Statistical analysis of small RNA-seq data

Hervé Seitz
(herve.seitz@igh.cnrs.fr)

IGH (CNRS), Montpellier, France

November 26, 2013

Small RNA-seq

Small RNA-seq

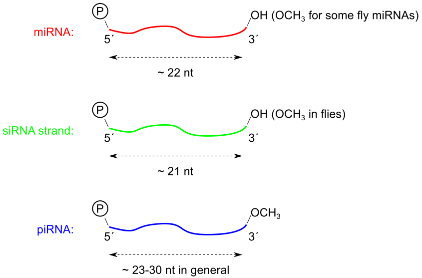
Three classes of small regulatory RNAs:

- microRNAs (“miRNAs”);
- small interfering RNAs (“siRNAs”);
- Piwi-interacting RNAs (“piRNAs”).

Small RNA-seq

Three classes of small regulatory RNAs:

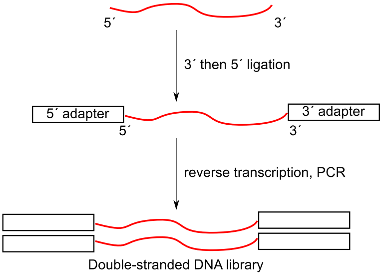
- microRNAs ("miRNAs");
- small interfering RNAs ("siRNAs");
- Piwi-interacting RNAs ("piRNAs").



Small RNA-seq

Three classes of small regulatory RNAs:

- microRNAs (“miRNAs”);
- small interfering RNAs (“siRNAs”);
- Piwi-interacting RNAs (“piRNAs”).

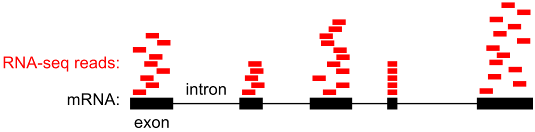


Small RNA-seq

Three classes of small regulatory RNAs:

- microRNAs ("miRNAs");
- small interfering RNAs ("siRNAs");
- Piwi-interacting RNAs ("piRNAs").

In a small RNA-seq experiment, the cDNA covers the full-length small RNA sequence (unlike RNA-seq mRNA fragments). Ligation biases are not averaged out.



Small RNA-seq

Three classes of small regulatory RNAs:

- microRNAs (“miRNAs”);
- small interfering RNAs (“siRNAs”);
- Piwi-interacting RNAs (“piRNAs”).

In a small RNA-seq experiment, the cDNA covers the full-length small RNA sequence (unlike RNA-seq mRNA fragments). Ligation biases are not averaged out.

Ligation biases were measured in Jayaprakash *et al.* (2011) *Nucleic Acids Research* 39(21): e141.

piRNA and siRNA diversity

miRNAs are moderately diverse (a few hundreds of known miRNA per genome) and can be highly expressed.

piRNA and siRNA diversity

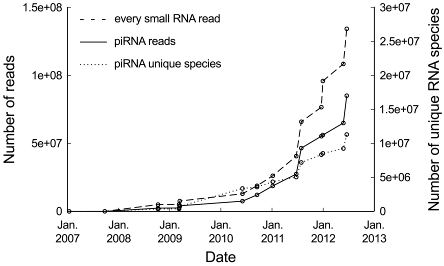
miRNAs are moderately diverse (a few hundreds of known miRNA per genome) and can be highly expressed.

siRNAs and piRNAs are much more diverse, but each individual siRNA or piRNA is poorly abundant.

piRNA and siRNA diversity

miRNAs are moderately diverse (a few hundreds of known miRNA per genome) and can be highly expressed.

siRNAs and piRNAs are much more diverse, but each individual siRNA or piRNA is poorly abundant.



(Sergeeva *et al.* (2013) *Biochemistry (Moscow)* 78(6): 613)

piRNA and siRNA diversity

The repertoire of piRNAs and siRNAs is currently unknown (and what is its biological significance?).

piRNA and siRNA diversity

The repertoire of piRNAs and siRNAs is currently unknown (and what is its biological significance?).

Biological functions for individual piRNAs or siRNAs are dubious (irreproducibility).

piRNA and siRNA diversity

The repertoire of piRNAs and siRNAs is currently unknown (and what is its biological significance?).

Biological functions for individual piRNAs or siRNAs are dubious (irreproducibility).

piRNAs and siRNAs are pooled (by targeted transposon, or by genomic locus) in order to get decent numbers to analyze.

Assessing differential expression

Assessing differential expression

Classical method: Fisher's exact test:

	Condition 1	Condition 2
Reads matching miRNA of interest	100	300
Reads not matching miRNA of interest	9,999,900	19,999,700

Assessing differential expression

Classical method: Fisher's exact test:

	Condition 1	Condition 2
Reads matching miRNA of interest	100	300
Reads not matching miRNA of interest	9,999,900	19,999,700

(p-value $= 0.00037$)

Assessing differential expression

Classical method: Fisher's exact test:

	Condition 1	Condition 2
Reads matching miRNA of interest	100	300
Reads not matching miRNA of interest	9,999,900	19,999,700

(p-value $= 0.00037$)

	Condition 1	Condition 2
Reads matching miRNA of interest	1	3
Reads not matching miRNA of interest	99,999	199,997

Assessing differential expression

Classical method: Fisher's exact test:

	Condition 1	Condition 2
Reads matching miRNA of interest	100	300
Reads not matching miRNA of interest	9,999,900	19,999,700

(p-value $= 0.00037$)

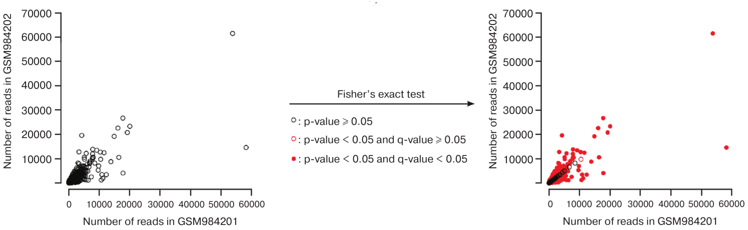
	Condition 1	Condition 2
Reads matching miRNA of interest	1	3
Reads not matching miRNA of interest	99,999	199,997

(p-value $= 1$)

Assessing differential expression

Assessing differential expression

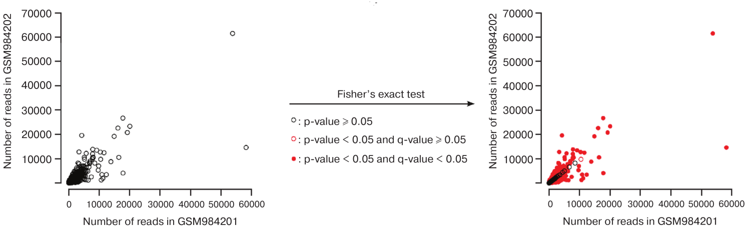
Comparison of two biological replicates (*Drosophila melanogaster* first instar larva small RNAs, modENCODE project):



(Sergeeva *et al.* (2013) *Biochemistry (Moscow)* 78(6): 613)

Assessing differential expression

Comparison of two biological replicates (*Drosophila melanogaster* first instar larva small RNAs, modENCODE project):

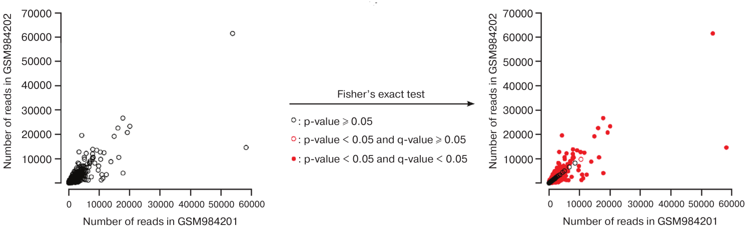


(Sergeeva *et al.* (2013) *Biochemistry (Moscow)* 78(6): 613)

Inter-individual variability is detected.

Assessing differential expression

Comparison of two biological replicates (*Drosophila melanogaster* first instar larva small RNAs, modENCODE project):



(Sergeeva *et al.* (2013) *Biochemistry (Moscow)* 78(6): 613)

Inter-individual variability is detected. Confounding effect, when comparing two genotypes.

Assessing differential expression

Detailed protocol for differential expression analysis:
Anders *et al.* (2013) *Nature Protocols* 8(9): 1765.
(developed for RNA-seq, not small RNA-seq)

Laboratory effects in small RNA-seq

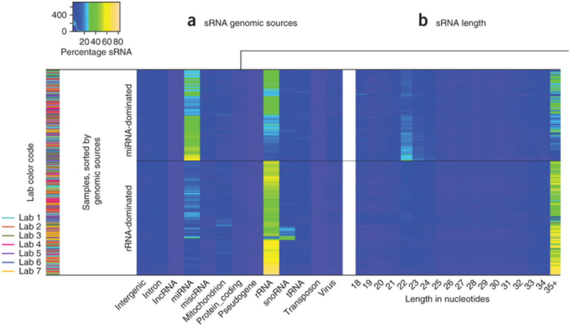
Laboratory effects in small RNA-seq

Sequencing the same collection of RNA samples in 7 sequencing facilities, to estimate laboratory effects: 't Hoen *et al.* (2013) *Nature Biotechnology* 31(11): 1015.

Laboratory effects in small RNA-seq

Sequencing the same collection of RNA samples in 7 sequencing facilities, to estimate laboratory effects: 't Hoen *et al.* (2013) *Nature Biotechnology* 31(11): 1015.

Results are consistent despite independent library preparation (most variation comes from the RNA extraction step; *e.g.*, rRNA fragment contamination). Main difference: % GC.



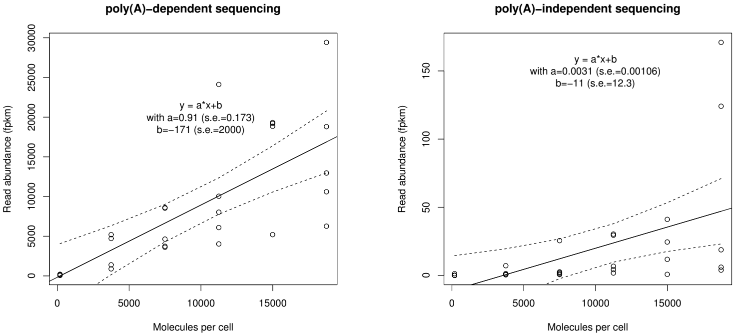
Absolute RNA quantification using RNA-seq

Absolute RNA quantification using RNA-seq

Goal: absolute measurement of mRNAs (number of molecules per cell).

(Natalia Pinzón Restrepo)

Absolute RNA quantification using RNA-seq



Conclusion

Ligation biases prevent RNA-to-RNA comparison in small RNA-seq experiments.

Conclusion

Ligation biases prevent RNA-to-RNA comparison in small RNA-seq experiments.

Individual piRNAs and siRNAs usually have low read counts (≈ 1 read per library), preventing direct statistical analysis.

Conclusion

Ligation biases prevent RNA-to-RNA comparison in small RNA-seq experiments.

Individual piRNAs and siRNAs usually have low read counts (≈ 1 read per library), preventing direct statistical analysis.

Novel statistical procedures are being developed to handle the extreme sensitivity of RNA-seq experiments (inter-individual variability cannot be neglected).

Conclusion

Ligation biases prevent RNA-to-RNA comparison in small RNA-seq experiments.

Individual piRNAs and siRNAs usually have low read counts (≈ 1 read per library), preventing direct statistical analysis.

Novel statistical procedures are being developed to handle the extreme sensitivity of RNA-seq experiments (inter-individual variability cannot be neglected).

When cautiously analyzed, small RNA-seq results are robust (even across laboratories).