Animal (2011), **5:6**, pp 861–866 © The Animal Consortium 2011
doi:10.1017/S1751731110002533

Use of the Affymetrix Human GeneChip array and genomic DNA hybridisation probe selection to study ovine transcriptomes

N. S. Graham[1], S. T. May[1], Z. C. T. R. Daniel[2], Z. F. Emmerson[1], J. M. Brameld[2†] and T. Parr[2]

[1]*Division of Plant and Crop Sciences, Nottingham Arabidopsis Stock Centre, School of Biosciences, University of Nottingham, Loughborough, LE12 5RD, UK;*
[2]*Division of Nutritional Sciences, School of Biosciences, University of Nottingham, Loughborough, LE12 5RD, UK*

(Received 6 August 2010; Accepted 12 October 2010; First published online 4 January 2011)

Affymetrix GeneChip microarrays are a powerful tool to study global gene expression profiles and have been used on many species. However, no sheep-specific Affymetrix GeneChip is currently available and the bovine array is fairly limited in its coverage and annotation. Previously, a probe-selection method based on hybridisation of genomic DNA (gDNA) was developed, which enables GeneChips to be used for species that they were not designed for. This approach can greatly increase the number of potential annotated genes that can be studied beyond that which is currently available, particularly for species that do not have comprehensively characterised genomes. In this study, we demonstrate that gDNA-based probe selection on the Affymetrix Human U133+2 GeneChip array can be used to study gene expression profiles in sheep tissues. More than 20 000 transcripts were detected in triplicate ovine skeletal muscle and liver samples, which is more than would be possible using the commercially available sheep-specific microarray. The majority of the top 15 differentially expressed genes for each tissue were known to either be expressed in a tissue-specific manner or relate to specific functions of that tissue. Gene ontology analysis of the differentially expressed genes revealed the expected differences in gene expression profiles between the two tissues. Besides demonstrating that the probe selection method can be used to study the ovine transcriptome, the benefits of this approach are that it can greatly increase the number of annotated and novel genes that can be studied beyond those currently possible using ovine- or bovine-specific microarrays. This same method also has the potential to allow the study of other species where species-specific microarrays are not available or whose genomes have not been comprehensively characterised.

Keywords: microarray, cross-species hybridisation, sheep, skeletal muscle, liver

Implications

This method has the potential to allow the study of global gene expression profiles in tissues/cells from species for which (i) there is no specific microarray available; (ii) the genome has not been sequenced; or (iii) the annotation is poor or at a very early stage. It utilises the genomic resources available for humans (or rodents) to allow researchers to study other species, without the need to sequence and annotate the whole genome for that species.

Introduction

A range of microarray platforms have been developed to investigate global gene expression profiles (Bar-Or *et al.*, 2007), including spotted arrays (using cDNAs, PCR products or oligonucleotides) and *in situ* synthesised arrays (e.g. Agilent SurePrint (Agilent Technologies, Stockport, Cheshire, UK) and Affymetrix (Santa Clara, CA, USA) GeneChip arrays). A few sheep microarray studies have been performed using in-house sheep-specific targeted arrays (Keane *et al.*, 2006; Galindo *et al.*, 2008; Watkins *et al.*, 2008) or the Affymetrix Bovine array (Fleming-Waddell *et al.*, 2007; Vuocolo *et al.*, 2007; Rowe *et al.*, 2008), with the latter assuming high homology between these species. Although a sheep-specific microarray is commercially available (Agilent), we chose to use the human Affymetrix GeneChip for sheep studies because it has (i) probes for more transcripts than the ovine and bovine arrays (45 K compared to 15 and 23 K, respectively); (ii) better coverage of the genome, including novel genes of unknown function; and (iii) more extensive gene annotation than the ovine or bovine arrays. In addition, this approach could potentially be used for other species where a specific array is unavailable. The design of Affymetrix GeneChips makes them ideal for use in cross-species experiments (Bar-Or *et al.*, 2007). On each GeneChip, a gene transcript is represented by a probe-set consisting of up to

[†] E-mail: john.brameld@nottingham.ac.uk

16 probe-pairs, each made up of a 25 base oligo perfect-match (PM) probe and the corresponding mis-match (MM) probe, with an MM at the 13th base, thereby measuring non-specific binding (Lipschutz *et al.*, 1999). In plants, a mixed molecular and bioinformatic approach based on probe selection via genomic DNA (gDNA) hybridisation demonstrated that the *Arabidopsis* ATH1 array could be used to study other plant species (Hammond *et al.*, 2005 and 2006; Broadley *et al.*, 2008; Morinaga *et al.*, 2008). The aim of this study was to test the use of the human Affymetrix GeneChip array to study ovine transcriptomes by comparing gene expression profiles in sheep liver and muscle using the probe selection method.

Material and methods

All animal procedures were approved by the University of Nottingham Ethics Committee and carried out in accordance with UK Home Office Guidelines.

gDNA hybridisation and probe selection

gDNA was extracted from whole sheep blood as per manufacturer's instructions using the Wizard Genomic DNA Purification kit (Promega, Southampton, UK), before being labelled, hybridised to the Affymetrix Human U133+2 array, scanned and gDNA cell intensity files (.cel files) generated, all as described previously (Hammond *et al.*, 2005). The next step was to identify probe-pairs in which the PM probe has a gDNA hybridisation intensity greater than the user defined threshold (Hammond *et al.*, 2005). Probe-pairs from the gDNA .cel files were selected using a .cel file parser script (http://affymetrix.arabidopsis.info/xspecies/), which produces a probe mask file (.cdf) compatible with a range of microarray analysis packages (e.g. Genespring). Probe mask files (.cdf) were produced using gDNA hybridisation intensity thresholds ranging from 0 to 1000.

Sheep tissue samples and RNA isolation

Samples of liver and *Longissimus dorsi* skeletal muscle were obtained from three male Mule × Charolais lambs (approximate age 120 days) within 5 min of death, quickly frozen in liquid nitrogen and stored at −80°C. Total RNA was isolated from all samples using Trizol as per the manufacturer's instructions including glycogen removal (Invitrogen, Paisley, UK), followed by an additional DNase treatment step (Promega). The isolated total RNA was resuspended in RNase-free water with yield and purity determined using an Agilent 2100 Bioanalyser (Agilent Technologies, Stockport, UK).

RNA hybridisation

Detailed methods were as described previously (Hammond *et al.*, 2005). Briefly, approximately 1 μg of total RNA was used to generate first strand cDNA by reverse transcription followed by synthesis of second strand cDNA. Double stranded cDNA products were purified and *in vitro* transcribed to generate biotinylated complementary RNAs (cRNAs), then purified and randomly fragmented before being hybridised on Affymetrix Human U133+2 GeneChip arrays and stained with streptavidin–phycoerythrin. Arrays were scanned and .cel raw data files generated. The DNA and RNA .cel files and the probe mask .cdf files are available to download from the NASC Xspecies website (http://affymetrix.arabidopsis.info/xspecies/).

Data analysis

Initially the RNA .cel files were loaded into Genespring GX 7.3 (Agilent technologies) using RMA normalisation algorithm (Irizarry *et al.*, 2003). The RNA files were reanalysed using the .cdf files generated from the gDNA hybridisation. Genes were selected as differentially expressed if they were more than twofold different and had a *P*-value <0.05 from an ANOVA test (using no false discovery rate correction). Gene Ontology (GO) analysis was performed using GO browser function in Genespring, which calculates a hypergeometric *P*-value.

Results

Probe selection using gDNA hybridisation

For probe selection to allow the use of the Human GeneChip array to study global gene expression in sheep tissues, sheep gDNA was hybridised to the Human U133 +2 array. After hybridisation, a probe-pair was retained if its PM hybridisation value exceeded a series of thresholds (ranging from 0 to 1000) using a .cel file parser (Hammond *et al.*, 2005). The number of probe-pairs retained in the resulting probe-mask files reduced rapidly with increased threshold (Figure 1), while the number of probe-sets (i.e. genes) retained reduced at a slower rate, indicating that gene retention was good at increased gDNA hybridisation thresholds.

Analysis of sheep tissue samples

To demonstrate that probe selection can enable ovine RNA transcriptome quantification, triplicate muscle and liver RNA samples were labelled and hybridised to the Human U133 +2 array and analysed using the probe-mask files generated from the gDNA hybridisation results. The number of genes differentially expressed between muscle and liver (more than twofold difference, $P < 0.05$) was determined using the probe-mask files corresponding to gDNA hybridisation thresholds ranging from 0 to 1000. Without probe selection (threshold = 0), 709 genes were differentially expressed (Table 1), which increased as the threshold increased, reaching a maximum of 2815 differentially expressed genes at a threshold of 450. Above this threshold the number of differentially expressed genes reduced because the number of probe-sets (i.e. genes) retained in the probe mask files was reduced. Thus, a threshold of 450 was chosen for subsequent analysis, since it allowed good sensitivity to detect the maximum number of transcripts.

At a gDNA hybridisation threshold of 450, 83896 probe-pairs were retained, representing 37 863 probe sets (i.e. gene

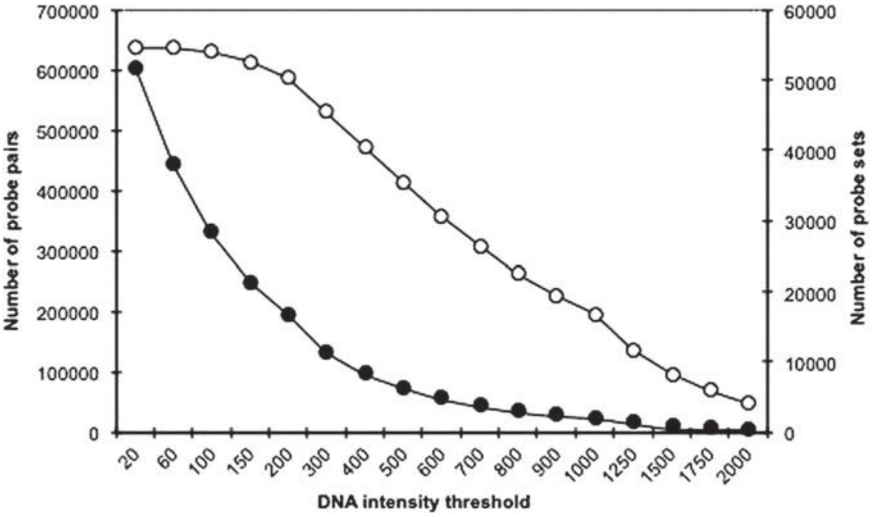

Figure 1 Number of probe-pairs (filled circles) and probe-sets (closed circles) retained from the Affymetrix U133+2 array as a function of the genomic DNA hybridisation intensity thresholds used to generate the probe mask files.

Table 1 *Number of genes differentially expressed between sheep liver and muscle using a range of probe mask thresholds*

Probe mask threshold[a]	Differentially expressed genes[b]
0	709
50	856
100	1300
150	1783
200	2235
250	2482
300	2665
350	2709
400	2779
450	2815
500	2755
550	2641
600	2539
700	2272
800	1997
900	1761
1000	1524

[a] gDNA hybridisation threshold used to generate the probe mask filter.
[b] Number of differentially expressed genes.

transcripts). The number of probe-pairs retained per probe-set varied between 1 and 14, with an average of 2.2. The RNA hybridisation data were analysed in more detail using the probe-mask generated at a gDNA hybridisation threshold of 450. In total, 20 699 gene transcripts (RMA normalised signal value >20) were detected in the two tissues (37.8% of total), with 18 693 gene transcripts detected in muscle and 18 787 detected in liver. Of the 2815 differentially expressed gene transcripts (more than twofold change, see supplementary data), 1051 were higher in muscle and 1764 higher in liver. As expected, the differentially expressed genes included many already known to be expressed in a tissue-specific manner or associated with tissue specific functions. For example, genes over-expressed in muscle compared to liver included ryanodine receptor 1, nebulin, myosin light and heavy polypeptides and creatine kinase (Table 2), all muscle-specific proteins; whereas genes over-expressed in liver compared with muscle included pantothenate kinase 1, squalene epoxidase, plasminogen and β-glucosidase (Table 3), all relating to known liver functions. Interestingly, the top 15 genes for each tissue also included some potentially novel 'Hypothetical proteins' with unknown functions in muscle or liver, which would not have been identified using the ovine or bovine arrays. Differentially expressed genes were further analysed by grouping them based on their GOs (http://www.geneontology.org/). As expected, the significant GO terms (ranked on *P*-value) for genes more than twofold higher in muscle compared to liver included 'muscle development', 'muscle contraction' and 'myoblast differentiation' (Table 4); while the GO terms of genes differentially expressed in liver included 'generation of precursor metabolites and energy', 'lipid metabolism' and 'cell redox homeostasis' (Table 5).

Discussion

The relative changes in probe pairs and probe sets with increasing threshold are consistent with previous studies in plants using the same technique (Hammond *et al*., 2005 and 2006; Graham *et al*., 2007). The results of the RNA analysis comparing skeletal muscle and liver transcriptomes demonstrate that the gDNA probe selection method is suitable for studying gene expression profiles in sheep tissues and produces biologically relevant data. The selection of the gDNA threshold used to perform the analysis has an effect on the number of differentially expressed genes identified. Increasing the threshold will increase the specificity (i.e. sequence similarity between sheep RNA and human probes) of the probes retained, but as this increases, the number of probe-sets and hence genes detected will decrease.

Table 2 *Top 15 genes over-expressed in sheep muscle compared to liver*

Probe set ID	Fold change (muscle to liver)	*P*-value	Gene name	Genbank accession no.	Description
236320_at	206.61	3.93E-05	CCDC17	AA010540	Coiled-coil domain containing protein 17
203861_s_at	181.16	0.00014	ACTN2	AU146889	Actin, alpha 2
1570569_at	180.83	2.51E-05	–	BC024156	cDNA clone
205485_at	169.2	4.15E-06	RYR1	NM_000540	Ryanodine receptor 1
205054_at	151.29	0.00157	NEB	NM_004543	nebulin
201539_s_at	141.04	0.000187	FHL1	U29538	Four and a half LIM domains 1
209742_s_at	134.59	0.00247	MYL2	AF020768	Myosin, light polypeptide 2
205295_at	134.41	0.00189	CKMT2	NM_001825	Creatine kinase
241332_at	124.07	1.62E-05	–	AW665136	Transcribed sequence
214121_x_at	102.88	0.000407	PDLIM7	AA086229	PDZ and LIM domain 7
204631_at	99.01	0.0058	MYH2	NM_017534	Myosin, heavy polypeptide 2
227951_s_at	99.01	0.00486	LOC147965	AW338561	Hypothetical protein
206394_at	99.01	0.0043	MYBPC2	NM_004533	Myosin binding protein C
235313_at	84.03	0.000144	NRAP	AA195854	Nebulin-related anchoring protein
214468_at	78.74	0.00112	MYH6	D00943	Myosin heavy peptide 6

Table 3 *Top 15 genes over-expressed in sheep liver compared to muscle*

Probe set ID	Fold change (liver to muscle)	*P*-value	Gene name	Genbank accession no.	Description
22145_s_at	82.01	0.0385	FZD5	NM_030804	Frizzled homolog 5
226649_at	47.51	0.0351	PANK1	AI373299	Pantothenate kinase 1
207958_at	44.25	0.0415	UGT2A1	NM_006798	UDP glycosyltranferase 2, polypeptide A1
213562_s_at	41.05	0.00742	SQLE	BF979497	Squalene epoxidase
234287_at	39.89	0.0439	HLCS	AJ001863	mRNA trapped exon b910
201120_s_at	37.56	0.0253	PGRMC1	AL547946	Progesterone receptor membrane component 1
1555441_at	34.64	0.0177	UBA6	BC031637	Ubiquitin-like 6
209978_s_at	29.26	0.0418	PLG	M74220	plasminogen
223798_at	27.46	0.00554	SLC41A2	AL136828	Solute carrier family 41, member 2
227395_at	27.03	0.0119	MGC10233	BE672045	Hypothetical protein
219954_s_at	26.07	0.0321	GBA3	NM_020973	Glucosidase, beta, acid 3
237292_at	23.49	0.0391	DPYSL3	BE669707	Dihydropyrimidinase-like 3
231225_at	22.83	0.0232	ZNF611	AI568622	Transcribed sequence
231783_at	21.3	0.0231	CHRM1	AI500293	Cholinergic receptor, muscarinic 1
231370_at	20.59	0.00966	PPM1A	AI701170	Protein phosphatase 1A, alpha isoform

Table 4 *Over-represented gene ontology (GO) terms for probe-sets over-expressed (more than twofold,* $P < 0.05$*) in sheep muscle compared to liver*

Category[a]	Genes in category[b]	Percent of genes in category[c]	Genes in list in category[d]	Percent of genes in list in category[e]	*P*-value[f]
GO:7517: muscle development	370	1.097	42	5.891	9.65E-19
GO:6936: muscle contraction	336	0.996	38	5.33	5.05E-17
GO:6941: striated muscle contraction	67	0.199	17	2.384	3.49E-14
GO:48741: skeletal muscle fiber development	122	0.362	18	2.525	1.11E-10
GO:48747: muscle fiber development	122	0.362	18	2.525	1.11E-10
GO:48637: skeletal muscle development	124	0.368	18	2.525	1.46E-10
GO:42692: muscle cell differentiation	124	0.368	17	2.384	1.17E-10
GO:45445: myoblast differentiation	97	0.288	15	2.104	2.02E-09
GO:7519: striated muscle development	195	0.578	20	2.805	7.28E-09
GO:8016: regulation of heart contraction rate	93	0.276	12	1.683	6.2E-07

[a] GO term.
[b] Number of probe-sets present on array within GO term.
[c] Percent of total probe-sets within GO term.
[d] Number of differentially expressed probe-sets within GO term.
[e] Percent of differentially expressed probe-sets within GO term.
[f] Hypergeometric *P*-value.

Table 5 *Over-represented gene ontology (GO) terms for probe-sets over-expressed (more than twofold,* P < 0.05*) in sheep liver compared to muscle*

Category[a]	Genes in category[b]	Percent of genes in category[c]	Genes in list in category[d]	Percent of genes in list in category[e]	*P*-value[f]
GO:6091: generation of precursor metabolites and energy	1144	3.391	70	5.604	3.11E-05
GO:6629: lipid metabolism	1364	4.043	78	6.245	0.000106
GO:45454: cell redox homeostasis	132	0.391	14	1.121	0.000398
GO:44255: cellular lipid metabolism	1121	3.323	64	5.124	0.000448
GO:8202: steroid metabolism	371	1.1	27	2.162	0.000728
GO:50875: cellular physiological process	21659	64.2	854	68.37	0.000861
GO:51234: establishment of localization	6334	18.77	278	22.26	0.000913
GO:51179: localization	6423	19.04	280	22.42	0.00131
GO:7626: locomotory behavior	972	2.881	55	4.404	0.00134
GO:43170: macromolecule metabolism	9541	28.28	401	32.11	0.00137

[a]GO term.
[b]Number of probe-sets present on array within GO term.
[c]Percent of total probe-sets within GO term.
[d]Number of differentially expressed probe-sets within GO term.
[e]Percent of differentially expressed probe-sets within GO term.
[f]Hypergeometric *P*-value.

Conversely, if a low threshold is used, probes will be retained that have low specificity, and hence the gene signals may be less accurate. In this study, the threshold that gave the greatest sensitivity (greatest number of differentially expressed genes) was used. This includes a number of transcripts (probe sets) with only one probe-pair being retained, but is equivalent to using a single oligonucleotide array. Since the number of probe-pairs retained per probe-set will vary with threshold, this needs to be considered when interpreting the data. Probe-sets with only a single probe retained may not be specific to a single gene and may be representative of a gene family. The advantages of using the human GeneChip array over available ovine or bovine arrays are: (i) it contains more transcripts (45 K compared with 15 and 23 K for ovine and bovine arrays); (ii) it has better coverage of the whole genome, including novel genes of unknown function; and (iii) it has more extensive gene annotation. In addition, labelling, hybridisation and processing of Affymetrix GeneChip arrays is highly automated, thereby producing very reproducible results, whereas cDNA arrays are less automated and therefore likely to be less reproducible. In addition, GeneChips have been shown to be more specific and have a greater dynamic range than cDNA arrays (Mah *et al.*, 2004; Woo *et al.*, 2004) and there are numerous publically available data sets for human GeneChips that could be directly compared with experiments performed using this technique. This method will allow transcriptome analysis of developmental and metabolic processes, as well as disease responses, in sheep and possibly other species for which specific microarrays are not available.

Acknowledgements

The authors gratefully acknowledge D. Surgay for technical assistance. N.G. was supported by a BBSRC grant (grant reference BBE01772×1).

References

Bar-Or C, Czosnek H and Koltai H 2007. Cross-species hybridizations: a developing tool for studying species diversity. Trends in Genetics 23, 200–207.

Broadley MR, White PJ, Hammond JP, Graham NS, Bowen HC, Emmerson ZF, Fray RG, Iannetta PPM, McNicol JW and May ST 2008. Evidence of neutral transcriptome evolution in plants. New Phytologist 180, 587–593.

Fleming-Waddell JN, Wilson LM, Olbricht GR, Vuocolo T, Byrne K, Craig BA, Tellam RL, Cockett NE and Bidwell CA 2007. Analysis of gene expression during the onset of muscle hypertrophy in callipyge lambs. Animal Genetics 38, 28–36.

Galindo RC, Munoz PM, de Miguel MJ, Marin CM, Blasco JM, Gortazar C, Kocan KM and de la Fuente J 2008. Differential expression of inflammatory and immune response genes in rams experimentally infected with a rough virulent strain of Brucella ovis. Veterinary Immunology and Immunopathology 127, 295–303.

Graham NS, Broadley MR, Hammond JP, White PJ and May ST 2007. Optimising the analysis of transcript data using high density oligonucleotide arrays and genomic DNA-based probe selection. BMC Genomics 8, 344.

Hammond JP, Broadley MR, Craigon DJ, Higgins J, Emmerson ZF, Townsend HJ, White PJ and May ST 2005. Using genomic DNA-based probe-selection to improve the sensitivity of high-density oligonucleotide arrays when applied to heterologous species. Plant Methods 1, 10.

Hammond JP, Bowen HC, White PJ, Mills V, Pyke KA, Baker AJM, Whiting SN, May ST and Broadley MR 2006. A comparison of Thlaspi caerulescens and Thlaspi arvense shoot transcriptomes. New Phytologist 170, 239–260.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U and Speed TB 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4, 249–264.

Keane OM, Zadissa A, Wilson T, Hyndman DL, Grere GJ, Baird DB, McCulloch AF, Crawford AM and McEwan JC 2006. Gene expression profiling of Naïve sheep genetically resistant and susceptible to gastrointestinal nematodes. BMC Genomics 7, 42.

Lipschutz RJ, Fodor SPA, Gingeras TR and Lockhart DJ 1999. High density synthetic oligonucleotide arrays. Nature Genetics 21, 20–24.

Mah N, Thelin A, Lu T, Nikolaus S, Kühbacher T, Gurbuz Y, Eickhoff H, Klöppel G, Lehrach H, Mellgård B, Costello CM and Schreiber S 2004. A comparison of oligonucleotide and cDNA-based microarray systems. Physiological Genomics 16, 361–370.

Morinaga S-I, Nagano AJ, Miyazaki S, Kubo M, Demura T, Fukuda H, Sakai S and Hasebe M 2008. Ecogenomics of cleistogamous and chasmogamous flowering: genome-wide gene expression patterns from cross-species microarray in Cardamine kokaiensis (Brassicaceae). Journal of Ecology 96, 1086–1097.

Rowe A, Gondro C, Emery D and Sangster N 2008. Genomic analyses of Haemonchus contortus infection in sheep: abomasal fistulation and two Haemonchus strains do not substantially confound host gene expression in microarrays. Veterinary Parasitology 154, 71–81.

Vuocolo T, Byrne K, White J, McWilliam S, Reverter A, Cockett NE and Tellam RL 2007. Identification of a gene network contributing to hypertrophy in callipyge skeletal muscle. Physiological Genomics 28, 253–272.

Watkins C, McKellar A, Jensen K, George A, Jones D, Sharp MJ, Stevenson K and Hopkins J 2008. Development and validation of an oligonucleotide microarray for immuno-inflammatory genes of ruminants. Veterinary Research Communications 32, 647–657.

Woo Y, Affourit J, Daigle S, Viale A, Johnson K, Naggert J and Churchhill G 2004. A comparison of cDNA, oligonucleotide, and Affymetrix GeneChip gene expression microarray platforms. Journal of Biomolecular Techniques 15, 276–284.