Differential gene expression in mammary gland epithelial cells between Holstein individuals with high or low α-lactalbumin milk protein content

Lars-Erik Holm[1], Nina Poulsen[2], Bo Thomsen[1], Frank Panitz[1], Lotte B. Larsen[2], Thao Thi Thu Le[2], Vivi R. Gregersen[1,3].

[1] *Dept. of Molecular Biology and Genetics, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark*
Larserik.Holm@mbg.au.dk (Corresponding author)
[2] *Dept. of Food Science, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark*
[3] *Current address: Qiagen Bioinformatics, Prismet, Silkeborgvej 2, 8000 Aarhus C, Denmark*

Introduction

α-lactalbumin is one of the key protein components in milk. It is a small, acidic and Ca^{2+} binding protein. It has several important functions, primarily as one of the two components in lactose synthase, which catalyzes the final step in lactose synthesis in the mammary gland. α-lactalbumin has a single strong Ca2+ binding site and several zinc binding sites. The binding of these ions can induce conformational changes and greater stability towards denaturing reagents (reviewed in Permyakov & Berliner, 2000). It has also been shown that α-lactalbumin can have apoptotic effects on specific tumors (Ho *et al*., 2017).

α-lactalbumin is also one of the key constituents of infant formula and in human milk it constitutes 20-25% of the proteins however in bovine milk it only constitutes 2-5% (Lönnerdal & Lien, 2003). This difference makes it necessary to complement bovine milk with additional α-lactalbumin in infant formulae especially as α-lactalbumin is rich in the essential amino acid tryptophan and is the key supplement for this amino acid (Lien, 2003).

The purpose of this study was to elucidate gene expression levels in individuals with high or low α-lactalbumin content in milk, respectively, in order to identify differentially expressed genes between these two phenotypes. This could aid in the understanding of the processes involved in the production of this protein in the mammary gland and could point towards pathways or genetic variants with an effect on production of α-lactalbumin. Likewise, it could also point to genes and/or pathways affected by changes in levels of α-lactalbumin.

Materials and methods

Animals sampled for RNAseq analysis consisted of 24 Danish Holstein originating from 4 different herds and all from 1st parity. RNA was extracted from mammary gland epithelial cells utilizing monoclonal anti-cytokeratin peptide 8 (Sigma-Aldrich, St. Louis, Missouri, USA) to capture cells from 0.4 liter fresh whole milk. RNA was precipitated using QIAzol (Qiagen, Hilden, Germany) following the “Purification of total RNA from animal tissue” protocol described in the miRNease Mini Handbook (Qiagen). The RNA was used to prepare strand-specific sequencing libraries (ScriptSeq, Illumina, San Diego) for each sample. The libraries were sequenced on an Illumina HiSeq2000 (100 bp paired-end).

Quality assessment was performed using FastQC (v. 0.11.3). SortMeRNA (v. 2.1) (Kopylova

et al. 2012) was used to remove rRNA prior to adapter clipping, quality trimming and length filtering with Timmomatic (v. 0.35) (Bolger *et al*., 2014). The cleaned paired-end sequences were mapped to the bovine genome UMD3.1 using STAR (v. 2.5.0a) (Dobin *et al*., 2013). STAR was also used to generate count data for the subsequent differential gene expression analysis. Gene expression differences between the 5 individuals with the highest and 5 with the lowest α-lactalbumin protein content in milk were analyzed using count data. Count data was normalized as suggested by Robinson & Oshlack (2010) prior to performing the likelihood ratio test implemented in EdgeR (Robinson *et al*., 2010; McCarthy *et al*., 2012). The list of differentially regulated genes was subjected to functional analysis using the DAVID (v. 6.8) functional annotation tool (Huang *et al.*, 2009a; 2009b). A total of 99 up-regulated and 21 down-regulated genes, respectively, were used for this analysis corresponding to all genes with $p<10^{-4}$ in the differential expression analysis.

Results

The identified differentially expressed genes (Table 1) did not reveal specific pathways or genes related to milk protein genes or known genes related to milk production phenotypes. The DAVID functional annotation analysis identified the GO terms “negative regulation of viral genome replication” (GO:0045071), “negative regulation of amyloid-beta formation” (GO: 1902430), “defence response to virus” (GO:0051607) and the KEGG pathway “MicroRNAs in cancer” (bta05206) as the most enriched with $p<0.01$, however none of these were significantly enriched after Bonferroni adjustment.

Discussion

The major milk protein genes was found to be expressed in RNAseq studies on mammary epithelial cells (Lemay *et al*., 2013) showing that the approach of working with this tissue does give a picture of the transcriptome in this cell type.

Table 1. List of the 25 most significant Differentially Expressed genes. All are upregulated in high α-lactalbumin milk protein animals except for one gene (underlined).

ensembl_gene_id	logFC	P-Value	FDR	Gene	Gene description
ENSBTAG00000042662	3.39	1.5E-13	2.0E-09	SNORD17	Small nucleolar RNA SNORD17
ENSBTAG00000042723	3.12	4.4E-11	2.9E-07	U11	U11 spliceosomal RNA
ENSBTAG00000007191	2.52	6.2E-09	2.8E-05	CCL5	C-C motif chemokine ligand 5
ENSBTAG00000007962	2.14	2.0E-08	6.7E-05	ATP9A	ATPase phospholipid transporting 9A (putative)
ENSBTAG00000043250	2.21	5.3E-08	1.4E-04	7SK	7SK RNA
ENSBTAG00000021452	2.62	8.6E-08	1.9E-04	TRANK1	tetratricopeptide repeat and ankyrin repeat containing 1
ENSBTAG00000045966	2.23	1.3E-07	2.5E-04	GPRIN3	GPRIN family member 3
ENSBTAG00000012371	1.74	1.9E-07	2.9E-04	CPD	carboxypeptidase D
ENSBTAG00000038698	2.45	2.0E-07	2.9E-04		
ENSBTAG00000028359	2.09	2.2E-07	3.0E-04	U3	Small nucleolar RNA U3
ENSBTAG00000022799	2.21	3.1E-07	3.4E-04	NOTCH1	notch 1
ENSBTAG00000048062	3.13	3.0E-07	3.4E-04	KDM6B	lysine demethylase 6B
ENSBTAG00000008154	1.85	4.0E-07	4.1E-04	LIMD2	LIM domain containing 2
ENSBTAG00000018804	2.60	4.9E-07	4.7E-04	CELSR2	cadherin EGF LAG seven-pass G-type receptor 2 precursor
ENSBTAG00000003857	2.99	5.3E-07	4.7E-04	SUSD6	sushi domain containing 6
ENSBTAG00000005923	2.22	8.4E-07	6.9E-04	ABTB2	ankyrin repeat and BTB domain containing 2
ENSBTAG00000020713	1.94	1.1E-06	8.6E-04	BACH2	BTB domain and CNC homolog 2
ENSBTAG00000014762	2.10	1.3E-06	9.5E-04	ISG20	interferon stimulated exonuclease gene 20
ENSBTAG00000003840	2.20	1.6E-06	1.2E-03	GUCY1B3	guanylate cyclase 1 soluble subunit beta
1ENSBTAG00000016061	2.22	1.9E-06	1.2E-03	RSAD2	radical S-adenosyl methionine domain containing 2
ENSBTAG00000045661	2.21	2.7E-06	1.6E-03	Metazoa_SRP	Metazoan signal recognition particle RNA
ENSBTAG00000023026	3.43	2.7E-06	1.6E-03		
ENSBTAG00000008040	-1.39	3.0E-06	1.7E-03	SPG20	spastic paraplegia 20 (Troyer syndrome)
ENSBTAG00000020053	1.42	3.2E-06	1.7E-03	ZEB1	zinc finger E-box-binding homeobox 1
ENSBTAG00000009785	1.57	3.1E-06	1.7E-03	SAMD4B	sterile alpha motif domain containing 4B

The list of differentially expressed genes do not give any indication of regulation of α-lactalbumin or processes regulated by α-lactalbumin levels although there are a number of small RNA genes which might be involved in regulation of other genes. The DAVID annotation does indicate a relationship towards viral defence or cancer related issues although not significant. α-lactalbumin in complex with fatty acids (HAMLET) are known to have an anti-tumor effect caused by specific folding properties (Mossberg *et al*., 2010) but an association with miRNA regulation has not been described. Association of specific folding variants of α-lactalbumin with bactericidal properties has also been described (Håkansson *et al*., 2000) but there are no studies showing anti-viral properties of α-lactalbumin as otherwise indicated by the DAVID annotation.

However, the most likely explanation for the observed differentially expressed genes are that effects on α-lactalbumin expression levels are not caused by genes expressed in the mammary epithelial cells and that α-lactalbumin levels do not affect the expression of other genes in the mammary epithelial cells. This leads to the conclusion that the genetic pathways affecting or being affected by α-lactalbumin levels are controlled in other tissues than the mammary epithelial cells.

List of References

Bolger, A.M., M. Lohse & B. Usadel, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30, 2114-2120.

Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleksi, S. Jha, P. Batut, M. Chaisson & T.R. Gingeras, 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29, 15-21.

Ho, J.C., A. Nadeem & C. Svanborg, 2017. HAMLET – a protein-lipid complex with broad tumoricidal activity. Biochem. Biophys. Res. Commun., 482, 454-458.

Huang, D.W., B.T. Sherman & R.A. Lempicki, 2009a. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc., 4, 44-57.

Huang, D.W., B.T. Sherman & R.A. Lempicki, 2009b. Bioinformatics enrichment tools: path toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res., 37, 1-13.

Håkansson, A., M. Svensson, A.-K. Mossberg, H. Sabharwal, S. Linse, I. Lazou, B. Lönnerdal & C. Svanborg, 2000. A folding variant of α-lactalbumin with bactericidal activity against *Streptococcus pneumoniae*. Mol. Microbiol., 35, 589-600.

Kopylova, E., L. Noé & H. Touzet, 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics, 28, 3211-3217.

Lemay, D.G., O.A. Ballard, M.A. Hughes, A.L. Morrow, N.D. Horseman & L.A. Nommsen-Rivers, 2013. RNA sequencing of the human milk fat layer transcriptome reveals distinct gene expression profiles at three stages of lactation. PLOSone, 8, e67531.

Lien, E.L., 2003. Infant formulas with increased concentrations of alpha-lactalbumin. Am. J. Clin. Nutr., 77, 1555S-1558S.

Lönnerdal, B. & E.L. Lien, 2003. Nutritional and physiological significance of alpha-lactalbumin in infants. Nutr. Rev., 61, 295-305.

Mossberg, A.K., K. Hun Mok, L.A. Morozova-Roche & C. Svanborg, 2010. Structure and function of human α-lactalbumin made lethal to tumor cells (HAMLET)-type complexes. FEBS J. 277, 4614-4625.

Robinson, M.D., D.J. McCarthy & G.K. Smyth, 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26, 139-140.

McCarthy, J.D., Y. Chen & K.G. Smyth, 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res., 40, 4288-4297.

Permyakov, E.A. & J.L. Berliner, 2000. α-lactalbumin: structure and function. FEBS Letters, 473, 267-274.

Robinson, M.D. & A. Oshlack, 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol., 11, R25.