A peer-reviewed version of this preprint was published in PeerJ on 8 March 2016.

Xiang R, Oddy VH, Archibald AL, Vercoe PE, Dalrymple BP. 2016. Epithelial, metabolic and innate immunity transcriptomic signatures differentiating the rumen from other sheep and mammalian gastrointestinal tract tissues. PeerJ 4:e1762 https://doi.org/10.7717/peerj.1762

Epithelial, metabolic and innate immunity transcriptomic signatures differentiating the rumen from other sheep and mammalian gastrointestinal tract tissues

Ruidong Xiang, Victor Hutton Oddy, Alan L. Archibald, Phillip E. Vercoe, Brian P. Dalrymple

Background. Ruminants are successful herbivorous mammals, in part due to their specialized forestomachs, the rumen complex, which facilitates the conversion of feed to soluble nutrients by micro-organisms. Is the rumen complex a modified stomach expressing new epithelial (cornification) and metabolic programs, or a specialised stratified epithelium that has acquired new metabolic activities, potentially similar to those of the colon? How has the presence of the rumen affected other sections of the gastrointestinal tract (GIT) of ruminants compared to non-ruminants? **Methods.** Transcriptome data from 11 tissues covering the sheep GIT, two stratified epithelial and two control tissues, was analysed using principal components to cluster tissues based on gene expression profile similarity. Expression profiles of genes along the sheep GIT were used to generate a network to identify genes enriched for expression in different compartments of the GIT. The data from sheep was compared to similar data sets from two non-ruminants, pigs (closely related) and humans (more distantly related). **Results.** The rumen transcriptome clustered with the skin and tonsil, but not the GIT transcriptomes, driven by genes from the epidermal differentiation complex, and genes encoding stratified epithelium keratins and innate immunity proteins. By analysing all of the gene expression profiles across tissues together 16 major clusters were identified. The strongest of these, and consistent with the high turnover rate of the GIT, showed a marked enrichment of cell cycle process genes (P=1.4E-46), across the whole GIT, relative to liver and muscle, with highest expression in the caecum followed by colon and rumen. The expression patterns of several membrane transporters (chloride, zinc, nucleosides, amino acids, fatty acids, cholesterol and bile acids) along the GIT was very similar in sheep, pig and humans. In contrast, short chain fatty acid uptake and metabolism appeared to be different between the species and different between the rumen and colon in sheep. The importance of nitrogen and iodine recycling in sheep was highlighted by the highly preferential expression of *SLC14A1*-urea (rumen), RHBG-ammonia (intestines) and *SLC5A5*-iodine (abomasum). The gene encoding a poorly characterized member of the maltase-glucoamylase family (MGAM2), predicted to play a role in the degradation of starch or glycogen, was highly expressed in the small and

large intestines. **Discussion.** The rumen appears to be a specialised stratified cornified epithelium, probably derived from the oesophagus, which has gained some liver-like and other specialized metabolic functions, but probably not by expression of pre-existing colon metabolic programs. Changes in gene transcription downstream of the rumen also appear have occurred as a consequence of the evolution of the rumen and its effect on nutrient composition flowing down the GIT.

Epithelial, metabolic and innate immunity transcriptomic signatures differentiating the rumen from other sheep and mammalian gastrointestinal tract tissues

Ruidong Xiang[1], V. Hutton Oddy[2], Alan L. Archibald[3], Phillip E. Vercoe[4], Brian P. Dalrymple[1,*]

[1]CSIRO Agriculture, 306 Carmody Road, St Lucia, QLD 4067, Australia

[2]NSW Department of Primary Industries, Beef Industry Centre, University of New England, Armidale, NSW 2351, Australia

[3]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian EH25 9RG, UK.

[4]School of Animal Biology and Institute of Agriculture, The University of Western Australia, 35 Stirling Highway Crawley WA 6009, Australia.

Email addresses:

Ruidong Xiang: ruidong.xiang@csiro.au

Hutton Oddy: hutton.oddy@dpi.nsw.gov.au

Alan Archibald: alan.archibald@roslin.ed.ac.uk

Philip E. Vercoe: philip.vercoe@uwa.edu.au

19 *Brian P. Dalrymple: brian.dalrymple@csiro.au

20 *: corresponding author

Abstract

Background. Ruminants are successful herbivorous mammals, in part due to their specialized forestomachs, the rumen complex, which facilitates the conversion of feed to soluble nutrients by micro-organisms. Is the rumen complex a modified stomach expressing new epithelial (cornification) and metabolic programs, or a specialised stratified epithelium that has acquired new metabolic activities, potentially similar to those of the colon? How has the presence of the rumen affected other sections of the gastrointestinal tract (GIT) of ruminants compared to non-ruminants?

Methods. Transcriptome data from 11 tissues covering the sheep GIT, two stratified epithelial tissues and two control tissues, was analysed using principal components to cluster tissues based on gene expression profile similarity. Expression profiles of genes along the sheep GIT were used to generate a network to identify genes enriched for expression in different compartments of the GIT. The data from sheep was compared to similar data sets from two non-ruminants, pigs (closely related) and humans (more distantly related).

Results. The rumen transcriptome clustered with the skin and tonsil, but not the GIT transcriptomes, driven by genes from the epidermal differentiation complex, and genes encoding stratified epithelium keratins and innate immunity proteins. By analysing all of the gene expression profiles across tissues together 16 major clusters were identified. The strongest of these, and consistent with the high turnover rate of the GIT, showed a marked enrichment of cell cycle process genes (P=1.4E-46), across the whole GIT, relative to liver and muscle, with highest expression in the caecum followed by colon and rumen. The expression patterns of several membrane transporters (chloride, zinc, nucleosides, amino acids, fatty acids, cholesterol

and bile acids) along the GIT was very similar in sheep, pig and humans. In contrast, short chain fatty acid uptake and metabolism appeared to be different between the species and different between the rumen and colon in sheep. The importance of nitrogen and iodine recycling in sheep was highlighted by the highly preferential expression of *SLC14A1*-urea (rumen), RHBG-ammonia (intestines) and *SLC5A5*-iodine (abomasum). The gene encoding a poorly characterized member of the maltase-glucoamylase family (MGAM2), predicted to play a role in the degradation of starch or glycogen, was highly expressed in the small and large intestines.

Discussion. The rumen appears to be a specialised stratified cornified epithelium, probably derived from the oesophagus, which has gained some liver-like and other specialized metabolic functions, but probably not by expression of pre-existing colon metabolic programs. Changes in gene transcription downstream of the rumen also appear have occurred as a consequence of the evolution of the rumen and its effect on nutrient composition flowing down the GIT.

INTRODUCTION

The ruminants, of which sheep, cattle, buffalo and goats are the major domesticated species, are now the most numerous large herbivores on earth. Their success is largely due to their specialized forestomachs, the rumen complex (the rumen, reticulum and omasum), and to rumination, the process of recycling the partially digested material via the mouth to reduce particle size and increase rate of fermentation (Hofmann, 1989). The forestomachs follow the oesophagus and precede the abomasum (the equivalent of the stomach of non-ruminants) (Hofmann, 1989). The evolutionary origin of the rumen is the subject of debate with out-pouching of the oesophagus, or of the stomach, as the most likely origins (Beck et al., 2009; Langer, 1988). The primary chambers of the rumen facilitate the action of a complex mixture of micro-organisms to ferment a portion of the plant polysaccharides (including starch, xylan and cellulose) and lipids to short chain volatile fatty acids (SCFAs), principally acetate, butyrate and propionate (Bergman, 1990). The SCFAs are the primary energy source in carbon of ruminants, and the rumen is the major site of their uptake.

From the rumen, partially processed plant material, nutrients, and micro-organisms pass through the omasum and enter the conventional gastrointestinal system: the abomasum, and the small and large intestines for further digestion and fermentation (in the large intestine). The abomasum is primarily a digestive organ lowering the pH of the rumen fluid and facilitating the first step of proteolysis prior to more extensive degradation in the duodenum and absorption of amino acids and small peptides. Pancreatic RNAses degrade microbial RNA in the small intestine contributing to nitrogen availability. On pasture, roughage or grass diets only small amounts of starch escape fermentation in the rumen and the remaining starch is generally digested in the small intestine, providing limited amounts of glucose (Deckardt et al., 2013).

Depending on the dietary source larger amounts of starch may escape fermentation in the rumen (Huntington, 1997). As a consequence glucose is not a major source of carbon in ruminants, and the liver is not a major site of (fatty acids) FA synthesis (Ingle et al., 1972). Biohydrogenation processes in the rumen (Van Nevel & Demeyer, 1996) increase the saturation of fatty acids (Jenkins et al., 2008; Van Nevel & Demeyer, 1996), and lipids that escape fermentation in the rumen are taken up in the small intestine. Fermentation of the remaining carbohydrates, lipid etc. occurs in the large intestine/hindgut. The hindgut is responsible for 5-10% of the total digestion of carbohydrates (Gressley et al., 2011) and for 8 to 17% of total production of SCFAs (Hoover, 1978). This contribution of hindgut fermentation may be altered on high grain diets (Fox et al., 2007; Mbanzamihigo et al., 1996). The overlap in functions of the rumen and the hindgut raises the question of whether the equivalent processes in the two tissues are undertaken by the same proteins and pathways; that is co-option of the hindgut program by the rumen, or by different proteins and pathways resulting from convergent evolution.

Unlike the stomach and subsequent segments of the GIT the rumen surface is a stratified squamous epithelium that is cornified and keratinized to protect the rumen from physical damage from the ingested plant material (Scocco et al., 2013). Due to the large numbers of microorganisms in the rumen it is also exposed to colonization of surfaces and potential attack from these organisms. The nature of the defences and the interaction between the surface of the rumen and the microbial populations has not been investigated in detail.

Herein, we utilised the latest sheep genome and transcriptome data (Jiang et al., 2014) to further dissect gene expression features of the ruminant GIT. We analyze the transcriptomes of six GIT tissue/cell types covering the majority of the sheep GIT in the context of reference samples from two other tissues with stratified squamous epithelium (skin and tonsil), another

component of the immune system (spleen), and two non-epithelial tissues (liver and muscle). Further, we systematically compared our results with existing transcriptome data from the human and pig gastrointestinal tracts and with relevant literature using candidate gene/protein based approaches. Our major aims were to identify: i) the distinctive features of ruminant GIT, ii) the common features shared between ruminant and mammalian GIT and iii) the developmental origin of the rumen.

METHODS

Data acquisition and statistical analysis

No new primary datasets were generated in this work, the major secondary datasets are included in the supplementary material. The sample preparation procedures and sequencing of the RNA are described in (Jiang et al., 2014) and experimental animal information were specified in Additional file 1: Table S1. Briefly, tissue samples were obtained from a trio of Texel sheep, i.e., ram (r), ewe (e) and their lamb (l). RNA was prepared and sequenced using stranded Illumina RNA-Seq with a yield of 70-150 million reads per tissue sample. 26 files of RNA sequence alignment data in the BAM format for 11 tissue/cell types, including skin (n=3), tonsil (n=1r), ventral rumen (n=3), abomasum (n=3), duodenum (n=1r), caecum(n=2, r and l), colon(n=3), rectum (n=3), spleen (n=2, r and l), liver (n=2, r and e) and muscle (n=3), was downloaded from the Ensembl sheep RNA sequencing archive, Oar_v3.1 (Huttenhower et al., 2009; Jiang et al., 2014). Detailed animal and gender distribution can be found in Supplementary Figure S1. Detailed raw RNA sequencing data from the same samples was also retrieved from the European Nucleotide Archive (ENA), study accession PRJEB6169. The raw mapping counts for each gene were calculated from the downloaded BAM files and the Ensembl sheep gene models (Ensembl), with additional gene models for genes at the EDC locus not included in the Ensembl sheep gene

models (Jiang et al., 2014), using HTSeq in the Python environment (Anders et al., 2015). The raw count data was normalized and clustered with DEseq2 (Love et al., 2013) to produce PCA plots and variance-stabilizing transformed gene expression values for network analysis described below. DEseq2 produced PCA sample clustering was further tested for significance using a k-means method and bivariate t-distributions based on the eigenvalues of the principle components. Calculation was performed using the stat_ellipse package (2012) and the raw outputs were presented in ggplot2 in R. EdgeR (Robinson et al., 2010) in Bioconductor in R v3.1.3 was used to analysis gene differential expression. After filtering for transcripts with at least 1 count per million in at least one of the 11 tissues, data was analysed using the Analysis of Variance-like procedure (special feature in EdgeR) and fitted to a simple model: $y = tissue_i + animal_j + e_{ij}$. Where y is raw transcript counts, $tissue_i$ (i=11) is 11 types of tissues and $animal_j$ (j=3) is the adjustment of types of animal (lamb, ram and ewe). Transcripts with significance levels (P) < 0.01 and false discovery rate (FDR) < 0.01 for tissue effects and differentially expressed in at least one of the 11 tissues were identified.

Co-expression network analysis

Variance-stabilizing transformed RNA sequencing expression values have properties similar to normalized microarray expression values in terms of network analysis (Giorgi et al., 2013) and raw counts of differentially expressed (FDR<0.01) transcripts were variance-stabilizing transformed (Durbin et al., 2002) using DEseq2. Transformed expression values were analyzed for co-expression using PCIT (Hudson et al., 2012; Reverter & Chan, 2008) in R v3.1.3 (Watson-Haigh et al., 2010). To reduce the complexity of the network the PCIT output was filtered for pairs of genes with a correlation coefficient > 0.9 and visualized in Cytoscape v3.1.2 (Shannon et al., 2003). The network cluster algorithm 'community cluster' within the GLay

plugin (Su et al., 2010) of cytoscape was used to subdivide the large network and identify explanatory sub-networks in an iterative manner until no obvious sub-network was observed in the large network. Pig genes assigned to 10 clusters showing differential expression in the pig GIT (Freeman et al., 2012) were mapped to sheep genes based on their gene symbols. The probability of over or under representation of pig GIT genes in a sheep GIT gene cluster was calculated using the hypergeometric distribution (Andrews et al., 1999). Functional enrichment of shared sets of genes within sheep clusters was analyzed using GOrilla (Eden et al., 2009) to identify biological pathways.

Gene expression pattern clustering

The transcripts present in the gene networks described above, and with an ANOVA $P < 0.01$ and a FDR <0.01, were included in k-mean clustering in R v3.1.3 based on $\log_2$ fold change across 11 tissues with abomasum being the reference. The k-mean analysis aimed to identify expression patterns to represent transcript groups showing elevated expression levels for the following sets of tissues v. the remaining tissues: 1) all GIT tissues, i.e., rumen, abomasum, duodenum, caecum, colon and rectum, 2) rumen and abomasum, 3) rumen and intestinal tissues, 4) abomasum and intestinal tissues, 5) rumen, 6) abomasum, 7) intestinal tissues, 8) rumen and skin, 9) rumen and tonsil, 10) rumen, skin and tonsil, 11) spleen, duodenum, caecum, and colon. The transcript names are determined based on the tissue(s) where included transcripts showed the highest expression. We filtered these identified transcript clusters with the criteria that 1) the average absolute expression of the transcript at the highest expressed tissue > 3 count per million, 2) the $\log_2$ expression fold difference of expression of the transcript from the tissue within the reference tissue group with the highest expression to the tissue within the elevated expressed tissue group with the highest expression, be > 0.5, and 3) from the tissue with the

highest expression to the tissue with the lowest expression within the elevated tissue group be < 0.5. The final expression of each transcript is presented in the format of log2 Fragments Per Kilobase of exon per Million fragments mapped (FPKM). Selected gene members and associated pathways were presented in heat maps based on their log2 FKPM values using GENE-E (Gould).

To understand the GIT associated SLC family genes, we performed a network analysis of expression as above. The PCIT output of network matrix was filtered for correlation coefficient > 0.7, clustered by GLay (Su et al., 2010) and visualized in Cytoscape v3.1.2 (Shannon et al., 2003).

Comprehensive transcript annotation

To complement the sheep genome annotation, we used multiple annotation sources and software to identify the function of the products encoded by the identified transcripts. Firstly, the transcripts of interest, both with or without a gene symbol, were validated for existence on the sheep genome, using comparisons of the sheep gene along within the locus with its ortholog(s) of human and bovine from Ensembl and NCBI. Secondly, GO was used to annotate genes. Thirdly, the functions and annotations of the genes were searched in Ensembl and NCBI, if no available description or gene information were identified, the biomedical literature was searched with GenCLiP 2.0 (Wang et al., 2014). When multiple biomedical functions were listed, functions related to gastrointestinal activity were prioritized for annotation. Fourthly, for a subset of genes Unigene (McGrath et al., 2010) and Genevestigator (Hruz et al., 2008) were used to identify transcript expression patterns in cattle and humans respectively. Protein sequences analysis was performed using Radar (Heger & Holm, 2000), to identify amino acid sequence repeats, and NetOGlyc 4.0 (Steentoft et al., 2013), to identify glycosylation sites.

Data access

No new primary datasets were generated in this work, the major secondary datasets are included in the supplementary material.

RESULTS AND DISCUSSION

Clustering of sheep GIT tissue transcriptomes

We performed principal component analysis (PCA) using RNA-Seq data from six GIT (rumen, abomasum, duodenum, caecum, colon and rectum), two epithelial (skin and tonsil), an immune (spleen) and two reference (liver and muscle) tissue/cell types from a trio of Texel sheep (ram, ewe and lamb (Jiang et al., 2014)). We included a total of 26 tissue samples, a similar tissue sample coverage to a previous transcriptomic study of the pig GIT (Freeman et al., 2012) to which the results of this analysis will be compared below. Three clusters of tissues were identified at the 95% confidence interval: cluster 1, skin, tonsil and rumen, cluster 2, muscle, and cluster 3, liver, spleen and the remaining GIT tissues (Figure 1A, Additional file: Figure S1A).

Identification of common and specific GIT and epithelial transcriptomic signatures

To identify the genes driving the clustering of the tissues we identified those transcripts with an ANOVA $P < 0.01$ and a false discovery rate (FDR) < 0.01, for differential expression in at least one tissue versus the other tissue types. This multi-tissue comparison reduced the impact of the small sample size for some tissues, in particular the duodenum (one tissue sample). Secondly, for a conservative gene network cluster analysis, the pair-wise gene correlation coefficient cut-off was set to 0.9 and we further filtered transcripts based on relative (fold change) and absolute (count per million) expression levels. We identified 16 major gene expression patterns,

representing common and specific transcriptomic signatures of the epithelial and GI tissues, accounting for 639 different transcripts (Figure 2A). A full list of the expression of the genes across the tissues with assignment to clusters is available (Additional file 1: Table S2, S3). Gene Ontology enrichment analysis of the clusters identified a number of significantly enriched terms (Table 1). A full list of the genes contributing to the enrichments is available (Additional file 1: Table S4). Most notable was the highly significant enrichment of the genes in the epithelia-intestine cluster for the GO-term, "cell cycle process". The higher expression of the majority of these genes in the epithelial and GIT tissues (Figure 2, Supplementary Table S2) is consistent with the much higher turnover rate of these tissues compared to liver and muscle (Milo et al., 2010) and may contribute to the structural adaptability of the rumen epithelia to different diets and health conditions (Dionissopoulos et al., 2012; Penner et al., 2011). Epithelia structure related pathways including 'cell junctions' showed significant enrichment in genes highly expressed in the rumen and the large intestine (Table 1). Gene members involved in cell junction functions have been reported to be important for the rumen epithelia to maintain pH homeostasis (Dionissopoulos et al., 2012; Steele et al., 2011a). Two other very significant enrichments were observed, "flavonoid biosynthetic process" in the rumen-intestine-liver cluster and "regulation of chloride transport" in the large intestine cluster (Table 1). The mammalian Epidermal Development Complex (EDC) locus is a cluster of up to 70 adjacent genes encoding proteins with roles in the development and the structure of stratified epithelia (Kypriotou et al., 2012). Although no significant enrichment of genes in the rumen cluster was identified by GO analysis several genes in the EDC region were very significantly overrepresented in the cluster (Table 1). This is consistent with our previous identification of several ruminant specific genes at the EDC locus highly preferentially expressed in the rumen (Jiang et al., 2014). The genes in the epithelia-

rumen-tonsil cluster were also very significantly enriched for EDC genes (Table 1). Thus the clustering of the rumen with the skin and tonsil appears to have been driven by genes involved in the development and structure of the stratified epithelium.

The stratified squamous rumen epithelium expression signature

The EDC locus genes are not the only genes encoding proteins involved in the synthesis of the cornified surface of the rumen and we looked for additional genes involved in cornification preferentially expressed in the rumen compared to skin and tonsil. The cross linking of the proteins of the cornified surface is mediated by transglutaminases (TGMs) (Eckert et al., 2005). Multiple TGMs are expressed in the rumen in this study, TGM1 and TGM3 appear to be the major rumen transglutaminases, but are also highly expressed in the skin (Figure 3). Keratins are major components of the cornified layers so we asked the question, are there keratin genes highly preferentially expressed in the rumen? Although no *KRT* genes showed expression as exclusive to the rumen as some of the EDC locus genes in our data, *KRT36* was grouped in the rumen expression cluster (Figure 3, Additional file 1: Table S3, Additional file 2: Figure S2), with significantly elevated expression in rumen, compared to the other studied tissues, and limited expression in skin. *KRT36* was previously identified as a novel keratin gene only expressed in sheep hair cortex (Yu et al., 2011) and its rumen expression showed significant responses to dietary changes in cattle (Li et al., 2015). However, in humans the highest expression of *KRT36* was in the tongue (Genevestigator (Hruz et al., 2008) analysis). Overall the transglutaminases and keratins do not appear to be as preferentially expressed in the rumen as some of the EDC locus genes.

Kallikrein-related peptidases are involved in the turnover of the cornified layers of the stratified epithelia, and deficiencies can lead to altered turnover of the surface layers of the epithelia (Hovnanian, 2013). In our study, *KLK12* is the only *KLK* family member preferentially expressed in the rumen (Figure 3, Additional file 1: Table S2). Members of the SPINK (serine peptidase inhibitor, Kazal type) family are inhibitors of the KLK family peptidases (Hovnanian, 2013), *SPINK5* is the only member of the family that is highly expressed in the rumen (Figure 3, Additional file 1: Table S2) in our data, but is also highly expressed in the tonsil and skin. KLK12 and SPINK5 may be involved in the regulation of the turnover and thickness of the cornified surface of the rumen epithelium, but may not form a rumen specific system.

Rumen micro-organism interactions

The rumen is the site of frequent interaction between the host and very dense populations of micro-organisms. In our study, *DUOX2* and *DUOXA2* encoding subunits of dual oxidase were preferentially expressed in the rumen (Figure 3), while *DUOX1* showed rumen-biased expression (Figure 3) and *DUOXA1* was highly expressed in all epithelia tissues (Figure 3). This observation is in line with the findings in the pig where the highest expression of *DUOXA1* and *DUOX1* was in the epithelial tissues. e.g., tongue and lower oesophagus (Freeman et al., 2012). In humans, the *DUOXA1* and *DUOX1* genes are also most highly expressed in epithelia tissues exposed to air, whilst *DUOX2* and *DUOXA2* are most highly expressed in a different set of tissues including the GIT (Genevestigator (Hruz et al., 2008) analysis). Thus, our findings suggest that the DUOX1s are active in general epithelial tissues, while DUOX2s are probably active specifically in rumen to play a major role in controlling microbial colonization. Previously in sheep, the highest expression levels of *DUOX1* and *DUOX2* were reported in the bladder and

abomasum, respectively, but the rumen and epithelial tissues were not included in the tissues surveyed (Lees et al., 2012).

PIP, encoding prolactin-induced protein (an aspartyl protease), was preferentially expressed in the rumen (Figure 3). In humans, *PIP* is also highly expressed in epidermal (Genevestigator (Hruz et al., 2008) analysis) and exocrine tissues, and in pigs in the salivary gland. Although PIP has been reported to be involved in regulation of the cell cycle in human breast epithelial cells (Cassoni et al., 1995; Naderi & Vanneste, 2014), its expression pattern in sheep (not part of the cell cycle cluster) is more consistent with a role in mucosal immunity (Hassan et al., 2009). Also highly expressed in the rumen were members of the *SERPINB* family of peptidase inhibitors (Figure 3), which are involved in the protection of epithelial surfaces in humans (Wang et al., 2012) and mice (Sivaprasad et al., 2011). EDC locus genes *PGLYRP3* and *PGLYRP4* encode peptidoglycan recognition proteins in the N-acetylmuramoyl-L-alanine amidase 2 family, which bind to the murein peptidoglycans of Gram-positive bacteria as part of the innate immune system. Additional EDC locus genes, *S100A8*, *S100A9* and *S100A12* (calgranulins A, B and C), encode key players in the innate immune function (Funk et al., 2015; Tong et al., 2014).

Rumen steroid metabolism

Amongst the genes preferentially expressed in the rumen (and often the liver) we identified a number of aldo/keto-reductases (Figure 4). *AKR1C1* can catalyze the conversion of progesterone to 20-alpha-hydroxy-progesterone (PGF2α) (Penning, 1997), retinals to retinols and bioactivates and detoxifies a range of molecules (El-Kabbani et al., 2011). Intravenous injection of PGF2α in goats has been shown to increase contraction of rumen smooth muscle, which leads to a reduction in the contraction rate of the rumen (van Miert & van Duin, 1991; Veenendaal et al., 1980). *AKR1C1* has also been reported to be preferentially expressed in the rumen of cattle (Kato

et al., 2015). The exact role of AKR1C1 in the rumen is unknown,. In addition, the gene encoding the related enzyme AKR1D1 (catalyzes the reduction of progesterone, androstenedione, 17-alpha-hydroxyprogesterone and testosterone to 5-beta-reduced metabolites) is highly expressed in the rumen and the liver and the gene encoding ARK1C4 in the rumen, liver and duodenum (Figure 4). The products of these genes are also likely to be involved in the metabolism of steroids in the rumen epithelium. In addition, we observed marked pathway enrichment of flavonoid biosynthetic process due to the identification of five members of the UDP-glucuronosyltransferase (UGT) gene family [29], with the highest expression levels in the rumen and liver (Additional file 1: Table S2). Flavonoids are only produced by plants, but UGT enzymes are highly active in mammals and catalyze the glucuronidation of a diverse chemical base including steroids, bile acids and opioids [29]. The functions of the products of these genes in the rumen require further investigation. However, results discussed here suggest important interactions between the rumen wall and activity of steroids.

Comparison of the sheep and pig GIT transcriptomes

To compare the ruminant and a closely related non-ruminant mammal GIT transcriptomes (Jiang et al., 2014), we mapped those transcripts previously reported to show specific expression patterns in the pig GIT (Freeman et al., 2012) to the sheep gene network (Figure 2B). Pig is the genomically closest non-ruminant to the ruminants (Groenen et al., 2012; Jiang et al., 2014) for which sufficient GIT transcriptome data is available. The overall overlap of the 639 genes in the sheep GIT network and the 2634 mappable pig GIT genes is 179, which is highly significant (Table 2). The smaller number of genes showing differential expression in our study versus the pig study is due to the application of stringent statistical filtering thresholds to minimize the

impact of the small number of samples per tissue. However, the overlap of 627 genes between the set of 2475 sheep genes identified using relaxed filtering criteria and the 2634 pig genes was also highly significant (P <10E-20), supporting the robustness of the approach. The set of 179 overlapping genes was highly significantly enriched for the GO-term "cell cycle process" (Table 2). The overlap of genes between the pig intestine clusters and the sheep epithelia-intestine cluster was highly significant and the overlap genes were again very highly significantly enriched for the GO-term "cell cycle process" (Table 2). A full list of the genes in the overlap and assignment to the pig and sheep gene clusters is available (Additional file 1: Table S5). Furthermore, pig genes preferentially expressed in the tongue and oesophagus have a highly significant overlap with sheep genes with high expression in the rumen and epithelial tissues (Figure 2B), enriched for the GO-term "epidermis development" (Table 2). Our results emphasises the contribution of cell cycle to the renewal of mammalian GIT epithelial surfaces (Crosnier et al., 2006).

Ruminant specific pathways for SCFA uptake and GIT metabolism?

SCFAs are the major source of energy in ruminants, with the primary sources of SCFAs being the rumen, and to a much lesser extent the large intestine. Carbonic anhydrases, which hydrate CO_2 to bicarbonate, are thought to play a significant role in the uptake of SCFAs by an SCFA/bicarbonate antiporter, and by providing protons at the rumen epithelium to neutralize the SCFAs and promote their diffusion into the ruminal epithelium (Bergman, 1990; Wang et al., 1996). There are many members of the carbonic anhydrase gene family (Tashian, 1989), several of which are expressed in mammalian gastrointestinal tissues (Freeman et al., 2012; Kivel et al., 2005; Parkkila et al., 1994; Tashian, 1989). In ruminants, *CA1* has previously been reported to encode a rumen specific carbonic anhydrase with low activities in the blood (unlike in other

353 mammals) and in the large intestines (Carter, 1971). Consistent with this, compared to all of the
354 other tissues in our dataset, *CA1* is highly expressed in the rumen and, albeit with lower but
355 significant expression, in the large intestine (Figure 4). *CA2* and CA*7* appear to encode the major
356 carbonic anhydrases in the large intestines (Figure 4). In humans *CA1*, *CA2* and *CA7* are highly
357 expressed in the colon (Genevestigator (Hruz et al., 2008) analysis). In contrast in pigs, whilst
358 *CA2* is highly expressed in the stomach, it is not highly expressed in the large intestine and *CA1*
359 and *CA7* were not reported to be differentially expressed across the GIT (Freeman et al., 2012).

360 The apical membrane SCFA/bicarbonate antiporter exchanges intracellular bicarbonate with
361 intra-ruminal SCFA and consistent with previous publications, *SLC4A9*, preferentially expressed
362 in the rumen in our dataset (Figure 4), encodes the most likely antiporter. The proposed
363 basolateral membrane SCFA/bicarbonate antiporter gene *SLC16A1* (exchanges intracellular
364 SCFA with blood bicarbonate), which has highest expression in the rumen in our dataset,
365 followed by the colon and rectum, has a much more general expression across the tissues than
366 *SLC4A9* (Figure 4). These expression patterns are consistent with previous findings in cattle
367 (Connor et al., 2010). SLC16A1 is also likely to be involved in the transport of ketone bodies
368 into the blood supply to the basolateral surface of the rumen epithelium (van Hasselt et al.,
369 2014).

370 HCO_3^--independent apical uptake of acetate in the rumen has also been observed (Aschenbach et
371 al., 2009). However, the transporter has not been identified, with candidates proposed in the
372 SLC4A, SLC16A, SLC21A, SLC22A and SLC26A families (Aschenbach et al., 2009). Members
373 of the SLC21A and SLC22A families showed generally low expression in the rumen in our study
374 (Additional file 1: Table S2). In addition to *SLC16A1* and *SLC4A9* discussed above, *SLC26A2*
375 and *SLC26A3* are highly expressed in the rumen in our dataset (Figure 4). Both genes encode

apical anion exchangers confirming them as candidates for encoding the apical HCO_3^--independent acetate uptake transporter. SLC26A3 is a Cl^-/ HCO_3^- exchanger (see fluid and electrolyte balance section below) and therefore is unlikely to be an HCO_3^--independent acetate transporter. However, SLC26A2 is a SO_4^{2-}/OH^-/Cl^- exchanger (Ohana et al., 2012) and remains a candidate for the proposed apical HCO_3^--independent acetate transporter. An HCO_3^--independent basolateral maxi-anion channel for $SCFA^-$ efflux to blood has also been proposed without an assigned transporter (Georgi et al., 2014). A survey of ABC (ATP-binding cassette) family transporters identified *ABCC3* as the most preferentially expressed in the rumen in our dataset and with the second highest expression in the large intestine (Figure 4). ABBC3 is an organic anion transporter with a possible role in biliary transport and intestinal excretion (Rost et al., 2002). Therefore, *ABCC3* may be involved in the efflux transport of $SCFA^-$ from the rumen epithelium to blood.

In most mammals, including humans, the liver is the major site of the synthesis of ketone bodies (acetoacetate and beta-hydroxybutyrate), but in ruminants the epithelium of the rumen is a major site of *de novo* ketogenesis (Lane et al., 2002). *HMGCS2* encodes an HMG-CoA synthase (3-hydroxy-3-methylglutaryl-CoA Synthase 2) in the ketogenesis pathway (Figure 5). This gene is significantly associated with bovine butyrate metabolism (Baldwin et al., 2012) and the encoded enzyme was predicted to be the rate limiting enzyme in sheep ruminal ketone body synthesis (Lane et al., 2002). As expected, in our data *HMGCS2* is highly expressed in the rumen compared to the other GIT tissues and the liver (Figure 4). *ACADS*, *HMGCL* and *BHD1*, which encode other enzymes involved in the ketone body pathway (Figure 5), are also highly expressed in the rumen relative to most of the other tissues studied (Figure 4). HMGCS1 and ACAT2 may also contribute to the ketone body pathway in the rumen, but their highest expression levels are

399 in the liver (Table S1). However, their expression in the rumen has been reported to actively
400 respond to different diets (Steele et al., 2011b) and acidosis conditions (Steele et al., 2012) in
401 cattle. Whilst *HMGCS2* is quite highly expressed in the colon, in contrast *ACADS*, *HMGCL* and
402 *BHD1* are not highly expressed (Figure 4), consistent with the colon not being a major
403 contributor to ketone body synthesis. Genes encoding enzymes for other steps in the pathways
404 from acetate and butyrate to ketone bodies are much more generally expressed across the tissues,
405 although expression of *ECHS1* and *ACAT1* are significantly higher in the rumen than in other
406 GIT tissues (Figure 4). In humans, in addition to the liver, *HMGCS2* also has high expression in
407 the intestine, including the jejunum and colon (Genevestigator (Hruz et al., 2008) analysis). In
408 contrast, the only enzyme in the pathway (Figure 4, 5) reported to be preferentially expressed in
409 the pig GIT was *BDH1*, in the fundus of the stomach (Additional file 1: Table S5). Thus the
410 rumen, abomasum, duodenum, caecum, colon and rectum in sheep all appear to have subtly
411 different SCFA transport and metabolism systems, and in the equivalent compartments of the
412 GIT appear to be different between sheep, humans and pigs.

413 **Long chain fatty acids (LCFAs) uptake, cholesterol homeostasis and bile acid recycling**
414 Due to the activity of the microbial populations of the rumen and the production of SCFAs
415 ruminants have less reliance on dietary LCFAs than non-ruminants. Does this reduced
416 importance lead to detectable differences in the transcriptome? The small intestine is the
417 principal site of uptake of LCFA and cholesterol homeostasis, and consistent with this the genes
418 encoding the well characterized components of the intestinal fatty acid uptake (*CD36*,
419 *SLC27A2/4/5* and *FABP2* (Wang et al., 2013)) and cholesterol homeostasis (*NPC1L1* and
420 *ABCG5/8* (Wang et al., 2013)) systems are expressed in the sheep small intestine (Figure 4), as
421 they are in humans and most are in the pig (Freeman et al., 2012). *FABP2* and *ABCG5* are

422 particularly preferentially expressed in the sheep small intestine relative to other GIT tissues
423 (Figure 4). However, it is thought that the major route of LCFA uptake at the apical membrane
424 of the GIT epithelium is by passive diffusion (Abumrad & Davidson, 2012).
425 Bile acids secreted by the liver and stored in the gall bladder before being released into the small
426 intestine play a major role in the uptake of LCFAs. Bile acids are recycled in the intestine.
427 SLC10A2 in the apical membrane and SLC51A and SLC51B in the basolateral membrane are
428 proposed to constitute the uptake systems in the human small intestine (Ballatori et al., 2013).
429 *SLC10A2* is also preferentially expressed in the small intestines of the pig, but preferential
430 expression of *SLC51A/B* has not been reported (Freeman et al., 2012). In sheep *SLC10A2* is
431 preferentially expressed in the small intestine, albeit it a low level (Figure 4). Whilst *SLC51B* is
432 highly expressed in the duodenum in sheep, the highest expression of the two subunits together
433 in sheep (*SLC51A/B*) is in the caecum and the colon (Figure 4), where they are also expressed in
434 humans and mice (Genevestigator (Hruz et al., 2008) analysis). Although described as subunits
435 of a complex, *SLC51A* and *SLC51B* have also been reported to be regulated differently (Ballatori
436 et al., 2013), thus the balance between expression of *SLC10A2* and *SLC51A* and *SLC51B* may
437 indicate differences in the bile acid uptake pathways in the duodenum, large intestines and liver
438 of sheep.

439 Overall despite the reduced importance of LCFAs sheep appear to have a very similar systems to
440 human and pigs for LCFA uptake and bile acid recycling.

441 **Saccharide metabolism**

442 Again as a consequence of the activity of the rumen microbes in mature ruminants the uptake of
443 dietary glucose may be less than 10% of glucose requirements (Young, 1977). The dietary

444 glucose comes primarily from the degradation of polysaccharides, in particular in the small
445 intestine of starch that has escaped degradation by the rumen microbial population. The primary
446 source of alpha-amylase required to digest the long polymers is the pancreas, which was not
447 investigated in this study. Genes encoding three enzymes likely to contribute to the digestion of
448 starch and other alpha-glycans, *MGAM* (maltase-glucoamylase), *MGAM2* (maltase-glucoamylase
449 2) and *SI* (sucrase-isomaltase) (Nichols et al., 2003), were preferentially expressed in the tissues
450 studied here. *SI* was preferentially expressed in the intestine-low in rectum gene cluster, *MGAM2*
451 was highly expressed in all intestinal tissues, while *MGAM* was also preferentially expressed in
452 the intestine (primarily the duodenum), but at a much lower level (Figure 4). In humans
453 (Genevestigator (Hruz et al., 2008) analysis) and pigs (Freeman et al., 2012), both *MGAM* and *SI*
454 are preferentially expressed in the small intestine. Expression of the orthologues of *MGAM2* has
455 not been reported in the GIT of humans (Genevestigator (Hruz et al., 2008) analysis) or
456 pigs(Freeman et al., 2012) .

457 The mammalian *MGAM* and *MGAM2* genes appear to have arisen by tandem duplication of a
458 single ancestral gene at the base of the mammals (Nichols et al., 2003; Nichols et al., 1998).
459 *MGAM2* genes are present in most mammals, but have been annotated as possible pseudogenes
460 in a number of species, including man (NCBI LOC93432). *MGAM2* is not well characterized in
461 any species. Comparative analysis of the protein sequences of MGAM and MGAM2 showed that
462 MGAM2 has additional sequence at the carboxy-terminus comprised of multiple copies of a 40
463 amino acid repeat not present in MGAM (Figure 6). The repeat unit is enriched in serine and
464 threonine, with similar sequences in the predicted sheep, cattle, pig and to a much lesser extent
465 human proteins (Figure 6). The repeat unit of MGAM2 is predicted to be heavily glycosylated
466 (Steentoft et al., 2013) to form a mucin-like domain. As in the rumen the microbial population in

the large intestine ferments plant material, contributing up to 10% of the total carbohydrate fermentation and conversion to SCFAs in the ruminant GIT (Gressley et al., 2011). Whilst the role MGAM2 is unclear it appears to represent a contribution from the host to the breakdown of plant polysaccharides by the bacterial population in the large intestine. MGAM produces glucose from maltose and MGAM2 may have a similar functionality, and therefore contribute to the uptake of the scarce supply of glucose in ruminants. Alternatively the high expression of MGAM2 and low expression of MGAM may reflect the reduced availability of glucose in the rumen GIT. Further investigation of this gene and the activity and function of its encoded protein will improve our understanding of carbohydrate metabolism in the large intestine of ruminants.

In humans the major uptake of glucose in the GIT occurs in the small intestine via SLC5A1 (aka SGLT1) in the apical membrane, and SLC2A2 (aka GLUT2) in the basolateral membrane (Roder et al., 2014). The expression pattern of these two genes in sheep (Figure 4) and pigs (Freeman et al., 2012) is consistent with a similar process in all three species.

Nitrogen acquisition and recycling

A high level of nitrogen recycling in the GIT is a characteristic of ruminants. Urea is the major input from the animal (primarily via the saliva and the rumen epithelium) and anabolic-N sources (in the small intestine) and ammonia (in the rumen, small and large intestines) are the major uptake molecules from the GIT (Lapierre & Lobley, 2001). *SLC14A1* (Figure 4), encoding SLC14A1 which mediates the basolateral cell membrane transport of urea, a key process in nitrogen secretion into the GIT (Abdoun et al., 2010), is highly preferentially expressed in the rumen in our dataset (Figure 4). However, in cattle expression of *SLC14A1* was not affected by differences in dietary N (Rojen et al., 2011) and doubts remain about the role of SLC14A1 in increasing rumen epithelial urea permeability at low dietary N. Urea is also thought to be

490 released by the epithelium of the small and large intestines (Lapierre & Lobley, 2001), but our
491 analysis did not identify a potential transporter.
492 Urea is converted to ammonia by microbial ureases and is used by rumen microorganisms to
493 synthesize microbial proteins (75-85% of microbial N) and nucleic acids (15-25% of microbial
494 N) (Fujihara & Shem, 2011) which are subsequently digested by the host in the intestines, thus
495 recovering the majority of the secreted nitrogen (Abdoun et al., 2006). Consistent with this,
496 *SLC3A1* (neutral and basic amino acid transporter) in our study is preferentially expressed in the
497 duodenum (Figure 4), as is *SLC28A2* (concentrative nucleoside transporter) the product of which
498 plays in an important role in intestinal nucleoside salvage and energy metabolism (Huber-Ruano
499 et al., 2010). Both genes were also highly expressed in the small intestine of pigs (Freeman et al.,
500 2012) and humans (Genevestigator (Hruz et al., 2008) analysis). *RHBG* (*SLC42A2*), an ammonia
501 transporter, is preferentially expressed in the sheep small and large intestines and the liver
502 (Figure 4) and is a candidate for an intestinal ammonia transporter. However, *RHBG* is not
503 expressed at particularly high levels in the human GIT (Genevestigator (Hruz et al., 2008)
504 analysis) relative to many other tissues, and was not reported to be preferentially expressed in the
505 pig GIT (Freeman et al., 2012). In humans uptake of ammonia in the large intestine is thought to
506 most likely occur (mainly) by passive non-ionic diffusion (Wrong & Vince, 1984). However,
507 *RHCG* (apical membrane) and *RHBG* (basolateral membrane) have also been proposed to
508 constitute an ammonium uptake pathway in the human GIT (Handlogten et al., 2005). The
509 expression profile of *RHCG* in sheep (Figure 4) is not consistent with such a pathway in sheep.

510 In addition to the secretion of urea into the rumen (a ruminant specific process) the increased
511 importance of nitrogen recycling in ruminants may have led to the apparent increased expression
512 of *RHBG* in the GIT of sheep.

513 **Iodine recycling**

514 *SLC5A5*, member 5 of solute carrier family 5, encoding a sodium iodide symporter is highly
515 preferentially expressed in the abomasum in our study (Figure 4). *SLC5A5* also has higher
516 expression in human (Genevestigator (Hruz et al. 2008) analysis) and rat stomach (Kotani et al.,
517 1998) than in other digestive tissues. The latter authors reported that the distribution of *SLC5A5*
518 transcripts in the stomach epithelium was consistent with a role of *SLC5A5* in the import or
519 export of iodine, from or to the stomach contents. In the rat, iodine is actively transported into
520 the gastric lumen and this transport is at least partly mediated by a sodium-iodide symporter
521 (Josefsson et al., 2006). In cattle the rate of iodine export by the abomasum epithelium into the
522 abomasum is much greater than the import of iodine from the abomasum (Miller et al., 1975),
523 suggesting that the role of SLC5A5 in sheep abomasum is to export iodine into the stomach
524 contents. In contrast, *SLC5A5* was not reported to be significantly more expressed in the pig
525 stomach versus other components of the GIT (Freeman et al., 2012). The specific physiological
526 role of iodine in the stomach/GIT is unknown, but a number of possibilities have been suggested:
527 iodine-conserving mechanisms to deal with low iodine concentrations in the diet (Miller et al.,
528 1975), antioxidative activity (Venturi & Venturi, 1999) and antimicrobial activity (Spitzweg et
529 al., 1999). The majority of the secreted iodine is thought to be recovered in the lower intestines.
530 Another member from the same transporter family SLC5A6, a sodium/multivitamin and iodide
531 co-transporter (de Carvalho & Quick, 2011), encoded by a gene showing expression in all
532 studied tissues, with the highest expression sheep large intestine (Figure 4) is a likely candidate
533 for the iodine importer. In humans, *SLC5A6* is also expressed in a wide range of tissues with
534 intestinal tissues being close to the top of the list (de Carvalho & Quick, 2011). In pigs, *SLC5A6*
535 is preferentially expressed in the small intestine (Freeman et al., 2012). The high expression of

SLC5A5 in the abomasum suggests that ruminants may have retained a higher dependence on iodine in the GIT than other mammals.

Zinc homeostasis

SLC39A4 encodes a transporter protein essential for zinc uptake in the mouse intestine (Dufner-Beattie et al., 2003) and stomach (Martin et al., 2013). *SLC39A4* is highly expressed in stomach and intestines in sheep (Figure 4) and humans (Genevestigator (Hruz et al., 2008) analysis), and showed the highest expression in pig small intestine (Freeman et al., 2012). Another zinc transporter encoding gene, *SLC39A5*, has a similar expression profile to *SLC39A4* in sheep (Figure 4), humans and pigs. However, SLC39A5 is located in the basolateral membrane and is involved in the secretion of zinc. In mouse gastrointestinal tract cells the two zinc transporters are reciprocally regulated (Weaver et al., 2007), together controlling the influx and efflux of zinc at the intestinal epithelium. It appears likely that sheep have a similar mechanism for zinc homeostasis to other mammals.

Fluid and electrolyte balance

Maintaining salt and water balance is an important function of the mammalian GIT. In the large intestine significant GO term enrichment was identified for regulation of chloride transport, due to the inclusion of *CA2*, *7* and *CFTR* (Table 1, Figure 4). This is in agreement with the reported critical chloride secretory mechanism in intestinal epithelial cells, associated with mucosal hydration (Barrett & Keely, 2000). SLC26A3, which is a Cl^-/HCO_3^- antiporter, imports Cl^- ions driven by bicarbonate, thus linking the activity of carbonic anhydrases and the leakage of Cl^- out of the cells by CFTR. *SLC26A3* is preferentially expressed in the large intestine of sheep (Figure

4) and the colon of pigs. Thus the expression of genes involved in fluid and electrolyte balance is similar between all three species.

Conclusions

As a significant event in the evolution of the true ruminants, the evolutionary origin of the rumen is the subject of debate, with out-pouching of the oesophagus, or of the stomach, as the two most likely origins (Beck et al., 2009; Langer, 1988). The cornification of the epithelia surface, tissue clustering analysis based on gene expression (driven by the epidermal structural proteins and innate immunity genes) and the relative lack of metabolic overlap with the abomasum strongly favours an oesophageal origin. Metabolically the rumen has many similarities with the liver, especially for SCFA metabolism and even though there are functional similarities with the large intestine, the complements of genes involved are not highly similar.

We have identified a small number of highly rumen specific metabolic processes, in particular the roles of SLC14A1 (urea secretion), SLC4A9 (SCFA uptake) and AKR1C1 (uncertain function). Overall our analysis has enabled gene expression data to be married up with decades of physiological and other research to link transport and enzymatic activities and the most likely genes encoding products with the activities. Nitrogen and iodine recycling have been identified as processes with a much greater importance in the sheep than in humans or pigs. These metabolic functions are protected by strong immune functions and stratified epidermis-like epithelium. The major rumen immune players are DUOX and SERPINB gene families and *DUOXA2*, *DUOX2s* and *SERPINB3/4-like 1* appear to be preferentially expressed in the rumen. These findings will bring novel insights into biomedical research on mammalian digestive and gastrointestinal systems.

Additional files

Additional file 1: Table S1. Sheep experimental design information. **Table S2.** Gene expression values with *P* and FDR <0.01 across tissues. **Table S3**. Assignment of genes to tissue clusters. **Table S4.** Gene members in each significant GO Term enrichment. **Table S5.** Genes overlapped between sheep and pig GIT clusters.

Additional file 2: Figure S1. Raw output of PCA clustering of tissues. **Figure S2.** Expression profile of GIT-related keratin genes. **Figure S3**. Clustering of transporter genes.

Abbreviations

GIT: gastrointestinal tissue

Acknowledgements

We would like to thank Richard Talbot for supervision of the generation of the RNA-Seq data.

References

2012. stat_ellipse https://github.com/JoFrhwld/FAAV/blob/master/r/stat-ellipse.R) (accessed 03072015.

Abdoun K, Stumpff F, and Martens H. 2006. Ammonia and urea transport across the rumen epithelium: a review. *Animal Health Research Reviews* 7:43-59. 10.1017/s1466252307001156

Abdoun K, Stumpff F, Rabbani I, and Martens H. 2010. Modulation of urea transport across sheep rumen epithelium in vitro by SCFA and CO2. *Am J Physiol Gastrointest Liver Physiol* 298:G190-G202. 10.1152/ajpgi.00216.2009

602 Abumrad NA, and Davidson NO. 2012. Role of the gut in lipid homeostasis. *Physiological*
603 *Reviews* 92:1061-1085. 10.1152/physrev.00019.2011
604 Anders S, Pyl PT, and Huber W. 2015. HTSeq--a Python framework to work with high-
605 throughput sequencing data. *Bioinformatics* 31:166-169. 10.1093/bioinformatics/btu638
606 Andrews GE, Askey R, and Roy R. 1999. *Special functions - Encyclopedia of Mathematics*.
607 Cambridge: Cambridge University Press.
608 Aschenbach JR, Bilk S, Tadesse G, Stumpff F, and Gabel G. 2009. Bicarbonate-dependent and
609 bicarbonate-independent mechanisms contribute to nondiffusive uptake of acetate in the
610 ruminal epithelium of sheep. *American Journal of Physiology Gastrointestinal and Liver*
611 *Physiology* 296:G1098-1107. 10.1152/ajpgi.90442.2008
612 Baldwin RLt, Wu S, Li W, Li C, Bequette BJ, and Li RW. 2012. Quantification of
613 Transcriptome Responses of the Rumen Epithelium to Butyrate Infusion using RNA-seq
614 Technology. *Gene Regulation and Systems Biology* 6:67-80. 10.4137/grsb.s9687
615 Ballatori N, Christian WV, Wheeler SG, and Hammond CL. 2013. The heteromeric organic
616 solute transporter, OSTalpha-OSTbeta/SLC51: a transporter for steroid-derived molecules.
617 *Molecular Aspects of Medicine* 34:683-692. 10.1016/j.mam.2012.11.005
618 Barrett KE, and Keely SJ. 2000. Chloride secretion by the intestinal epithelium: molecular basis
619 and regulatory aspects. *Annual Review of Physiology* 62:535-572.
620 10.1146/annurev.physiol.62.1.535
621 Beck DC, Jiang H, and Zhang L. 2009. Elucidating the Evolutionary Relationships among Bos
622 taurus Digestive Organs Using Unigene Expression Data. *International Journal of*
623 *Evolutionary Biology*:Article ID 803142. 10.4061/2009/803142
624 Bergman EN. 1990. Energy contributions of volatile fatty acids from the gastrointestinal tract in
625 various species. *Physiological Reviews* 70:567-590.
626 Carter MJ. 1971. The carbonic anhydrase in the rumen epithelial tissue of the ox. *Biochimica et*
627 *Biophysica Acta* 235:222-236.
628 Cassoni P, Sapino A, Haagensen DE, Naldoni C, and Bussolati G. 1995. Mitogenic effect of the
629 15-kDa gross cystic disease fluid protein (GCDFP-15) on breast-cancer cell lines and on
630 immortal mammary cells. *International Journal of Cancer* 60:216-220.
631 Connor EE, Li RW, Baldwin RL, and Li C. 2010. Gene expression in the digestive tissues of
632 ruminants and their relationships with feeding and digestive processes. *animal* 4:993-1007.
633 10.1017/s1751731109991285
634 Crosnier C, Stamataki D, and Lewis J. 2006. Organizing cell renewal in the intestine: stem cells,
635 signals and combinatorial control. *Nature Reviews Genetics* 7:349-359. 10.1038/nrg1840
636 de Carvalho FD, and Quick M. 2011. Surprising substrate versatility in SLC5A6: Na^+-coupled I^-
637 transport by the human Na^+/multivitamin transporter (hsmvt). *Journal of Biological*
638 *Chemistry* 286:131-137. 10.1074/jbc.M110.167197
639 Deckardt K, Khol-Parisini A, and Zebeli Q. 2013. Peculiarities of enhancing resistant starch in
640 ruminants using chemical methods: opportunities and challenges. *Nutrients* 5:1970-1988.
641 10.3390/nu5061970
642 Dionissopoulos L, Steele M, AlZahal O, and McBride B. 2012. Adaptation to high grain diets
643 proceeds through minimal immune system stimulation and differences in extracellular
644 matrix protein expression in a model of subacute ruminal acidosis in nonlactating dairy
645 cows. *American Journal of Animal and Veterinary Sciences* 7:84-91.
646 Dufner-Beattie J, Wang F, Kuo Y-M, Gitschier J, Eide D, and Andrews GK. 2003. The
647 Acrodermatitis Enteropathica Gene ZIP4 Encodes a Tissue-specific, Zinc-regulated Zinc

Transporter in Mice. *Journal of Biological Chemistry* 278:33474-33481. 10.1074/jbc.M305000200

Durbin BP, Hardin JS, Hawkins DM, and Rocke DM. 2002. A variance-stabilizing transformation for gene-expression microarray data. *Bioinformatics* 18:S105-S110. 10.1093/bioinformatics/18.suppl_1.S105

Eckert RL, Sturniolo MT, Broome A-M, Ruse M, and Rorke EA. 2005. Transglutaminase Function in Epidermis. *J Investig Dermatol* 124:481-492.

Eden E, Navon R, Steinfeld I, Lipson D, and Yakhini Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10:48. 10.1186/1471-2105-10-48

El-Kabbani O, Dhagat U, and Hara A. 2011. Inhibitors of human 20alpha-hydroxysteroid dehydrogenase (AKR1C1). *Journal of Steroid Biochemistry and Molecular Biology* 125:105-111. 10.1016/j.jsbmb.2010.10.006

Ensembl. Sheep Genome v3.1 http://www.ensembl.org/Ovis_aries/Info/Index (accessed 03072015.

Fox JT, Depenbusch BE, Drouillard JS, and Nagaraja TG. 2007. Dry-rolled or steam-flaked grain-based diets and fecal shedding of Escherichia coli O157 in feedlot cattle. *Journal of Animal Science* 85:1207-1212. 10.2527/jas.2006-079

Freeman T, Ivens A, Baillie JK, Beraldi D, Barnett M, Dorward D, Downing A, Fairbairn L, Kapetanovic R, Raza S, Tomoiu A, Alberio R, Wu C, Su A, Summers K, Tuggle C, Archibald A, and Hume D. 2012. A gene expression atlas of the domestic pig. *BMC Biology* 10:90.

Fujihara T, and Shem MN. 2011. Metabolism of microbial nitrogen in ruminants with special reference to nucleic acids. *Animal Science Journal* 82:198-208. 10.1111/j.1740-0929.2010.00871.x

Funk S, Mark R, Bayo P, Flechtenmacher C, Grabe N, Angel P, Plinkert PK, and Hess J. 2015. High S100A8 and S100A12 protein expression is a favorable prognostic factor for survival of oropharyngeal squamous cell carcinoma. *International Journal of Cancer* 136:2037-2046. 10.1002/ijc.29262

Georgi MI, Rosendahl J, Ernst F, Gunzel D, Aschenbach JR, Martens H, and Stumpff F. 2014. Epithelia of the ovine and bovine forestomach express basolateral maxi-anion channels permeable to the anions of short-chain fatty acids. *Pflügers Archiv European Journal of Physiology* 466:1689-1712. 10.1007/s00424-013-1386-x

Giorgi FM, Del Fabbro C, and Licausi F. 2013. Comparative study of RNA-seq- and Microarray-derived coexpression networks in Arabidopsis thaliana. *Bioinformatics* 29:717-724. 10.1093/bioinformatics/btt053

Gould J. http://www.broadinstitute.org/cancer/software/GENE-E/2015).

Gressley TF, Hall MB, and Armentano LE. 2011. Ruminant Nutrition Symposium: Productivity, digestion, and health responses to hindgut acidosis in ruminants. *Journal of Animal Science* 89:1120-1130. 10.2527/jas.2010-3460

Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens H-J, Li S, Larkin DM, Kim H, Frantz LAF, Caccamo M, Ahn H, Aken BL, Anselmo A, Anthon C, Auvil L, Badaoui B, Beattie CW, Bendixen C, Berman D, Blecha F, Blomberg J, Bolund L, Bosse M, Botti S, Bujie Z, Bystrom M, Capitanu B, Carvalho-Silva D, Chardon P, Chen C, Cheng R, Choi S-H, Chow W, Clark RC, Clee C, Crooijmans RPMA, Dawson HD, Dehais P, De Sapio F, Dibbits B, Drou N, Du

694 Z-Q, Eversole K, Fadista J, Fairley S, Faraut T, Faulkner GJ, Fowler KE, Fredholm M, Fritz
695 E, Gilbert JGR, Giuffra E, Gorodkin J, Griffin DK, Harrow JL, Hayward A, Howe K, Hu Z-
696 L, Humphray SJ, Hunt T, Hornshoj H, Jeon J-T, Jern P, Jones M, Jurka J, Kanamori H,
697 Kapetanovic R, Kim J, Kim J-H, Kim K-W, Kim T-H, Larson G, Lee K, Lee K-T, Leggett
698 R, Lewin HA, Li Y, Liu W, Loveland JE, Lu Y, Lunney JK, Ma J, Madsen O, Mann K,
699 Matthews L, McLaren S, Morozumi T, Murtaugh MP, Narayan J, Truong Nguyen D, Ni P,
700 Oh S-J, Onteru S, Panitz F, Park E-W, Park H-S, Pascal G, Paudel Y, Perez-Enciso M,
701 Ramirez-Gonzalez R, Reecy JM, Rodriguez-Zas S, Rohrer GA, Rund L, Sang Y,
702 Schachtschneider K, Schraiber JG, Schwartz J, Scobie L, Scott C, Searle S, Servin B,
703 Southey BR, Sperber G, Stadler P, Sweedler JV, Tafer H, Thomsen B, Wali R, Wang J,
704 Wang J, White S, Xu X, Yerle M, Zhang G, Zhang J, Zhang J, Zhao S, Rogers J, Churcher
705 C, and Schook LB. 2012. Analyses of pig genomes provide insight into porcine demography
706 and evolution. *Nature* 491:393-398.
707 http://www.nature.com/nature/journal/v491/n7424/abs/nature11622.html#supplementary-
708 information
709 Handlogten ME, Hong SP, Zhang L, Vander AW, Steinbaum ML, Campbell-Thompson M, and
710 Weiner ID. 2005. Expression of the ammonia transporter proteins Rh B glycoprotein and Rh
711 C glycoprotein in the intestinal tract. *American Journal of Physiology Gastrointestinal and*
712 *Liver Physiology* 288:G1036-1047. 10.1152/ajpgi.00418.2004
713 Hassan MI, Waheed A, Yadav S, Singh TP, and Ahmad F. 2009. Prolactin inducible protein in
714 cancer, fertility and immunoregulation: structure, function and its clinical implications.
715 *Cellular and Molecular Life Sciences* 66:447-459. 10.1007/s00018-008-8463-x
716 Heger A, and Holm L. 2000. Rapid automatic detection and alignment of repeats in protein
717 sequences. *Proteins* 41:224-237.
718 Hofmann R. 1989. Evolutionary steps of ecophysiological adaptation and diversification of
719 ruminants: a comparative view of their digestive system. *Oecologia* 78:443-457.
720 Hoover WH. 1978. Digestion and absorption in the hindgut of ruminants. *Journal of Animal*
721 *Science* 46:1789-1799.
722 Hovnanian A. 2013. Netherton syndrome: skin inflammation and allergy by loss of protease
723 inhibition. *Cell & Tissue Research* 351:289-300. 10.1007/s00441-013-1558-1
724 Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W, and
725 Zimmermann P. 2008. Genevestigator v3: a reference expression database for the meta-
726 analysis of transcriptomes. *Adv Bioinformatics* 2008:420747. 10.1155/2008/420747
727 Huber-Ruano I, Pinilla-Macua I, Torres G, Casado FJ, and Pastor-Anglada M. 2010. Link
728 between high-affinity adenosine concentrative nucleoside transporter-2 (CNT2) and energy
729 metabolism in intestinal and liver parenchymal cells. *Journal of Cellular Physiology*
730 225:620-630. 10.1002/jcp.22254
731 Hudson N, Dalrymple B, and Reverter A. 2012. Beyond differential expression: the quest for
732 causal mutations and effector molecules. *BMC Genomics* 13:356.
733 Huttenhower C, Haley EM, Hibbs MA, Dumeaux V, Barrett DR, Coller HA, and Troyanskaya
734 OG. 2009. Exploring the human genome with functional maps. *Genome Research* 19:1093-
735 1106. 10.1101/gr.082214.108
736 Ingle DL, Bauman DE, and Garrigus US. 1972. Lipogenesis in the Ruminant: in vivo Site of
737 Fatty Acid Synthesis in Sheep. *Journal of Nutrition* 102:617-623.

738 Jenkins TC, Wallace RJ, Moate PJ, and Mosley EE. 2008. Recent advances in biohydrogenation
739 of unsaturated fatty acids within the rumen microbial ecosystem. *Journal of Animal Science*
740 86:397-412. 10.2527/jas.2007-0588
741 Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, Wu C, Muzny DM, Li Y, Zhang W,
742 Stanton J-A, Brauning R, Barris WC, Hourlier T, Aken BL, Searle SMJ, Adelson DL, Bian
743 C, Cam GR, Chen Y, Cheng S, DeSilva U, Dixen K, Dong Y, Fan G, Franklin IR, Fu S,
744 Fuentes-Utrilla P, Guan R, Highland MA, Holder ME, Huang G, Ingham AB, Jhangiani SN,
745 Kalra D, Kovar CL, Lee SL, Liu W, Liu X, Lu C, Lv T, Mathew T, McWilliam S, Menzies
746 M, Pan S, Robelin D, Servin B, Townley D, Wang W, Wei B, White SN, Yang X, Ye C,
747 Yue Y, Zeng P, Zhou Q, Hansen JB, Kristiansen K, Gibbs RA, Flicek P, Warkup CC, Jones
748 HE, Oddy VH, Nicholas FW, McEwan JC, Kijas JW, Wang J, Worley KC, Archibald AL,
749 Cockett N, Xu X, Wang W, and Dalrymple BP. 2014. The sheep genome illuminates
750 biology of the rumen and lipid metabolism. *Science* 344:1168-1173.
751 10.1126/science.1252806
752 Josefsson M, Evilevitch L, Westrom B, Grunditz T, and Ekblad E. 2006. Sodium-iodide
753 symporter mediates iodide secretion in rat gastric mucosa in vitro. *Experimental Biology*
754 *and Medicine (Maywood, NJ)* 231:277-281.
755 Kato D, Suzuki Y, Haga S, So K, Yamauchi E, Nakano M, Ishizaki H, Choi K, Katoh K, and
756 Roh S-G. 2015. Utilization of digital differential display to identify differentially expressed
757 genes related to rumen development. *Animal Science Journal*:n/a-n/a. 10.1111/asj.12448
758 Kivel AJ, Kivel J, Saarnio J, and Parkkila S. 2005. Carbonic anhydrases in normal
759 gastrointestinal tract and gastrointestinal tumours. *World Journal of Gastroenterology*
760 11:155-163.
761 Kotani T, Ogata Y, Yamamoto I, Aratake Y, Kawano JI, Suganuma T, and Ohtaki S. 1998.
762 Characterization of gastric Na+/I- symporter of the rat. *Clinical Immunology and*
763 *Immunopathology* 89:271-278.
764 Kypriotou M, Huber M, and Hohl D. 2012. The human epidermal differentiation complex:
765 cornified envelope precursors, S100 proteins and the 'fused genes' family. *Experimental*
766 *Dermatology* 21:643-649. 10.1111/j.1600-0625.2012.01472.x
767 Lane MA, Baldwin RLt, and Jesse BW. 2002. Developmental changes in ketogenic enzyme gene
768 expression during sheep rumen development. *Journal of Animal Science* 80:1538-1544.
769 Langer P. 1988. *The mammalian herbivore stomach: comparative anatomy, function and*
770 *evolution*: Gustav Fischer.
771 Lapierre H, and Lobley GE. 2001. Nitrogen Recycling in the Ruminant: A Review. *Journal of*
772 *Dairy Science* 84:E223-E236. 10.3168/jds.S0022-0302(01)70222-6
773 Lees MS, H. Nagaraj S, Piedrafita DM, Kotze AC, and Ingham AB. 2012. Molecular cloning and
774 characterisation of ovine dual oxidase 2. *Gene* 500:40-46.
775 Li Y, Carrillo JA, Ding Y, He Y, Zhao C, Zan L, and Song J. 2015. Ruminal Transcriptomic
776 Analysis of Grass-Fed and Grain-Fed Angus Beef Cattle. *PLoS ONE* 10:e0116437.
777 10.1371/journal.pone.0116437
778 Love M, Anders S, and Huber W. 2013. Differential analysis of count data–the DESeq2 package.
779 Martin AB, Aydemir TB, Guthrie GJ, Samuelson DA, Chang SM, and Cousins RJ. 2013. Gastric
780 and colonic zinc transporter ZIP11 (SLC39A11) in mice responds to dietary zinc and
781 exhibits nuclear localization. *Journal of Nutrition* 143:1882-1888. 10.3945/jn.113.184457

Mbanzamihigo L, van Nevel CJ, and Demeyer DI. 1996. Lasting effects of monensin on rumen and caecal fermentation in sheep fed a high grain diet. *Animal Feed Science and Technology* 62:215-228.

McGrath JA, Bolling MC, and Jonkman MF. 2010. Lethal Acantholytic Epidermolysis Bullosa. *Dermatologic Clinics* 28:131-135.

Miller JK, Swanson EW, and Spalding GE. 1975. Iodine Absorption, Excretion, Recycling, and Tissue Distribution in the Dairy Cow. *Journal of Dairy Science* 58:1578-1593. 10.3168/jds.S0022-0302(75)84753-9

Milo R, Jorgensen P, Moran U, Weber G, and Springer M. 2010. BioNumbers--the database of key numbers in molecular and cell biology. *Nucleic Acids Research* 38:D750-753. 10.1093/nar/gkp889

Naderi A, and Vanneste M. 2014. Prolactin-induced protein is required for cell cycle progression in breast cancer. *Neoplasia* 16:329-342 e321-314. 10.1016/j.neo.2014.04.001

Nichols BL, Avery S, Sen P, Swallow DM, Hahn D, and Sterchi E. 2003. The maltase-glucoamylase gene: common ancestry to sucrase-isomaltase with complementary starch digestion activities. *Proc Natl Acad Sci U S A* 100:1432-1437. 10.1073/pnas.0237170100

Nichols BL, Eldering J, Avery S, Hahn D, Quaroni A, and Sterchi E. 1998. Human small intestinal maltase-glucoamylase cDNA cloning. Homology to sucrase-isomaltase. *Journal of Biological Chemistry* 273:3076-3081.

Ohana E, Shcheynikov N, Park M, and Muallem S. 2012. Solute carrier family 26 member a2 (SLC26A2) protein functions as an electroneutral SOFormula/OH-/Cl- exchanger regulated by extracellular Cl. *Journal of Biological Chemistry* 287:5122-5132. 10.1074/jbc.M111.297192

Parkkila S, Parkkila AK, Juvonen T, and Rajaniemi H. 1994. Distribution of the carbonic anhydrase isoenzymes I, II, and VI in the human alimentary tract. *Gut* 35:646-650. 10.1136/gut.35.5.646

Penner GB, Steele MA, Aschenbach JR, and McBride BW. 2011. Ruminant Nutrition Symposium: Molecular adaptation of ruminal epithelia to highly fermentable diets. *Journal of Animal Science* 89:1108-1119. 10.2527/jas.2010-3378

Penning TM. 1997. Molecular endocrinology of hydroxysteroid dehydrogenases. *Endocrine Reviews* 18:281-305. 10.1210/edrv.18.3.0302

Reverter A, and Chan EKF. 2008. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 24:2491-2497. 10.1093/bioinformatics/btn482

Robinson MD, McCarthy DJ, and Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139-140. 10.1093/bioinformatics/btp616

Roder PV, Geillinger KE, Zietek TS, Thorens B, Koepsell H, and Daniel H. 2014. The role of SGLT1 and GLUT2 in intestinal glucose transport and sensing. *PLoS ONE* 9:e89977. 10.1371/journal.pone.0089977

Rojen BA, Poulsen SB, Theil PK, Fenton RA, and Kristensen NB. 2011. Short communication: Effects of dietary nitrogen concentration on messenger RNA expression and protein abundance of urea transporter-B and aquaporins in ruminal papillae from lactating Holstein cows. *Journal of Dairy Science* 94:2587-2591. 10.3168/jds.2010-4073

826 Rost D, Mahner S, Sugiyama Y, and Stremmel W. 2002. Expression and localization of the
827 multidrug resistance-associated protein 3 in rat small and large intestine. *Am J Physiol*
828 *Gastrointest Liver Physiol* 282:G720-G726. 10.1152/ajpgi.00318.2001
829 Scocco P, Mercati F, Brusaferro A, Ceccarelli P, Belardinelli C, and Malfatti A. 2013.
830 Keratinisation degree of rumen epithelium and body condition score in sheep grazing on
831 Brachypodium rupestre. *Veterinaria Italiana* 49:211-217.
832 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and
833 Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular
834 interaction networks. *Genome Research* 13:2498-2504. 10.1101/gr.1239303
835 Sivaprasad U, Askew DJ, Ericksen MB, Gibson AM, Stier MT, Brandt EB, Bass SA, Daines
836 MO, Chakir J, Stringer KF, Wert SE, Whitsett JA, Le Cras TD, Wills-Karp M, Silverman
837 GA, and Khurana Hershey GK. 2011. A nonredundant role for mouse SERPINB3A in the
838 induction of mucus production in asthma. *Journal of Allergy and Clinical Immunology*
839 127:254-261, 261.e251-256. 10.1016/j.jaci.2010.10.009
840 Spitzweg C, Joba W, Schriever K, Goellner JR, Morris JC, and Heufelder AE. 1999. Analysis of
841 human sodium iodide symporter immunoreactivity in human exocrine glands. *Journal of*
842 *Clinical Endocrinology and Metabolism* 84:4178-4184. 10.1210/jcem.84.11.6117
843 Steele MA, Croom J, Kahler M, AlZahal O, Hook SE, Plaizier K, and McBride BW. 2011a.
844 Bovine rumen epithelium undergoes rapid structural adaptations during grain-induced
845 subacute ruminal acidosis. *American Journal of Physiology Regulatory Integrative and*
846 *Comparative Physiology* 300:R1515-1523. 10.1152/ajpregu.00120.2010
847 Steele MA, Dionissopoulos L, AlZahal O, Doelman J, and McBride BW. 2012. Rumen epithelial
848 adaptation to ruminal acidosis in lactating cattle involves the coordinated expression of
849 insulin-like growth factor-binding proteins and a cholesterolgenic enzyme. *Journal of Dairy*
850 *Science* 95:318-327. 10.3168/jds.2011-4465
851 Steele MA, Vandervoort G, AlZahal O, Hook SE, Matthews JC, and McBride BW. 2011b.
852 Rumen epithelial adaptation to high-grain diets involves the coordinated regulation of genes
853 involved in cholesterol homeostasis. *Physiological Genomics* 43:308-316.
854 10.1152/physiolgenomics.00117.2010
855 Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT,
856 Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L, Gupta R, Bennett EP, Mandel U,
857 Brunak S, Wandall HH, Levery SB, and Clausen H. 2013. Precision mapping of the human
858 O-GalNAc glycoproteome through SimpleCell technology. *EMBO Journal* 32:1478-1488.
859 10.1038/emboj.2013.79
860 Su G, Kuchinsky A, Morris JH, States DJ, and Meng F. 2010. GLay: community structure
861 analysis of biological networks. *Bioinformatics* 26:3135-3137.
862 10.1093/bioinformatics/btq596
863 Tashian RE. 1989. The carbonic anhydrases: Widening perspectives on their evolution,
864 expression and function. *Bioessays* 10:186-192. 10.1002/bies.950100603
865 Tong L, Lan W, Lim RR, and Chaurasia SS. 2014. S100A proteins as molecular targets in the
866 ocular surface inflammatory diseases. *Ocul Surf* 12:23-31. 10.1016/j.jtos.2013.10.001
867 van Hasselt PM, Ferdinandusse S, Monroe GR, Ruiter JP, Turkenburg M, Geerlings MJ, Duran
868 K, Harakalova M, van der Zwaag B, Monavari AA, Okur I, Sharrard MJ, Cleary M,
869 O'Connell N, Walker V, Rubio-Gozalbo ME, de Vries MC, Visser G, Houwen RH, van der
870 Smagt JJ, Verhoeven-Duif NM, Wanders RJ, and van Haaften G. 2014. Monocarboxylate

transporter 1 deficiency and ketone utilization. *New England Journal of Medicine* 371:1900-1907. 10.1056/NEJMoa1407778

van Miert AS, and van Duin CT. 1991. Feed intake and rumen motility in dwarf goats. Effects of some alpha 2-adrenergic agonists, prostaglandins and posterior pituitary hormones. *Veterinary Research Communications* 15:57-67.

Van Nevel CJ, and Demeyer DI. 1996. Influence of pH on lipolysis and biohydrogenation of soybean oil by rumen contents in vitro. *Reproduction Nutrition Development* 36:53-63.

Veenendaal GH, Nijnanten FMAW-V, Duin CTMVAN, and Miert ASJPAMV. 1980. Role of circulating prostaglandins in the genesis of pyrogen (endotoxin)-induced ruminal stasis in conscious goats. *Journal of Veterinary Pharmacology and Therapeutics* 3:59-68. 10.1111/j.1365-2885.1980.tb00409.x

Venturi S, and Venturi M. 1999. Iodide, thyroid and stomach carcinogenesis: evolutionary story of a primitive antioxidant? *European Journal of Endocrinology* 140:371-372. 10.1530/eje.0.1400371

Wang G, Xu Z, Wang R, Al-Hijji M, Salit J, Strulovici-Barel Y, Tilley A, Mezey J, and Crystal R. 2012. Genes associated with MUC5AC expression in small airway epithelium of human smokers and non-smokers. *BMC Medical Genomics* 5:21.

Wang J-H, Zhao L-F, Lin P, Su X-R, Chen S-J, Huang L-Q, Wang H-F, Zhang H, Hu Z-F, Yao K-T, and Huang Z-X. 2014. GenCLiP 2.0: a web server for functional clustering of genes and construction of molecular networks based on free terms. *Bioinformatics* 30:2534-2536. 10.1093/bioinformatics/btu241

Wang LQ, Baldwin RL, and Jesse BW. 1996. Isolation and characterization of a cDNA clone encoding ovine type I carbonic anhydrase. *Journal of Animal Science* 74:345-353.

Wang TY, Liu M, Portincasa P, and Wang DQ. 2013. New insights into the molecular mechanism of intestinal fatty acid absorption. *European Journal of Clinical Investigation* 43:1203-1223. 10.1111/eci.12161

Watson-Haigh NS, Kadarmideen HN, and Reverter A. 2010. PCIT: an R package for weighted gene co-expression networks based on partial correlation and information theory approaches. *Bioinformatics* 26:411-413. 10.1093/bioinformatics/btp674

Weaver BP, Dufner-Beattie J, Kambe T, and Andrews GK. 2007. Novel zinc-responsive post-transcriptional mechanisms reciprocally regulate expression of the mouse SLC39A4 and SLC39A5 zinc transporters (Zip4 and Zip5). *Biological Chemistry* 388:1301-1312. 10.1515/bc.2007.149

Wrong OM, and Vince A. 1984. Urea and ammonia metabolism in the human large intestine. *Proceedings of the Nutrition Society* 43:77-86.

Young JW. 1977. Gluconeogenesis in Cattle: Significance and Methodology. *Journal of Dairy Science* 60:1-15. 10.3168/jds.S0022-0302(77)83821-6

Yu Z, Wildermoth JE, Wallace OAM, Gordon SW, Maqbool NJ, Maclean PH, Nixon AJ, and Pearson AJ. 2011. Annotation of sheep keratin intermediate filament genes and their patterns of expression. *Experimental Dermatology* 20:582-588. 10.1111/j.1600-0625.2011.01274.x

Table 1(on next page)

Gene Ontology enrichments of clusters

Table 1 Gene Ontology enrichments of clusters

Cluster	GO-term	FDR corrected *P*-value[1]
Rumen	EDC locus[2]	7.1E-13[3]
Epithelia-rumen-tonsil	EDC locus[2]	8.6E-15[3]
	Defense response to fungus	8.6E-03
Epithelia-rumen bias	Keratinization	2.4E-04
Epithelia-all	-	-
Epithelia-large intestine	Desmosome organization	4.7E-03
Epithelia-GI-liver	Cell junction organization	6.3E-03
Abomasum-intestine	-	-
Intestine-low in rectum	-	-
Large intestine	Regulation of chloride transport	4.5E-05
Intestine	-	-
Epithelia-intestine	Cell cycle process	1.4E-46
Abomasum	Digestion	3.8E-02
Small intestine	-	-
Rumen-abomasum	Platelet aggregation	2.2E-04
Rumen-intestine-liver	Flavonoid biosynthetic process	5.5E-10
Intestine-spleen	Humoral immune response	4.5E-02

[1]Top significantly enriched pathway selected from GORILLA analysis (see methods) for each input gene cluster
[2]Genes in the EDC locus of the sheep genome.
[3]Enrichment of EDC locus genes was calculated using the hypergeometric distribution.

Table 2(on next page)

Representation of the pig GIT gene clusters in the sheep GIT network

Table 2 Representation of the pig GIT gene clusters in the sheep GIT network

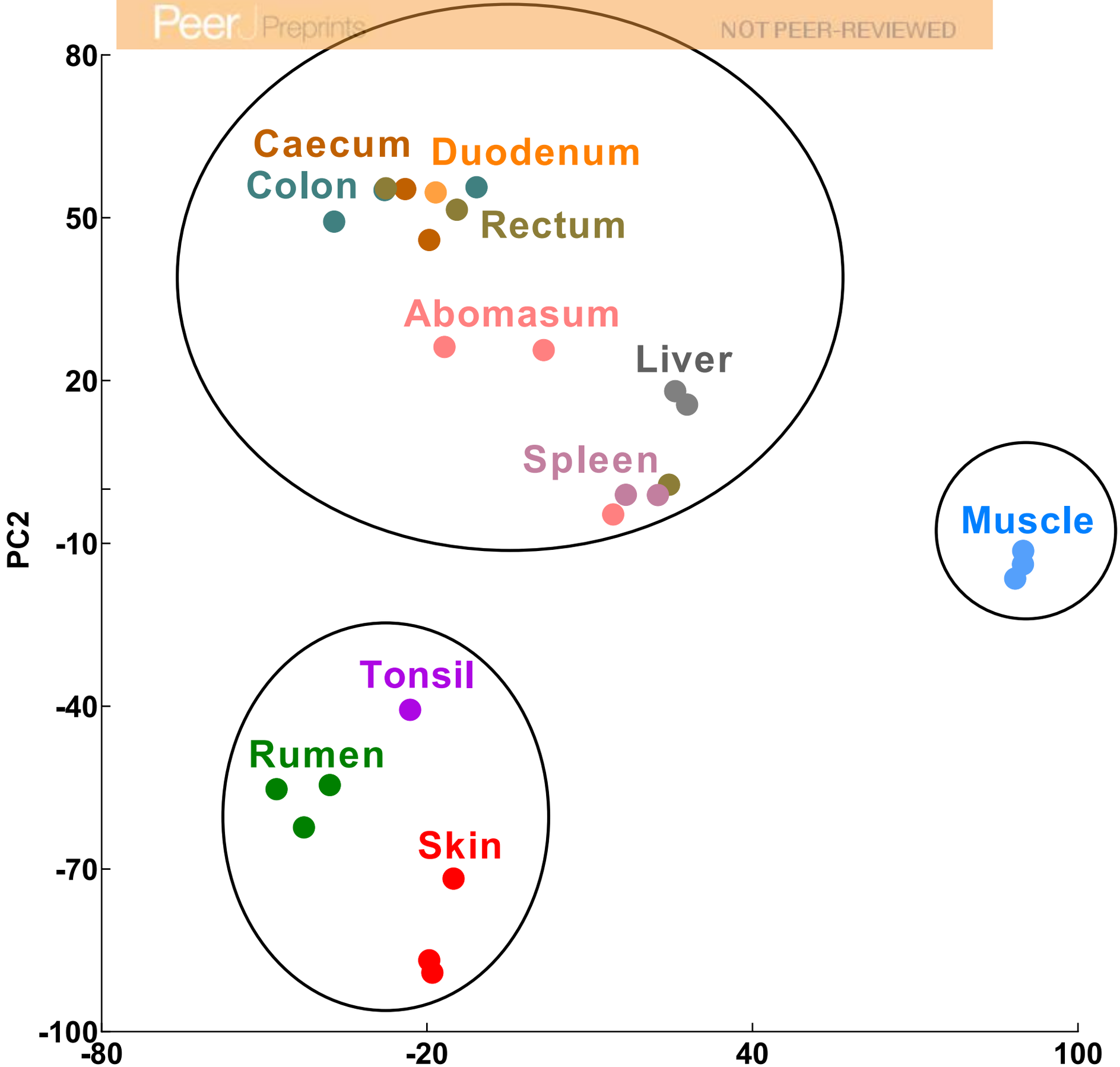
Pig cluster[1]	Pig tissues[1]	Pig cell type of origin[2]	Overlap	*P*-value[2]	Representation	Sheep tissues	Go term enrichment	*P*-value[3]
Overall			179	8.1E-31	Over		Cell cycle process	2.0E-13
1, 7	Intestine -	Immune cells/cell cycle	58	2.4E-11	Over	Epithelia, intestine	Cell cycle process	1.5E-33
3, 8	Tongue-oesophagus	Stratified squamous epithelia	73	1.3E-34	Over	Rumen, epithelia, abomasum, large intestine	Epidermis development	2.9E-05
2, 4, 9	Oesophagus-stomach	Muscle	9	0.0002[4]	Under[4]	Rumen, abomasum	na	
6, 13, 15	Salivary gland	Stratified columnar epithelia	4	0.1777	None		na	
5, 12, 14, 16	Stomach-intestine	Ciliate/glandular epithelia	35	5.4E-09	Over	Stomach intestine	na	
10	Stomach	Neuronal	0	na	na		na	

[1]Numbers, names and grouping of pig gene clusters by cell type of origin are according to (Freeman et al. 2012).
[2]Calculated hypogeometric *P* values, representing the significance of representation of pig genes in sheep gene network.
[3]FDR corrected GO term enrichment *P* values.
[4]If overlap with just the rumen and rumen-abomasum clusters, significant (*P*=8E-05) over representation

Figure 1 (on next page)

Transcriptomic sample clustering.

Each dot represents one tissue sample from a single animal. Circles indicate significant clusters (confidence interval = 95%). Raw PCA plots are available (Additional file 2: Figure S1).

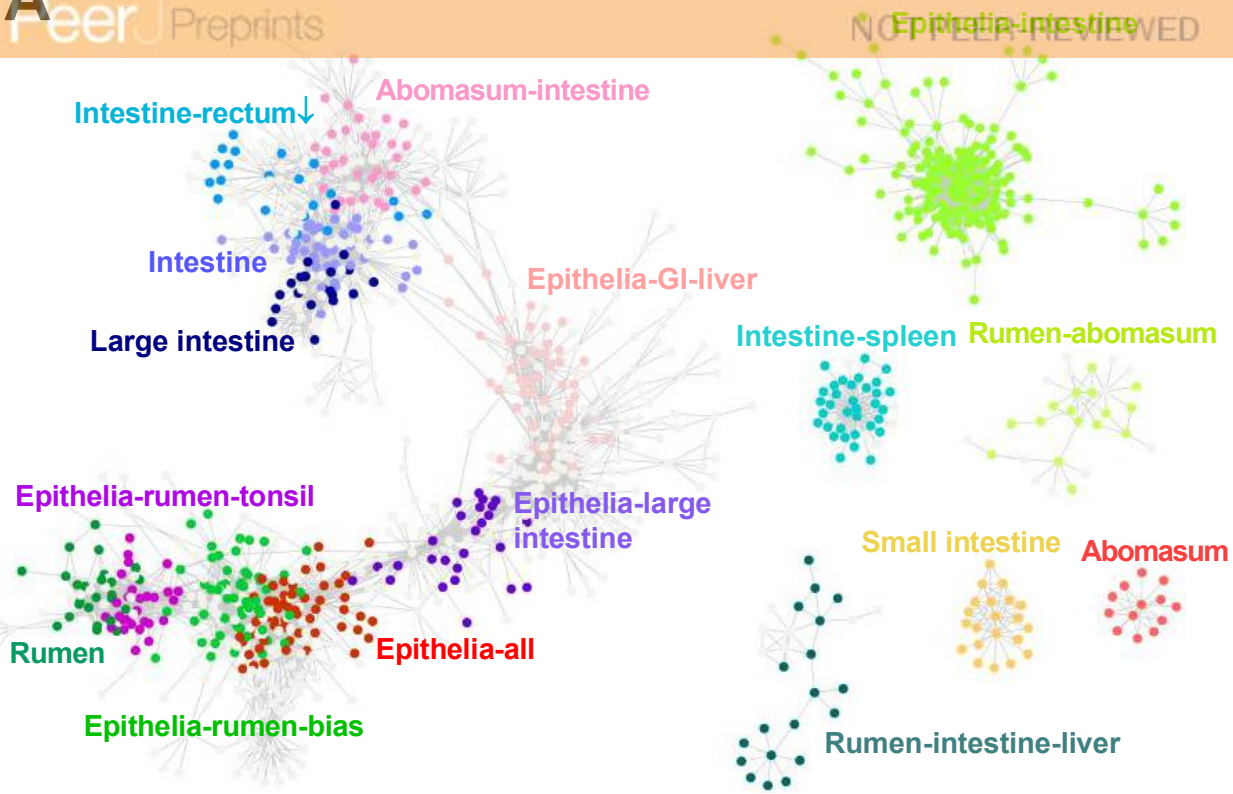


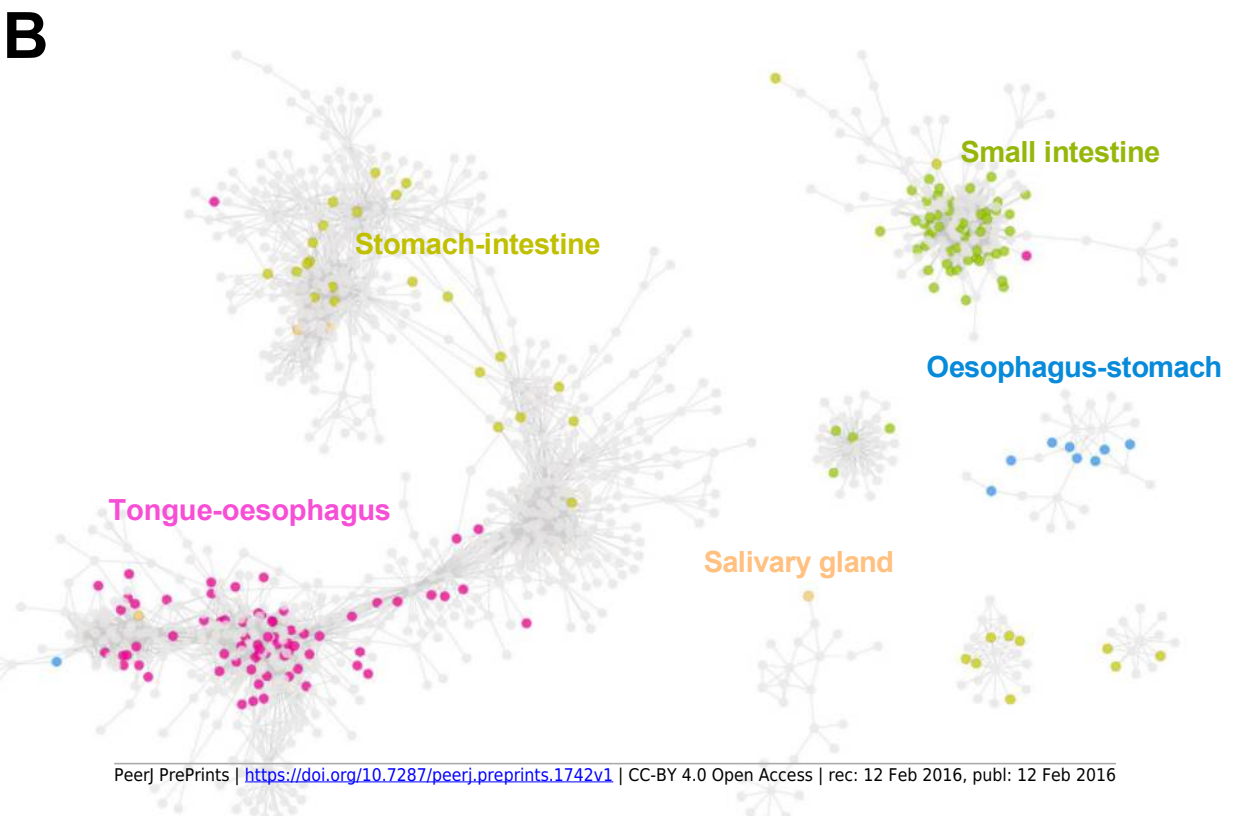
PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1742v1 | CC BY 4.0 Open Access | rec: 12 Feb 2016, publ: 12 Feb 2016

Figure 2(on next page)

Gene co-expression network.

(**A**) Each dot represents a sheep transcript and different colors represent the tissue(s) where the transcript showed high expression, compared to the other tissues. Rectum↓: low in rectum. (**B**) The same gene co-expression network with only the orthologous genes present in specific pig GIT clusters (Freeman et al. 2012) highlighted (Additional file 1). The names and colors of pig cluster were determined according to the tissues where genes showed the highest and the second highest expression level in the pig GI gene network (Freeman et al. 2012) .



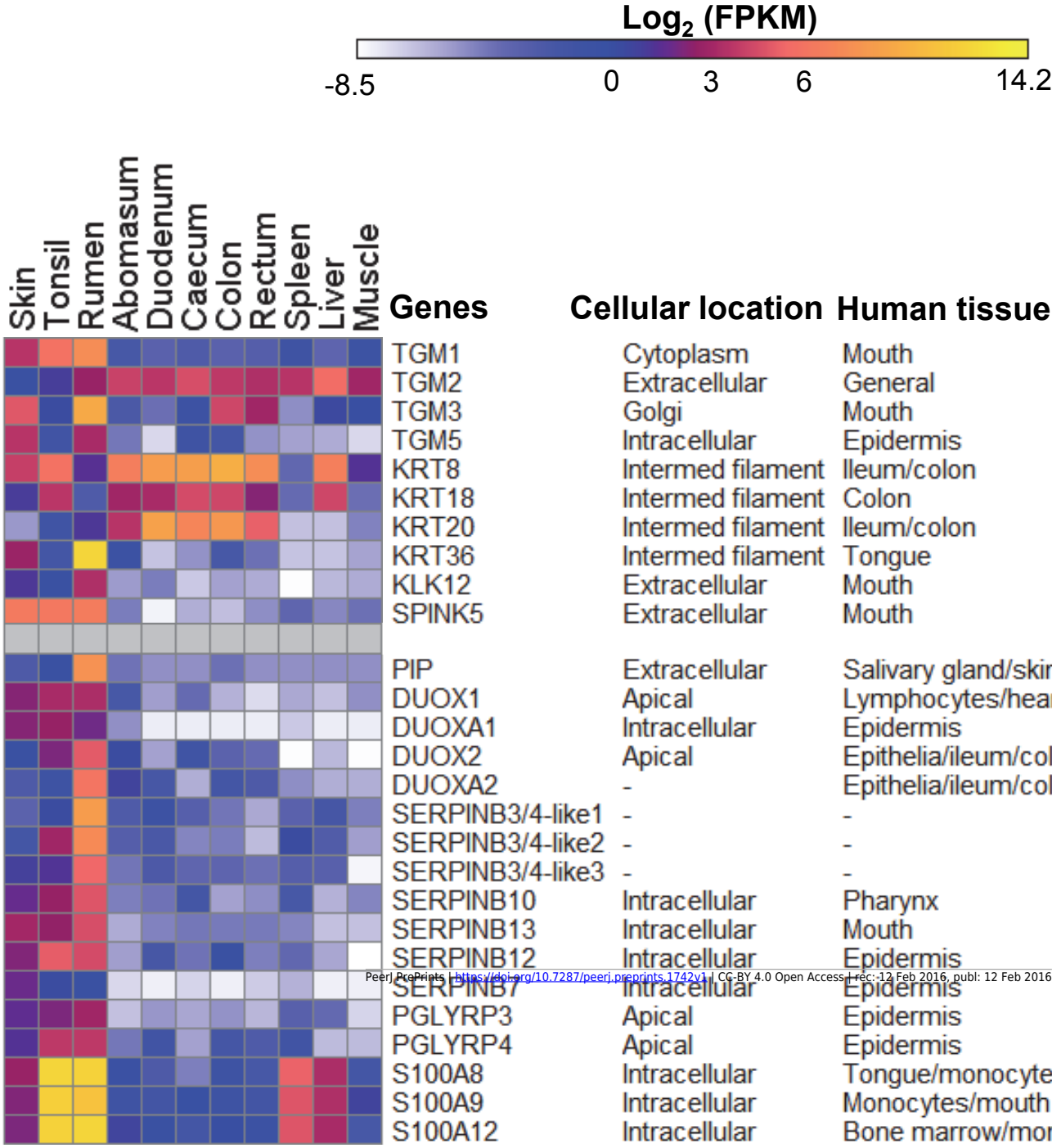


PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1742v1 | CC-BY 4.0 Open Access | rec: 12 Feb 2016, publ: 12 Feb 2016

Figure 3(on next page)

Expression profiles of innate immunity and epithelial development genes in sheep.

Data are presented with $\log_2$ Fragments Per Kilobase of exon per Million fragments mapped (FPKM) values along with the subcellular locations and/or tissues of pig (Freeman et al. 2012) and human (Genevestigator (Hruz et al. 2008) analysis) where these genes showed high expression. Cellular location information were derived from GENATLAS database (Frezal 1998) .



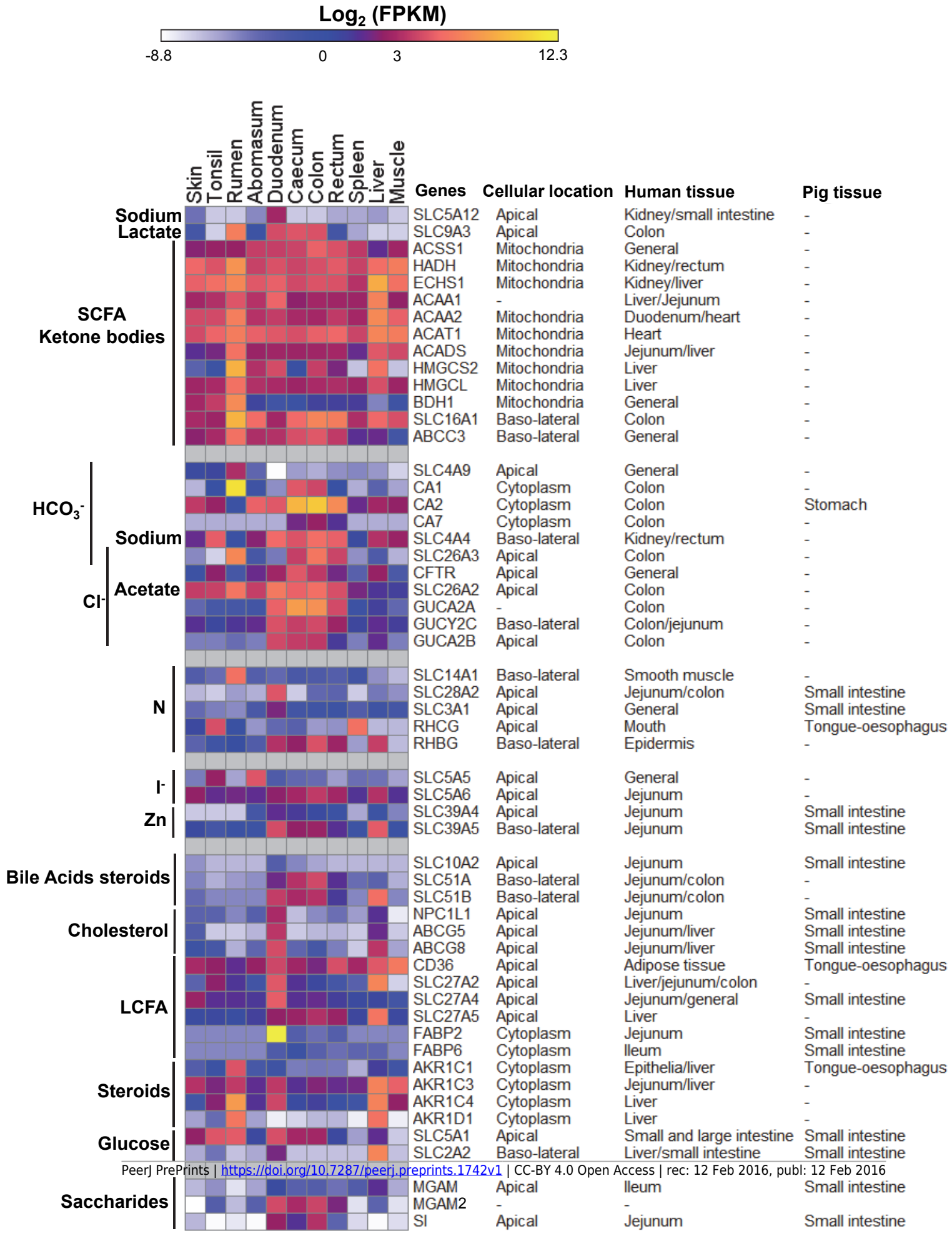
Genes	Cellular location	Human tissue	Pig tissue
TGM1	Cytoplasm	Mouth	Tongue-oesophagus
TGM2	Extracellular	General	-
TGM3	Golgi	Mouth	-
TGM5	Intracellular	Epidermis	Tongue-oesophagus
KRT8	Intermed filament	Ileum/colon	Tongue-oesophagus
KRT18	Intermed filament	Colon	-
KRT20	Intermed filament	Ileum/colon	-
KRT36	Intermed filament	Tongue	Tongue-oesophagus
KLK12	Extracellular	Mouth	-
SPINK5	Extracellular	Mouth	-
PIP	Extracellular	Salivary gland/skin	Salivary gland
DUOX1	Apical	Lymphocytes/heart	Tongue-oesophagus
DUOXA1	Intracellular	Epidermis	Tongue-oesophagus
DUOX2	Apical	Epithelia/ileum/colon	-
DUOXA2	-	Epithelia/ileum/colon	-
SERPINB3/4-like1	-	-	-
SERPINB3/4-like2	-	-	-
SERPINB3/4-like3	-	-	-
SERPINB10	Intracellular	Pharynx	Tongue-oesophagus
SERPINB13	Intracellular	Mouth	-
SERPINB12	Intracellular	Epidermis	-
SERPINB7	Intracellular	Epidermis	Tongue-oesophagus
PGLYRP3	Apical	Epidermis	-
PGLYRP4	Apical	Epidermis	Tongue-oesophagus
S100A8	Intracellular	Tongue/monocytes	Tongue-oesophagus
S100A9	Intracellular	Monocytes/mouth	Tongue-oesophagus
S100A12	Intracellular	Bone marrow/monocytes	Tongue-oesophagus

Figure 4(on next page)

Gene expression profiles of metabolic processes discussed in the text.

Data are presented with $\log_2$ Fragments Per Kilobase of exon per Million fragments mapped (FPKM) values along with the subcellular locations and/or tissues of pig (Freeman et al. 2012) and human (Genevestigator (Hruz et al. 2008) analysis) where these genes showed high expression. Texts and bars on the left side of the heatmap indicate involved pathways for covered genes described in the article. Cellular location information were derived from GENATLAS database (Frezal 1998).

Figure 4

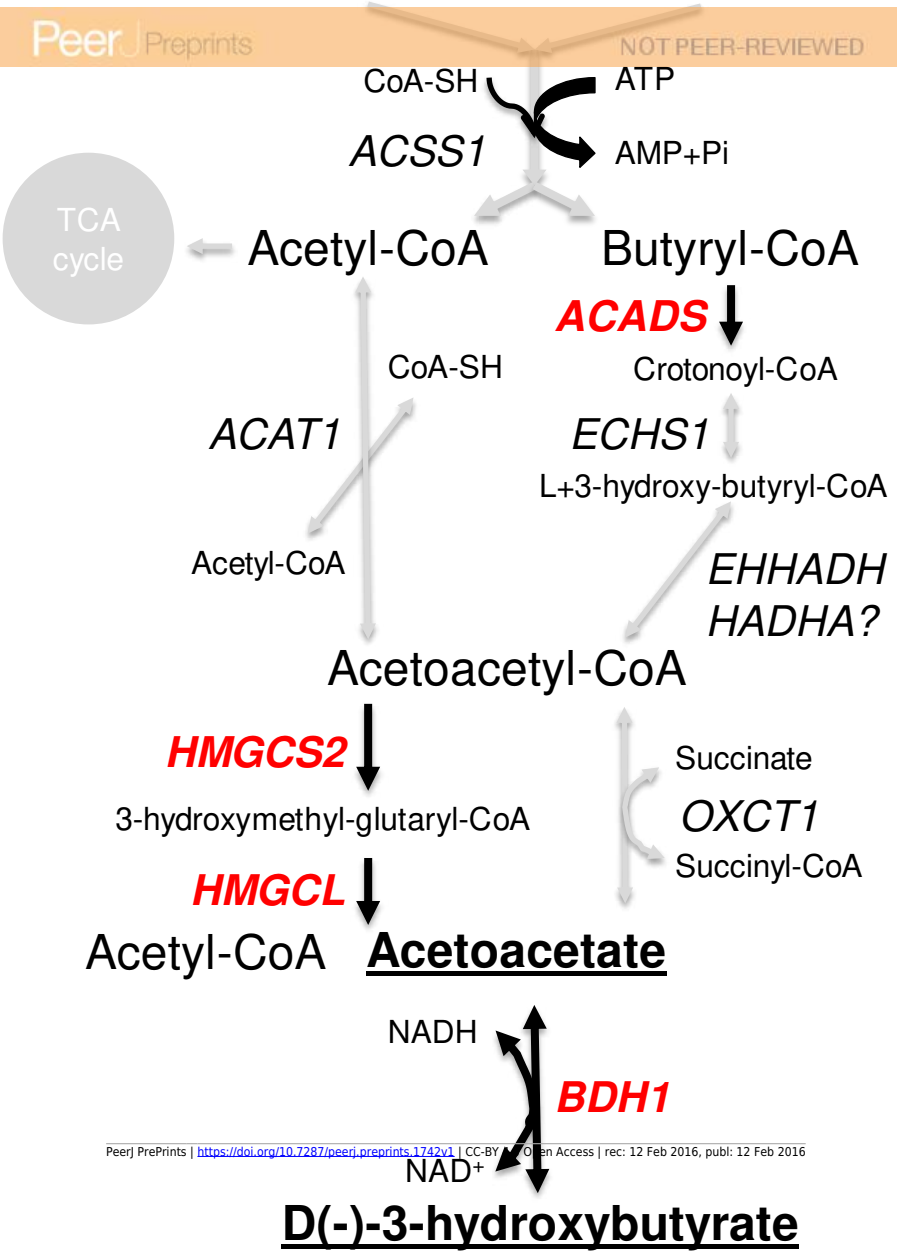


PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1742v1 | CC-BY 4.0 Open Access | rec: 12 Feb 2016, publ: 12 Feb 2016

Figure 5(on next page)

Ruminant ketone body metabolism pathways.

Key enzyme encoding genes (red text) and pathways (black arrow) are highlighted.



PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1742v1 | CC-BY Open Access | rec: 12 Feb 2016, publ: 12 Feb 2016

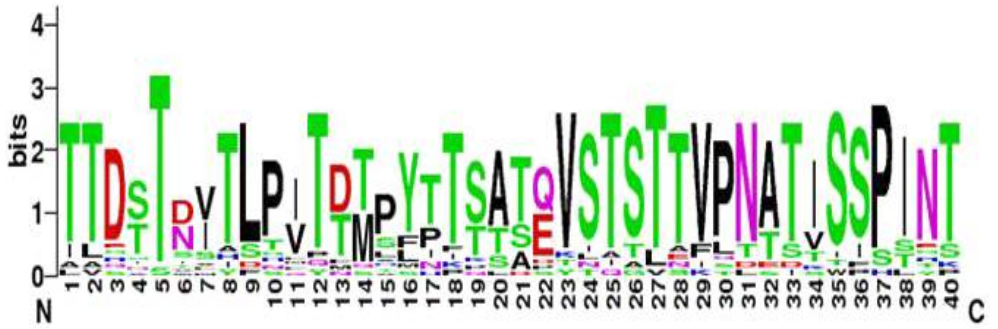
Figure 6(on next page)

Organization of the MGAM2 carboxy-teminus.

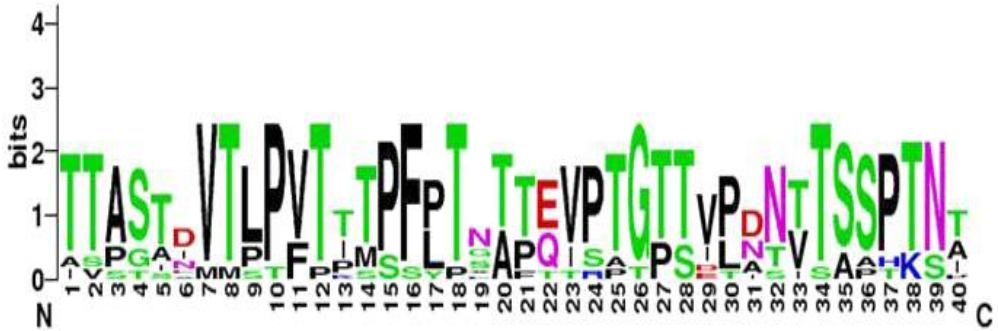
Consensus motifs of the serine/threonine rich 40 amino acid repeats at the carboxy-terminus of predicted MGAM-like proteins. **(A)** sheep. **(B)** cattle. **(C)** Pig. **(D)** human.

A Sheep MGAM-like protein amino acid repeats

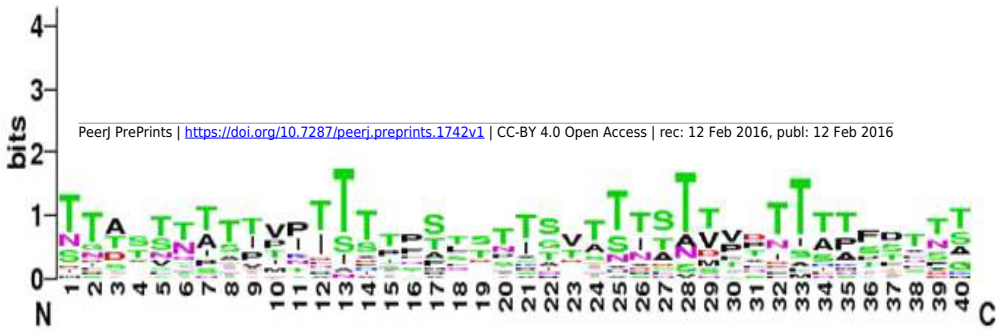
B Cattle MGAM-like protein amino acid repeats



C Pig MGAM-like protein amino acid repeats



D Human MGAM-like protein amino acid repeats



PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1742v1 | CC-BY 4.0 Open Access | rec: 12 Feb 2016, publ: 12 Feb 2016