data "RNA-seq data generated from SEQC (MAQC-III) study", by Yang Liao and Wei Shi with contributions from Steve Lianoglou.

# Results

The Sequencing Quality Control (SEQC) project is coordinated by the US FDA. Examining Illumina HiSeq, Life Technologies SOLiD and Roche 454 platforms at multiple laboratory sites using reference RNA samples with built-in controls, authors

- Assess RNA sequencing (RNA-seq) performance for junction discovery and differential expression profiling.
- Using complementary metrics, compare it to microarray data,

# Results

The Sequencing Quality Control (SEQC) project is coordinated by the US FDA. Examining Illumina HiSeq, Life Technologies SOLiD and Roche 454 platforms at multiple laboratory sites using reference RNA samples with built-in controls, authors

- Assess RNA sequencing (RNA-seq) performance for junction discovery and differential expression profiling.
- Using complementary metrics, compare it to microarray data,
- and compare it to quantitative PCR (q-PCR) data.

# Results

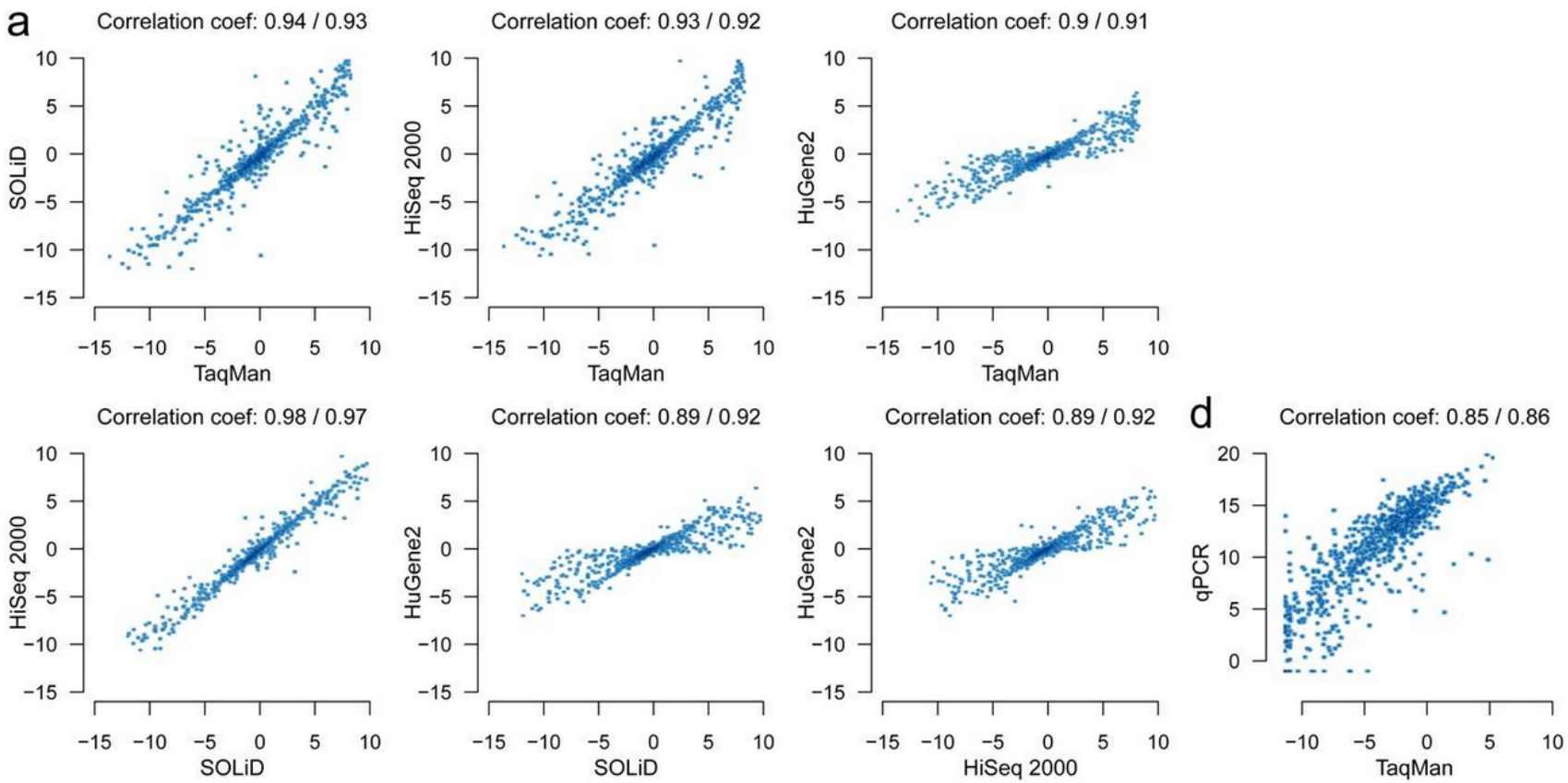
The Sequencing Quality Control (SEQC) project is coordinated by the US FDA. Examining Illumina HiSeq, Life Technologies SOLiD and Roche 454 platforms at multiple laboratory sites using reference RNA samples with built-in controls, authors

- Assess RNA sequencing (RNA-seq) performance for junction discovery and differential expression profiling.
- Using complementary metrics, compare it to microarray data,
- and compare it to quantitative PCR (q-PCR) data.

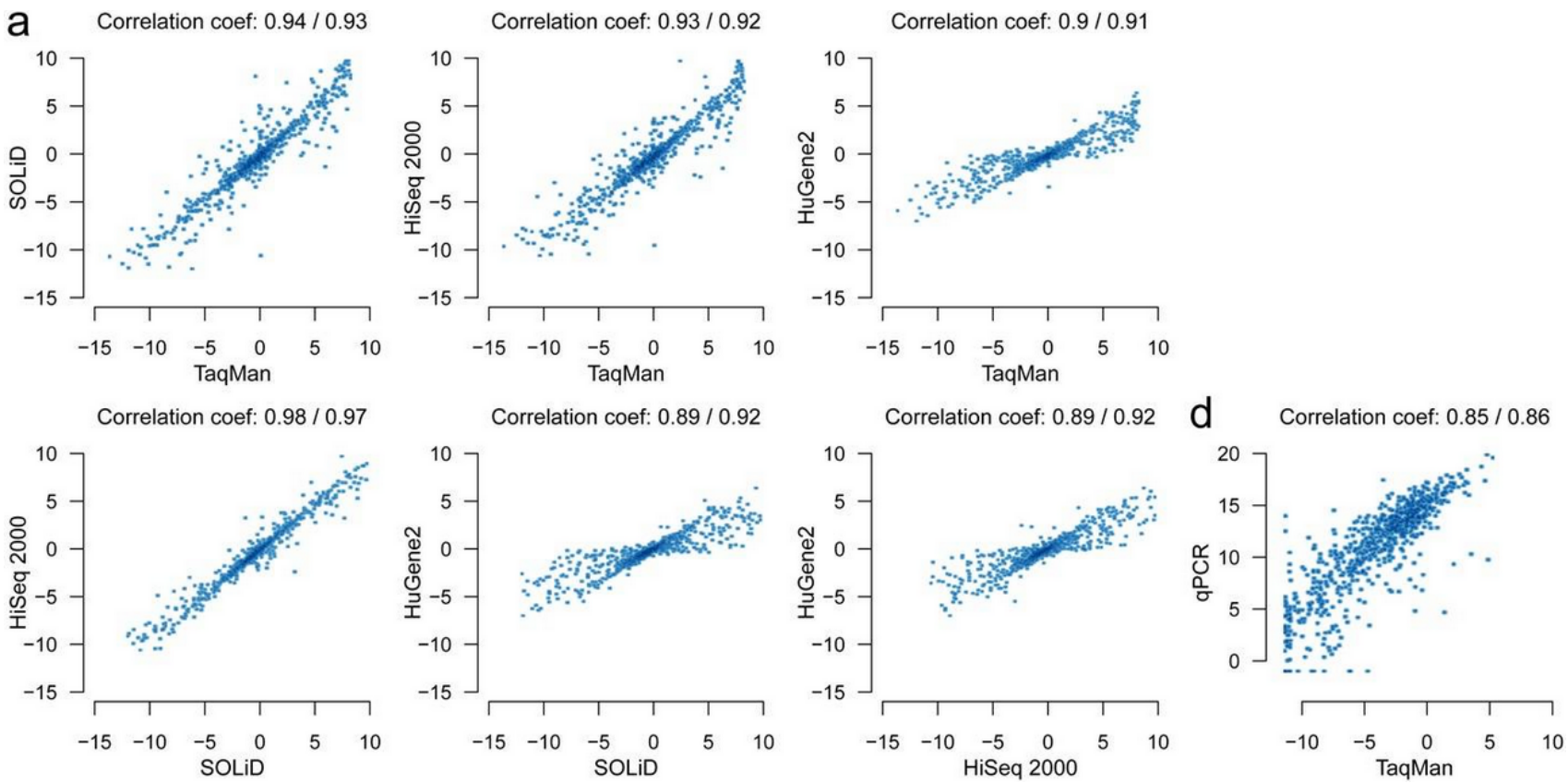
# Cross-platform agreement of expression levels

(a) Comparison of log2 fold-change estimates for 843 selected genes. Good and similar concordances were observed between relative expression measures from the MAQC-III HiSeq 2000 and SOLiD sequencing platforms, MAQC-I TaqMan and the MAQC-III Affymetrix HuGene2 arrays (Pearson and Spearman correlation) (d) Comparison of TaqMan and

PrimePCR for 843 selected genes.




Expression estimates vary considerably for individual genes, with some genes showing high expression in one platform but are not detected at all by the other.




Expression estimates vary considerably for individual genes, with some genes showing high expression in one platform but are not detected at all by the other.

# "Gold" standard

**q-PCR-based methods**:

- high sensitivity;
- and dynamic range.

# "Gold" standard

**q-PCR-based methods**:

- high sensitivity;
- and dynamic range.
- q-PCR traditionally been used as a reference "gold" standard.

# "Gold" standard

**q-PCR-based methods**:

- high sensitivity;
- and dynamic range.
- q-PCR traditionally been used as a reference "gold" standard.

- One challenge of q-PCR data is the presence of non-detects (those reactions failing to attain the expression threshold).

# "Gold" standard

**q-PCR-based methods**:

- high sensitivity;
- and dynamic range.
- q-PCR traditionally been used as a reference "gold" standard.

- One challenge of q-PCR data is the presence of non-detects (those reactions failing to attain the expression threshold).
- While most current software replaces these non-detects with the maximum possible Ct value, recent work has shown that this introduces large biases in estimation of both absolute and differential expression.[2]

# "Gold" standard

**q-PCR-based methods**:

- high sensitivity;
- and dynamic range.
- q-PCR traditionally been used as a reference "gold" standard.

- One challenge of q-PCR data is the presence of non-detects (those reactions failing to attain the expression threshold).
- While most current software replaces these non-detects with the maximum possible Ct value, recent work has shown that this introduces large biases in estimation of both absolute and differential expression.[2]
- Considerable differences in expression level measurements from different PCR-based assays can be observed.

# "Gold" standard

**q-PCR-based methods**:

- high sensitivity;
- and dynamic range.
- q-PCR traditionally been used as a reference "gold" standard.

- One challenge of q-PCR data is the presence of non-detects (those reactions failing to attain the expression threshold).
- While most current software replaces these non-detects with the maximum possible Ct value, recent work has shown that this introduces large biases in estimation of both absolute and differential expression.[2]
- Considerable differences in expression level measurements from different PCR-based assays can be observed.

# Ideas

- Validation has been an important part in RNA-seq publication. The differentially expressed genes (at least some) identified using RNA-seq are often validated using q-PCR.
- I am going to identify specific genes, where results of RNA-seq and q-PCR disagree.

# Ideas

- Validation has been an important part in RNA-seq publication. The differentially expressed genes (at least some) identified using RNA-seq are often validated using q-PCR.
- I am going to identify specific genes, where results of RNA-seq and q-PCR disagree.
- If the disagreement caused by low detection in q-PCR experiment, it may be possible to estimate non-detected Ct values, and repeat the validation procedure.

# Ideas

- Validation has been an important part in RNA-seq publication. The differentially expressed genes (at least some) identified using RNA-seq are often validated using q-PCR.
- I am going to identify specific genes, where results of RNA-seq and q-PCR disagree.
- If the disagreement caused by low detection in q-PCR experiment, it may be possible to estimate non-detected Ct values, and repeat the validation procedure.
- The claim of the project is q-PCR results may be validated using RNA-seq data.

# Ideas

- Validation has been an important part in RNA-seq publication. The differentially expressed genes (at least some) identified using RNA-seq are often validated using q-PCR.
- I am going to identify specific genes, where results of RNA-seq and q-PCR disagree.
- If the disagreement caused by low detection in q-PCR experiment, it may be possible to estimate non-detected Ct values, and repeat the validation procedure.
- The claim of the project is q-PCR results may be validated using RNA-seq data.

# References


SEQC/MAQC-III Consortium *A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium* Nature Biotechnology 32, 903914, 2014.

M.N. McCall, H.R. McMurray, H. Land and A. Almudevar *On non-detects in Quantitative real-time PCR data.* Bioinformatics V. 30 no. 16, 2310-2316, 2014.

# Thank you for your attention!

Questions?

# Thank you for your attention!
# Questions?