FMSC The Finnish Microarray and Sequencing Centre
Biocity, 5th floor • Tykistökatu 6, 20520 Turku, FINLAND
Tel: +358 2 333 8634 • E-mail: fmsc@btk.fi • http://fmsc.btk.fi
BF Biocenter Finland Core Lab Program
illumina CSPro CERTIFIED SERVICE PROVIDER
Agilent Certified Services Provider Microarray-Based Genomic Analysis

The Finnish Microarray and Sequencing Centre

Quality Assurance of Microarray Data

Miina Nurmi 21.5.2013

FMSC The Finnish Microarray and Sequencing Centre
Biocity, 5th floor • Tykistökatu 6, 20520 Turku, FINLAND
Tel: +358 2 333 8634 • E-mail: fmsc@btk.fi • http://fmsc.btk.fi
BF Biocenter Finland Core Lab Program
illumina CSPro CERTIFIED SERVICE PROVIDER
Agilent Certified Services Provider Microarray-Based Genomic Analysis

Our main tasks

- **Services:** Provide full service from experimental design to bioinformatics
- **Instrumentation:** Identify needs, coordinate purchases, set up new instruments and maintain them
- **Research:** Develop original research approaches and methods
- **Education:** Educate the research community and encourage use of modern technologies

Miina Nurmi 21.5.2013

FMSC The Finnish Microarray and Sequencing Centre
Biocity, 5th floor • Tykistökatu 6, 20520 Turku, FINLAND
Tel: +358 2 333 8634 • E-mail: fmsc@btk.fi • http://fmsc.btk.fi

FMSC in the Biocenter Finland Genome-Wide Method Network

Members:

- Institute of Biotechnology (BI, Helsinki)
- Biocentrum Helsinki (BCH)
- Institute for Molecular Medicine Finland (FIMM)
- Biocenter Kuopio (BCK)
- Biocity Turku (BCT)

FMSC responsibilities within BF Genome-Wide Methods Network:

- Gene expression microarrays
- RNA sequencing
- Sequencing of immunoprecipitated DNA/RNA (ChIP-seq, CLIP-seq)
- Developing advanced techniques and optimizing reagents for studies on epigenetics and chromatin structure exploiting both next generation sequencing (NGS) and next-next generation platforms

FMSC The Finnish Microarray and Sequencing Centre
Biocity, 5th floor • Tykistökatu 6, 20520 Turku, FINLAND
Tel: +358 2 333 8634 • E-mail: fmsc@btk.fi • http://fmsc.btk.fi

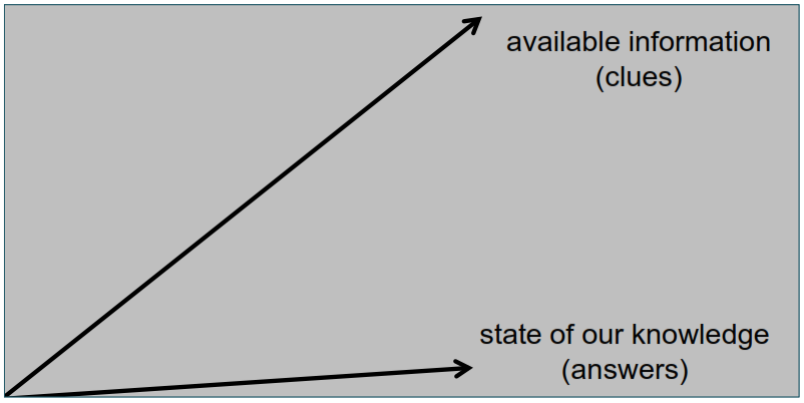
Important issues related to microarrays and next generation sequencing applications

- Sample quality
- Standardization of conditions
- Choosing the correct time point
- Pooling of samples should be avoided
- Number of replicates
- The choice of control samples
- Real-time PCR validation of results often needed
- Massive amounts of data (storage/management)
- Interpretation of data/results
- Incomplete understanding of the source of errors and bias in data

- Data analysis issues should be considered when planning the experiments!

Miina Nurmi 21.5.2013

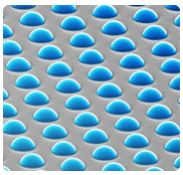
Massive amounts of data



FMSC The Finnish Microarray and Sequencing Centre
Biocity, 5th floor • Tykistökatu 6, 20520 Turku, FINLAND
Tel: +358 2 333 8634 • E-mail: fmsc@btk.fi • http://fmsc.btk.fi
BF Biocenter Finland
Affymetrix Core Lab Program
illumina CSPro
Agilent Certified Services Provider Microarray-Based Genomic Analysis

Example: Illumina Gene Expression Microarrays

- 23 000 probes/stripe
- 1 – 2 stripes/sample
- 23 000 – 46 000 transcipts
- Human & Mouse
  - Human HT-12 v.4 with 46 000 transcripts
  - Mouse WG-6 with 46 000 transcripts
  - Mouse Ref-8 with 23 000 transcripts



QC before beginning: RNA quantity and purity

- NanoDrop 2000 (Thermo Scientific)
  - Spectrophotometer
  - Minimum amount of sample: 0.5 µl
  - Pathlength 1 mm
- RNA quantity
  - Accurate quantification is crucial for standardized sample preparation
  - Quantity measured using absorbance at 260 nm
  - Concentration range 0.8–12,000 ng/µl
- RNA purity
  - Chemical residuals may inhibit sample processing
  - Purity measured using absorbance ratios between absorbances at 260 nm, 280 nm and 230 nm

FMSC The Finnish Microarray and Sequencing Centre
Biocity, 5th floor • Tykistökatu 6, 20520 Turku, FINLAND
Tel: +358 2 333 8634 • E-mail: fmsc@btk.fi • http://fmsc.btk.fi
BF Biocenter Finland
Affymetrix Core Lab Program
illumina CSPro CERTIFIED SERVICE PROVIDER
Agilent Certified Services Provider Microarray-Based Genomic Analysis

QC before beginning: RNA integrity

- Bioanalyzer 2100 (Agilent)
  - Microfluidics-based gel electrophoresis
  - Concentration range ≥50 pg/µl
- Importance of RNA integrity
  - Accurate analysis of gene expression requires good RNA quality
  - Fragmented RNA causes false negative results

Miina Nurmi 21.5.2013

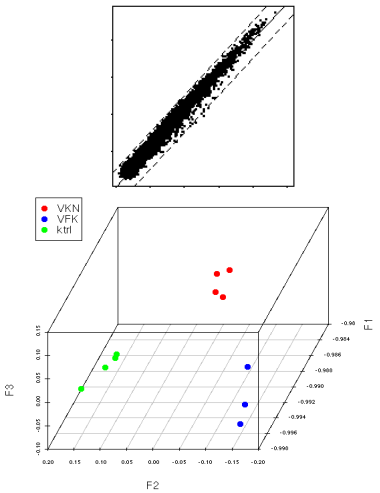
QC during sample processing

- Standardized methods
  - Documentation: protocol, deviations, reagent lot numbers, instruments
- Measurement of labeled cRNA product quality
  - Bioanalyzer 2100 (Agilent)
  - Size distribution among samples
  - Optional control samples can be included

FMSC The Finnish Microarray and Sequencing Centre
Biocity, 5th floor • Tykistökatu 6, 20520 Turku, FINLAND
Tel: +358 2 333 8634 • E-mail: fmsc@btk.fi • http://fmsc.btk.fi
BF Biocenter Finland
Affymetrix Core Lab Program
illumina CSPro CERTIFIED SERVICE PROVIDER
Agilent Certified Services Provider Microarray-Based Genomic Analysis

Microarray results QC

- Different QC methods in different applications
- Comparison between samples
  - Acceptable variation between samples
  - In data analysis, normalization methods can reduce the effects of variation
- Clustering
  - E.g. PCA, principal component analysis
  - Replicates should cluster close to each other
- Result validation using other methods
  - “Second opinion”: other microarray platforms
  - “Golden standard”: RT-QPCR



Miina Nurmi 21.5.2013

FMSC The Finnish Microarray and Sequencing Centre
Biocity, 5th floor • Tykistökatu 6, 20520 Turku, FINLAND
Tel: +358 2 333 8634 • E-mail: fmsc@btk.fi • http://fmsc.btk.fi
BF Biocenter Finland
Affymetrix Core Lab Program
illumina CSPro CERTIFIED SERVICE PROVIDER
Agilent Certified Services Provider Microarray-Based Genomic Analysis

How to Manage Complex Methods and Large Amounts of Data?

- The more we know, the better we can assess the accuracy of results
- Best practices being formed

Miina Nurmi 21.5.2013

Next Generation Sequencing: Guiding the translation from research to clinical applications

A.S. Gargis[1], L.V. Kalman[1], M.W. Berry[2], D.P. Bick[3], D.P. Dimmock[3], T. Hambuch[4], F. Lu[2], E. Lyon[5], K.V. Voelkerding[5], B.A. Zehnbauer[1], I.M. Lubin[1], and Working Group*
[1] Centers for Disease Control and Prevention, Atlanta, GA, [2] SeqWright, Inc., Houston, TX, [3] Medical College of Wisconsin, Milwaukee, WI, [4] Illumina Clinical Services, San Diego, CA, [5] ARUP Laboratories, Salt Lake City, UT

Workgroup Recommendations: Defining analytical performance characteristics for NGS

Performance Characteristics	Workgroup established definitions for NGS applications	Workgroup established metrics and processes for evaluation of NGS analytic performance
Accuracy	The closeness of agreement between a measured value and the true value, which for NGS is the accepted reference sequence.	• Coverage - The number of independent overlapping base calls made at a given position •Depth of coverage • Average coverage •Uniformity or distribution of coverage • Quality scores - The confidence in a base or variant call
Precision	The degree to which repeated measurements give the same result (repeatability and reproducibility).	•Monitor performance for: • Library variability: independent library preparations • Intra- run variability: same sample, same library, same run • Inter-run variability: same sample, same library, different runs • Inter-operator variability
Analytic Sensitivity	The likelihood that the assay will detect a sequence variation, if present.	Depth of coverage must be sufficient to minimize a loss of sensitivity and specificity. The depth of coverage achieved with NGS will vary across the genome and therefore should be established across all regions of the sequence targeted for the clinical application. Analysis of RMs possessing comparable types of sequence variations across the targeted region by an orthogonal technique can provide a useful comparator.
Analytic Specificity	The probability that the assay will not detect a sequence variation, if not present.	
Reportable Range	The regions of the genome for which the NGS technology can accurately produce sequence information (e.g. multiple genes, exome, large genomic regions)	Define areas of difficulty (e.g. repeat regions, insertions and deletions, allele dropouts) near the regions of interest. Biases introduced by capture-based or enrichment methods should be identified.
Reference Range	Establishment of reportable sequence variations expected to occur in the target population that the assay can detect.	Materials containing the type of sequence variation(s) appropriately distributed within the target sequence may establish the capacity of the test to detect similar disease-associated mutations.

Poster presentation at the 2011 American Society of Human Genetics annual meeting, Montreal, Canada

Next Generation Sequencing: Guiding the translation from research to clinical applications

A.S. Gargis[1], L.V. Kalman[1], M.W. Berry[2], D.P. Bick[3], D.P. Dimmock[3], T. Hambuch[4], F. Lu[2], E. Lyon[5], K.V. Voelkerding[5], B.A. Zehnbauer[1], I.M. Lubin[1], and Working Group*

[1] Centers for Disease Control and Prevention, Atlanta, GA, [2] SeqWright, Inc., Houston, TX, [3] Medical College of Wisconsin, Milwaukee, WI, [4] Illumina Clinical Services, San Diego, CA, [5] ARUP Laboratories, Salt Lake City, UT

Workgroup Recommendations: Defining analytical performance characteristics for NGS

Performance Characteristics	Workgroup established definitions for NGS applications	Workgroup established metrics and processes for evaluation of NGS analytic performance
Accuracy		ls made at a given position
Precision		s ne run erent runs
Analytic Sensitivity	present.	sitivity and specificity. The depth of coverage achieved with NGS will vary across the genome and therefore should be established across all regions of the sequence targeted for the clinical application. Analysis of RMs possessing comparable types of sequence variations across the targeted region by an orthogonal technique can provide a useful comparator.
Analytic Specificity	The probability that the assay will not detect a sequence variation, if not present.	
Reportable Range	The regions of the genome for which the NGS technology can accurately produce sequence information (e.g. multiple genes, exome, large genomic regions)	Define areas of difficulty (e.g. repeat regions, insertions and deletions, allele dropouts) near the regions of interest. Biases introduced by capture-based or enrichment methods should be identified.
Reference Range	Establishment of reportable sequence variations expected to occur in the target population that the assay can detect.	Materials containing the type of sequence variation(s) appropriately distributed within the target sequence may establish the capacity of the test to detect similar disease-associated mutations.

http://www.cdc.gov/osels/lspppo/pdf/NextGenSequencing_Guiding_the_translation_from_research_to_clinical_applications.pdf

Poster presentation at the 2011 American Society of Human Genetics annual meeting, Montreal, Canada

FMSC The Finnish Microarray and Sequencing Centre
Biocity, 5th floor • Tykistökatu 6, 20520 Turku, FINLAND
Tel: +358 2 333 8634 • E-mail: fmsc@btk.fi • http://fmsc.btk.fi
BF Biocenter Finland Core Lab Program
Affymetrix illumina CSPro CERTIFIED SERVICE PROVIDER
Agilent Certified Services Provider Microarray-Based Genomic Analysis

Future Challenges

- "Real-life data"
  - Diagnostic use: real people, real situations, real responsibility
- Reliable interpretation of results

FMSC The Finnish Microarray and Sequencing Centre
Biocity, 5th floor • Tykistökatu 6, 20520 Turku, FINLAND
Tel: +358 2 333 8634 • E-mail: fmsc@btk.fi • http://fmsc.btk.fi

BF Biocenter Finland

Affymetrix Core Lab Program

illumina CSPro CERTIFIED SERVICE PROVIDER

Agilent Certified Services Provider Microarray-Based Genomic Analysis

FMSC Personnel

Director Riitta Lahesmaa

Attila Gyenesei
Group Leader
Bioinformatics
Core Facility Services

Juha-Pekka Pursiheimo
Senior Scientist
Life Science
Applications

Riikka Lund
Senior Scientist
Life Science
Applications

David Hawkins
Group Leader
Epigenomics
Emerging Technologies

Asta Laiho
Project Engineer
Bioinformatics

Leni Kauko
Researcher
NGS

Päivi Junni
Lab. Technician
NGS, Epigenomics
Illumina arrays

Cristina Valensisi
Postdoctoral Fellow
Hawkins Group
ChIP-Seq

Sini Junttila
Graduate Student
Administration,
Bioinformatics

Eveliina Virtanen
Project Engineer
Affymetrix, NGS

Bogata Fezazi
Lab. Technician
Epigenomics,
Affymetrix arrays

Kalyan Pasumarthy
Postdoctoral Fellow
Hawkins Group
RRBS

Leena Kytomäki
Project Engineer
Bioinformatics,
Illumina arrays

Sanna Vuorikoski
Researcher
NGS, Epigenomics

Oso Rissanen
Lab. Technician
DNA seq, RT-PCR

Miina Nurmi
Lab. Technician
Illumina, RT-PCR,
DNA seq

Reija Venho
Lab. Technician
Agilent, Illumina

http://www.btk.fi/microarray-and-sequencing/

Miina Nurmi 21.5.2013