

Annotation and *in silico* localization of the Affymetrix GeneChip Porcine Genome Array

WATCHARAPONG NARABALLOBH^{1,3}, SIRIWADEE CHOMDEJ³, EDUARD MURANI¹,
KLAUS WIMMERS¹ and SIRILUCK PONSUKSILI²

¹Research Unit Molecular Biology, Leibniz Institute for Farm Animal Biology, Dummerstorf, Germany, ²Research Group Functional Genomics, Leibniz Institute for Farm Animal Biology, Dummerstorf, Germany, ³Department of Biology, Faculty of Science, Chiang Mai University, Thailand

Abstract

Expression microarrays including the Affymetrix GeneChip Porcine Genome Arrays are valuable tools for studying genes and functional networks relevant for the expression of complex traits and the responsiveness of the organism to various treatments. An updated annotation and, for the first time, localization on the porcine physical genome map of »Affymetrix GeneChip Porcine Genome Array probe sets« was made through a workflow of 3 pipelines of comparisons addressing various NCBI (National Center for Biotechnology Information) and Ensembl (Ensembl project) databases. »BLAST« (Basic Local Alignment Search Tool) comparisons of Affymetrix probe set consensus sequences with the Ensembl Sscrofa 9 cDNA database provided 23 799 probe sets with hits. After annotation 19 730 gene symbols were obtained using the data management system BioMart. Comparison of the Affymetrix probe set consensus sequences with the porcine genome sequence (Ensembl Sscrofa 9 LatestGP database) revealed 23 298 probe sets with BLAST hits. In the third pipeline in addition to Ensembl Sscrofa 9 cDNA and genomic sequence databases also human, mouse and pig NCBI reference sequence RNA databases were interrogated in an integrated approach where also a threshold of bit score >50 or >90% identity over >100 bp was applied in order to filter questionable annotations and localizations. Gene symbols and gene names were queried from HGNC (human genome organization (HUGO) gene nomenclature committee), EASE (EASE: the Expression Analysis Systematic Explorer) and Entrez Gene revealing 20 269 annotated probe sets. 20 467 probe sets were *in silico* mapped addressing various sources: Ensembl Sscrofa 9 LatestGP, pre-Ensembl Sscrofa 8.52 LatestGP, NCBI pig reference sequence RNA and genomic databases and PigQTLdb (Pig Quantitative Trait Locus [QTL] database). Using the new annotation and localization data in functional genomics studies will facilitate improving the understanding of the control of quantitative traits in pigs.

Keywords: microarray, expression, genomics, annotation, mapping

Zusammenfassung

Annotation und *in silico* Kartierung des Affymetrix GeneChip Porcine Genome Arrays

Mikroarrays für Expressionsanalysen einschließlich der Affymetrix GeneChip Porcine Gene Arrays sind wertvolle Werkzeuge, um Gene und funktionelle Netzwerke zu analysieren, die für die Ausprägung komplexer Merkmale und für die Reaktion des Organismus auf verschiedene Behandlungen relevant sind. Eine aktualisierte Annotation und erstmals eine Kartierung der »Affymetrix GeneChip probe sets« (Sonden) im Schweinegenom wurde auf drei Wegen unter Einbeziehung von verschiedenen NCBI (National Center for Biotechnology Information) und Ensembl (Ensembl Projekt) Datenbanken vorgenommen. »BLAST«-Vergleiche (Basic Local Alignment Search Tool) der Konsensussequenzen, von denen die Affymetrix Sonden abgeleitet wurden, mit der Ensembl Sscrofa 9 cDNA Datenbank lieferte 23 799 Sonden mit Treffern. Mittels BioMart konnten 19 730 Sonden Gensymbole zu geordnet werden. Vergleiche mit Ensembl Sscrofa 9 genomischen Sequenzinformationen (LatestGP) zeigten für 23 298 Sonden BLAST-Treffer. Im dritten Ansatz wurden NCBI RNA-Referenz-Sequenzen von Mensch, Maus und Schwein sowie Ensembl Sscrofa 9 Datenbanken gemeinsam abgefragt und eine Schwelle für die Annotation angewandt (bit score >50; >90% Identität über >100 bp). Gensymbole und Gennamen wurden von HGNC, EASE und Entrez Datenbanken übernommen und ergaben 20 269 annotierte Sonden. 20 467 Sonden wurden *in silico* kartiert durch Vergleich mit Ensembl Sscrofa 9 LatestGP, pre-Ensembl Sscrofa 8.52 LatestGP, NCBI pig reference sequence RNA, NCBI pig genomic Datenbanken und PigQTLdb (Pig Quantitative Trait Locus [QTL] database). Die Verwendung der neuen Annotations- und Lokalisationsdaten wird Untersuchungen zur Funktionalen Genomik beim Schwein hinsichtlich der Steuerung quantitativer Merkmale unterstützen.

Schlüsselwörter: Mikroarray, Expression, Genomik, Annotation, Kartierung

Introduction

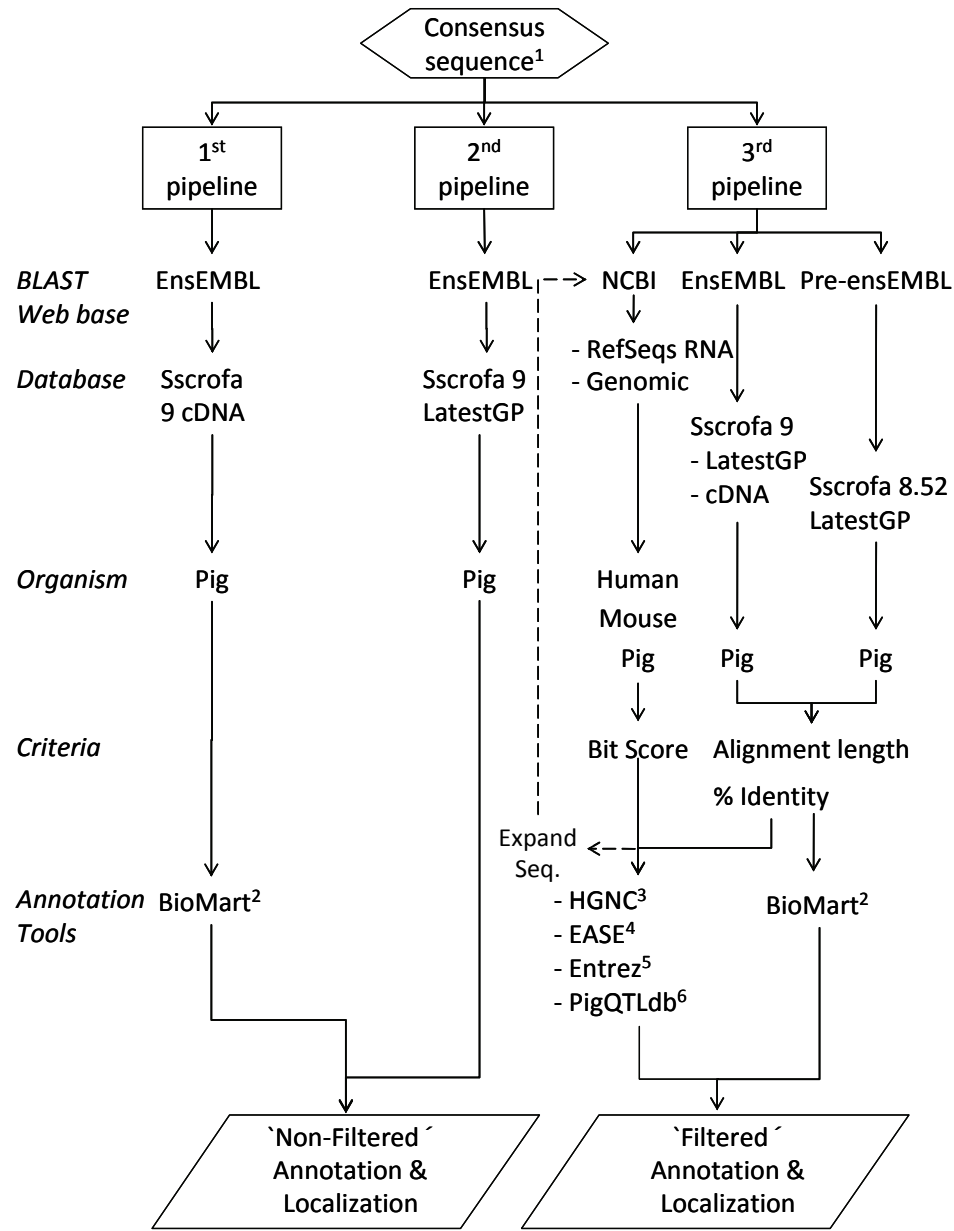
Recently, microarray technologies have enabled the measurement of gene expression levels for thousands of genes in parallel. Functional genomics provides new opportunities for determining the molecular processes underlying phenotypic variation towards developmental and reproduction traits, growth, immune response and host-pathogen interaction (BAI *et al.* 2003, BAND *et al.* 2002, BERNARD *et al.* 2007, COGBURN *et al.* 2003, MOSER *et al.* 2004, PONSUKSILI *et al.* 2007, 2008a, 2008b, SUCHYTA *et al.* 2003, TUGGLE *et al.* 2007). Currently several whole genome expression microarrays for pigs are available including consortium arrays like the »Pigoligoarray« (<http://www.pigoligoarray.org>) or commercial microarrays like the Agilent Porcine Gene Expression Microarray (Agilent Technologies; <http://www.chem.agilent.com>) and the Affymetrix GeneChip Porcine Genome Array (Affymetrix, <http://www.affymetrix.com>). Also application-specific microarrays have been developed by several research groups. The microarray design of Affymetrix GeneChips has been proven to provide highly reproducible expression profiles and is available for many different species (HARTMANN *et al.* 2009). Originally, at the time point

of its release the Affymetrix GeneChip porcine genome array was moderately annotated (1st release based on UniGene Build 28 [August 2004]). Less than 10% of the probe sets on this array were described with gene names, posing a challenge to biological interpretation of data (TSAI *et al.* 2006). 19 675 of 24 123 probe sets on the Affymetrix Porcine microarray, representing 11 265 unique genes, were annotated based on BLAST (Basic Local Alignment Search Tool, ALTSCHUL *et al.* 1990) comparison of EnsEMBL human cDNA and genomic sequences (Ensembl project, <http://www.ensembl.org/info/about/credits.html>; HUBBARD *et al.* 2002, 2007) with the Affymetrix porcine target sequences, which were extended with porcine sequence information of the Pig Gene Index (The Institute for Genome Research, TIGR; <http://www.jcvi.org>; LEE *et al.* 2005), (TSAI *et al.* 2006). In an 2010 update 19992 probes were annotated (TSAI *et al.* 2006; <http://www4.ncsu.edu/~stsai2/annotation>). However, no information about the position of the respective genes in the porcine genome was provided.

The main objective of this study was to update and improve the annotation of the probe sets represented on the Affymetrix GeneChip Porcine Genome Array and at the same time to provide a localization of the probe sets onto the recent porcine genomic sequence (Sscrofa 9) (http://www.ensembl.org/Sus_scrofa/Info/Index), which will be highly valuable in current efforts to use functional and genetical genomics in pigs. We provide a public accessible annotation and localization data table facilitating the application of the data in pig genomic studies. For annotation of the probe sets homology searches of their corresponding consensus sequences were done to various NCBI (National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov>; PRUITT *et al.* 2007) databases, HGNC (human genome organization (HUGO) gene nomenclature committee, <http://www.genenames.org>; WAIN *et al.* 2002, BRUFORD *et al.* 2008), EASE (The Expression Analysis Systematic Explorer, <http://david.abcc.ncifcrf.gov/ease/ease1.htm>; DOUGLAS *et al.* 2003), EnsEMBL and pre-EnsEMBL (pre build version of EnsEMBL) (Ensembl project, <http://www.ensembl.org/info/about/credits.html>; HUBBARD *et al.* 2002, 2007, BIRNEY *et al.* 2004), and PigQTLdb (Pig Quantitative Trait Locus [QTL] database, <http://www.animalgenome.org/cgi-bin/QTLdb/SS/index>; HU *et al.* 2005).

Material and methods

FASTA format sequences of the Affymetrix GeneChip Porcine Genome Array were retrieved from the array library file (<http://www.affymetrix.com/support/index.affy>) including 23 935 sequences (consensus sequence). The workflow was made up of three pipelines (Figure). BLAST was used as the main method to predict similarity of the sequences in all three pipelines, however different databases were targeted. The highest bit score of each BLAST result was selected for the annotation of each individual probe set. The first pipeline provided annotation data which were queried from EnsEMBL Sscrofa 9 cDNA (http://www.ensembl.org/Sus_scrofa/Info/Index; last accessed 20 Nov. 2009) database. Extracted raw data were stored in a Microsoft Excel spread sheet annotation file. After that, EnsEMBL identifiers were queried via BioMart (<http://www.biomart.org>; SMEDLEY *et al.* 2009) to gain more information especially regarding gene symbols and gene names. The second pipeline also addressed EnsEMBL



¹Consensus Sequence from Affymetrix GeneChip Porcine Genome Array (<http://www.affymetrix.com/support/index.affy>) ²BioMart (<http://www.biomart.org>), ³Database of the HUGO Gene Nomenclature Committee (http://www.genenames.org/cgi-bin/hgnc_downloads.cgi), ⁴The Expression Analysis Systematic Explorer (<http://david.abcc.ncifcrf.gov/ease/ease1.htm>), ⁵Entrez Gene database (<http://www.ncbi.nlm.nih.gov/gene>) ⁶Pig QTLdb (<http://www.animalgenome.org/cgi-bin/QTLdb/SS/index>)

Figure
Workflow of annotation and localization of Affymetrix GeneChip Porcine Genome Array probe sets
Diagramm der Arbeitsschritte zur Annotation und Lokalisation von Affymetrix GeneChip Porcine Genome Array Sonden

SsCrofa 9 database, but LatestGP (latest version of genomic sequence; http://www.ensembl.org/Sus_scrofa/Info/Index; last accessed 20 Nov. 2009). This database provided genomic and chromosome position which served to obtain localization data. The third pipeline designated »custom pipeline« combined queries of multiple data sources of homologous porcine, and heterologous human or mouse genomic and transcriptomic information, the application of thresholds regarding obtained identities, and the use of expanded sequences in heterologous BLAST search. For comparisons with NCBI databases (<http://www.ncbi.nlm.nih.gov/RefSeq>; last accessed 1 Jun 2009) BLAST hits were sorted by bit score and priority was given for identities to human, mouse and pig. For probe set consensus sequences with bit scores <50 (for comparisons with NCBI databases) in heterologous comparisons but high identity to SsCrofa 9 sequences, the porcine consensus sequence was expanded by 2 500 bp up-stream and down-stream obtained from SsCrofa 9 LatestGP (»expand« in table). Expanded sequences were BLASTed again against human, mouse and pig NCBI sequences. Similarly for comparison with EnSEMBL databases thresholds used to filter were by alignment length (≥ 100 bp) and % identity (≥ 90). Gene identity numbers, GI numbers, and GenBank accession numbers were used as primary key to obtain gene symbols, gene names and other data from HGNC and EASE. The assignments to chromosomes and positions were retrieved in priority of EnSEMBL SsCrofa 9 LatestGP and cDNA, pre-EnSEMBL SsCrofa 8.52 LatestGP, NCBI pig reference sequence RNA, NCBI pig genomic and PigQTL databases. Results from PigQTLdb were used to search the alignment of pig chromosomes to the human genome where Affymetrix microarray elements are virtually mapped.

Results

Three pipelines were used to obtain and integrate available information of porcine, human and mouse genome and transcriptome sequence information and to provide an up-to-date annotation of Affymetrix porcine probe sets. The first and second pipeline provided annotation and localization data obtained without applying thresholds to filter BLAST hits of low reliability (»non-filtered annotation and localization«). In the third pathway in particular also heterologous sequence information from human and mouse was used and thresholds for alignment length and % identity were introduced to obtain the most reliable annotations and localizations (»filtered annotation and localization«).

Comparison of consensus sequences used to derive probe sets represented on the Affymetrix GeneChip Porcine Genome Array with EnSEMBL SsCrofa 9 cDNA database revealed 23 799 probe sets with any BLAST hit. Results are presented in 26 columns summarizing raw BLAST result (Probe set ID, Blast header result, Blast pair-wise result, High Score, Smallest Sum Probability P(N), N, Query start, Query stop, Subject start, Subject stop, Alignment length, Identity value, Identity value full, % Identity, Chromosome position) and results obtained after querying EnSEMBL Transcript ID with BioMart (SMEDLEY *et al.* 2009) (EnSEMBL Transcript ID, EnSEMBL Gene ID, Gene symbol, Gene name, Chromosome, Gene start (bp), Gene end (bp), Transcript start (bp), Transcript end (bp), EntrezGene ID). The non-filtered data contained elements with % identity at 79-100% summing up to 19 730 gene symbols.

BLASTing EnSEMBL Sscrofa 9 LatestGP database revealed 23 298 probe sets with any BLAST hit. Outputs are summarized in columns showing the same as raw BLAST results as described above (here not results referring to EnSEMBL Transcript ID). All % identities ranged between 80 and 100%.

Integration of data retrieved by querying human, mouse and pig genomic and cDNA sequence databases and gene localization and annotation information and application of filters on degree of similarity led to results shown in 13 columns about annotation and 8 columns about localization of genes represented by the Affymetrix probe sets. NCBI BLAST to the reference RNA database (Probe Set Name, BLAST Result Name, % Identity, Alignment Length, Start Position, Stop Position, e-Value, Bit Score, Organism, Type of Sequence, Gene Symbol, Gene Name, Pathway related and Number of pathway related) was completed in June 2009 and revealed 20 689 probe sets meeting our threshold for filtering (18 177 human IDs, 1 023 mouse IDs and 1 489 pig IDs). Localization data are shown in columns of the datasheet referring to alignment length, % identity, high score (from EnSEMBL) or bit score (from NCBI), smallest sum probability P(N) (from EnSEMBL) or e-value (from NCBI), start position, stop position and chromosome. In total 20 439 probe sets were above the threshold of ≥ 100 bp of alignment length and $\geq 90\%$ identity (last accessed 20 Nov. 2009) (including data from EnSEMBL Sscrofa 9 LatestGP database: 16 513 probe sets, EnSEMBL Sscrofa 9 cDNA database: 874, pre-EnSEMBL Sscrofa 8.52 database: 773, NCBI Pig genomic database: 359, NCBI pig reference mRNA database: 70, Pig QTL database: 1 850). Annotation and localization data obtained from the three pipelines are available via the journal's webpage (<http://archanimbreed.com>) or from the authors.

Discussion

Using three pipelines annotations and localizations of probe sets of the Affymetrix GeneChip Porcine Genome Array were obtained querying EnSEMBL and NCBI databases including non-filtered BLAST results and results meeting thresholds regarding the degree of identity. Gene annotation provided by EnSEMBL includes automatic annotation, manual curation and other important databases (BIRNEY *et al.* 2004). The recent version of porcine genome sequence, the high-coverage Sscrofa 9 assembly (April 2009) including chromosomes 1 to 18 and X of the pig genome is based on the integrated highly contiguous physical map of the pig genome as a template for sequencing (HUMPHRAY *et al.* 2007). Querying the EnSEMBL cDNA database gained slightly more probe sets with any hit than LatestGP database and it is more suitable for obtaining annotation data via BioMart. However, EnSEMBL LatestGP database is beneficial for obtaining localization information of the probe sets which relies on whole genomic data. Non-filtered data of BLAST provided putative localization of nearly all probe sets (23 709 and 23 298 for EnSEMBL cDNA and LatestGP, respectively) however with some uncertainty in terms of annotation and localization. In contrast, the application of the thresholds of bit score > 50 for NCBI database analysis and of ≥ 100 bp of alignment length and $\geq 90\%$ identity for EnSEMBL queries in the third pipeline, revealed high degrees of concordance: of 16 513 and 9 313 probe sets for either genomic DNA or cDNA database queries

meeting these thresholds only 137 (0.8 and 1.4%, respectively) had localizations on different chromosomes depending on the source of data. The thresholds used here are comparable to the limits mentioned in the annotation provided by TSAI *et al.* (2006). In the third pipeline, providing the most reliable annotation and localization based on application of thresholds, the RefSeqs were used as primary key for querying gene names and gene symbols according to the new nomenclature from HGNC, which is more specific and well organized. Further we referred to EASE and Gene Entrez. The annotation of the Affymetrix GeneChip Porcine Genome Array (version 2009.03.12 build 28; <https://www.affymetrix.com/analysis/downloads/na28/ivt/Porcine.na28.annot.csv.zip>) showed 5 665 gene symbols. These gene data match the new annotation for 4 834 probe sets (85.33%). TSAI *et al.* (2006) used extended porcine sequences for BLAST against EnsEMBL human database for comparison. Most of the probe sets (23 946) were annotated when omitting any thresholds regarding degree of identity; 19 675 probe sets were annotated with bit scores >50 (2010 update 19 992). In the third pipeline we used expanded sequences for annotation in those cases only, when heterologous BLAST of the Affymetrix consensus sequence revealed no annotation (1 529 cases). According to the average gene length of 2 500 bp expanded sequences were used covering 2 500 bp up and down stream of the consensus sequence. These sequences were checked for their mapping position in the human-pig comparative map. Localization and chromosome position were primarily obtained from EnsEMBL Sscrofa 9. Further mapping information was taken from pre-EnsEMBL Sscrofa 8.52, NCBI pig reference sequence RNA and genomic database and PigQTLdb.

In summary, 20 689 probe sets were assigned to a known gene, 20 439 probe sets were *in silico* mapped onto the porcine genome sequence assembly, and for 18 561 we obtained both kind of information. This study shows how improvable *in silico* porcine annotation and localization is thanks to the current efforts of the Pig Genome Sequencing project. The annotation and localization of Affymetrix GeneChip Porcine Genome Array probe sets facilitates the application of this tool in functional genomics and genetical genomics studies in pigs.

Acknowledgements

This work was supported by the EC-funded FP6 Project »SABRE« (FOOD-CT-2006-01625) and by internal funds of the FBN and the Graduate School Chiang Mai University.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215, 403-10
- Bai Q, McGillivray C, da Costa N, Dornan S, Evans G, Stear MJ, Chang KC (2003) Development of a porcine skeletal muscle cDNA microarray: analysis of differential transcript expression in phenotypically distinct muscles. *BMC Genomics* 4, 8
- Band MR, Olmstead C, Everts RE, Liu ZL, Lewin HA (2002) A 3800 gene microarray for cattle functional genomics: comparison of gene expression in spleen, placenta, and brain. *Anim Biotechnol* 13, 163-72
- Bernard C, Cassar-Malek I, Le Cunff M, Dubroeuq H, Renand G, Hocquette, JF (2007) New indicators of beef sensory quality revealed by expression of specific genes. *J Agric Food Chem* 55, 5229-37

- Birney E, Andrews TD, Bevan P, Caccamo M, Chen Y, Clarke L, Coates G, Cuff J, Curwen V, Cutts T, Down T, Eyras E, Fernandez-Suarez XM, Gane P, Gibbins B, Gilbert J, Hammond M, Hotz HR, Iyer V, Jekosch K, Kahari A, Kasprzyk A, Keefe D, Keenan S, Lehtvaslaiho H, McVicker G, Melsopp C, Meidl P, Mongin E, Pettett R, Potter S, Proctor G, Rae M, Searle S, Slater G, Smedley D, Smith J, Spooner W, Stabenau A, Stalker J, Storey R, Ureta-Vidal A, Woodward KC, Cameron G, Durbin R, Cox A, Hubbard T, Clamp M (2004) An overview of Ensembl. *Genome Res* 14, 925-28
- Bruford EA, Lush MJ, Wright MW, Sneddon TP, Povey S, Birney E (2008) The HGNC Database in 2008: a resource for the human genome. *Nucleic Acids Res* 36 (Database issue), 445-8
- Cogburn LA, Wang X, Carre W, Rejto L, Porter TE, Aggrey SE, Simon J (2003) Systems-wide chicken DNA microarrays, gene expression profiling, and discovery of functional genes. *Poult Sci* 82, 939-51
- Douglas A H, Glynn D Jr, Brad TS, Lane HC, Lempicki RA (2003) Identifying Biological Themes within Lists of Genes with EASE. *Genome Biol* 4, 70
- Hartmann A, Nuernberg G, Repsilber D, Janczyk P, Walz C, Ponsuksili S, Souffrant W-B, Schwerin M (2009) Effects of threshold choice on the results of gene expression profiling, using microarray analysis, in a model feeding experiment with rats. *Arch Tierz* 52, 65-78
- Hintermair V (2007) Comparative Analysis of *Bos taurus* and *Homo sapiens* DNA Microarrays. Master Thesis, Citation key: bioinflmu-528, LMU München, Germany
- Hu Z, Dracheva S, Jang W, Maglott D, Bastiaansen J, Rothschild MF, Reecy JM (2005) A QTL resource and comparison tool for pigs: PigQTLDB. *Mamm Genome* 16, 792-800
- Hubbard TJ, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyras E, Gilbert J, Hammond M, Huminiński L, Kasprzyk A, Lehtvaslaiho H, Lijnzaad P, Melsopp C, Mongin E, Pettett R, Pocock M, Potter S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, Clamp M (2002) The Ensembl genome database project. *Nucleic Acids Res* 30, 38-41
- Hubbard TJ, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, Down T, Dyer SC, Fitzgerald S, Fernandez-Banet J, Graf S, Haider S, Hammond M, Herrero J, Holland R, Howe K, Howe K, Johnson N, Kahari A, Keefe D, Kokocinski F, Kulesha E, Lawson D, Longden I, Melsopp C, Megy K, Meidl P, Ouverdin B, Parker A, Prlic A, Rice S, Rios D, Schuster M, Sealy I, Severin J, Slater G, Smedley D, Spudich G, Trevanion S, Vilella A, Vogel J, White S, Wood M, Cox T, Curwen V, Durbin R, Fernandez-Suarez XM, Flicek P, Kasprzyk A, Proctor G, Searle S, Smith J, Ureta-Vidal A, Birney E (2007) Ensembl 2007. *Nucleic Acids Research* 35 (Database issue), D610-7
- Humphray SJ, Scott CE, Clark R, Marron B, Bender C, Camm N, Davis J, Jenks A, Noon A, Patel M, Sehra H, Yang F, Rogatcheva MB, Milan D, Chardon P, Rohrer G, Nonneman D, de Jong P, Meyers SN, Archibald A, Beever JE, Schook LB, Rogers J (2007) A high utility integrated map of the pig genome. *Genome Biol* 8, 139
- Lee Y, Tsai J, Sunkara S, Karamycheva S, Pertea G, Sultana R, Antonescu V, Chan A, Cheung F, Quackenbush J (2005) The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. *Nucleic Acids Res* 33 (Database Issue), 71-4
- Moser RJ, Reverter A, Kerr CA, Beh KJ, Lehnert SA (2004) A mixed-model approach for the analysis of cDNA microarray gene expression data from extreme-performing pigs after infection with *Actinobacillus pleuropneumoniae*. *J Anim Sci* 82, 1261-71
- Ponsuksili S, Jonas E, Murani E, Phatsara C, Srikanthai T, Walz C, Schwerin M, Schellander K, Wimmers K (2008a) Trait correlated expression combined with expression QTL analysis reveals biological pathways and candidate genes affecting water holding capacity of muscle. *BMC Genomics* 9, 367
- Ponsuksili S, Murani E, Phatsara C, Jonas E, Walz C, Schwerin M, Schellander K, Wimmers K (2008b) Expression profiling of muscle reveals transcripts differentially expressed in muscle that affect water-holding capacity of pork. *J Agric Food Chem* 56, 10311-17
- Ponsuksili S, Murani E, Walz C, Schwerin M, Wimmers K (2007) Pre- and postnatal hepatic gene expression profiles of two pig breeds differing in body composition: insight into pathways of metabolic regulation. *Physiol Genomics* 29, 267-79
- Pruitt KD, Tatusova T, Maglott DR (2007) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 35 (Database issue), 61-5
- Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G, Kasprzyk A (2009) BioMart - biological queries made easy. *BMC Genomics* 10, 22
- Suchyta SP, Sipkovsky S, Kruska R, Jeffers A, McNulty A, Coussens MJ, Tempelman RJ, Halgren RG, Saama PM, Bauman DE, Boisclair YR, Burton JL, Collier RJ, DePeters EJ, Ferris TA, Lucy MC, McGuire MA, Medrano JF, Overton TR, Smith TP, Smith GW, Sonstegard TS, Spain JN, Spiers DE, Yao J, Coussens PM (2003) Development and testing of a high-density cDNA microarray resource for cattle. *Physiol Genomics* 15, 158-64

- Tsai S, Cassady JP, Freking BA, Nonneman DJ, Rohrer GA, Piedrahita JA (2006) Annotation of the affymetrix porcine genome microarray. *Anim Genet* 37, 423-24.
- Tuggle CK, Wang Y, Couture O (2007) Advances in swine transcriptomics. *Int J Biol Sci* 3, 132-52
- Wain HM, Bruford EA, Lovering RC, Lush MJ, Wright MW, Povey S (2002) Guidelines for Human Gene Nomenclature. *Genomics* 79, 464-70

Received 6 July 2009, accepted 7 December 2009.

Corresponding author:

SIRILUCK PONSUKSILI

email: s.wimmers@fhn-dummerstorf.de

Research Group Functional Genomics, Leibniz Institute for Farm Animal Biology, Wilhelm-Stahl-Allee 2,
18196 Dummerstorf, Germany
