

Identification of *Cis*-regulatory Elements of Butyrophilin Gene of the Mammary Gland

Demie Aschalew L^{1*}, Zegeye Abiy² and Qayyum Khan A³

¹Department of Biology, College of Natural and Computational Sciences, Wollega University, Post Box No: 395, Nekemte, Ethiopia

²Biotechnology Program, School of Postgraduate Studies, Addis Ababa University, Addis Ababa, Ethiopia

³Department of Biology, Faculty of Natural Sciences, Arba Minch University, Arba Minch, Ethiopia

Abstract

Butyrophilin subfamily 1 member A1 is a highly expressed gene in mammary gland of all mammals during lactation. It is found to be the major integral protein in the milk fat globule membrane. Its interactions with other membrane elements and soluble proteins of the mammary epithelial cells regulate the secretion of milk fat. In order to reveal any shared *cis*-acting elements, the human butyrophilin subfamily 1 member A1 5'-flanking genomic sequence was analyzed and compared with four mammals comprising of a rodent and three primates, *viz.*, *Mus musculus*, and *Pan troglodytes*, *Pongo obelli*, and *Macaca mulata*, and three other mammary specific human milk genes using publicly available bioinformatics tools. Prior to a multiple sequence analysis, low complexity DNA sequences were masked using CENSOR. The multiple sequence analysis revealed nine highly conserved regions of similarities in the 5'-flanking region of butyrophilin genes across species. Consensus putative transcription factor binding sites were identified using MatInspector and compared with SiteGA results. They were subsequently examined for the expression in the mammary gland as well as for their occurrence in the previously identified region of homology. Finally, CCCTC-binding factor (CTCF) and nuclear receptor subfamily 2 - factor (NR2F) were identified at a similar distance from the transcription start site in the 5'-flanking region of butyrophilin gene across the species. However, the exact interaction of these transcription factors with the butyrophilin gene is not known, which needs further investigation.

Article Information

Article History:

Received : 17-07-2013

Revised : 27-09-2013

Accepted : 29-09-2013

Keywords:

Milk fat globule
Butyrophilin gene
Transcription factor
Binding sites

*Corresponding Author:

Demie Aschalew L

E-mail:

aschusmart2006@gmail.com

INTRODUCTION

Milk is the principal source of food that provides all the required nutritional components for the neonates and thus is complex in its composition (Lamey *et al.*, 2007). Among various components of milk, the regulation of milk fat production has been the subject of interest for investigation. Milk fat is released as micro lipid droplets surrounded by a special membrane composed of lipid bi-layer, the milk fat globule membrane (MFGM). According to Mather and Keenan (1998) MFGM originates from the plasma membrane of the mammary epithelial cells (MEC) during the extrusion of lipid droplets from the cells. Similarly, it interacts with different peripheral and membrane proteins (Mather, 2000).

The genes, butyrophilin subfamily 1 member A1 (BTN1A1), Xanthine oxidoreductase (XOR) and adipophilin (ADPH) are described for their potential role in the secretion of milk fat from MEC by the 'tripartite model of milk fat secretion' in McManaman *et al.* (2007).

BTN1A1 is described as quantitative trait loci (QTL) with valuable economic importance among various milk genes in Abiy *et al.* (1999). Gene knockout experiments have confirmed the importance of BTN1A1 in lactating mammary gland for the proper secretion of milk fat in Ogg *et al.* (2004) and Robenek *et al.* (2006). Until its recent

finding in other tissues including spleen and thymus by applying reverse –transcription PCR technique and sensitive biochemical survey by Smith *et al.* (2010), butyrophilin is an exclusively expressed gene in the secretory epithelium of the mammary gland during lactation (Frank *et al.*, 1981).

Milk secretion is specified temporally and spatially in all mammals. This indicates the presence of precise and well established control system for milk secretion. As mentioned in Cartharius *et al.* (2005) promoter analysis is an essential step in the eukaryotes to identify regulatory networks for particular genes. Therefore, finding out *cis*-regulatory motifs of BTN1A1 gene that confer its spatial and temporal specificity will be an important step towards determining the quantity and quality of milk fat expression and production at gene level. The present investigation is undertaken to identify and compare conserved transcription factor binding sites (TFBSs) that lie in 5'- regions of the DNA sequences of BTN1A1 gene in human, a rodent and three primates, and three other human milk genes.

MATERIALS AND METHODS

DNA sequences of BTN1A1 gene of five mammalian orthologs, comprising *Homo sapiens*, *Mus musculus*, *Pan troglodytes*, *Pongo obelli* and *Macaca mulata* were retrieved from GenBank. The length of the DNA sequences included 1000 bp upstream from the Transcription start site (TSS) and the downstream sequence until the CDS, the codon ATG. All the sequences were stored in FASTA format. DNA sequences of three other human milk genes, namely, beta-casein (CNS2), adipophilin (ADPH) and alpha-lactalbumine (LALBA) genes were collected in the same manner.

Prior to multiple sequence alignment (MSA), DNA repeat elements were identified and screened using CENSOR, a service of the Genetic Information Research Institute (Kohany *et al.* 2006). MSA was conducted for the two categories of genes, viz., the BTN1A1 genes across five species, and the human BTN1A1 and three other milk genes using ClustalW2 of Larkin *et al.* (2007). The ATG, the first codon, and 34 randomly selected nucleotides were included and utilized in the MSA as the anchor sequence for the analysis across species, and the four human milk genes, respectively.

A search for TATA box was performed by submitting sequence to Transcription Regulatory Element Database (TRED) of Zhao *et al.* (2005), and the McPromoter program (version MM:II) was used for TSS prediction (Ohler, 2000).

MatInspector (version 8.5) of Cartharius *et al.* (2005) was utilized for the identification of putative promoter elements other than the core promoter elements. Both groups, the BTN1A1 genes across five species and the four human milk genes were submitted to MatInspector for the prediction of the TFBS separately. The outputs generated by MatInspector prediction were compared with another TFBS prediction program, SiteGA. Among TFBSs predicted by MatInspector, the common ones were further examined by DiAlign, an additional program used to scan for the presence of TFBSs in the conserved regions of the MSA. The expression of the identified TFs in mammary gland was also examined.

RESULTS

The details of 5-flanking sequences for milk genes, BTN1A1 in five mammalian species, mRNA transcription start sites and the number of base pairs with repeat and without repeat sequences are presented in Table 1 A and B. A minimum and maximum of 1724 and 2090 bp were used for comparison of BTN1A1 genes across the species. For the second category i.e. for the analysis among the human milk genes a minimum of 1016 and a maximum of 2152 bp were used for the comparative analysis.

McPromoter generated output that doesn't meet the threshold value for both categories. TRED did not find TATA box in any of the five BTN1A1 and ADPF genes. TATA box was found -75 and -25 bp upstream from TSS of CNS2 and LALBA respectively. The MSA revealed some sort of similarity in both analyses. The MSA for the five BTN1A1 resulted in nine regions with long stretch of conserved DNA sequences. However, the alignment level for the four milk gene seems poor i.e. interrupted with many longer gaps.

MatInspector predicted a total of 1755 and 959 matches of putative TFBS for the five species and the four milk genes respectively out of which 68 TFBS were conserved across the species. On the other hand, SiteGA predicted only three TFBSs, namely; HNF4, IRF1 and SREBP to be conserved across the species (Table 3). MatInspector displays the outputs in a graphical view panel, which can be used for visual inspection (Figure 1).

TFs that are common for the two groups were selected as consensus TFs. Then, the consensus TFBS were examined for their expression in the mammary gland. This was performed through the Stanford's 'SOURCE' public database that is accessible through its official website. The examination on the expression of the consensus TF

Table 1: The 5'-flanking DNA sequences used for the comparative analysis and their genomic features.

A) BTN1A1 genes across species.

Organisim	CHR ¹	Accession/ Gene ID Regions	TSS*	Base Pairs with/ Without Repeats
<i>H. sapiens</i>	6	>gi/224589818:26499428-26501517	617	2090/1171
<i>M. musculus</i>	13	>gi 149292733:c23558770-23557047	872	1724/1508
<i>P. troglodytes</i>	6	>gi 291061370:26767393-26769342	693	1950/1243
<i>P. obelii</i>	6	>gi 241864936:27621082-27623024	724	1943/1300
<i>M. mulata</i>	4	>gi 109156893:26532708-26534650	714	1943/1261

B) The four human milk genes.

Gene (<i>H.sapeins</i>)	CHR ¹	Accession/Gene ID Regions	TSS*	Base Pairs With/ Without Repeats
BTN1A1	6	>gi/224589818:26499428-26501517	617	2090/1171
CSN2	4	>gi/224589816:c70826711-70827726	350	1016/0365
ADPH	9	>gi/2245898221:c19126422-19128573	474	2152/1346
LALBA	12	>gi/224589803:c48963801-48964829	686	1629/1112

*TSS refers to the nucleotide sequence position at which transcription of the mRNA starts after the repeat sequences are removed. ¹ the chromosome number where the gene is located. CENSOR detected different types of low complexity elements and screened them all in both groups. Pair-wise alignments among the five mammalian BTN1A1 revealed different levels of similarity (Table 2A). The pair-wise similarity level within the four human milk genes is shown in Table 2B.

Table 2: The pair-wise similarity indices of the query sequences using Genomix © DiAlign professional TF Release 3.1.3 August 2008. The two most similar sequences are marked as 1.000 (not necessarily perfect match); percents are the proportion of identical nucleic acid residues of the shorter sequence.

A) Across the five species.

	<i>M. musculus</i> (1508 bp)	<i>P. troglodytes</i> (1243 bp)	<i>P. obelii</i> (1300 bp)	<i>M. mulata</i> (1261 bp)
<i>H. sapiens</i> (1171 bp)	0.051 33 %	1.000 95 %	0.935 92 %	0.833 90 %
<i>M. musculus</i> (1508 bp)		0.054 33 %	0.047 31 %	0.054 32 %
<i>P. troglodytes</i> (1243 bp)			0.976 91 %	0.881 90 %
<i>P. obelii</i> (1300 bp)				0.871 88 %

B) Across the four human milk gene.

	ADPH (1346 bp)	LALBA (714 bp)	BTN1A1 (1171 bp)
CSN2 (365 bp)	0.354 10 %	0.554 17 %	0.596 29 %
ADFP (1346 bp)		1.000 24 %	0.441 10 %
LALBA (714 bp)			0.503 11 %

Table 3: Common TFBSs in the regions of similarity of the BTN1A1 multiple sequence analysis across the species. Common putative TFBS identified by MatInspector and SiteGA softwares in the conserved regions. Cells with 'None' indicate the absence of common TFBS in the region. * Region designate the nucleotide position in human BTN1A1.

Conserved Region*	MatInspector- TFBS (matrix) in 100% of the Species	SiteGA- TFBS in 100% of species
1-42	V\$AP1F, V\$ARID, V\$BRNF, V\$FKHD, V\$MEF2, V\$PARF, V\$SORY	None
372-435	O\$INRE, V\$BRN5, V\$BRNF, V\$CAAT, V\$CLOX, V\$HOXC, V\$OCT1, V\$SORY	None
479-572	V\$AP2F, V\$CTCF, V\$EBOX, V\$HAND, V\$NEUR, V\$NR2F, V\$STAF	HNF4
580-619	V\$AIRE, V\$BNCF, V\$BRN5, V\$NR2F, V\$PPAR	None
727-796	V\$ARID, V\$BRNF, V\$CDXF, V\$CHRF, V\$FKHD, V\$GCMF, V\$OCT1	None
837-874	V\$CLOX & V\$HBOX	None
889-954	V\$MYT1, V\$GATA, V\$HOXF	HNF4
971-1044	V\$FKHD, V\$HAND, V\$RXRF, V\$ZFHX	IRF1
1074-1128	V\$AP2F, V\$CAAT, V\$CREB, V\$SPZ1	SREBP

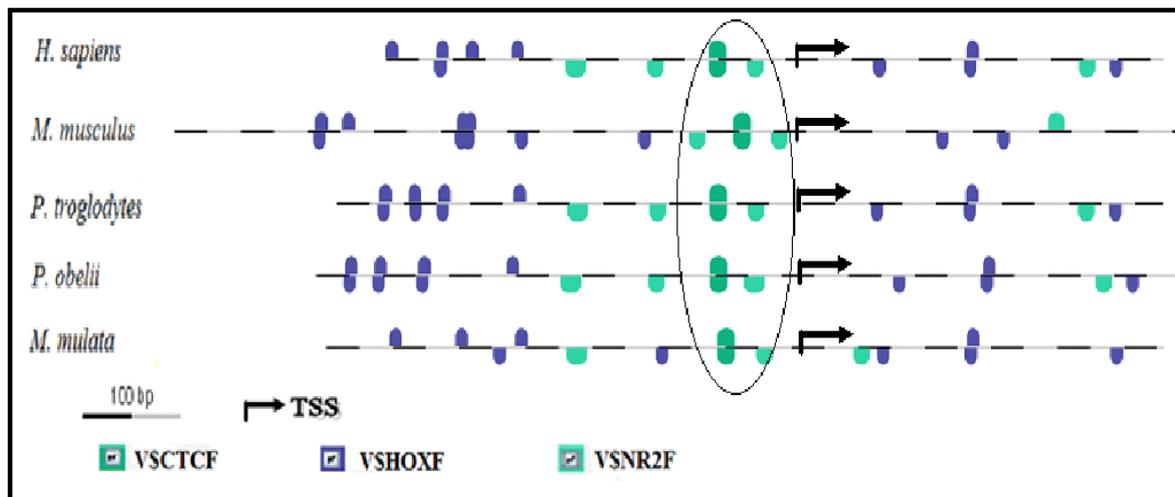


Figure 1: MatInspector graphical visualization of CTCF and NR2F in BTN1A1 genes across the five mammals. The upper panel shows the position of the TFBS in the sequences and the lower colored boxes are the catalog for the respective matrix families predicted. The arrows indicate the transcription start sites.

in the mammary gland revealed seven candidate TFBS; these are BRN5, CEBP, CTCF, ETSF, FKHD, GATA and NR2F. According to the DiAlign prediction, out of the TFs expressed in the mammary gland only four (CTCF, ETSF, GATA and NR2F) were found to lie in the conserved sequences of BTN1A1 genes across species.

However, among these TFBSs, only CTCF and NR2F were considered for further analysis because of their relatively short conserved distance from the

TSS and within themselves in the MatInspector's graphic panel of the BTN1A1 genes (Figure1). DiAlign did not find any TFBS that is conserved for the three other milk genes compared against BTN1A1 in human.

DISCUSSION

Cis- regulatory elements are usually confined upstream and/or downstream near to the TSS. Particularly, the 5' flanking region was preferred because it is described to have an array of potential

TFs important in gene regulation (Marinescu *et al.* 2005). According to Mather & Keenan (1998) this region consists of a number of binding sites for transcription factors. As stated in Ogg *et al.* (2004) BTN1A1 gene is expressed during lactation in MEC. The general argument is that promoter regions have common consensus sequences which are binding sites for potential TFs in BTN1A1 gene across species and other milk gene which can determine their periodic and tissue specific expression.

Before the MSA repeat sequences were scanned from the DNA sequences since they have skewing effect on the alignment level. In consistent with the existing fact, the MSA confirmed that the four primates are closely related to each other than to the rodent. Among the species incorporated, the primates are evolutionarily related thus, their BTN1A1 genes are aligned well; however, *Mus musculus* is an outlier species. The three other human milk genes CSN2, ADPH and LALBA were chosen for a comparison with butyrophilin, because all genes are highly expressed in mammary gland during lactation (Mather, 2000).

Prediction of TSS using McPromoter has predicted no TSS above the threshold. This might be due to the 39.5% sensitivity and 56.3% specificity of the program attempting to exactly predict the location of TSS (Ohler, 2000). The TSS is therefore retrieved from the GenBank sequence features annotation, based on the transcription start site of the mRNA. The TATA box of CSN2 and LALBA are located in a relatively short distance from the TSS though butyrophilin has no defined TATA box.

Almost every part of the input sequences were potential binding sites for TFs. In order to reduce the number of spurious TFs, consensus TFBSs were identified and examined for the expression in mammary gland. This filtered out seven TFBSs namely BRN5, CEBP, CTCF, ETSF, FKHD, GATA and NR2F. These TFBS were further analyzed to see whether they lay in the previously identified similarity regions of the MSA. Hence, CTCF, ETSF, GATA and NR2F were filtered out. In order to make our search finer, relative distances of the four TFBS from the TSS were inspected visually which indeed reduced the false candidates.

Visual inspection revealed us only CTCF and NR2F laying in relatively short distance from the TSS of BTN1A1 genes across species (figure 1). At this point, the two group i.e. the BTN1A1 genes across species and the four human milk genes lost intersection. The two TFBSs were not found in relatively similar distance from the TSS in the other milk genes which might indicate the presence of distinct pathway for the expression of BTN1A1as stated in Bauman *et al.* (2006). The binding sites for

both TFs lay between -73 and -131 upstream from the TSS which is an ideal region where *cis*-acting regulatory elements reside. Although we didn't find any information about the presence of cooperation involving CTCF and NR2F in gene regulation, the relatively short conserved distances from the TSS and within themselves will magnify their alleged role in BTN1A1 expression.

According to Li *et al.* (2004) CTCF regulates the expression of various genes important in development via chromatin insulation or enhancer blocking. It is a ubiquitously expressed vertebrate nuclear protein with numerous functions. It was first detected in chicken *c-myc* gene ~200 upstream of the TSS (Lobanenkov *et al.* 1990). More recently, the regulatory role of CTCF in the expression of insulin-like growth factor II (Igf2) has been demonstrated by Li *et al.* (2008). On the other hand, Igf2 was discovered as a mediator of prolactin (PRL) action in the mouse mammary gland during ductal branching and alveolar development in Russel *et al.* (2003) and in addition, the role of prolactin in milk production was demonstrated before two decades by Borelline and Oka (1989). This indirect correlation between CTCF and prolactin support our finding thus making it good candidate *cis*-regulatory element for BTN1A1.

Nuclear receptor subfamily 2 factors (NR2F) consist of various transcription factors important in the regulation of the expression of different genes. Some of the TFs classified under this subfamily include hepatocyte nuclear factors (HNFs), Chicken ovalbumin upstream promoter (COUP) and nuclear hormone receptor TR2 (TR2). NR2F (HNF4 in our case which was predicted by siteGA too) is found in close proximity to both CTCF binding site and TSS of BTN1A1 genes of species investigated which certainly took our attention. Thus, it is considered as a better candidate regulatory element. In addition, the participation of HNF4 together with various transcription factors such as SREBP in the regulation of fatty acid synthesis (lipogenesis) was mentioned in Laliotis *et al.* (2010). According to the findings of the current study the two TFs and their respective binding sites deserve further experimental investigation to ensure their actual regulatory role during the expression of BTN1A1 gene.

CONCLUSIONS

The spatial and temporal expression of BTN1A1 is determined by *cis*-acting element confined in the promoter region. The TFBS predictor tool revealed typical proximal promoter elements. However, TFBS predictor tools give a huge number of binding sites (false positives) and filtering TFBSs of interest from such quantities is not as easy task and needs a more detailed and broader mechanism to be closer to the goal. Although this analysis has identified

candidate *cis*-regulatory elements by means of comparative analysis, the prediction of such programs can infer only the binding probability for a site, however not the actual functionality. Their functionality can ultimately be proven by a wet-lab experiment with defined settings because the potential binding sites in a promoter might be functional in specific tissues or developmental stages and non-functional under a different circumstance. Thus, CTCF and NR2F (HNF4) are identified as candidate BTN1A1 regulatory elements; however, the exact mechanism by which these transcription factors interact with butyrophilin remains to be investigated. In order to understand the functional cooperation between these *cis*-acting elements additional experimental analysis is mandatory. Therefore, among regulatory elements that could be investigated for their anticipated regulatory role on butyrophilin gene expression, CTCF and NR2F (HNF4) can be considered as potential candidates.

ACKNOWLEDGMENTS

We would like to thank Ministry of Education, Ethiopia for funding this research work.

REFERENCES

Abiy, Z., Ashwell, M., Ogg, S., Rexroad, C. and Mather, I. H. (1999). RFLP markers in the bovine butyrophilin gene. *Animal Genetics* 30:382-405.

Bauman, D. E., Mather, I. H., Wall, R. J., and Lock, A. L. (2006). Major Advances Associated with the Biosynthesis of Milk. *Journal of Dairy Science* 89: 1235-1243.

Borellini, F. and Oka T. (1989). Growth control and differentiation in mammary epithelial cells. *Environmental health perspectives* 80: 85-99.

Cartharius, K., Frech K., Grote K., Klocke B., Haltmeier M., Klingenhoff, A., Frinch M., Bayerlein M. and werner T. (2005). MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* 21(13): 2933-2942.

Franke, W.W., Heid, H.W., Grund, C., Winter, S., Freudenstein, C., Schmid, E., Jarasch, E.D., Keenan, T.W. (1981). Antibodies to the major insoluble milk fat globule membrane-associated protein: specific location in apical regions of lactating epithelial cells. *Journal of Cell Biology* 89:485-494.

Kohany, O., Gentles, A., Hankus, L., and Jurka, J. (2006). Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 7: 474.

Laliotis, G.P., Bizelis I. and Rogdakis E. (2010). Comparative Approach of the de novo Fatty Acid Synthesis (Lipogenesis) between Ruminant and Non Ruminant Mammalian Species: From Biochemical Level to the Main Regulatory Lipogenic Genes. *Current Genomics* 11:168-183.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson,

T.J. and Higgins, D.G. (2007). Clustal, W and Clustal X version 2.0. *Bioinformatics* 23(21): 2947-2948.

Lemay D. G., Neville M. C., Rudolph M. C., Pollard K. S. and German J. B. (2007). Gene regulatory networks in lactation: identification of global principles using bioinformatics. *BMC Systems Biology* 1:56.

Li, T., Hu, J., Qiu, X., Ling, J., Chen, H., Wang, S., Hou, A., Vu, T., and Hoffman, A. (2008). CTCF Regulates Allelic Expression of *Igf2* by Orchestrating a Promoter-Polycomb Repressive Complex 2 Intrachromosomal Loop. *Molecular and cellular biology* 28(20): 6473-82.

Li, T., Lu, Z. and Lus, L. (2004). Regulation of eye development by transcription control of CCCTC binding factor. *Journal of BioChemistry* 279:27575-83.

Lobanenkov V. V., Nicolas R. H., Adler V. V., Paterson H., Klenova E. M., Polotskaja A. V. and Goodwin G. H. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene* 5:1743-1753.

Marinescu, V. D., Kohane, I.S and Riva, A. (2005). MAPPER: a search engine for the computational identification of putative transcription factor binding sites in multiple genomes. *BMC Bioinformatics* 6:79.

Mather, I.H. and Keenan, T.W. (1998). Origin and secretion of milk lipids. *Journal of Mammary Gland Biology Neoplasia* 3:259-273.

Mather, I. (2000). A review and proposed nomenclature for major proteins of the milk-fat globule membrane. *Journal of Dairy Science*, 83: 203-247.

McManaman J.L., Russell T. D., Schaack J., Orlicky D. J. & Robenek H. (2007). Molecular Determinants of Milk Lipid Secretion. *Journal of Mammary Gland Biology Neoplasia* 12:259-268.

Ogg, S., Weldon, A., Dobbie, L., Smith, A., and Mather, I.H. (2004). Expression of butyrophilin (Bt1a1) in lactating mammary gland is essential for the regulated secretion of milk-lipid droplets. *Proceedings of the National Academy of Sciences* 101(27):10084-10089.

Ohler, U. (2000). Promoter Prediction on a Genomic Scale. The Adh Experience. *Genome Research* 10: 539-542.

Robenek H., Hofnagel O., Buers I., Lorkowski S., Schnoor M., Robenek M.J., Heid H., Troyer D., Severs N.J. (2006). Butyrophilin controls milk fat globule secretion. *Proceedings of the National Academy of Sciences* 103: 10385-10390.

Russell C. H., Harris J., Hadsell D. L., Lee A. V., Ormandy C. J., and Vonderhaar B. K. (2003). Local Insulin-Like Growth FactorII Mediates Prolactin Induced Mammary Gland Development. *Molecular Endocrinology* 17(3): 460-471.

Smith, I. A., Knezevic B. R., Ammann J. U., Rhodes D. A., Aw D., Palmer D. B., Mather I. H. and Trowsdale J. (2010). BTN1A1, the Mammary Gland Butyrophilin, and BTN2A2 Are Both Inhibitors of T Cell Activation. *The Journal of immunology* 184: 3514-3525.

Zhao, F., Xuan, Z., Liu, L. and Zhang, M. (2005). TRED: a Transcriptional Regulatory Element Database and a platform for *in silico* gene regulation studies. *Nucleic Acids Research* 33: D103-D107.