Speeding towards transcriptomics

Kristin Sæbø Pettersen
Norwegian Veterinary Institute
"Speeding towards –omics ... - chemical and microbiological food analysis"
AOAC Europe - NMKL - NordVal International symposium
4 th of June, 2019
Overview

- Transcriptomics in a food safety omics perspective
- *L. monocytogenes* – a case study using transcriptomics to analyse the effect of visible light
  - Applied transcriptomics on other foodborne pathogenic bacteriae
- Take home messages
Incidence and cost of foodborne illness

1 in 6 Americans each year suffer a foodborne disease (CDC)

2 Million Annual deaths in emerging areas due to foodborne infections

3,000 Annual foodborne disease deaths in U.S. (CDC)

$80 Billion Annual cost of losses and illness caused by foodborne disease

Veterinærinstituttet
Norwegian Veterinary Institute

Cook et al 2018
Foodomics and Food Safety: Where we are

Foodomic analyses of food from the production to the consumption

(Andielkovic et al 2017)
Use of omics methods for the advancement of food quality and food safety

Word cloud (graphical representation of word frequency) showing various state-of-the-art “omics” technologies

(Cook et al 2018)
Foodomics: A novel approach for food microbiology

Fig. 1. (A) Foodolomics-related papers published between 2001-2016 (based on ISIS web of science). (Xu et al 2017)
The central biological dogma

DNA is transcribed to mRNA and mRNA is translated to protein. New technologies (e.g., “omics” technologies) have been developed to study changes in DNA, mRNA, and protein under varying physiological and environmental conditions (Cook et al 2018).

Each of the different omics data types provide a different layer and can confirm assumptions made at prior levels.
Transcriptomics

- Transcriptomics is the study of all RNA in one cell or a population of cells
  - precise measurement of transcription,
  - full picture of the extent and complexity of transcriptomes.

- Two techniques in transcriptomics:
  - Microarrays
    - can't be used for unknown RNA characterization since microarray was designed for known sequences
  - RNA-sequencing (RNA-Seq).
    - can be used for qualitative and quantitative analysis on any RNA type, including messenger RNAs (mRNAs), microRNAs, small interfering RNAs, and long noncoding RNAs.

- Foodomics: transcriptomic analysis reveals potentially “food safety targets” in complex physiological pathways
  - Ex; find synergies of hurdle technology effects to benefit food safety
Transcriptomics technologies

RNA sequencing
RNA microarray
Expressed sequence tag
Serial/cap analysis of gene expression

(Lowe et al 2017)
Google search topics trend “Microarray” and “RNA-Seq”
DNA gene in genome

In vivo

Transcription
Pre-mRNA
Intron splicing
Mature mRNA

In vitro

Reverse transcription
ds-cDNA
Fragmentation
ds-cDNA fragments
Fluorescent labelling
Labelled fragments
Array binding
Ordered microarray

In silico

Array fluorescence intensity
Array fluorescence intensity
Gene 1 2 3 4

(Lowe et al, 2017)
DNA gene in genome

In vivo

Transcription
Pre-mRNA
Intron splicing
Mature mRNA

In vitro

Fragmentation
RNA fragments
Reverse transcription
ds-cDNA fragments
High-throughput sequencing

Sequences
TATGAGACGCATGCTA
ACCCCGCC
GCGATATATA
GCGACGATGACT
ATATAGC
TCGACTGCCAT

In silico

Sequence processing
Alignment

Genome sequence

Splice variant A

Splice variant B

(Lowe et al, 2017)
Comparison of contemporary methods

<table>
<thead>
<tr>
<th>Method</th>
<th>RNA-Seq</th>
<th>Microarray</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input RNA amount</td>
<td>Low ~ 1 ng total RNA [25]</td>
<td>High ~ 1 μg mRNA [26]</td>
</tr>
<tr>
<td>Labour intensity</td>
<td>High (sample preparation and data analysis) [10][23]</td>
<td>Low [10][23]</td>
</tr>
<tr>
<td>Prior knowledge</td>
<td>None required, though genome sequence useful [23]</td>
<td>Reference transcripts required for probes [23]</td>
</tr>
<tr>
<td>Quantitation accuracy</td>
<td>~90% (limited by sequence coverage) [27]</td>
<td>&gt;90% (limited by fluorescence detection accuracy) [27]</td>
</tr>
<tr>
<td>Sequence resolution</td>
<td>Can detect SNPs and splice variants (limited by sequencing accuracy of ~99%) [27]</td>
<td>Dedicated arrays can detect splice variants (limited by probe design and cross-hybridisation) [27]</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>$10^{-5}$ (limited by sequence coverage) [27]</td>
<td>$10^{-3}$ (limited by fluorescence detection) [27]</td>
</tr>
<tr>
<td>Dynamic range</td>
<td>$&gt;10^5$ (limited by sequence coverage) [28]</td>
<td>$10^3$–$10^4$ (limited by fluorescence saturation) [28]</td>
</tr>
<tr>
<td>Technical reproducibility</td>
<td>&gt;99% [29][30]</td>
<td>&gt;99% [31][32]</td>
</tr>
</tbody>
</table>

RNA-Seq, RNA Sequencing

https://doi.org/10.1371/journal.pcbi.1005457.t001
Sequencing technology platforms commonly used for RNA-Seq.

<table>
<thead>
<tr>
<th>Platform (Manufacturer)</th>
<th>Commercial release</th>
<th>Typical read length</th>
<th>Maximum throughput per run</th>
<th>Single read accuracy</th>
<th>RNA-Seq runs deposited in the NCBI SRA (Oct 2016) [74]</th>
</tr>
</thead>
<tbody>
<tr>
<td>454 (Roche, Basel, Switzerland)</td>
<td>2005</td>
<td>700 bp</td>
<td>0.7 Gbp</td>
<td>99.9%</td>
<td>3548</td>
</tr>
<tr>
<td>Illumina (Illumina, San Diego, CA, USA)</td>
<td>2006</td>
<td>50–300 bp</td>
<td>900 Gbp</td>
<td>99.9%</td>
<td>362903</td>
</tr>
<tr>
<td>SOLiD (Thermo Fisher Scientific, Waltham, MA, USA)</td>
<td>2008</td>
<td>50 bp</td>
<td>320 Gbp</td>
<td>99.9%</td>
<td>7032</td>
</tr>
<tr>
<td>Ion Torrent (Thermo Fisher Scientific, Waltham, MA, USA)</td>
<td>2010</td>
<td>400 bp</td>
<td>30 Gbp</td>
<td>98%</td>
<td>1953</td>
</tr>
<tr>
<td>PacBio (Pacbio, Menlo Park, CA, USA)</td>
<td>2011</td>
<td>10,000 bp</td>
<td>2 Gbp</td>
<td>87%</td>
<td>160</td>
</tr>
</tbody>
</table>

NCBI, National Center for Biotechnology Information; SRA, Sequence Read Archive; RNA-Seq, RNA sequencing.

https://doi.org/10.1371/journal.pcbi.1005457.t002
Transcriptomics on *L. monocytogenes*: a case study
Cost of various foodborne pathogens

- **Listeria monocytogenes**: 18.8% ($2.7 billion)
- **Salmonella (nontyphoidal)**: 23.4% ($3.4 billion)
- **Norovirus**: 14.2% ($2.1 billion)
- **Campylobacter spp.**: 12.4% ($1.8 billion)
- **Toxoplasma gondii**: 21.1% ($3.1 billion)
- **Remaining 9 pathogens**: 10.1% ($1.5 billion)

Other pathogens:
- *Clostridium perfringens*: 2.2%
- *Vibrio vulnificus*: 2.0%
- *E. coli O157:H7*: 1.9%
- *Yersinia enterocolitica*: 1.8%
- *Shigella spp.*: 0.9%
- *Vibrio spp. other*: 0.8%
- *Cryptosporidium parvum*: 0.3%
- *STEC non-O157*: 0.2%
- *Cyclospora cayetanensis*: 0.01%

(USDA Economic Research Service, Cook et al, 2018)
Visible light – why should *L. monocytogenes* care?

http://www.nature.com/ng/journal/v48/n3/full/ng.3515.html

https://ing.dk/sites/ing/files/topillustration/2017/09/fordampningenergi1.jpg
Cyanobacteria use micro-optics to sense light direction

Nils Schuergers, Tchern Lenn, Ronald Kampmann, Markus V Meissner, Tiago Estevés, Maja Temerinac-Ott, Jan G Korvink, Alan R Lowe, Conrad W Mullineaux, and Annegret Wilde

Fred Rieke, Howard Hughes Medical Institute, University of Washington, United States;

Author information > Article notes > Copyright and License information Disclaimer
Effect of visible light

• Reactive Oxygen Species (ROS) produced in illuminated bacteria - phototoxic effect
• Production of antioxidants – light sensitivity
• NB! Low amounts of ROS that has been found to induce cell growth.
• Intense blue light, preferably at 415nm, is better than red light for bacteria killing.

Lubart et al 2011
Initially experiment
What are the deeper mechanisms? Which genes are activated? (Transcriptomics....)
How do we find which genes the bacterium express when exposed to visible light?

- Sequencing
- Processing raw sequences and align with a reference genome
- Counting sequences and statistical analysis

2,409 significant regulated genes (ca 80%)

<table>
<thead>
<tr>
<th>Temp (°C)</th>
<th>Time (min)</th>
<th>Total radiant exposure (Joule/cm²)</th>
<th>Blue light radiant exposure (Joule/cm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>20</td>
<td>0.8</td>
<td>0.3</td>
</tr>
<tr>
<td>20</td>
<td>180</td>
<td>6.1</td>
<td>2.4</td>
</tr>
<tr>
<td>37</td>
<td>20</td>
<td>0.8</td>
<td>0.3</td>
</tr>
</tbody>
</table>

20 min = ca 8 s
180 min = ca 60 s
Considerations for experimental design

• Include a bioinformatician in early phase of planning experimental design and budget!
• Snaphsot! Versus fingerprinting....
• Stationary versus exponential growth phase
• Time of storage (RNA quality)
• rRNA – mRNA -> depletion
• Reference level
  • Global transcriptomics: you assume no difference
  • Single gene expression: reference gene and control group
• Depth
• Read length
• Single – paired –end reads
Applied transcriptomics on other foodborne pathogenic bacteria

5. Conclusion

425 DGEs were identified responding to cold stresses in *E. coli* O157:H7 by using throughput transcriptome sequencing technology. The results suggested that gene rpoSwas referred to be the critical switch for the whole regulation to resist prolonged cold stress in *E. coli* O157:H7. In addition, *E. coli* O157:H7 would protect itself from prolonged cold harm mainly by increasing the expression of genes involved in the synthesis of [unsaturated fatty acids](https://) and the accumulation of [trehalose](https://) as well as betaine. Furthermore, cold shock genes that well-known in other reports were not the key factor in prolonged cold tolerance. More likely, these genes expressed in the early acclimation phase other than long storage phase.
In summary, this is the first report of the complete transcriptome of \textit{C. jejuni} IA3902 during exposure to an important and relevant natural host environment, the sheep gallbladder. We have demonstrated that the transcriptional “landscape” during direct interaction within the host, as displayed by utilizing \textit{in vivo} inoculation of and RNA recovery from the sheep gallbladder environment, provides a more robust picture of the complexity of gene regulation required for survival when compared to \textit{in vitro} exposure to ovine bile alone. A subset of genes were identified that are believed to play an important role in survival within bile, as well as survival in the host environment, including two highly expressed hypothetical proteins that warrant further study. In addition to the identification of important protein coding genes that are differentially expressed, seven previously identified non-coding RNAs were also confirmed to be differentially expressed within our data, suggesting that they may also play a key role in rapid regulation of gene expression upon exposure to bile and the host environment.
Applied transcriptomics on other foodborne pathogenic bacteria

Transcriptomics analysis on ampicillin-induced and non-ampicillin-induced biofilms were performed by RNA-sequencing, differentially expressed genes identification and annotation, GO functional and KEGG pathway enrichment. The viability and biomass of ampicillin-induced biofilm showed dramatical increase compared to the non-ampicillin-induced biofilm. A total of 530 differentially expressed genes (DEGs) with 167 and 363 genes showing up- and down-regulation, respectively, were obtained. Upon GO functional enrichment, 183, 252, and 21 specific GO terms in biological process, molecular function and cellular component were identified, respectively. Eight KEGG pathways including “Microbial metabolism in diverse environments”, “S. aureus infection”, and “Monobactam biosynthesis” were significantly enriched. In addition, “beta-lactam resistance” pathway was also highly enriched. In ampicillin-induced biofilm, the significant up-regulation of genes encoding multidrug resistance efflux pump AbcA, penicillin binding proteins PBP1, PBP1a/2, and PBP3, and antimicrobial resistance proteins VraF, VraG, Dlt, and Aur indicated the positive response of S. aureus to ampicillin. The up-regulation of genes encoding surface proteins ClfB, IsdA, and SasG and genes (cap5B and cap5C) which promote the adhesion of S. aureus in ampicillin induced biofilm might explain the enhanced biofilm viability and biomass.
Take home message

• Transcriptomics and food safety
  • Mooving towards transcriptomics
  • Not yet at the «applied stage», such as genomics, maybe more suitable for revealing deeper mechanisms and new «targets»

• Broad spectrum visible light has a significant impact on several of the foodborne bacterial pathogens
  • To be included in preventive strategies for contamination?
  • Experimental design in the lab when simulating host conditions.....
Acknowledgement

• Funding: Pathogens in the Food Chain, NFR 221663 – F40

• Co-authors: «Shining Light on L. monocytogenes – a transcriptome analysis»
  • Arvind YM Sundaram*, Taran Skjerdal **, Yngvild Wasteson ***, Toril Lindbäck ***, Anne Kijewski ***, Marina Aspholm ***

  • *Norwegian Sequencing Centre, Department of Medical Genetics, Oslo University Hospital, Norway
  • **Norwegian Veterinary Institute, Ullevålsveien 68, N-0450 Oslo, Norway
  • ***Norwegian University of Life Sciences, Campus Adamstuen, Ullevålveien 72, 0454 Oslo, Norway

• Terje Christensen, Statens Strålevern
• Ellen Christensen, Veterinærinstituttet
• Gro Johannessen, Veterinærinstituttet

Food Micro, Berlin, September 2018; Posterprice