Novel research approaches in lactation biology

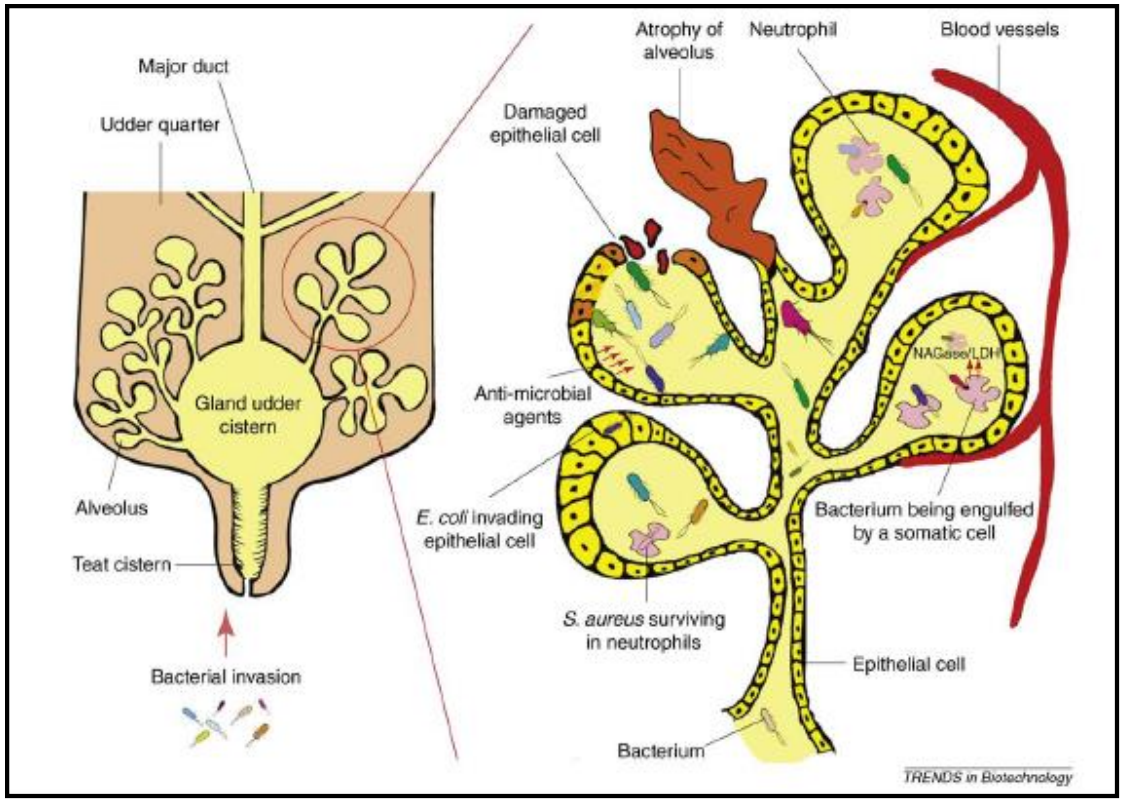
Peter DOVČ

University of Ljubljana, Biotechnical Faculty,
Department of Animal Science

Content:

- Problem definition
- *In silico* procedures
- Sequence analysis
- Comparative functional analysis
- Integration of different types of data
- Gene onthology
- Data interpretation
- Future perspectives

Lactation and mastitis



Vigieur et al., 2009

Problem definition

- Phenotypic data are heterogeneous and incomplete
- Genomic data are limited to several species, for the majority of species on one “model” sequence is available
- Haplotype data is seldom available
- Causal loci are not identified, the biochemical and physiological understanding of complex traits is still poor

Sequence comparison between species illustrate common evolutionary history

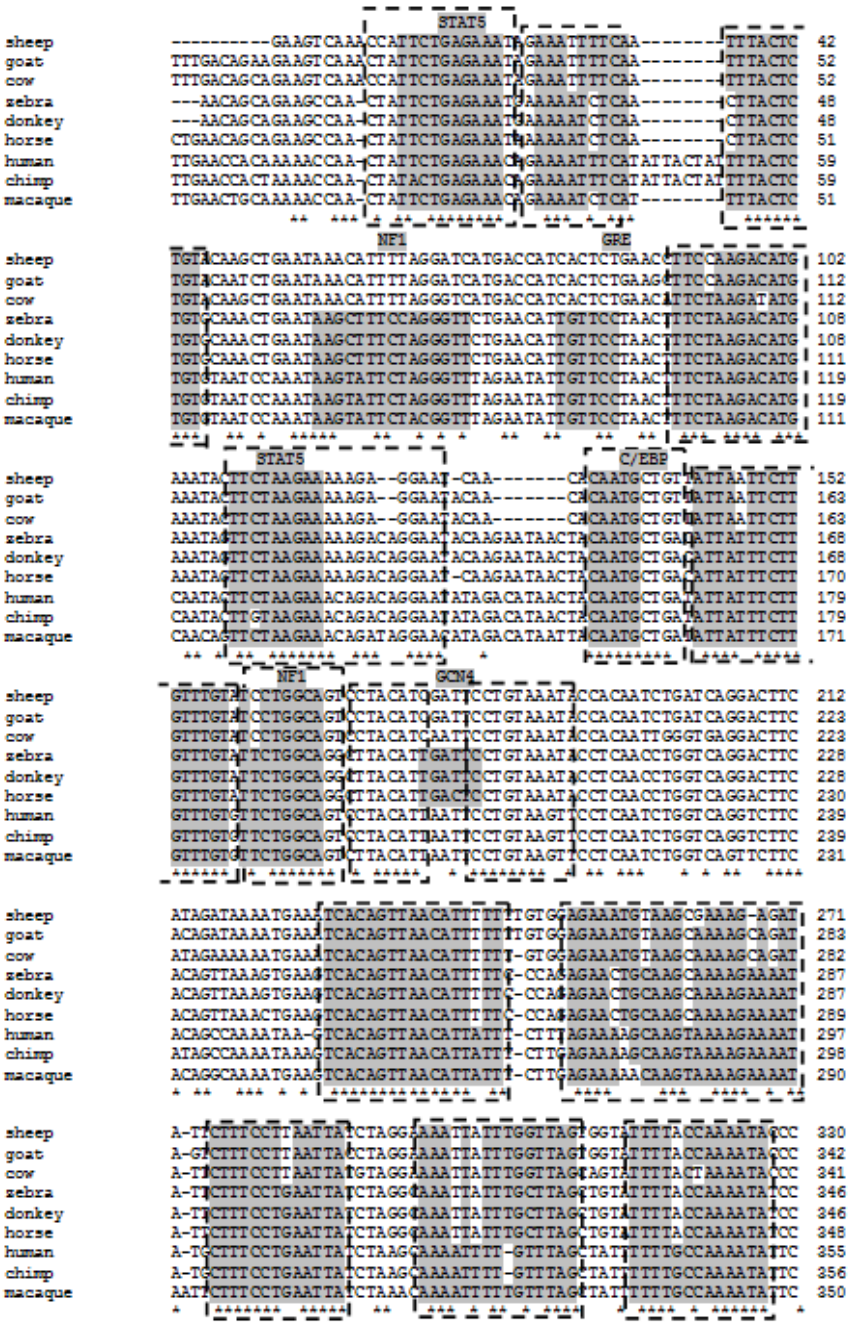
```
                        c.205
donkey    ACCAACATAGACTAGCTGTACTAATTAATAATCAACATATGCCTTACCAATATTATGCAA
zebra     ACCAACATAGACTAGCTGTACTAATTAATAATCAACATATGCCTTACCAATATTATGCAA
horse     ACCAACATAGACTAGCTTTACTAATTAATAATCAACATATGCCTTACCAATATTATGCAA 247
          ***************** ****************************************
                H  R  L  A L>V L  I  N  N  Q  H  M  P  Y  Q  Y  Y  A

donkey    GACCAGCTGCAGTTAGGCCACATGTCCAAATTCCTCAATGGCAAGTCCTGCCAAATATCT
zebra     GACCAGCTGCAGTTAGGCCACATGTCCAAATTCCTCAATGGCAAGTCCTGCCAAATATCT
horse     GACCAGCTGCAGTTAGGCCACATGTCCAAATTCCTCAATGGCAAGTCCTGCCAAATATCT 307
          ************************************************************
          R  P  A  A  V  R  P  H  V  Q  I  P  Q  W  Q  V  L  P  N  I
                                  c.334              c.354C>T
donkey    ACCCATCCACTGTGGTACGTCATCCACGCCCGCATCCATCATTTATCGCCATTCCTCCAA
zebra     ACCCATCCACTGTGGTACGTCATCCACGCCCGCATCCATCATTTATTGCCATTCCTCCAA
horse     ACCCATCCACTGTGGTACGTCATCCATGCCCGCATCCATCATTTATTGCCATTCCTCCAA 367
          ************************** ******************* *************
          Y  P  S  T  V  V  R  H  P C>R P  H  P  S  F  I  A  I  P  P
                     c.383A>T
donkey    AGAAACTTCAGGAGAAAACAGTCATCCCTAAGATCAATACTATTGCTACTGTTGAGCCTA
zebra     AGAAACTTCAGGAGAAAACAGTCATCCCTAAGATCAATACTATTGCTACTGTTGAGCCTA
horse     AGAAACTTCAGGAGATAACAGTCATCCCTAAGATCAATACTATTGCTACTGTTGAGCCTA 427
          *************** ******************************************
          K  K  L  Q  E I>K T  V  I  P  K  I  N  T  I  A  T  V  E  P
            c.433G>A                                     c.481G>T
donkey    CACCTATTCCTACCCCTGAACCAACAGTGAACAATGCAGTCATTCCAGACGCCTCCTCAG
zebra     CACCTGTTCCTACCCCTGAACCAACAGTGAACAATGCAGTCATTCCAGACGCCGCCTCAG
horse     CACCTATTCCTACCCCTGAACCAACAGTGAACAATGCAGTCATTCCAGACGCCTCCTCAG 487
          ***** ************************************************ ******
          T  P I>V P  T  P  E  P  T  V  N  N  A  V  I  P  D  A S>A  S
                                   c.517A>G c.525
donkey    AATTCATCATCGCCAGCACACCTGAGACTACCACAGTTCCAGTTACTTCAC---------
zebra     AATTCATCATCGCCAGCACACCTGAGACTACCACAGTTCCAGTTACTTCACCG-------
horse     AATTCATCATCGCCAGCACACCTGAGACTGCCACAGTCCCAGTTACTTCACCTGTGGTCC 547
          ****************************** ******* *************
          E  F  I  I  A  S  T  P  E  T T>A T  V  P  V  T  S  P  V  V
```

Sequence comparison of **Exon 4** of **CSN3** gene from donkey, zebra and horse. Numbering is based on horse sequence (AY040863). The chymosin-sensitive region is underlined.

Hobor et al., 2008

Promoter regions of homologous genes share conserved sequence regions



Sequence comparison of promoter regions of the CSN3 Gene.

Hobor et al., 2008

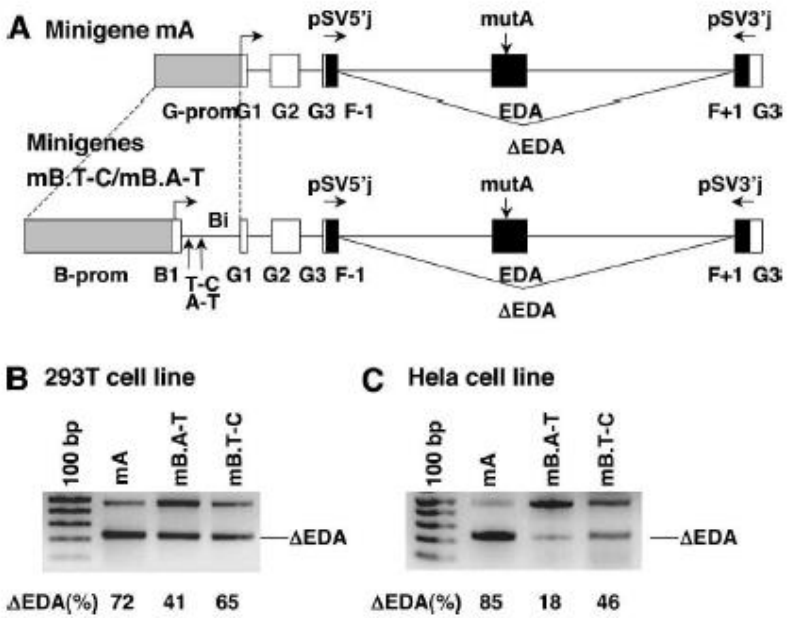
Conserved character of the regulatory mechanisms

Position of putative TFBS[1]	Transcription factor	Conservation of putative TFBS
−1115 to −1107	STAT5	Ruminants/Equidae/Primates
−1071 to −1057	NF1	Equidae/Primates
−1048 to −1043	GRE	Equidae/Primates
−1019 to −1011	STAT5	Ruminants/Equidae/Primates
−986 to −982	C/EBP	Ruminants/Equidae/Primates
−959 to −952	NF1	Ruminants/Equidae/Primates
−942 to −937	GCN4	Equidae
−784 to −779	C/EBP	Ruminants/Equidae
−741 to −736	GRE	Ruminants/Equidae
−739 to −732	STAT6	Ruminants/Equidae
−725 to −721	C/EBP	Ruminants/Equidae/Primates
−714 to −710	C/EBP	Equidae/Primates
−709 to −705	C/EBP	Equidae/Primates
−429 to −422	STAT6 partial	Ruminants/Equidae/Primates
−340 to −336	TBP	Equidae
−325 to −319	YY1	Equidae
−316 to −312	C/EBP	Equidae
−253 to −244	C/EBP beta	Equidae
−241 to −236	TFIID	Ruminants/Equidae/Primates
−153 to −145	SP1	Equidae
−84 to −78	YY1	Equidae

[1]Positions are numbered relative to the transcription start site.

Conserved TFBSs in the promotor region of the CSN3 gene in nine mammalian species.

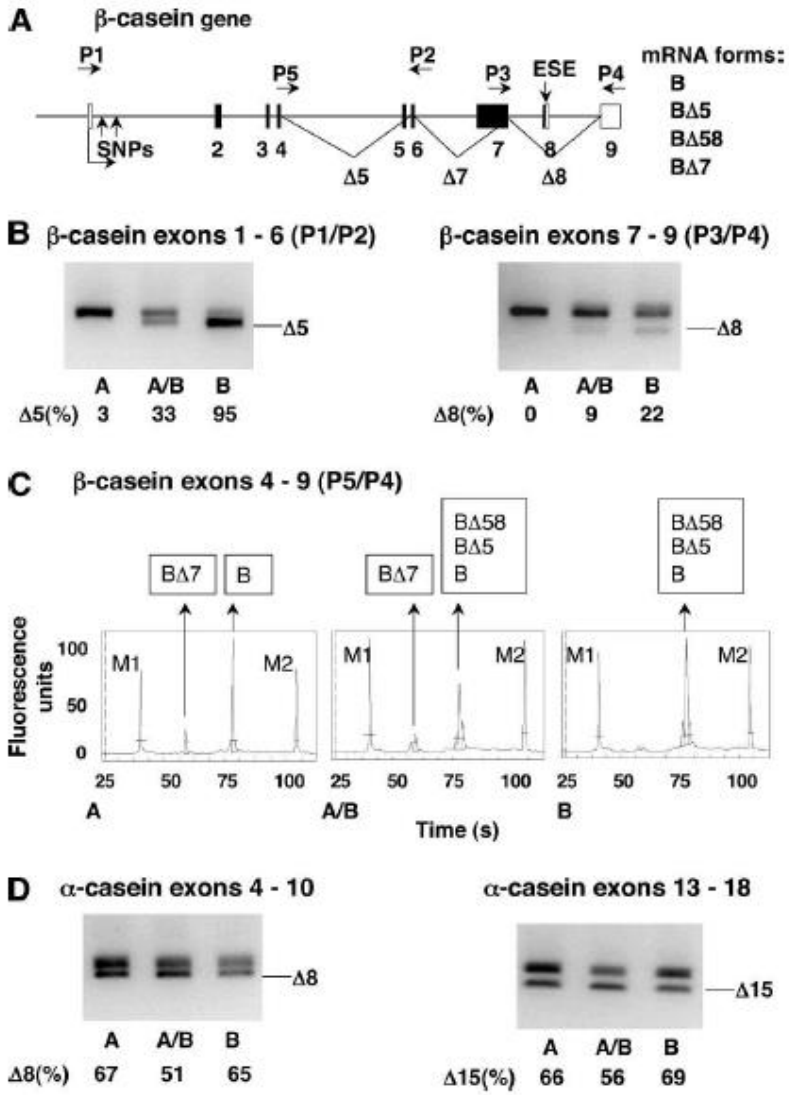
Splicing signals are not unique for each species



Effect of the ISE1 in b-casein intron 1 in concert with the b-casein promoter on alternative splicing of pre-mRNA from pSVEDA Tot minigene. (A) Partial structure of hybrid minigenes (minigenes mA, mB.T-C, and mB.A-T) depicted to scale. Promoters, a-globin (G) + b-casein (B) exons and fibronectin exons (F-1, EDA, F+1) are indicated in gray, white, and black boxes, respectively. In all constructs, ESE (element A) in the EDA exon was mutated (mutA) (vertical arrow). In the construct with the a-globin promoter (mA), the promoter was replaced by the b-casein promoter, exon 1, and part of intron 1 (Bi) (dashed lines). (B) 293T and (C) HeLa cells were transfected with the constructs. Upper fragments represent forms with the included EDA exon; lower fragments represent forms with the skipped exon and are indicated with DEDA. The numbers below each lane are the percentage of skipping of the EDA exon.

Lenasi et al., RNA (2006)

Function of the RNA splicing mechanism across the species

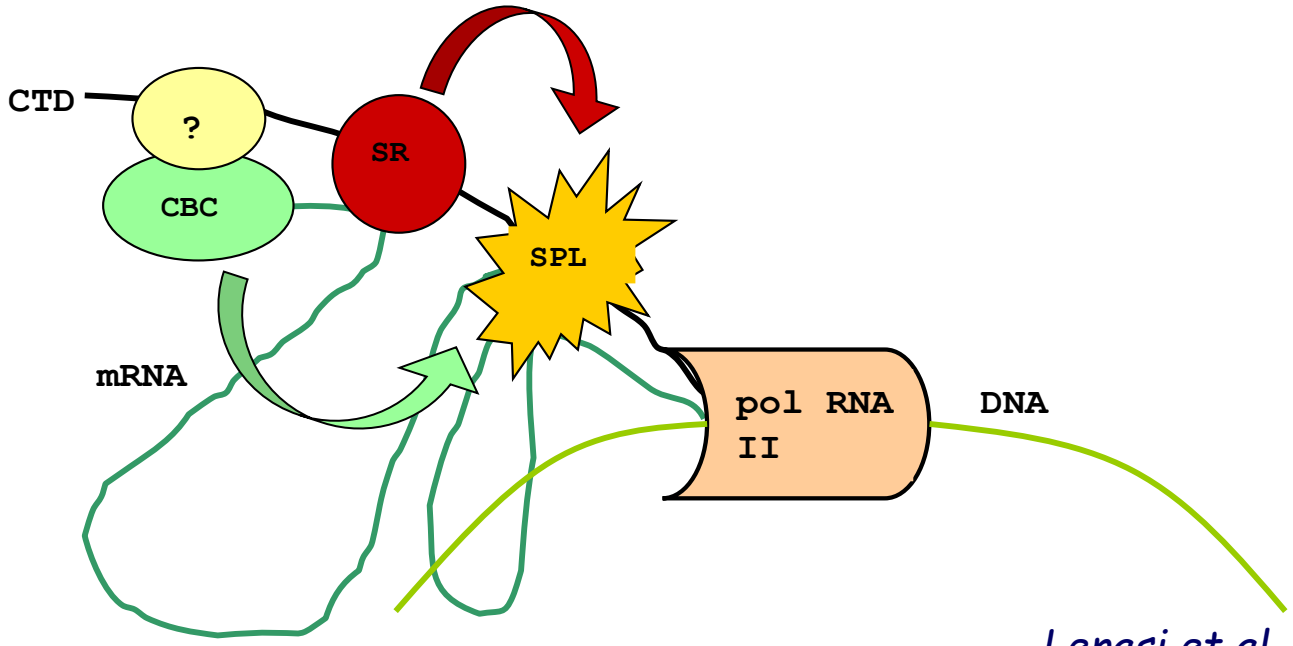


Splicing patterns of equine b- and aS1-casein mRNA. (A) Structure of b-casein gene depicted to scale; for intronic regions, the length of bovine introns is considered, and noncoding and coding exons are indicated in white and black boxes, respectively. Alternative splicing of exons 5, 8, and partially 7 is marked with lines. Alternative splicing results in four mRNA forms (right), the form with all exons is marked with B, other forms are marked with Bdelta, followed by numbers of skipped exons. Arrows above the gene indicate primers P1–P5. Two SNPs in intron 1 and ESE in exon 8 are indicated by arrows. (B,C,D) RNA was isolated from mare's milk; splicing variants were detected by RT–PCR and analyzed by electrophoresis. A, A/B, and B are representative samples of three b-casein splicing patterns. (B,D) Alternative inclusion of one weak exon at a time; upper fragments represent forms with included exon; lower fragments represent forms with skipped exon and are indicated with D followed by the exon number. The numbers below each lane are the percentage of skipping of the weak exon obtained by capillary electrophoresis of PCR products (mean value of analyzed samples). (B) Alternative inclusion of the b-casein exons 5 (left, primers P1/P2) and 8 (right, primers P3/P4). (D) Alternative inclusion of the aS1- casein exons 8 (left) and 15 (right). (C) Alternative inclusion of the bcasein exons 5, 8, and partially 7 (primers P5/P4); analysis of fragments by capillary electrophoresis. Peaks marked with M1 and M2 represent size markers (100 and 10,000 bp). Analyzed fragments representing four mRNA forms (detection limit was set above six fluorescent units) are indicated in frames above the peaks.

Proposed model of co-transcriptional splicing of β-casein mRNA

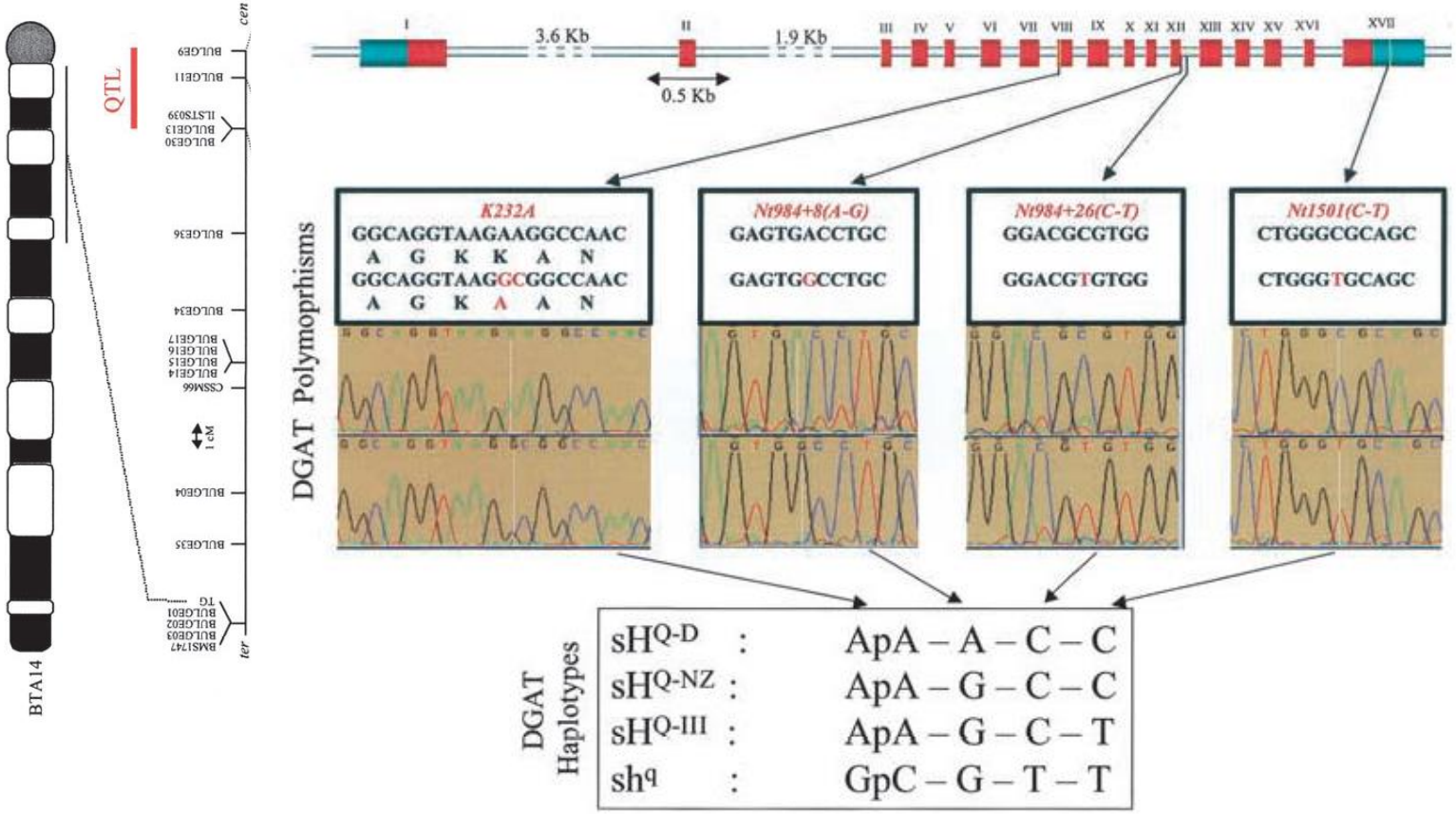
Polymorphism in the intron 1 is part of the nucleotide sequence, typical for ESE element

GAAGAA → GAUGAA ; AAGAAAAU → AUGAAAAU



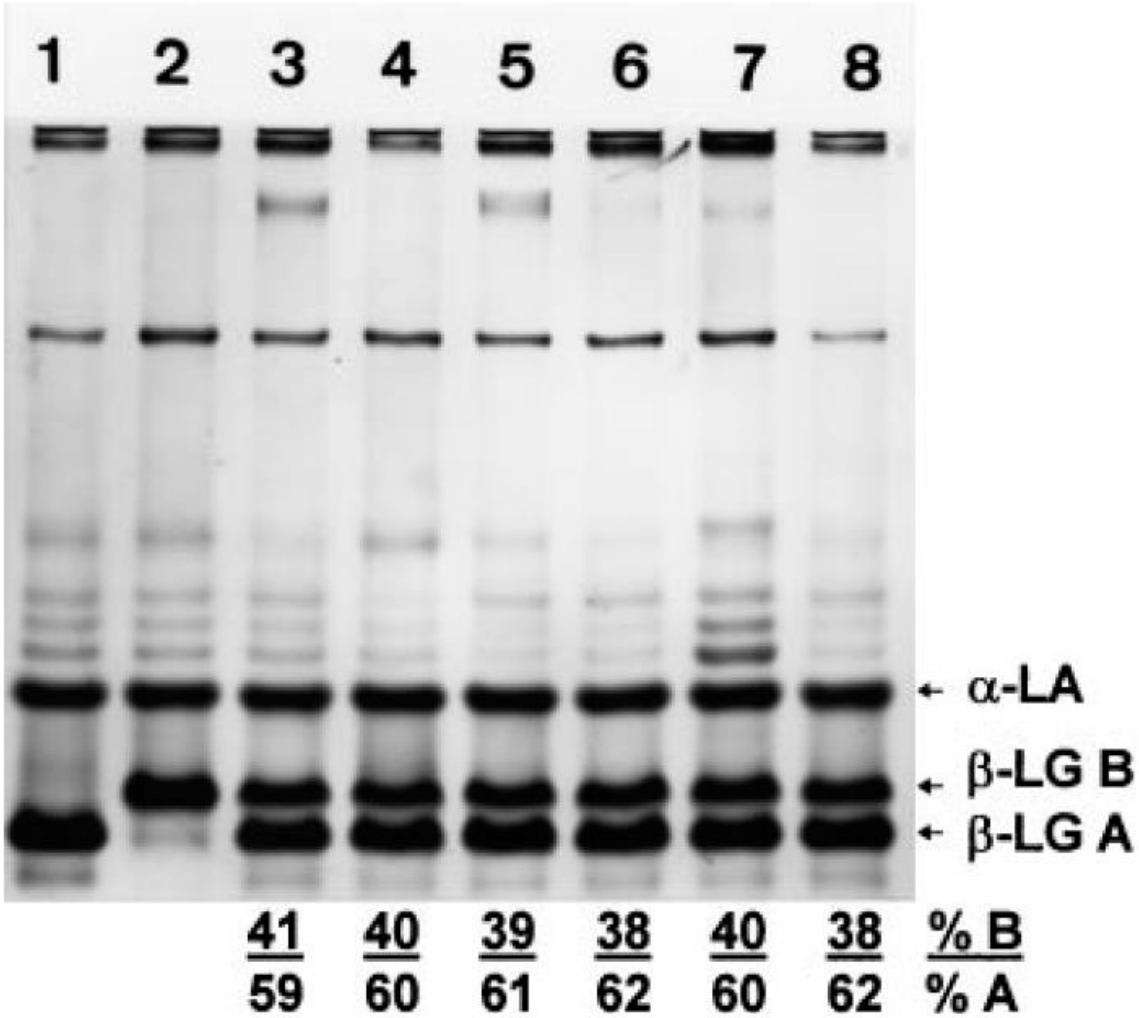
Lenasi et al., RNA (2006)

DGAT1 was the first milk trait QTL confirmed at molecular level



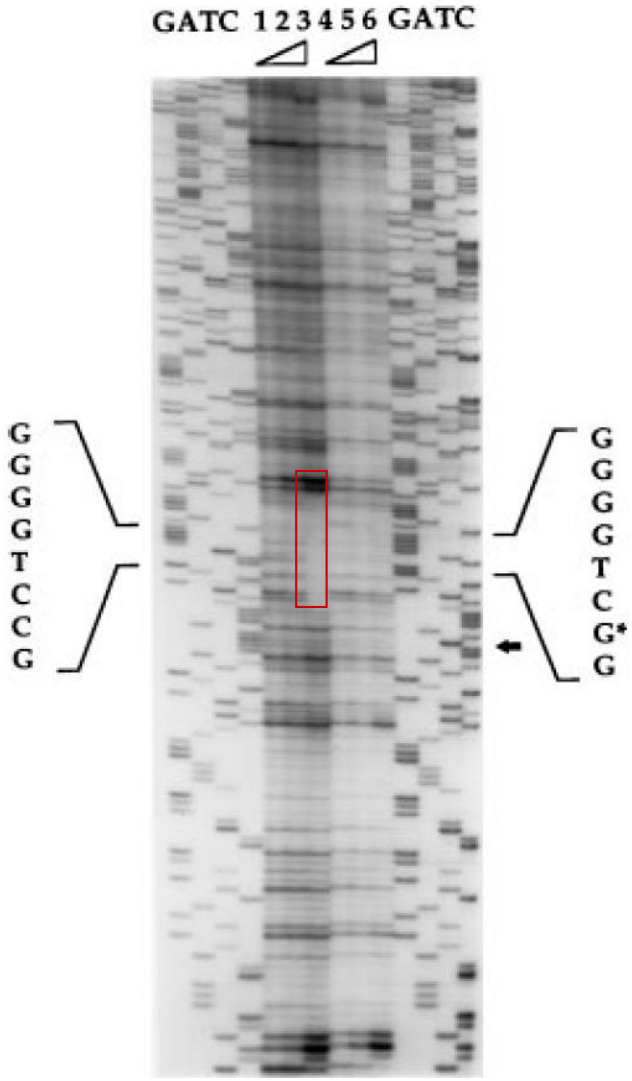
Grisart et al., 2002

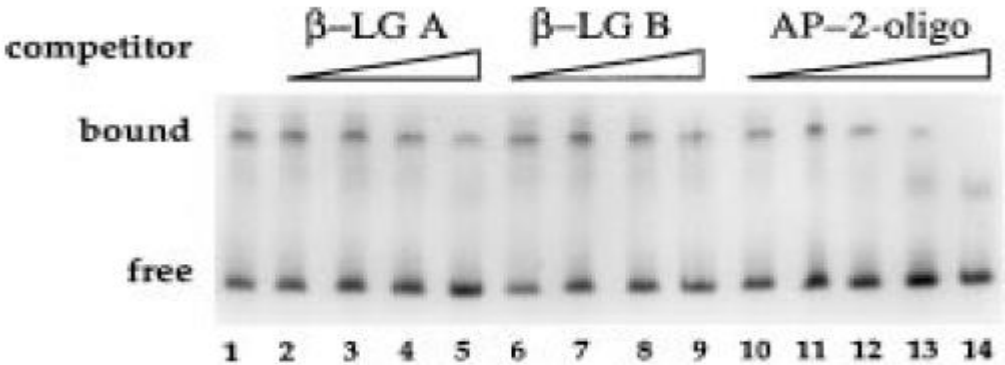
AP-2 story…

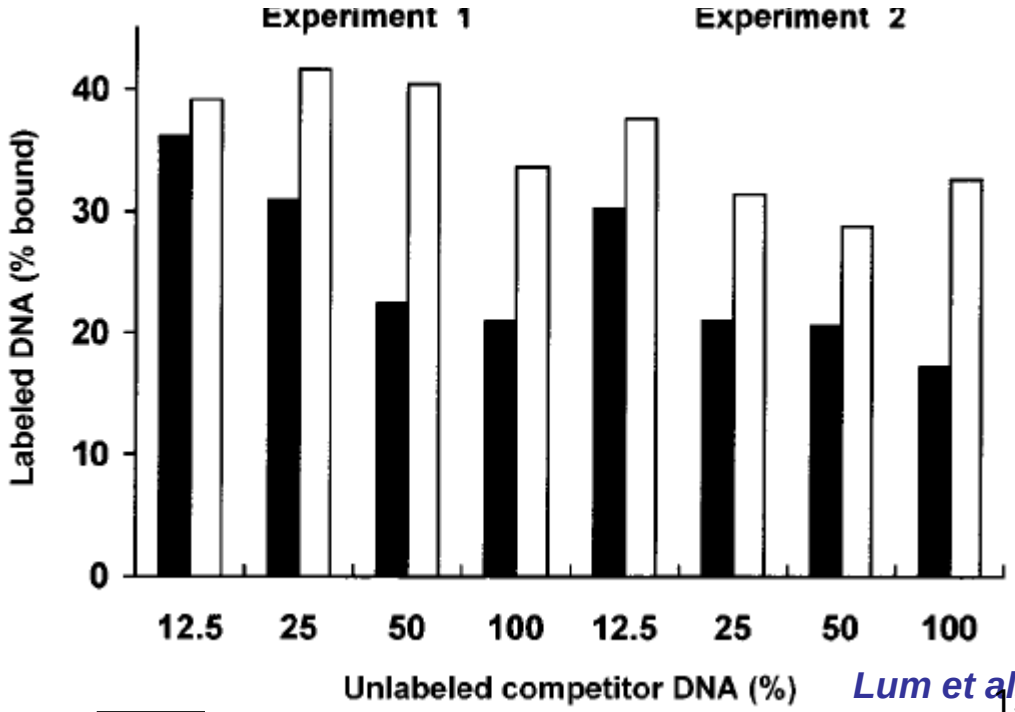


Lum et al., 1997

C to G transition in the β-LG promoter increases AP-2 binding *in vitro*

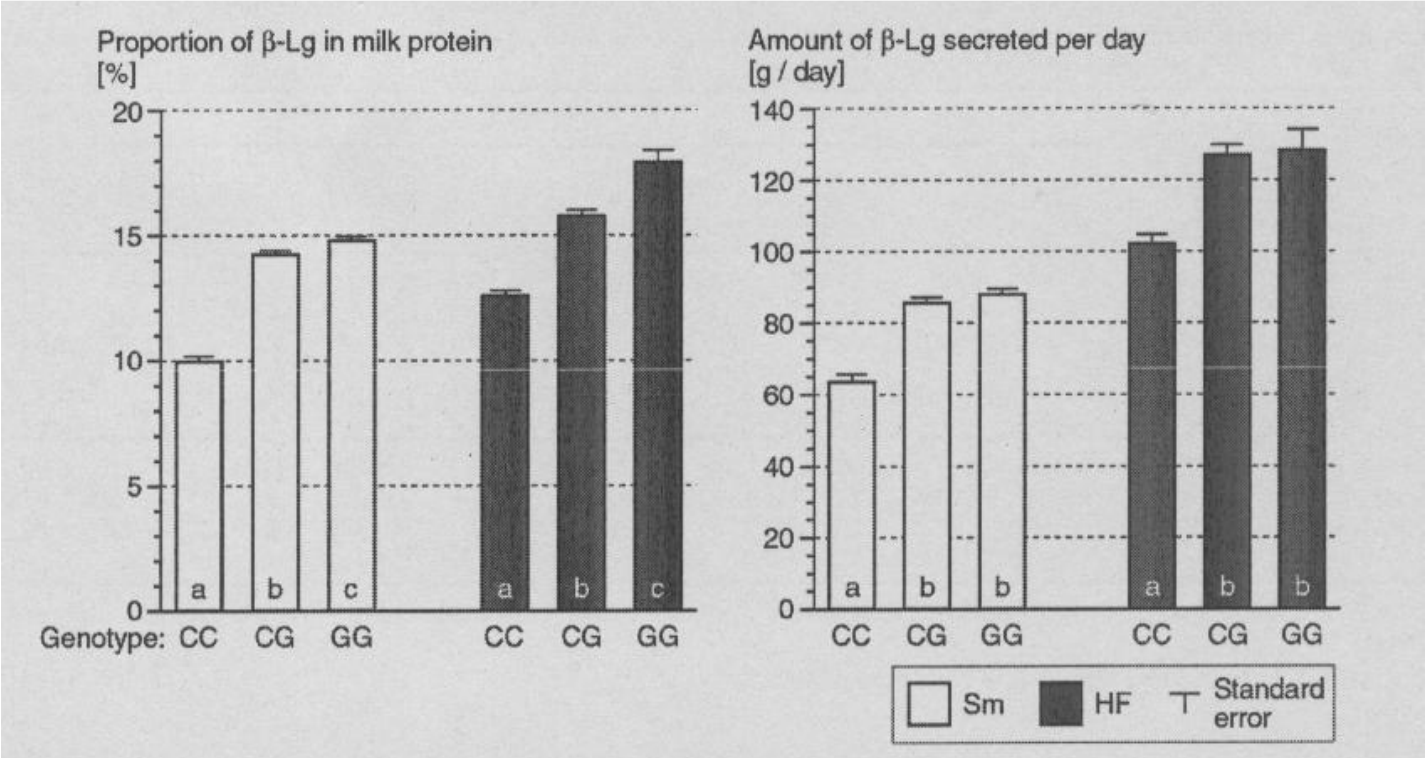






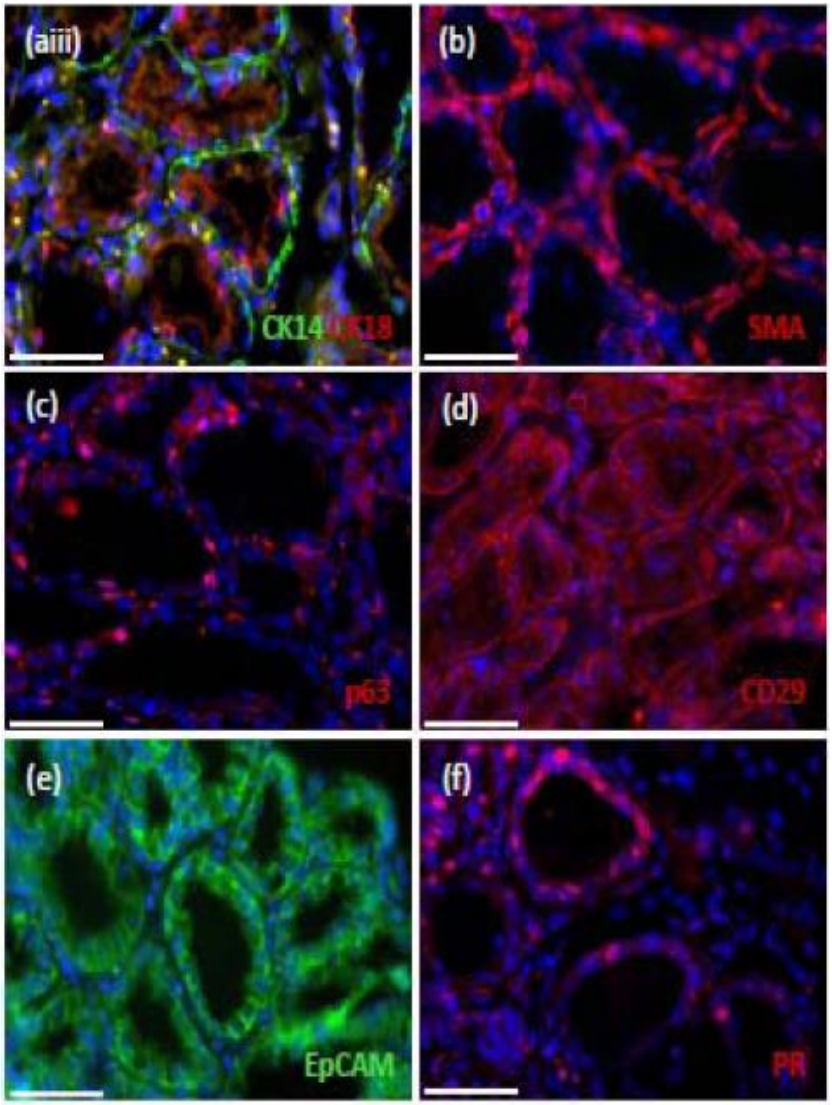
Lum et al. 1997

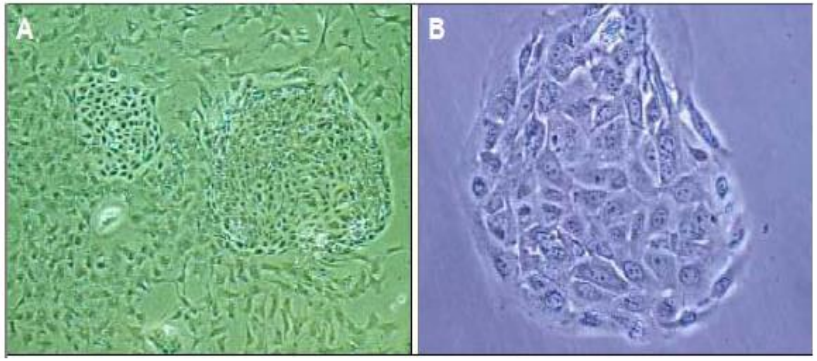
C to G transition increases proportion of β-LG in milk



Kuss et al., 2003

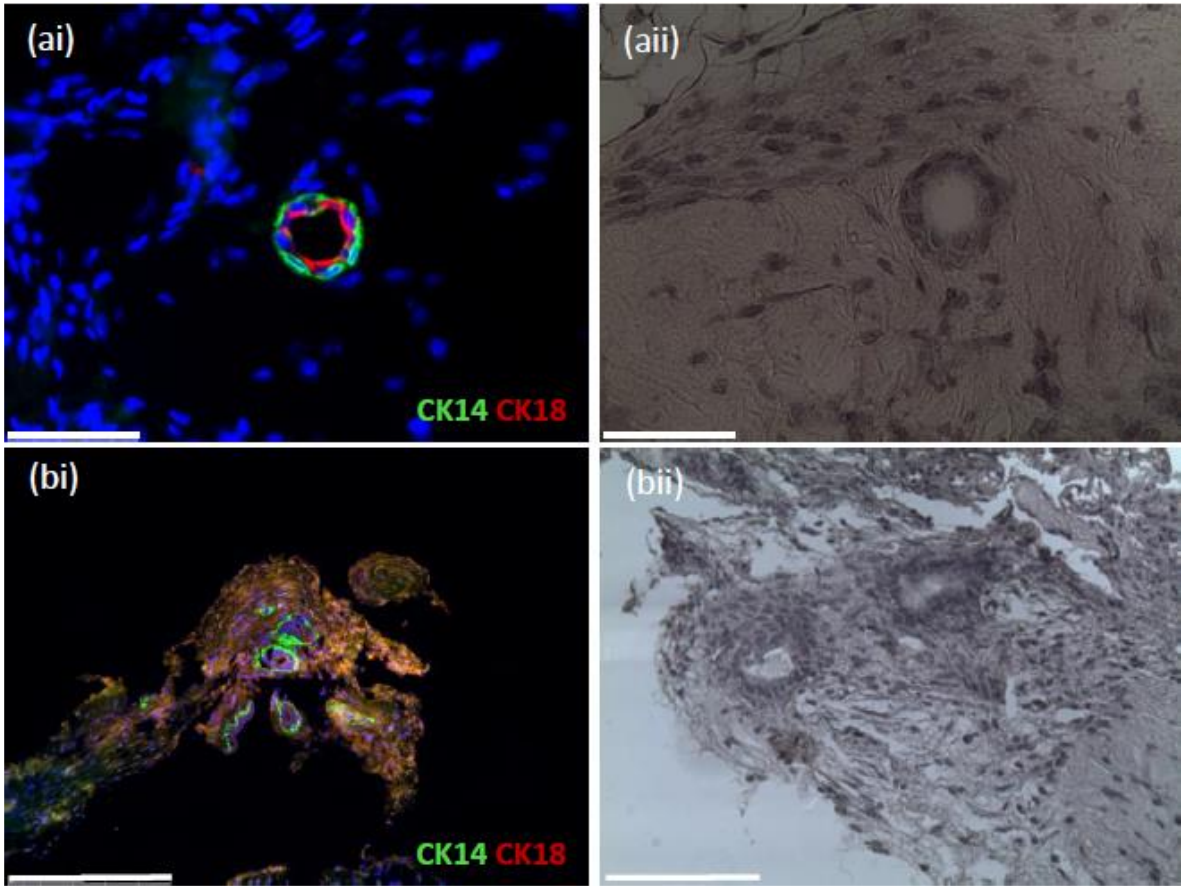
gMEC model for mammary gland function





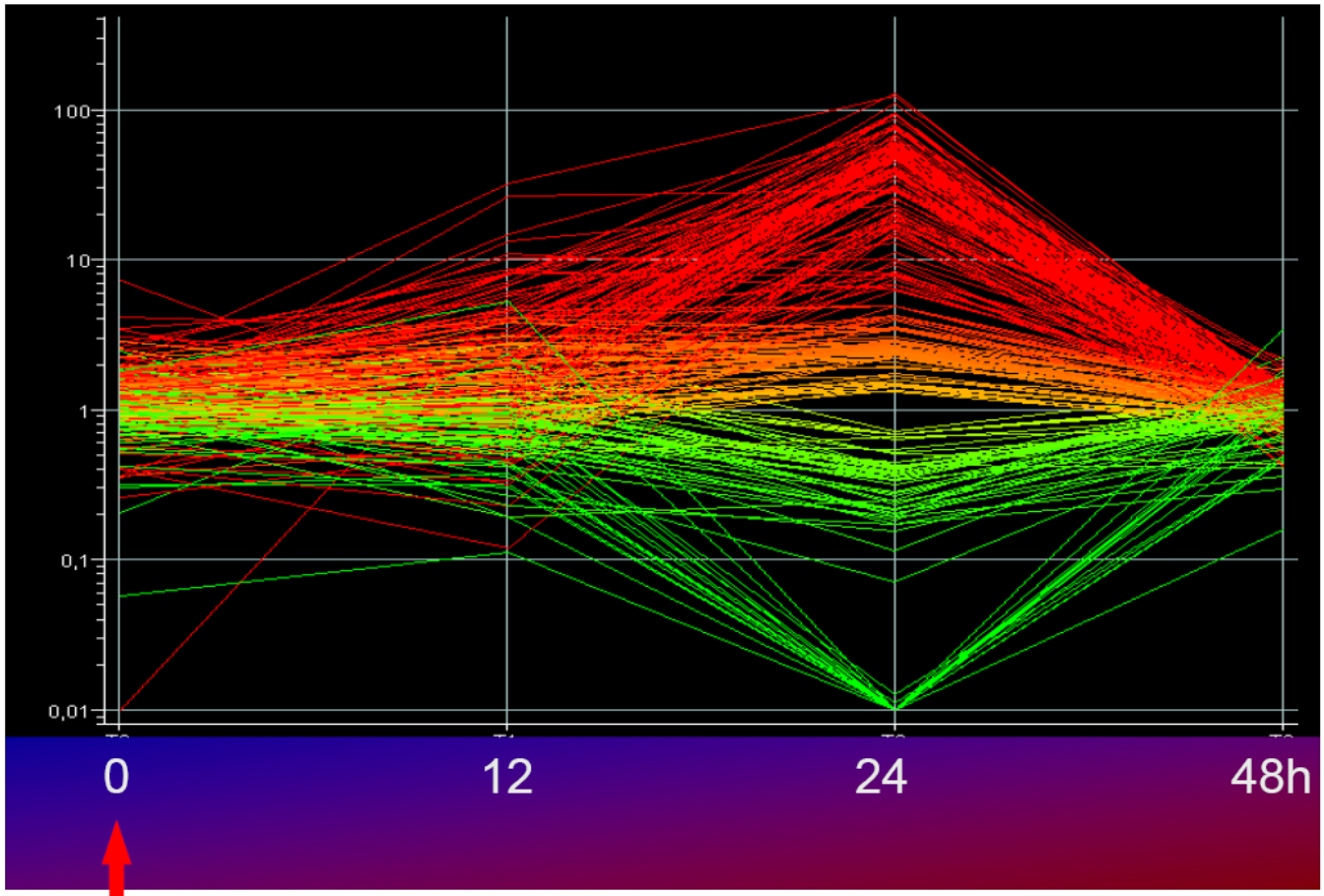
- Cytokeratin 14/18, α-smooth muscle actin, CD29 and p63 staining of goat MEC. Luminal cells were marked by antibodies against cytokeratin 18 and progesterone receptor. Epithelial cell adhesion molecule (EpCAM) staining was also located in the luminal compartment.
- Morphology and proliferative potential of primary mammary gland cells. Epithelial cells build islands surrounded by myo-epithelial cells (A). Morphology of epithelial cell colony resembles alveolar structures in the mammary gland (B).

Regenerative potential of gMEC fraction shows stem cell capacity



Expression and spatial localization of immunofluorescently detected markers in regenerated structures. Cytokeratin 14 and 18 expression in xenografts (ai, bi) and haematoxylin and eosin (H&E) staining of xenograft sections (aii, bii) after 5 weeks.

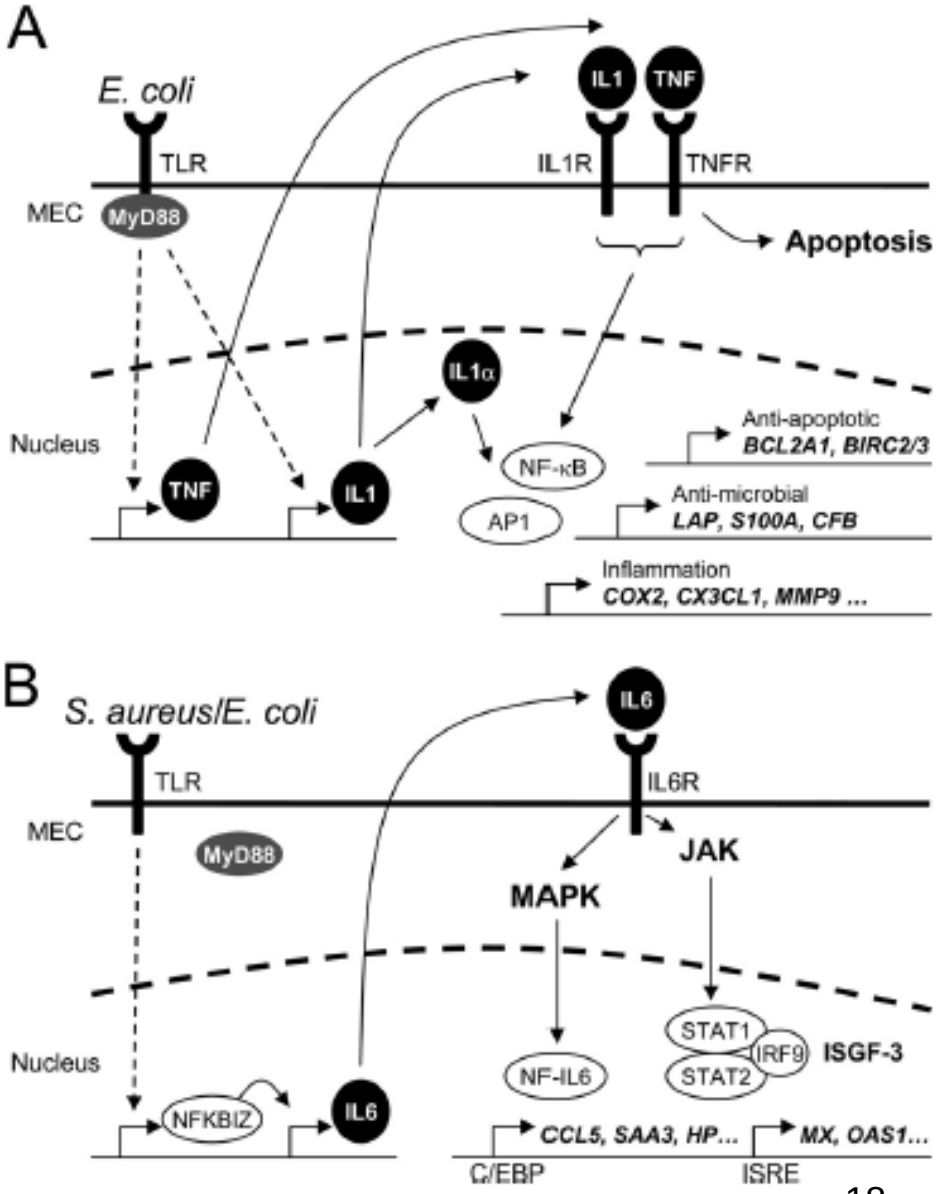
Gene expression of 400 genes in the goat mammary gland after infection with *S. aureus.*



infection

Pascal Rainard, 2009

Mastitis and immune function of the mammary gland



(A) Proinflammatory master cytokines IL-1 and IL-1 alpha, TNF. Activation of NF-B and AP1, which induces the expression of the secondary response genes (e.g., BCL2A1, BIRC1/2, LAP, S100A8/9, CFB, COX2, CX3CL1, and MMP9).

(B) Early activation of NFKBIZ - TLR-mediated induction of IL-6 expression. JAK cascade leads to the formation of an active ISGF3 complex, which may induce the expression of genes harboring ISRE sites in their promoters. MAPKs activated through IL-6 signaling induce the activation of NF-IL-6 - transcription factor C/EBP.

Günther et al., 2011

Integration of data from micro array experiments with *E. coli* and *S. aureus* in the *bovine mammary database* (BMD)

E.coli (RI, RIBFA)

(789 differentially expressed genes)

S. aureus (RI, RIBFA)

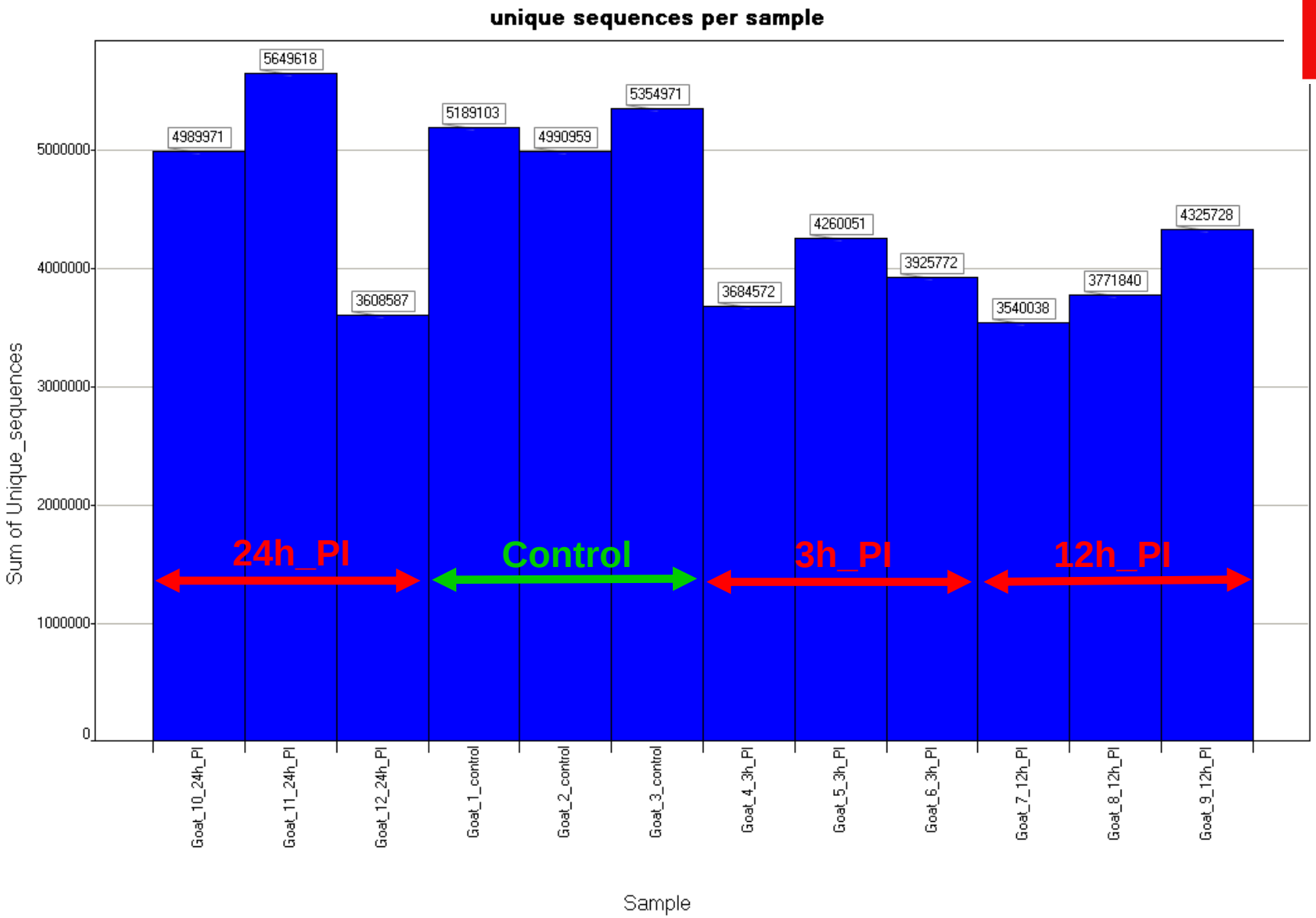
(270 differentially expressed genes)

(*H-M- Seyfert, RIBFA*)

Genes which have been identified by both approaches

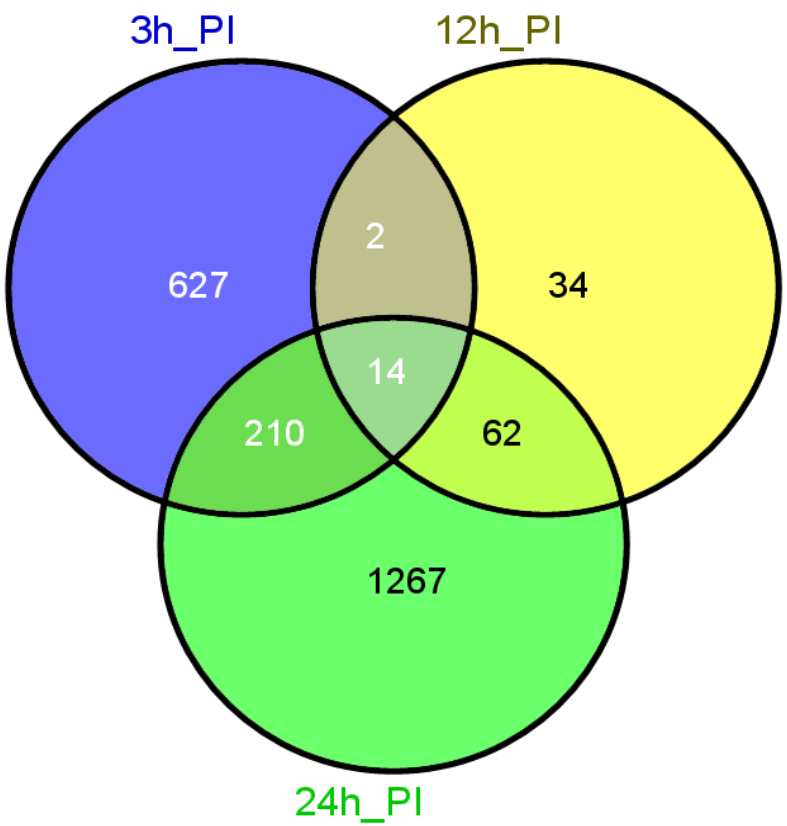
- ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2
- XDH	xanthine dehydrogenase
- TNF	tumor necrosis factor
+ LTF	lactoferrin
+ LBP	lipopolysaccharide binding protein
+ IL1B	interleukin 1 beta
+ IL6	interleukin 6
+ IL8	interleukin 8
+ CCL2	chemokine (C-C motif) ligand 20
+ CD14	CD14 antigen
+ B4GALT1	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 1

RNA_Seq Analysis of MEC

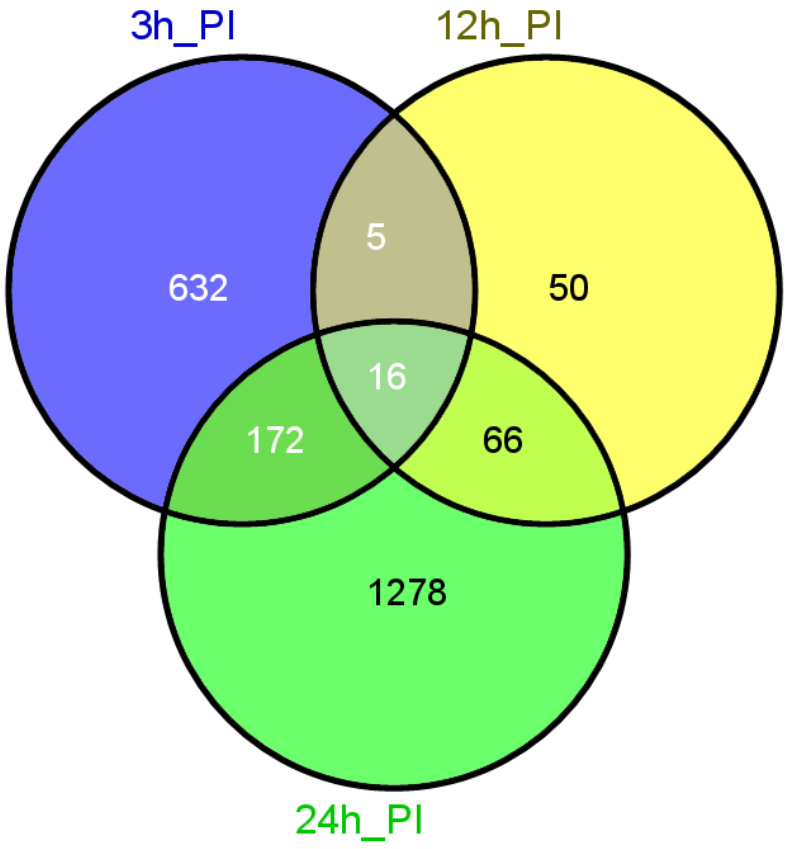


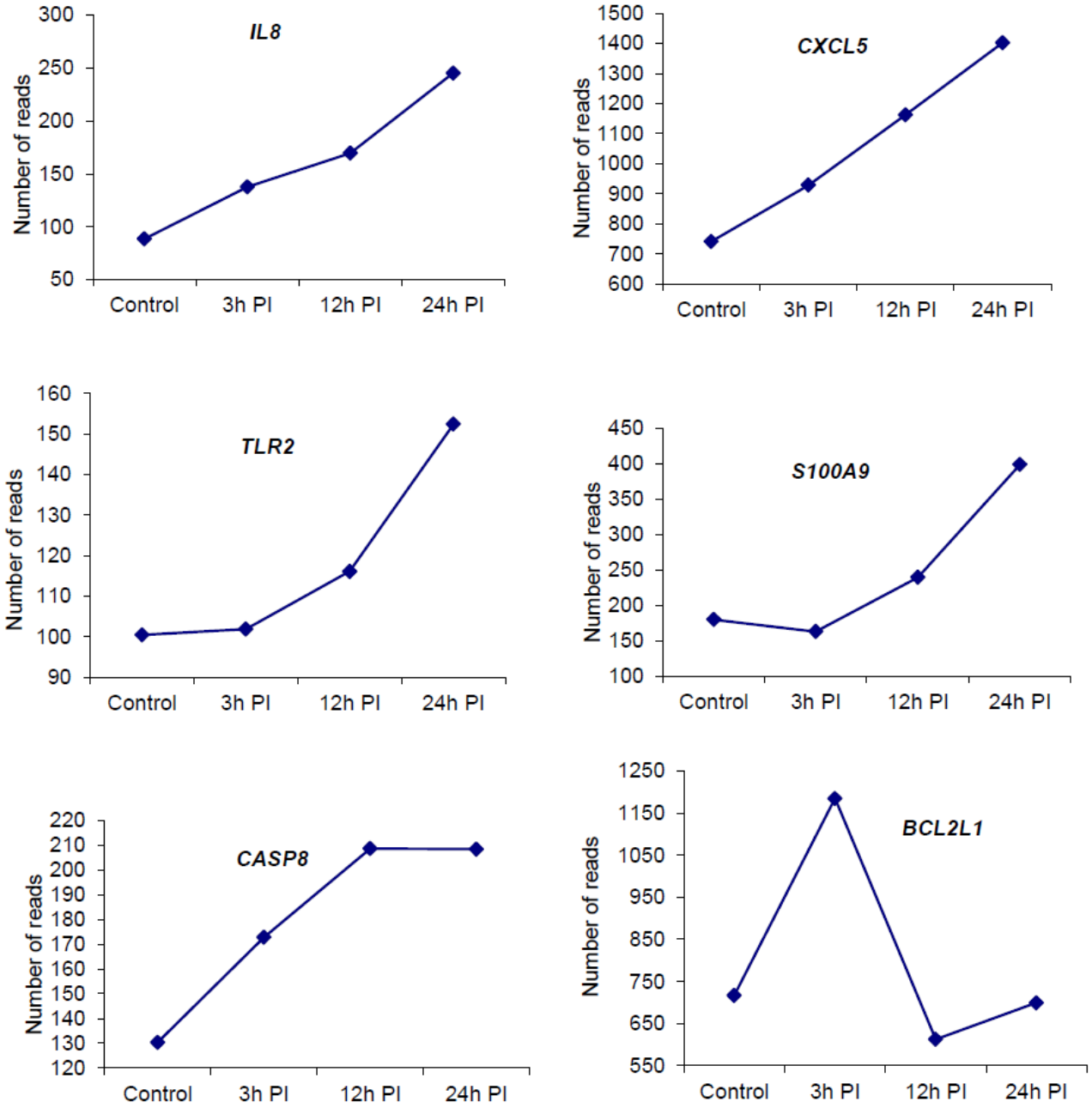
DEG after infection with *Mycoplasma agalactiae*

"**NCBI** *Bos Taurus* RefSeq RNA sequences"



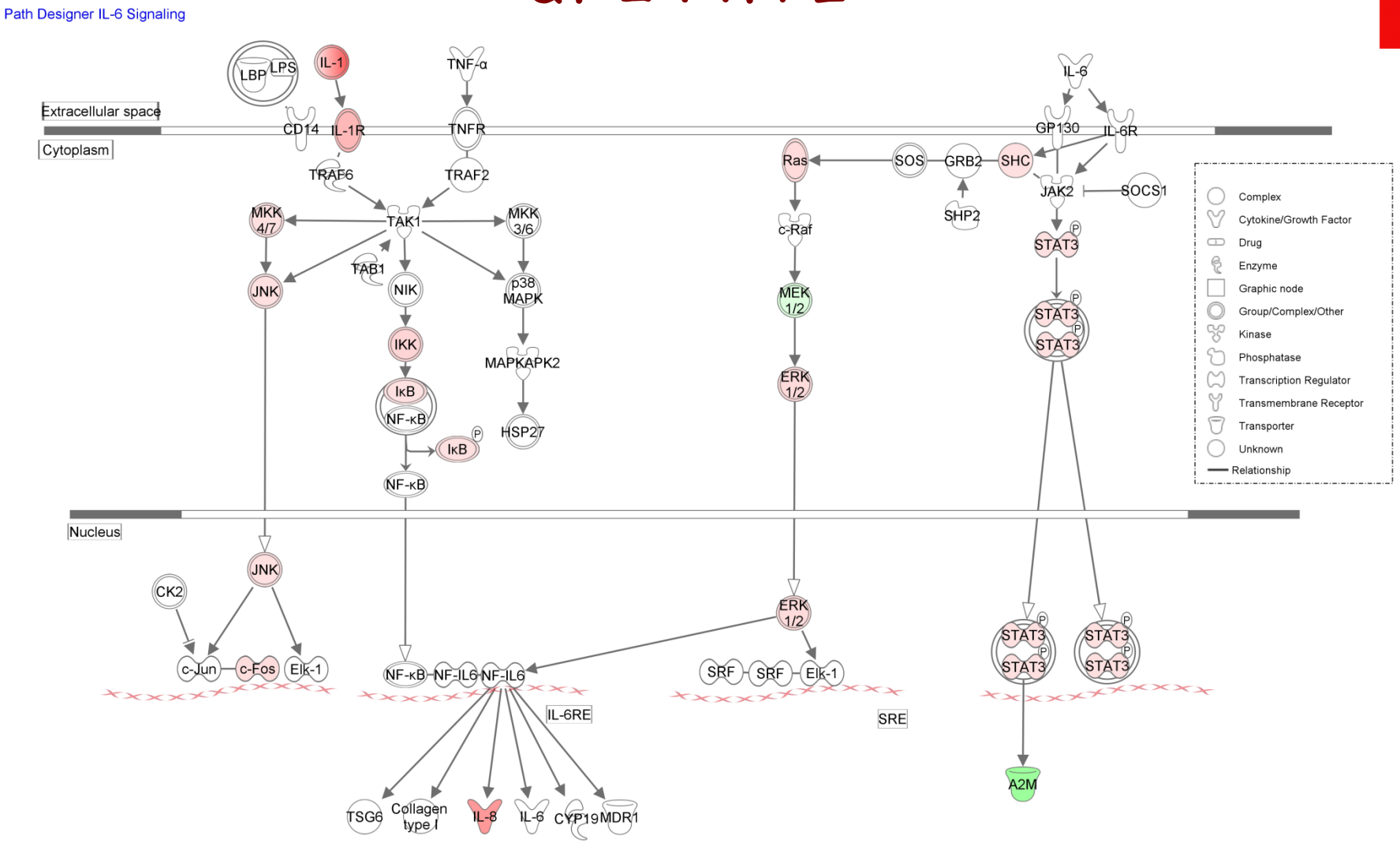
"**ENSEMBL** *Bos Taurus* 4.0.58_cDNA"





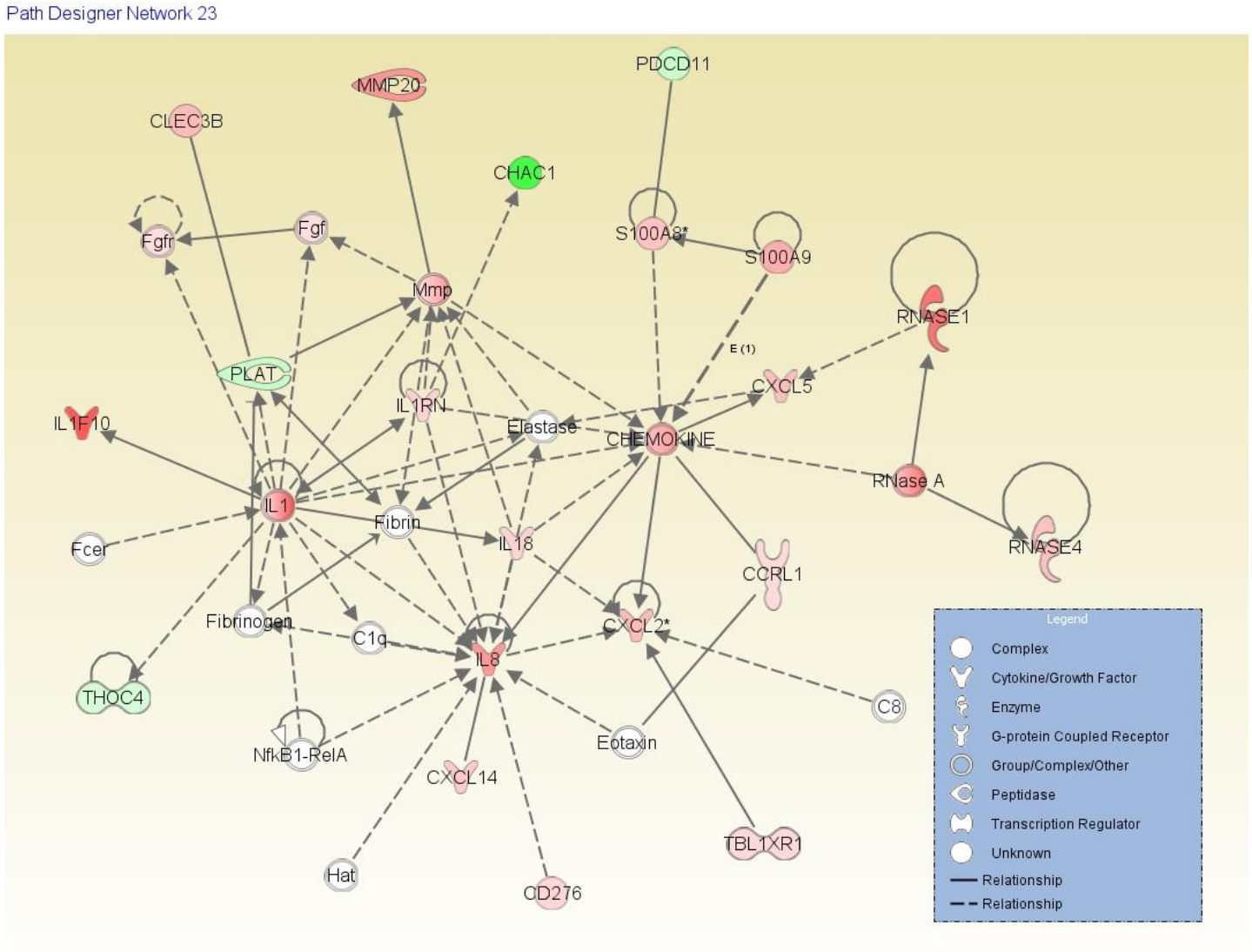
Expression patterns of representative immune response associated candidate genes at different time-points post Ma infection

IL-6 and IL-1 signalling pathways at 24 h PI



© 2000-2010 Ingenuity Systems, Inc. All rights reserved.

»Cell movement, development and hematologic functions, immune system signalling«



FDR ≤ 0,05

Revealing the whole genome sequence opened new horizon for genomic research in lactation biology

Special accent was given to genomic regions associated with

- Lactation
- Immune system
- Growth traits

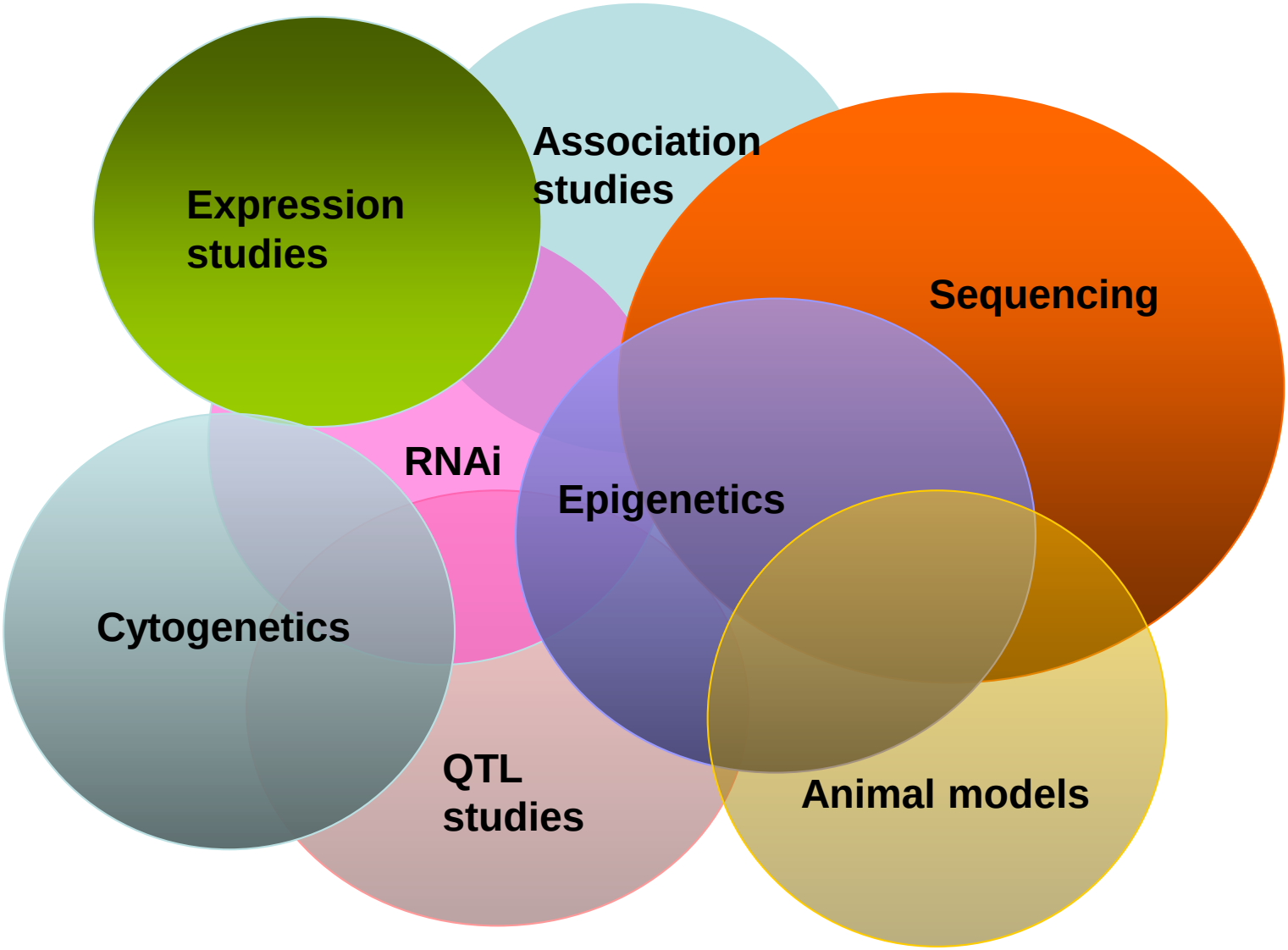
Epigenetics and miRNA

Polymorphism		Ancestral allele				Derived allele(s)				
SNPId	Alleles	Octamer	Expr	Source	miRNA	Octamer	Expr	Source	miRNA	Gene Id
rs41255722	C/G					ACACTGGA	n/a	L X	miR-199a-5p miR-199b	IL1B_BOVIN
rs41255717	A/G	TCTGATAA	n/a	L	miR-361					IL1B_BOVIN

Polymorphic miRNA targets in the bovine *IL1B* gene

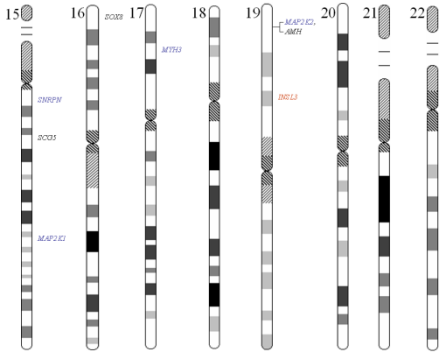
We found 359 putative target sites for mammary gland expressed miRNAs within the candidate genes, including polymorphic miRNA targets. Among them we identified using the Patrocles database polymorphic miRNA target sites for bta-miR-199b, -199a-5p, and -361 in the *IL1B* gene and for -miR-126 in the *CYP11B1* gene. The expression of miR-199b, -199a-5p, and -126 in the bovine mammary gland has already been experimentally confirmed

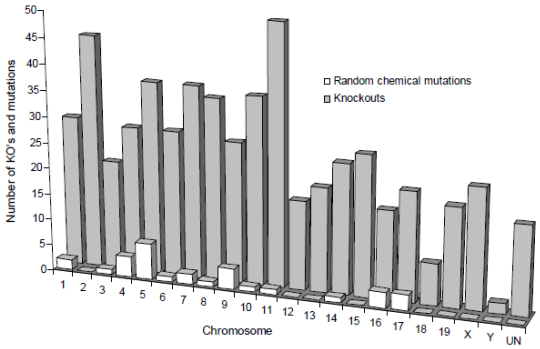
Complexity of data integration



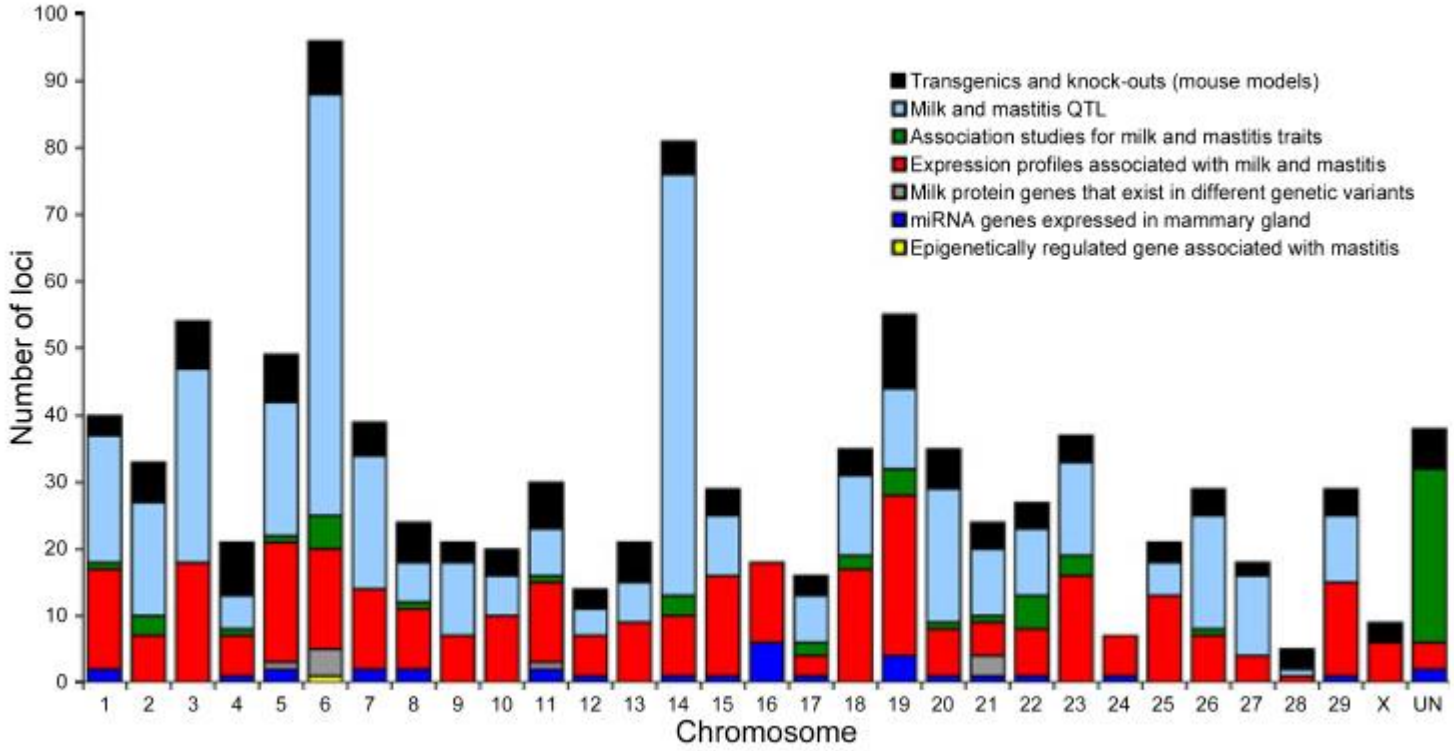
Data integration

- Different approaches (QTL, SNP, KO…)
- Different sources (Datenbanken, Publikationen…)
- Co-localisation of loci (Visualisation, Map based approach)
- Causal relationship among genes (IPA)



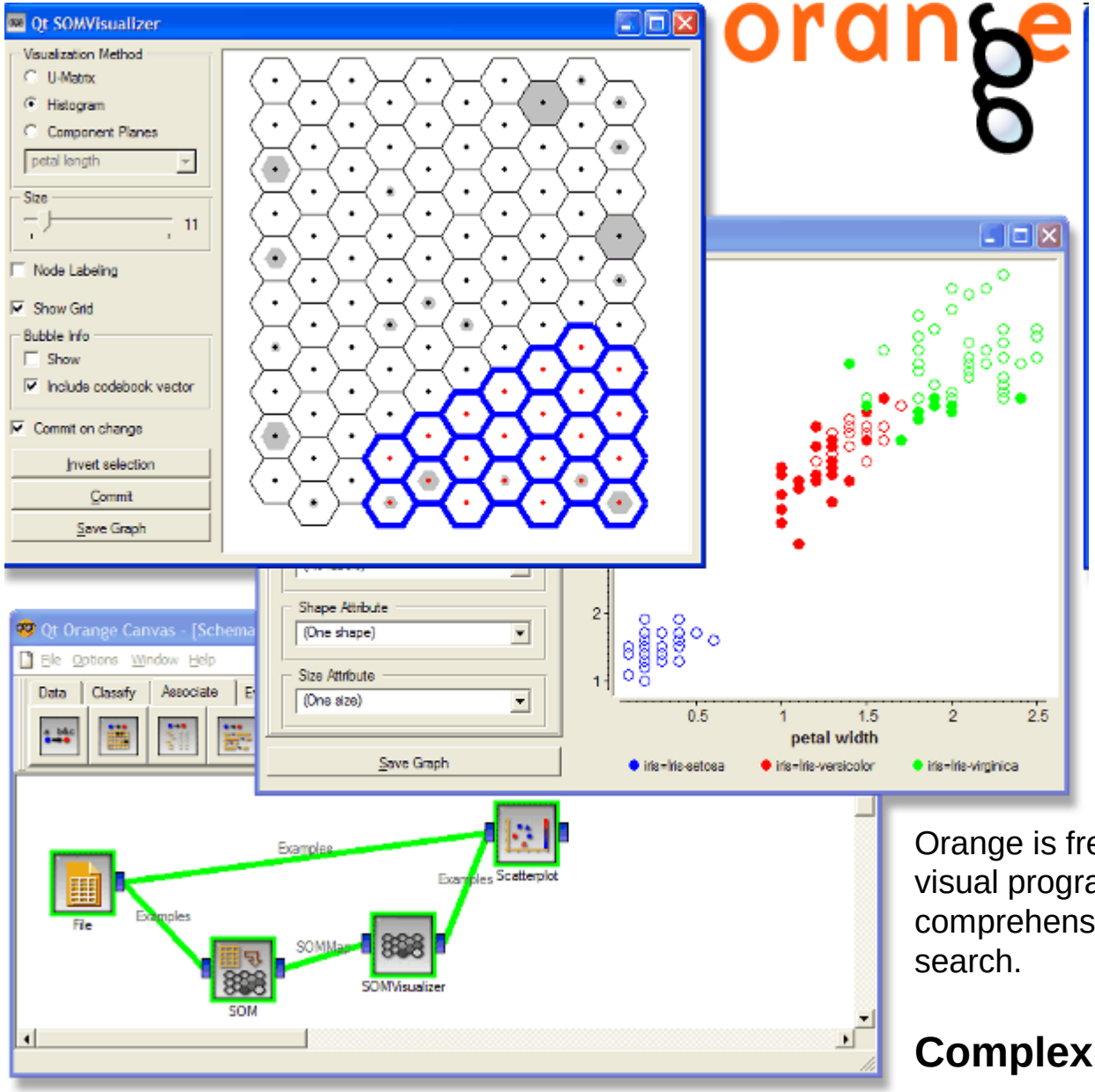


Number of genes related to lactation and mastitis



Selection of candidate genes for milk traits and mastitis

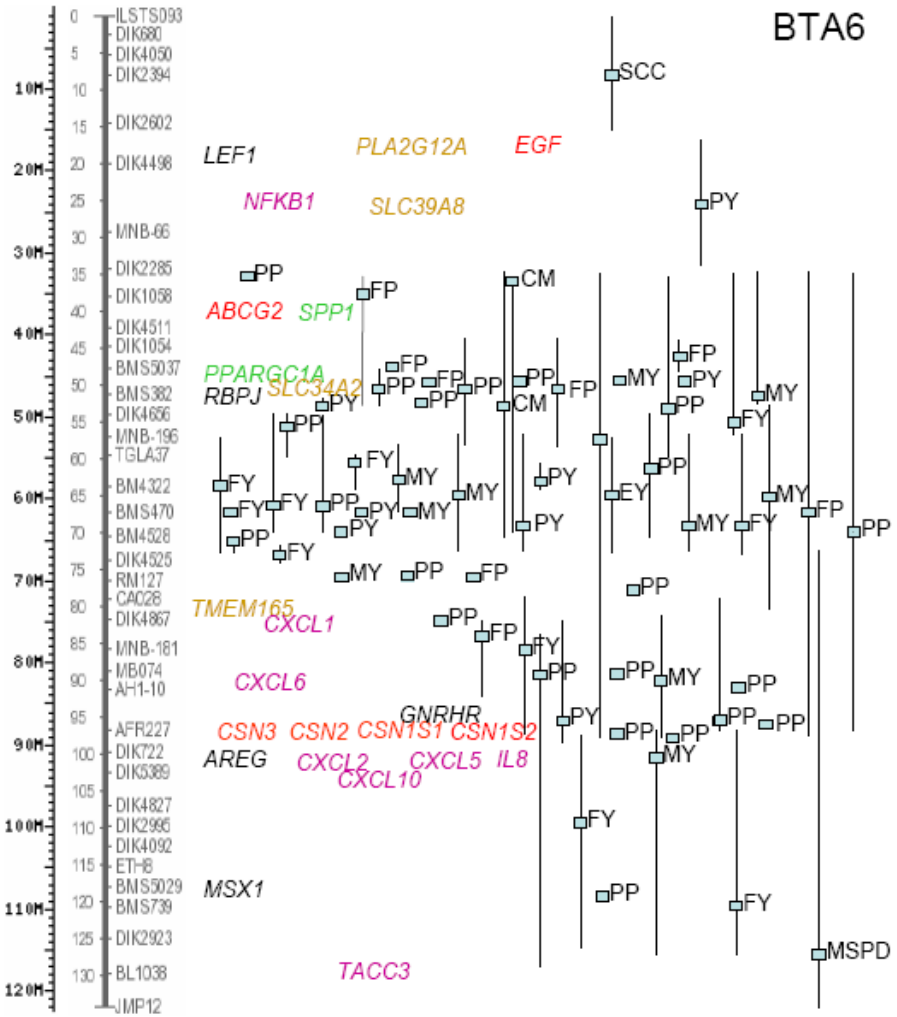
List of the most promising candidate genes and number of independent studies by approach									*In silico* analysis						
			Association studies		Expression studies					Number of SNPs					
Gene	Gene name	Mouse KOs and transgenic experiments	Milk traits	Mastitis traits	Milk traits[1]	Mastitis traits	Milk protein genetic variants	Epigenetic studies	Expression in lactating mammary gland[2] (mouse)	Promoter (5 kb upstream)	5'UTR	Exons[3]	Introns (context 100 bp)	3'UTR	Functions[4]
Associated with milk production															
ABCG2	*ATP-binding cassette, sub-family G (WHITE), member 2*		+ + +		+				***	0	0	1	1	0	E
ATP2B2	*ATPase, Ca++ transporting, plasma membrane 2*	+			+				*	1	0	1	3	0	D
B4GALT1	*UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 1*	+			+				**	0	0	0	1	0	A
BTN1A1	*Butyrophilin, subfamily 1, member A1*	+			+				***	0	0	3 (2)	0	0	C
CSN1S1	*Casein alpha s1*		+ + +		+		+	+	***	0	0	1 (1)	0	0	C
CSN1S2	*Casein alpha s2*				+		+		NA	0	0	0	0	0	NA
CSN2	*Casein beta*	+					+		***	0	0	3 (3)	0	0	B
CSN3	*Casein kappa*	+	+ + +		+		+		***	0	0	0	3	0	B
DGAT1	*Diacylglycerol O-acyltransferase 1*	+	+ + + + + + +						**	NA	NA	NA	NA	NA	A
EGF	*Epidermal growth factor (beta urogastrone)*	+			+				NA	0	0	0	0	0	B
GHR	*Growth hormone receptor*		+ + + +						**	0	0	0	0	0	B
ID2	*Inhibitor of DNA binding 2, dominant negative helix-loop-helix protein*	+			+				*	0	0	0	1	0	B
LALBA	*Lactalbumin, alpha*	+			+		+		***	0	0	0	0	1	A
LEP	*Leptin*	+	+ + +		+				**	0	0	8 (3)	4	8	B
LGB	*Lactoglobulin, beta*		+ +				+		NA	0	1	1	0	0	NA
MFGE8	*Milk fat globule-EGF factor 8 protein*	+			+				***	0	0	0	0	1	A
NME1	*Non-metastatic cells 1, protein (NM23A) expressed in*	+			+				NA	0	0	3 (2)	0	1	D
PRL	*Prolactin*	+	+ +						NA	0	0	0	0	0	B
PTHLH	*Parathyroid hormone-like peptide*	+			+				**	0	0	0	0	0	C
STAT5A	*Signal transducer and activator of transcription 5A*	+	+ +						**	0	0	0	0	0	B
XDH	*Xanthine dehydrogenase*	+			+				***	0	0	2 (1)	1	0	C
Associated with mastitis resistance															
ACTB	*Actin, beta, cytoplasmic*	+				+			*	0	0	0	0	1	D
C5AR1	*Complement component 5a receptor 1*					+ +			**	0	0	1	0	0	B
CD14	*CD14 antigen*					+ + + +			***	0	1	1	0	0	A

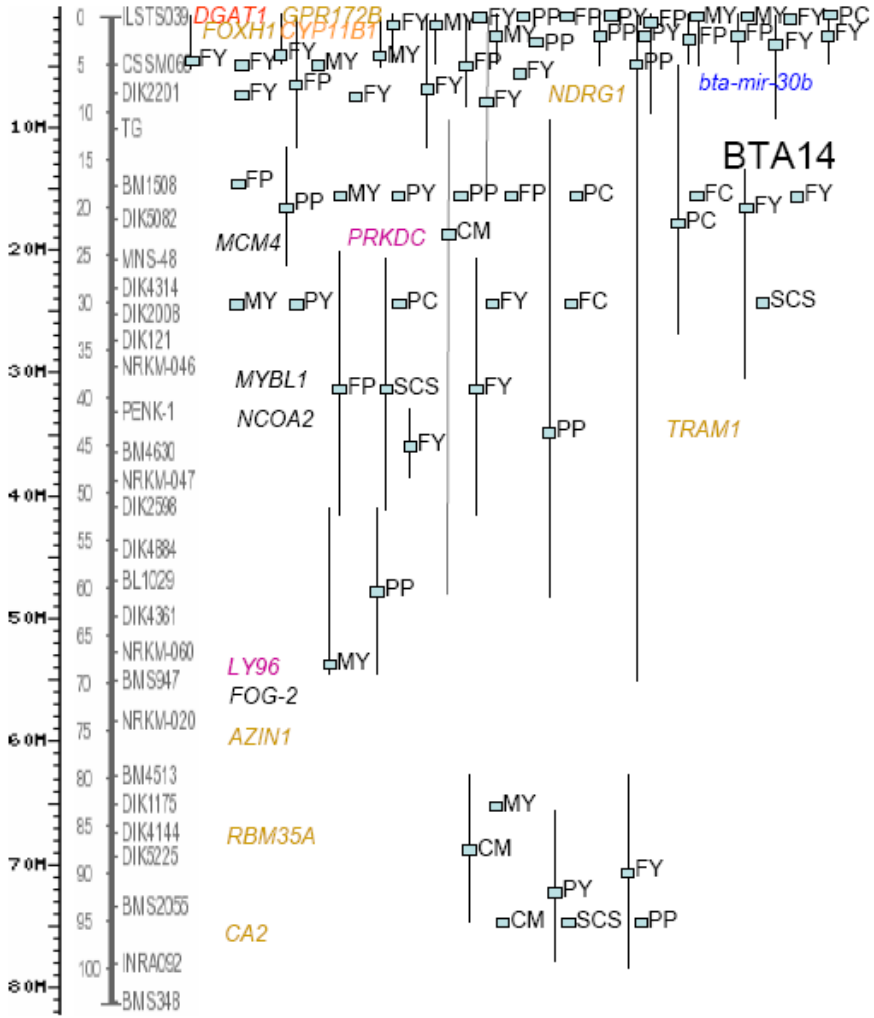


Orange is free software for data visualisation and visual programming. Orange allaws comprehensive bioinformatics analysis and text search.

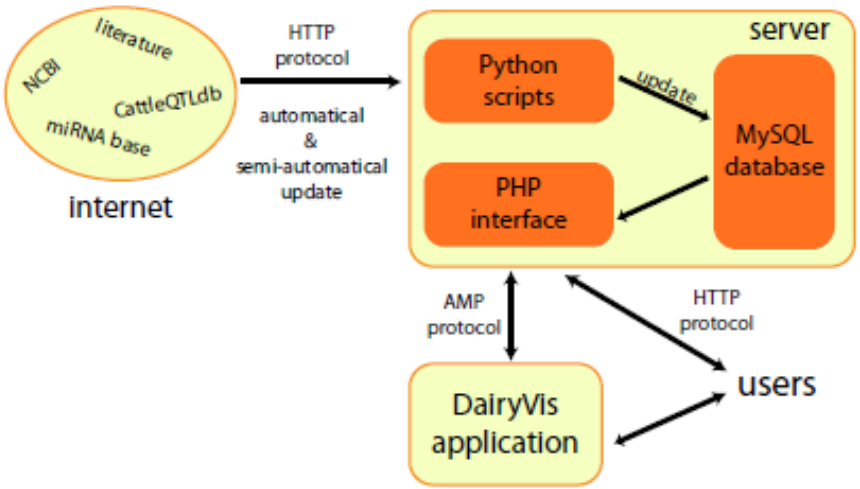
Complex, flexible, fast.

Gene maps can be used for integration of data



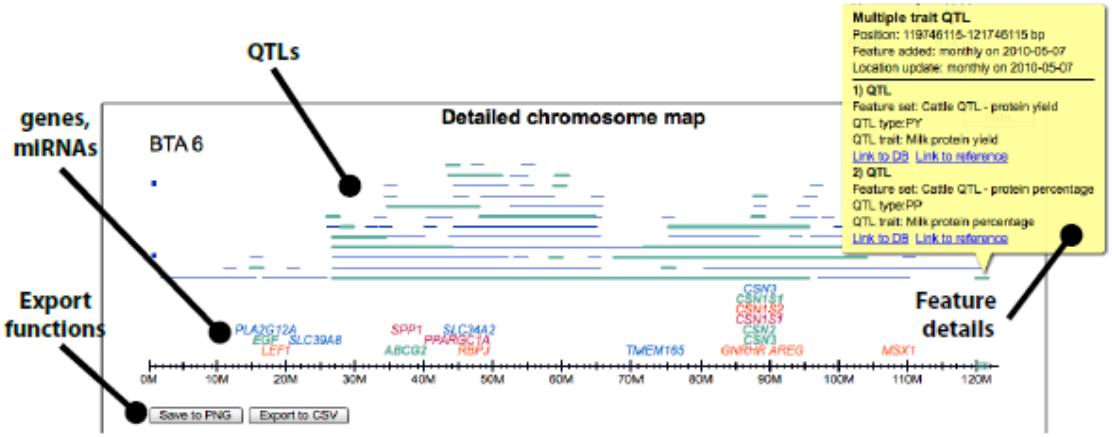


DairyVis - a specialized web application

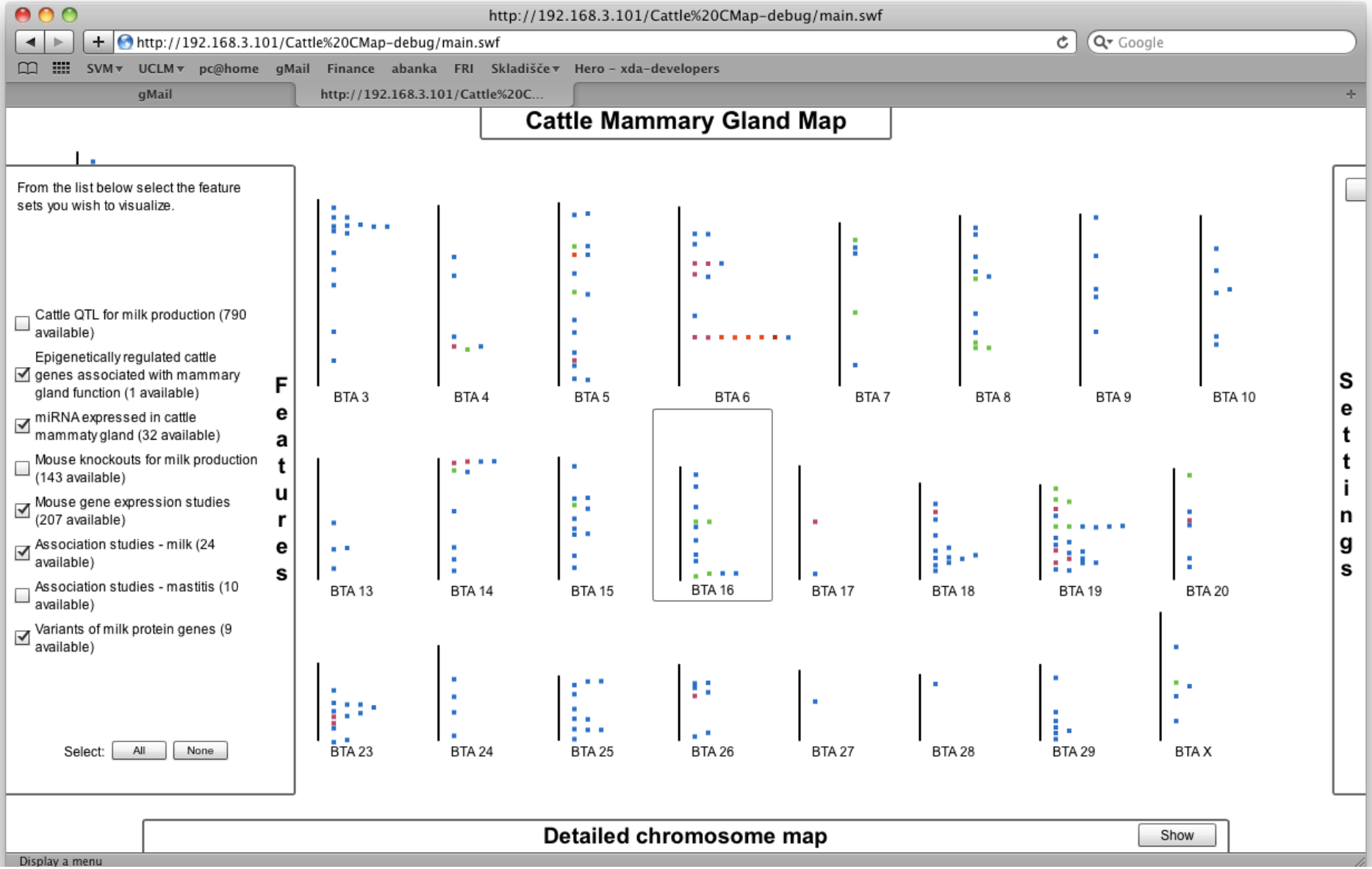


- Integrates public data from different databases
- Automatical and semi-automatical update
- Interactive user surface

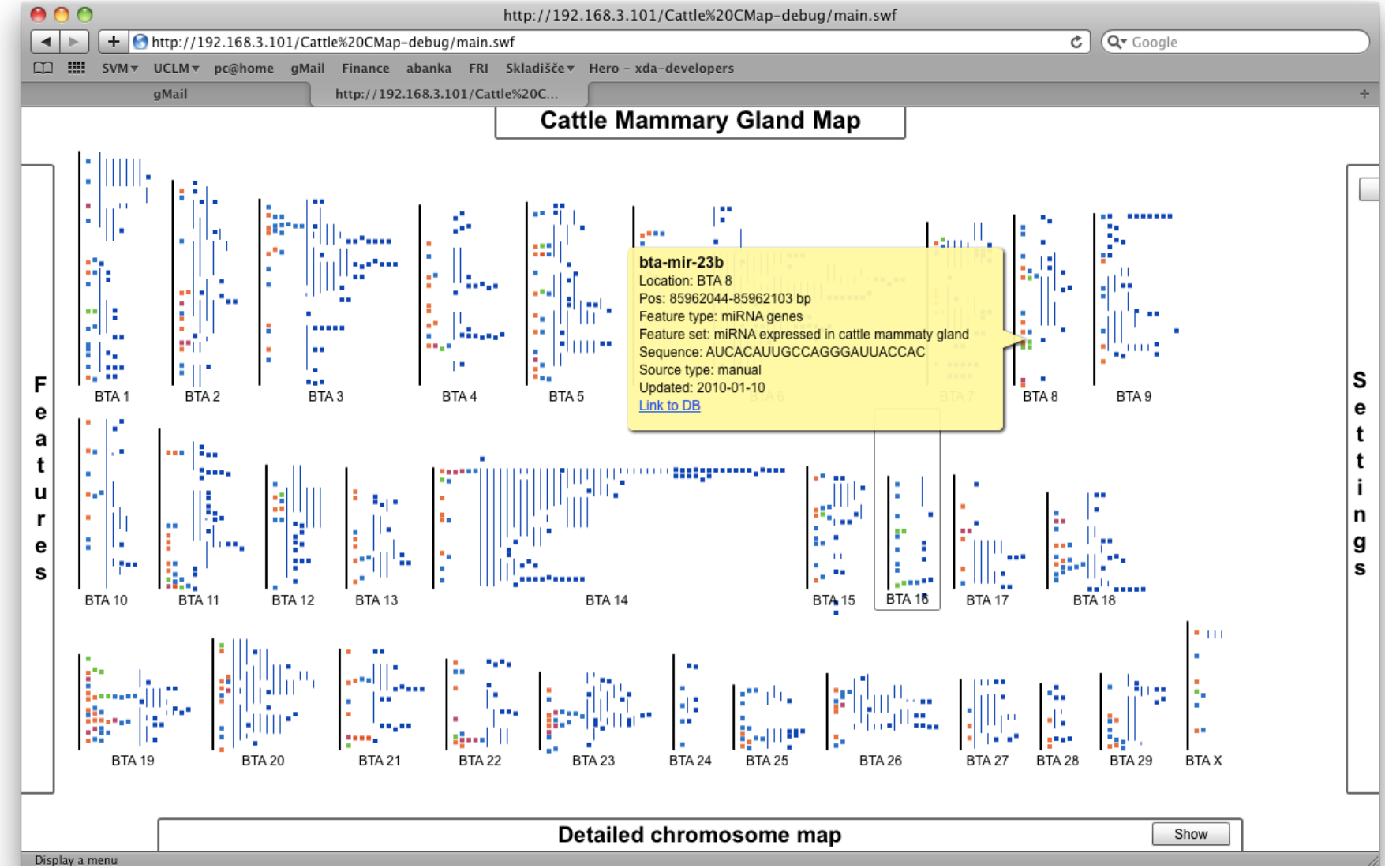
- Detailed browsing through chromosomal regions
- Selection of features
- Different output formats



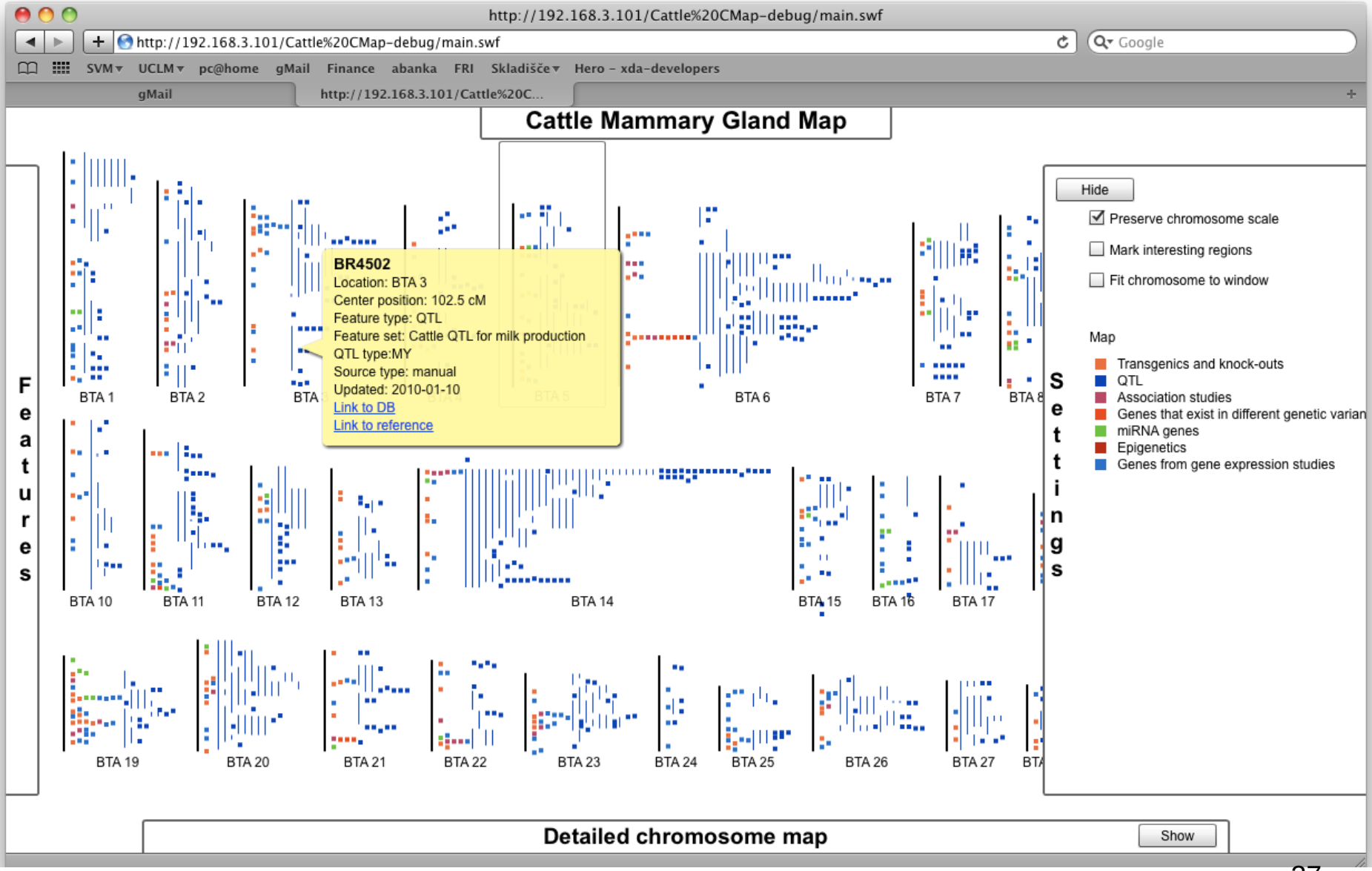
Mammary Gland Database

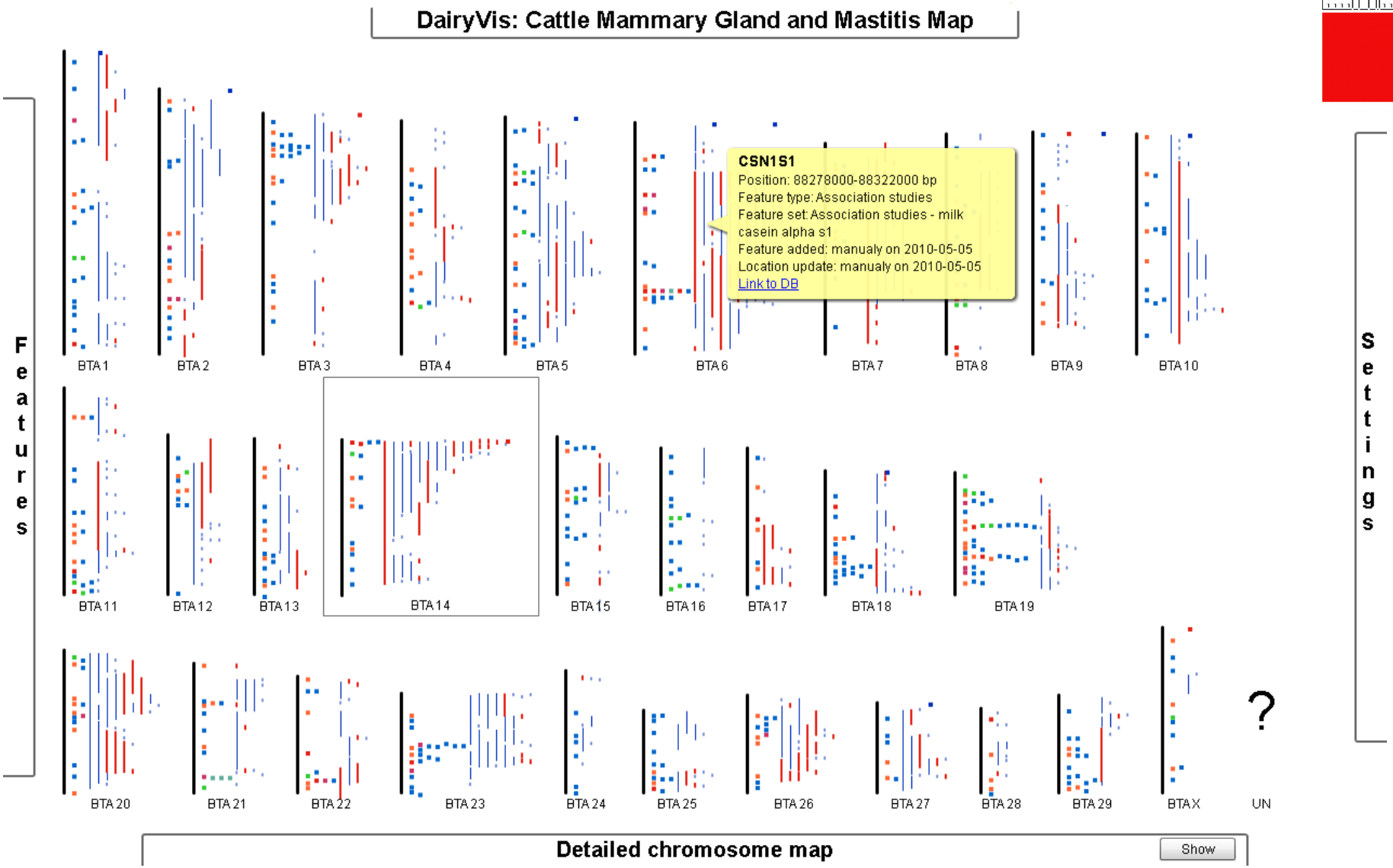


Mammary Gland Database



Mammary Gland Database





http://dairyvis.fri.uni-lj.si/DairyVis/DairyVis.html

In silico analysis of candidate genes (male reproduction)

Gene	BioGPS (GNF)	Ensembl SNP's (cattle)					miRBase (Sanger)	Ingenuity Pathways Analysis
	Expression in testis[1] (mouse / human)	Promoter (5kb upstream)	5'UTR	Exon[2] (non-synonymous SNPs)	Intron (context 100bp)[3]	3'UTR	miRNAs with putative target sites in candidate genes (bta-miR)	Top functions[4]
ACE	**/**	0	0	0	1	0	21*, 214, 199a-5p, 193a, 621	A
ACR	***/**	0	0	0	0	0	/	B
ACTB	*/*	0	0	0	0	1	/	A
AKT1	*/*	0	0	0	1	0	127, 425-3p	C
APOB	*/-	0	0	1 (1)	0	0	200a	B
APRT	**/*	0	0	0	0	0	210, 20b, 214, 106, 20a, 17-5p, 93, 425-3p	B
AR	*/*	0	0	0	0	0	30a-5p, 30d, 30e-5p, 214, 124a	A
BRCA2	**/*	NA	NA	NA	NA	NA	320, 499, 29b, 29c, let-7a*	A
CAMK4	*/**	NA	NA	NA	NA	NA	/	A
CREM	**/*	0	0	0	0	0	/	A
DAZ family	NA	NA	NA	NA	NA	NA	/	B
DAZL	***/***	0	0	1	2	5	/	B
DDX25	***/***	0	0	0	2	0	186, 499, 142, 342, 369-5p	A
ESR1	*/*	0	0	0	1	0	128a	A
FHL5	***/*	0	0	0	0	0	/	A
FSHR	*/*	0	0	2 (1)	6 (1)	0	138, 148b, 30d, 363	A
GNRHR	**/*	0	0	1	0	0	142*, 380-3p, 30b, 30c	A
GSTM1	**/**	0	0	0	0	0	214, 151	B
HSPA2	***/***	0	0	0	0	0	/	B
KIT	*/*	0	0	0	0	0	7, 425-3p, 200a, 192, 215	C
KITL (KITLG)	*/*	0	0	0	0	0	7, 425-3p, 200a, 192, 215	C

Top functions:

A – Cellular development, reproductive system development and function, cardiovascular disease;

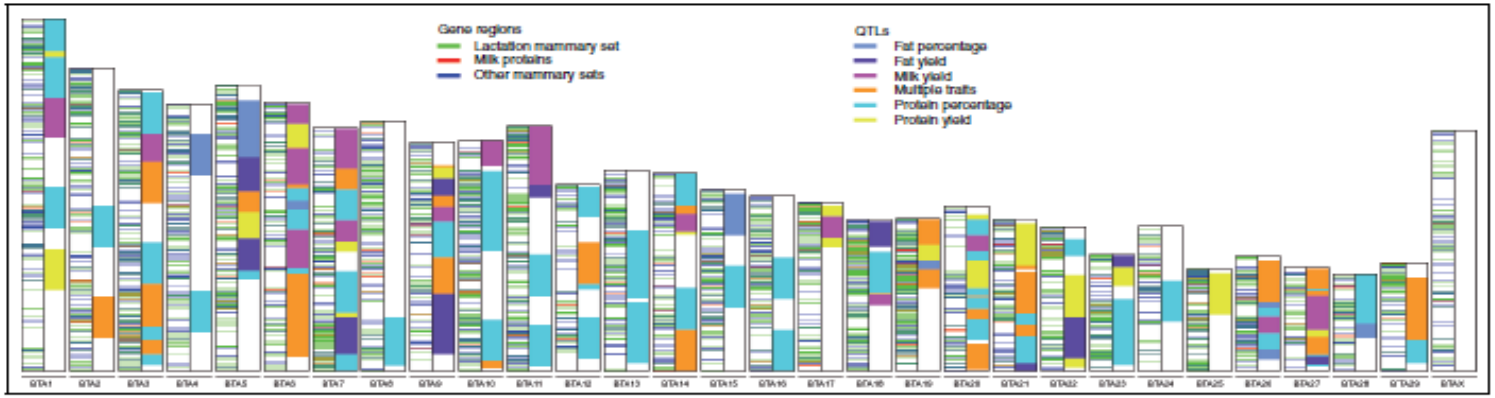
B – Cell cycle, Cell-to-cell signaling and interaction, cellular function and maintenance;

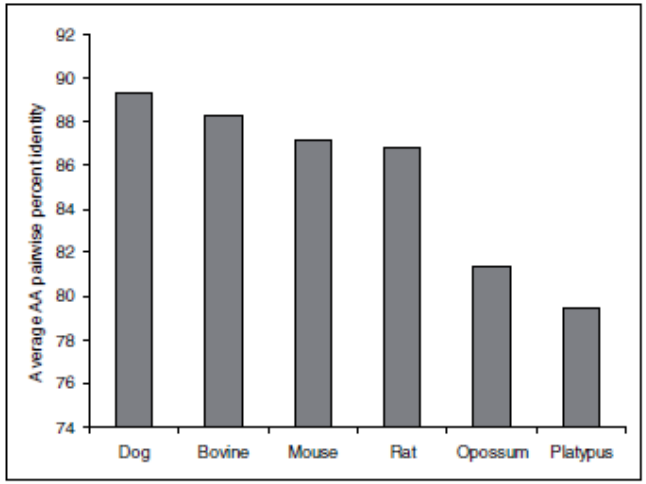
C – Cell death, cellular development, embryonic development;

D – RNA damage and repair, cell cycle, cell death;

E – Cell cycle, cellular assembly and organization, DNA replication, recombination, and repair.

The gene onthology is still leaky and incomplete





Highly conserved milk proteins

Protein name	GLEAN ID[a]	Gene symbol	Average PID
Rab 11A	GLEAN_20537	*RAB11A*	100
GTP binding protein G I G S G T beta subunit 1	GLEAN_21827	*GBB1*	100
GTP binding regulatory protein beta 2 chain	GLEAN_22534	*GNB2*	100
Rho C	GLEAN_13128	*RHOC*	99.8
Rab 11B	GLEAN_03051	*RAB11B*	99.6
Rap 1b	GLEAN_10112	*RAP1B*	99.6
GTP binding protein Sara	GLEAN_16602	*SAR1A*	99.5
Rab 3A	GLEAN_02763	*RAB3A*	99.5
Histone 2, H2ab	GLEAN_23583	*HIST2H2AB*	99.2
SAR1B protein	GLEAN_08536	*SAR1B*	98.9
14-3-3 protein beta alpha	GLEAN_04527	*YWHAB*	98.9
Leucine-rich repeat containing protein 8	GLEAN_26477	*LRC8A*	98.8
Rab 18	GLEAN_21462	*RAB18*	98.6
Rho GDP-dissociation	GLEAN_11194	*ARHGDIA*	98.5
Rab 5C	GLEAN_17381	*RAB5C*	98.5
AD158	GLEAN_08550	*LRC8C*	98.1
Rab 3C	GLEAN_20950_P10949	*RAB3C*	98.0
Eukaryotic translation initiation factor 4, gamma 2	GLEAN_19671	*EIF4G2*	97.9
ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit	GLEAN_03119	*ATP5A1*	97.7
ARP3 (actin-related protein 3, yeast) homolog	GLEAN_25161	*ACTR3*	97.5
Vimentin	GLEAN_20783	*VIM*	97.3
GTP binding protein alpha 14	GLEAN_08781	*GNA14*	97.2
Endoplasmin precursor (GRP94/GP96)	GLEAN_20794	*HSP90B1*	97.1
Lymphocyte cytosolic protein 1 (65 K macrophage protein/L-plastin)	GLEAN_05236	*LCP1*	97.0
Retinal short-chain dehydrogenase/reductase	GLEAN_03662	*SDR1*	96.9

Lemay et al., 2009