

Research article

Open Access

## Novel transcripts discovered by mining genomic DNA from defined regions of bovine chromosome 6

Rosemarie Weikard\*, Tom Goldammer, Annett Eberlein and Christa Kuehn

Address: Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere (FBN), 19196, Dummerstorf, Germany

Email: Rosemarie Weikard\* - [weikard@fbn-dummerstorf.de](mailto:weikard@fbn-dummerstorf.de); Tom Goldammer - [tomgoldammer@fbn-dummerstorf.de](mailto:tomgoldammer@fbn-dummerstorf.de); Annett Eberlein - [eberlein@fbn-dummerstorf.de](mailto:eberlein@fbn-dummerstorf.de); Christa Kuehn - [kuehn@fbn-dummerstorf.de](mailto:kuehn@fbn-dummerstorf.de)

\* Corresponding author

Published: 24 April 2009

Received: 1 September 2008

*BMC Genomics* 2009, **10**:186 doi:10.1186/1471-2164-10-186

Accepted: 24 April 2009

This article is available from: <http://www.biomedcentral.com/1471-2164/10/186>

© 2009 Weikard et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Linkage analyses strongly suggest a number of QTL for production, health and conformation traits in the middle part of bovine chromosome 6 (BTA6). The identification of the molecular background underlying the genetic variation at the QTL and subsequent functional studies require a well-annotated gene sequence map of the critical QTL intervals. To complete the sequence map of the defined subchromosomal regions on BTA6 poorly covered with comparative gene information, we focused on targeted isolation of transcribed sequences from bovine bacterial artificial chromosome (BAC) clones mapped to the QTL intervals.

**Results:** Using the method of exon trapping, 92 unique exon trapping sequences (ETS) were discovered in a chromosomal region of poor gene coverage. Sequence identity to the current NCBI sequence assembly for BTA6 was detected for 91% of unique ETS. Comparative sequence similarity search revealed that 11% of the isolated ETS displayed high similarity to genomic sequences located on the syntenic chromosomes of the human and mouse reference genome assemblies. Nearly a third of the ETS identified similar equivalent sequences in genomic sequence scaffolds from the alternative Celera-based sequence assembly of the human genome. Screening gene, EST, and protein databases detected 17% of ETS with identity to known transcribed sequences. Expression analysis of a subset of the ETS showed that most ETS (84%) displayed a distinctive expression pattern in a multi-tissue panel of a lactating cow verifying their existence in the bovine transcriptome.

**Conclusion:** The results of our study demonstrate that the exon trapping method based on region-specific BAC clones is very useful for targeted screening for novel transcripts located within a defined chromosomal region being deficiently endowed with annotated gene information. The majority of identified ETS represents unknown noncoding sequences in intergenic regions on BTA6 displaying a distinctive tissue-specific expression profile. However, their definite regulatory function has to be analyzed in further studies. The novel transcripts will add new sequence information to annotate a complete bovine genome sequence assembly, contribute to establish a detailed transcription map for targeted BTA6 regions and will also be helpful to dissect of the molecular and regulatory background of the QTL detected on BTA6.

## Background

Independent linkage analyses consistently indicate several QTL for production, health and conformation traits on bovine chromosome 6 (BTA6) [1-3]. Prerequisite to elucidate the molecular background causing the genetic variation at the QTL is a detailed, well-annotated gene and transcript sequence map of the chromosomal regions covering the QTL intervals. Particularly, functional genomic studies in chromosomal regions containing QTL rely on transcript sequences to create expression microarrays and perform QTL expression studies. Through the efforts of the bovine genome projects the currently identified genes are primarily derived from large-scale sequencing efforts in combination with sequence comparison approaches to available gene sequences from sequence-ready genomes as mice and human. Comparative mapping of genes from the syntenic human chromosome 4 (HSA4) on BTA6 by high-resolution radiation hybrid mapping was performed [4-10] helping to identify genes with orthologous counterparts on BTA6. However, it is not clear, which proportions of species-specific genes could have been missed by such efforts. Missing genes may be due to the fact that there are a number of genes in individual mammalian genomes, which are expressed at a very low level or only in a time-limited or tissue-specific manner at defined developmental stages, and which are difficult to be isolated by conventional cloning techniques.

Moreover, there is increasing indication for the existence of lineage- or species-specific transcripts from numerous studies (e.g., [11]). Recent studies [12-15] provided conclusive evidence for the abundance of lineage-, tissue-, development-, or spatial-temporal-specific transcripts encoded in the cattle genome. For instance, Kumar and colleagues [12] discovered a number of novel transcripts predominantly expressed in the cattle placenta, which not had been found previously in primates or rodents. These results indicate towards the molecular basis of the relative anatomical variation in placental architecture of ruminants and suggest that loss and gain of genes and transcript variants are important mechanisms of genome evolution in mammals. Especially in ruminants, there is already unequivocal evidence for the occurrence of highly divergent genes encoding proteins having adaptive functions in the biology and metabolism of reproduction (e.g., IFN tau, [16]).

Given the unique adaptations for the ruminant reproductive system, it is conceivable that also other highly divergent processes associated with rumination and dramatic adaptive changes in ruminant metabolism, for instance, at the onset of lactation and continuous maintenance of a high milk production performance [17-21] could be caused by the action of individual species-specific genes or transcripts. This would provide important rationale for

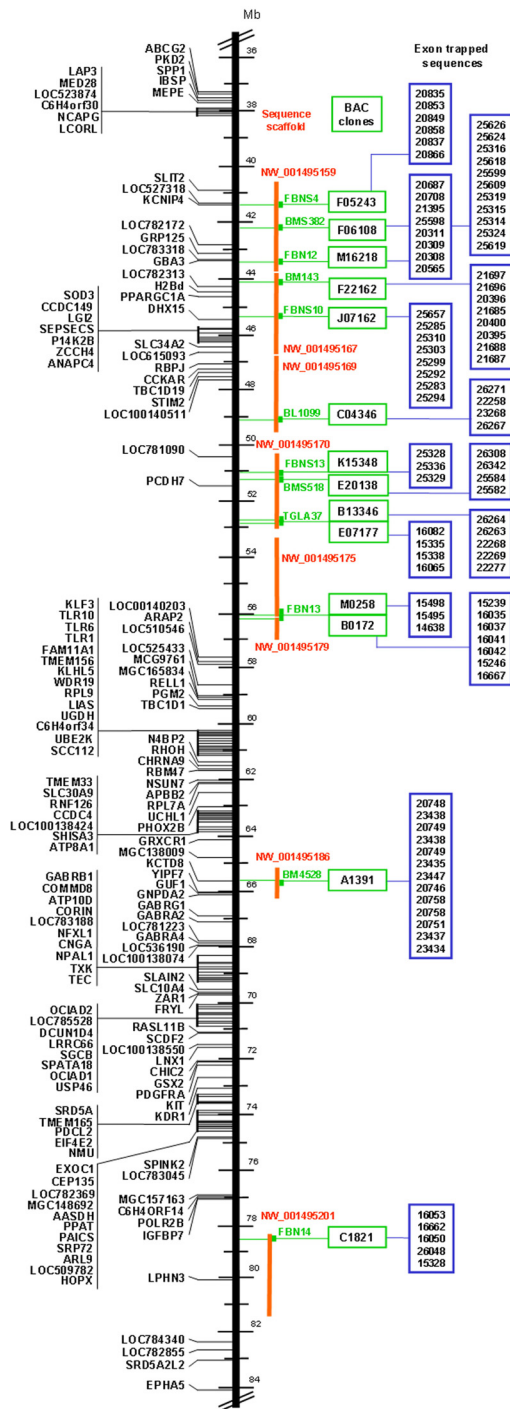
the identification of genes or transcripts being located in QTL intervals associated with milk production traits.

To complete the bovine gene catalogue on the targeted BTA6 region carrying the QTL, alternative methods will be required to identify novel transcripts and genes independently of their individual tissue-specific or spatial-temporal expression pattern. In our study, we focused on targeted isolation of transcribed sequences from defined subchromosomal regions on BTA6 poorly covered with gene information. Therefore, we exploited bovine bacterial artificial chromosome (BAC) clones mapped to the QTL interval using exon trapping, a technique that has been developed to identify genes in cloned eukaryotic DNA (e.g., [22-24]). This technique enables the experimental identification of potential exons in a fragment of eukaryotic DNA of unknown gene content or intron-exon structure based on the presence of functional splice sites; thus making the exon trapping assay independent of the expression pattern of the transcripts to be identified. This feature represents the main advantage of the exon trapping method compared to other methods of gene identification (cDNA selection [25-27] or cDNA hybridization approaches [28-30]).

Identified putative transcript sequences isolated from BTA6-specific BAC clones *in vitro* using the exon trapping technique were subjected to a comprehensive sequence comparison analysis by alignment to sequence assemblies of the bovine, human and mouse genomes. The genuine presence of the identified exon trapping sequences (ETS) in the bovine transcriptome was verified by multi-tissue expression analysis using RT-PCR.

## Results

As an initial step towards systematic transcript analysis in the targeted BTA6 region we performed exon trapping on fourteen selected chromosome region-specific BAC clones that mainly mark the segment between BTA6q16 and BTA6q24 and correspond to defined subchromosomal regions between 41.4 and 78.5 Mb on the sequence assembly for BTA6 (NCBI, Btau4.0) as shown in Figure 1. For each BAC clone two exon trapping libraries were generated based on both, completely *Bam*HI/*Bgl*II and partially *Sau*3AI restriction enzyme digested BAC DNA. To eliminate the majority of sequences due to vector self splicing or cryptic splicing events, we employed PCR screening of clones from the exon trapping libraries followed by size selection of PCR products prior to sequencing. From each library 30 randomly selected clones were subjected to PCR with insert-flanking vector-derived primers and a medium of 14 clones from each exon trapping library were selected for sequencing. Thus, a total of 396 putative exon trapping sequences (ETS) with a size varying from 35 to 349 bp were identified.



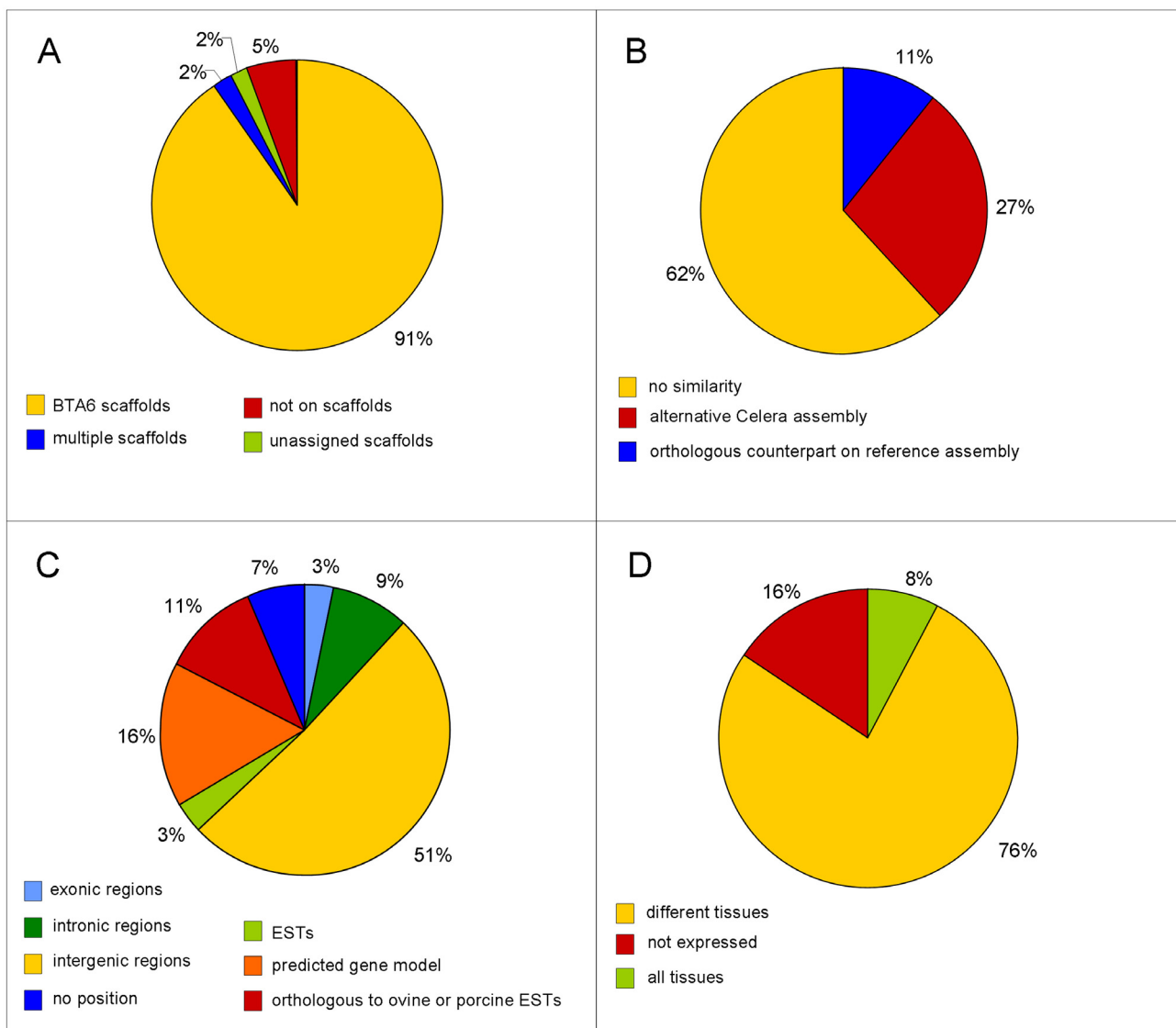
**Figure 1**  
**Chromosomal localization of exon trapping sequences on bovine chromosome 6 (BTA6).** Left: Protein-coding genes assigned to BTA6 (bovine genome sequence assembly Btau4.0). Right: Markers used for BAC library screening (green). BAC clones (green-framed boxes). Bovine genomic scaffolds (red). Exon trapping sequences (blue-framed boxes).

Sequence comparison identified 92 sequences as being derived from the trapping vector caused by vector self-splicing and from the BAC vector sequence or due to cloning artefacts, which indicated 23% false positive splicing events. Of the remaining 304 putative ETS, 92 (30%) were unique sequences. The number of unique sequences detected per BAC clone varied between three and fifteen. There was a relatively high percentage (70%) of ETS observed redundantly. Some sequences were found with a frequency up to fourteen-fold. This redundancy is due to the fact that for each BAC clone two separate exon trapping libraries based on two different restriction enzyme digestion of BAC DNA were constructed. However, the set of ETS detected in the two single libraries constructed per BAC clone was not completely identical. Furthermore, it has also to be considered that the exon trapping technique itself includes several steps based on PCR amplification, a feature that intrinsically should be associated with the abundance of multiple copies of specific DNA targets.

The size of unique putative ETS ranged from 35 to 317 bp with a medium of about 125 bp. Locus-specific primers were derived for unique putative ETS with a size exceeding 60 bp (see Additional file 1). These primers were applied to remapping of ETS to their parent BAC clones by PCR. The chromosomal localization of exons on BTA6 could be inferred from physical mapping of the BACs carrying the trapped sequences by FISH or by *in silico* sequence similarity mapping of the corresponding BAC end sequences on the genome sequence assembly of BTA6 in combination with our high-resolution RH map [9].

All 92 unique ETS were further analyzed by screening the bovine genome sequence assembly (NCBI, Btau4.0) for similarity to known sequences, genes or transcripts. The results are summarized in Figure 2A and presented in detail in Additional file 2. Sequence identity to sequence scaffolds on BTA6 was detected for the majority of unique ETS (91%). Two ETS were identified on sequence scaffolds not yet assigned. A further two sequences contained repetitive sequence motifs and matched to more than one scaffold located on different chromosomes. We found that five ETS (5%) could not be assigned to the bovine genome sequence resources by *in silico* sequence similarity search.

As illustrated in Figure 2B and presented in Additional file 2 in more detail, for a number of ETS (28%) sequence similarity was found to genomic sequence scaffolds from alternative sequence assembly of the human genome (based on Celera assembly, for instance ETS\_25316, ETS\_25598 etc.). Furthermore, a few ETS (11%) revealed similarity to genomic sequences located on the syntenic chromosomes HSA4 and MMU5 of the human and mouse reference genome assemblies. The remaining ETS



**Figure 2**  
**Analysis of transcripts isolated from gene-poor regions on bovine chromosome 6 (BTA6) by exon trapping.** A: Assignment of exon trapping sequences (ETS) to the bovine genome assembly Btau4.0. B: Sequence similarity of ETS to genome sequence assemblies from human and mice. C: Distribution of ETS to the annotated bovine genome. D: Expression of ETS analyzed in a multi-tissue panel of a lactating cow.

did not detect comparative genomic counterparts in human and mouse.

According to Additional file 2 and Figure 2C, we found that only a small number of ETS showed identity to known bovine cDNA sequences or expressed sequence tagged sites (ESTs). On BAC clone BBI\_750F0243 carrying ETS\_20835, ETS\_20849 and ETS\_20866, we identified four exons of the bovine KCNIP4 gene (NM\_001076935), which has an orthologous counterpart on the syntenic HSA4 region. The ETS\_20866 sequence was not yet iden-

tified on scaffolds of the current bovine sequence assembly, although the mRNA of the bovine KCNIP4 gene revealed 100% identity to this sequence, which is also supported by sequence identity of ETS\_20866 to several bovine ESTs. Strong sequence similarity of ETS\_20866 was also identified to the adjacent exons 11 and 12 of the orthologous human gene. This result indicates that the bovine KCNIP4 gene is not completely annotated in the current bovine sequence assembly because at least two exons are missed in the predicted gene model. Besides to human, high comparative sequence similarity of these

ETS was also observed to exons of the KCNIP4 gene from other species (rat, mouse, chicken and horse).

Sequence identity to known bovine ESTs was detected for ETS\_25609 on BAC BBI\_750F06108, ETS\_15246 on BAC BBI\_750B0172, and ETS\_25294 on BAC BBI\_750J07162. The ESTs were all assigned to sequence scaffolds on BTA6 (NW\_001495159, NW\_001495179, and NW\_001495167, respectively), where the corresponding parental BAC clones were assigned to. In addition, ten ETS captured from BAC clone BBI\_75006108 revealed sequence similarity to two ovine mRNAs and one ETS from BAC clone BBI\_750B0172 showed similarity to a porcine transcript.

To analyze their potential protein-coding capacity, the unique ETS were screened for existing open reading frames (ORFs) using the program ORF Finder at NCBI. For most ETS, the program could predict ORFs. However, no ORF spanned the sequence length of the analyzed ETS completely. Protein coding sequences predicted by the program were subjected to sequence alignment with TBLASTN in the nucleotide collection and the ESTs databases at NCBI. Again, only the sequences ETS\_20835, ETS\_20849 and ETS\_20866 showed high similarity matches pointing to the KCNIP4 gene from several species.

Including the *ab initio* map at the NCBI *Bos taurus* MapViewer [31] into our analysis we identified fifteen ETS with sequence similarity to bovine gene models predicted by the gnomon gene prediction program. One of them (ETS\_16667) pointed to a hypothetical locus (LOC100140872) predicted to be a pseudogene similar to the PHYHIPL gene (see Additional file 2, Figure 2C).

The results obtained by the different sequence similarity analyses suggested that the majority of the isolated transcripts in our study presumably represent bovine noncoding transcripts located in intergenic regions. By screening the comprehensive mammalian noncoding RNA database (RNAdb 2.0) we could not identify similarity to known noncoding RNAs in the available data sets [32].

To elucidate, if the identified ETS were really expressed in the bovine transcriptome, multi-tissue expression analysis by RT-PCR using locus-specific primers on total RNA from nine different bovine reference tissues (liver, small intestine, kidney, thyroid gland, mammary gland, pituitary gland, skeletal muscle, subcutaneous and intestinal fat) was performed. For each BAC clone we exemplarily analyzed at least three ETS, if available and specific primers could be designed.

The results are summarized in Figure 2D and are displayed in detail in Figure 3. They show that 43 ETS out of 51 analyzed (84%) were expressed at least in one tissue isolated from a lactating cow. Specifically, 39 ETS displayed a different expression pattern in the tissues analyzed, and four revealed expression in all tissues examined.

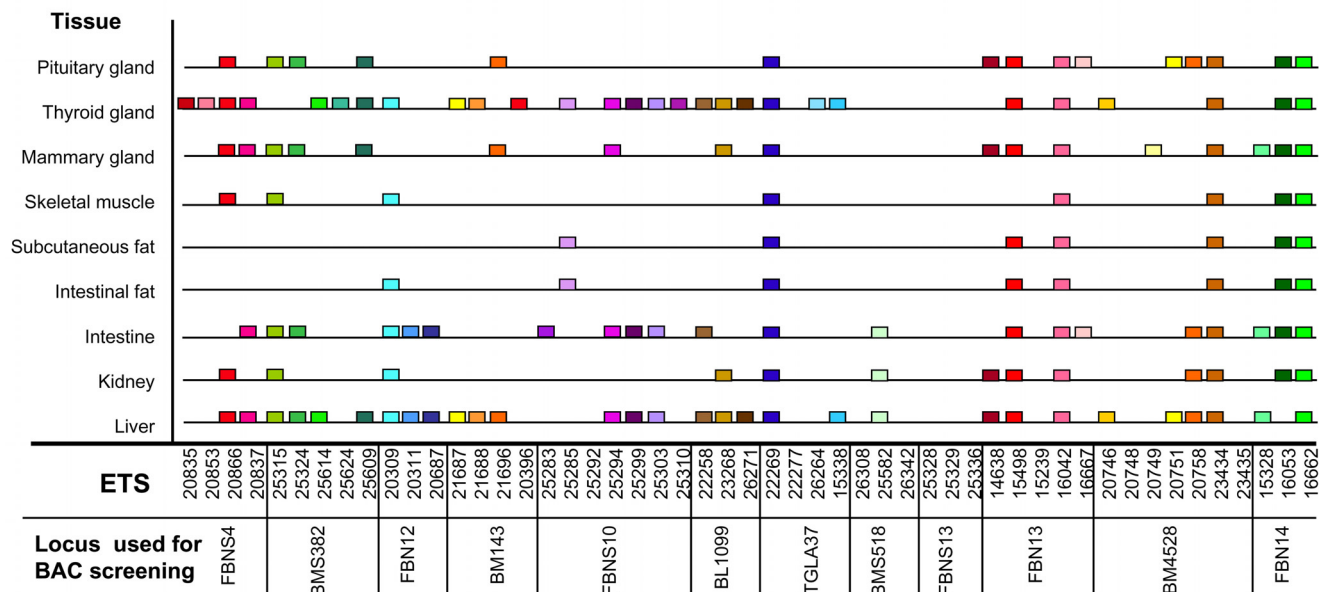
In Figure 4(A–D), examples of different expression patterns of ETS ranging from expression in only a few tissues (A, C, D) to ubiquitous expression (B) observed in the experiment are shown. Figure 4E displays expression of the house-keeping gene GAPDH in the tissues analyzed in our study. Generally, ETS revealed amplification of the expected amplicon size in bovine genomic DNA (see Figure 4A–C, lane BT). However, in the case of ETS\_20866 amplification in bovine genomic DNA failed (Figure 4D, lane BT), which is due to the fact that this ETS consists of two adjacent exons. This was supported by sequence comparison to the human KCNIP4 gene sequence revealing that the primers of ETS\_20866 span an intronic region on genome level. There were a further 6 unique ETS captured by exon trapping (ETS\_21696, ETS\_25582, ETS\_16037, ETS\_15328, ETS\_25303, ETS\_26263), which also seem to contain at least two adjacent transcripts and are interrupted by intergenic DNA sequence on genomic sequence level. This result is supported by both alignment of the identified sequences to sequence scaffolds and experimentally, by amplification of the corresponding longer amplicons on bovine genomic DNA.

## Discussion

The main goal of the Bovine Genome Project, the identification of the whole bovine genome sequence, is almost achieved [33], but still there are several regions in the bovine genome not sufficiently annotated and characterized. Genome assemblies also rely on the existence of transcript sequences to merge contigs together, verify the assembly of whole genome shotgun reads, and annotate genes. Further analysis of QTL regions of interest may include physical and transcription mapping, identification of positional and functional candidate genes and isolation of the corresponding full-length cDNA as well as association studies on the selected genes.

This study represents a further step in the ongoing molecular and genetic analysis of complex traits and annotation of genes and transcripts localized in the region on BTA6 containing QTL for milk and meat production, health and conformation traits.

As an initial step towards systematic analysis of transcripts and genes in this region, we carried out exon trapping using selected bovine BAC clones previously mapped to QTL intervals on BTA6. Mining a genomic interval comprising about 1 Mb for transcribed sequences using this



**Figure 3**  
**Tissue-specific expression pattern of sequences isolated by exon trapping on BAC clones specific to bovine chromosome 6 (BTA6).** Y-axis: Tissues of a lactating cow. X-axis: Exon trapped sequences (ETS) ordered to the marker loci used for BAC library screening. Expression in a specific tissue was analyzed by RT-PCR and is displayed by a coloured box specific for each ETS.

technique, we identified a total of 92 unique exon trapping sequences. Genome similarity searches revealed sequence identity matches to sequence scaffolds on BTA6 for most unique ETS (91%). With two ETS, which were identified on sequence scaffolds not yet assigned to the current NCBI sequence assembly of the bovine genome, gaps could be closed on the genome sequence level. There were about 2% of ETS, which could not be unambiguously assigned to the bovine genome sequence, because of matches to multiple chromosomes due to repetitive sequence motifs. We found only 5% of ETS without any hits to known sequences contained in the archive of bovine sequence databases. These ETS can provide new additional sequence information to complete the current genome sequence assembly. This result further indicates that targeted deep sequencing within the corresponding genomic regions would be required to improve the accuracy of the BTA6 sequence assembly.

Comparative sequence similarity search to human and mouse genome sequences revealed that 11% of the isolated ETS displayed high similarity to genomic sequences located on the syntenic chromosomes HSA4 and MMU5 of the human and mouse reference genome assemblies pointing to highly conserved genome regions in these species. Almost a third of the ETS identified similar equivalent sequences in genomic sequence scaffolds from the

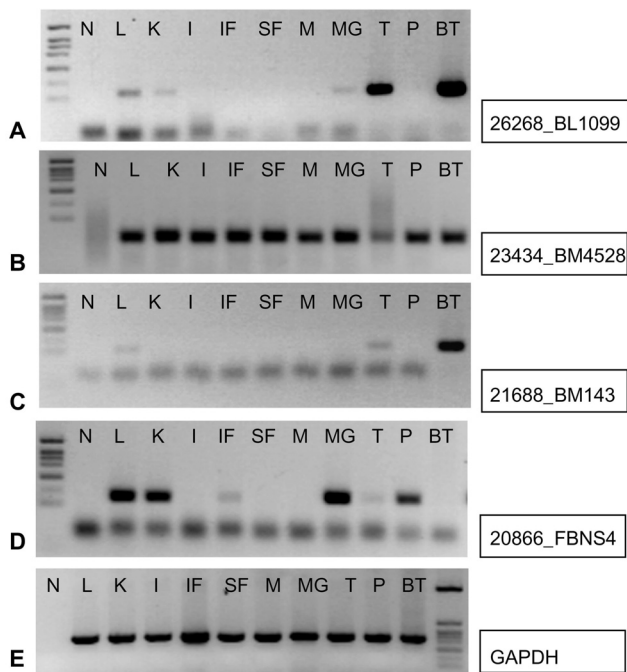
alternative Celera-based sequence assembly of the human genome. The residual 62% of ETS without comparative genomic sequence counterparts in human and mouse refer to presumably species-specific genomic regions in the bovine genome.

Screening the gene, ESTs and protein databases at NCBI detected only a few known transcribed sequences revealing identity to the ETS isolated by exon trapping in our study (17%, Figure 2C).

Whereas 6% of all ETS identified known bovine transcripts, a further 16% of ETS pinpointed to bovine gene models predicted *ab initio*.

For the evaluation of the relatively low proportion of ETS identifying known transcripts, we have to consider, that the BAC clones subjected to exon trapping in our study had been selected from regions on BTA6 poorly covered with protein-coding genes, which could still be noted on the current sequence assembly Btau4.0 (Figure 1). Eleven BACs were assigned to gene desert regions [9], which have been found conserved in mammals and birds [34,35] and were thought to be transcriptionally silent. Hence, we could not expect *a priori* that many sequences would have been annotated as known genes. It should also to be considered that even the current annotation of the bovine





**Figure 4**  
**Expression of selected exon trapped sequences (ETS) on a bovine multi-tissue panel analyzed by RT-PCR.** A, C, D: examples for different tissue-specific expression, B: example for ubiquitous tissue expression. E: expression of GAPDH as control house-keeping gene. Sequences analyzed are given right to each electrophoresis profile. L: liver, K: kidney, I: small intestine, IF: intestinal fat, SF: subcutaneous fat, M: skeletal muscle, MG: mammary gland, T: thyroid gland, P: pituitary gland, BT: bovine genomic DNA, N: no template control.

genome is still limited and consequently, the set of functional elements is not completely identified to date. Additionally, transcripts from the most lowly expressed genes, or genes specifically expressed in important but relatively minor cell types may very likely be under-represented in the ESTs database predominantly established by large scale ESTs projects.

At the beginning of our experiments it was known that only one BAC clone (BBI\_750F0243) was assigned near a human gene (KCNIP4) on the syntenic region on HSA4 by *in silico* comparative mapping. Thus, this BAC clone could serve as a proof for the efficacy of the exon trapping procedure. Indeed, we identified ETS in this BAC clone pinpointing to four exons of the bovine KCNIP4 gene, which underlined and validated the usefulness of the exon trapping method for targeted mining of transcribed sequences in defined chromosomal regions based on genomic DNA from BAC clones. In addition, by identifying two additional exons of the bovine KCNIP4 gene,

which are not present in the current bovine genome assembly Btau4.0, it was exemplarily demonstrated that this experimental approach is a useful complement for the annotation of the bovine genome sequence.

Because there was no identity detected to known genes and ESTs by *in silico* sequence comparison for the majority of ETS identified in our study, these sequences are assumed to be novel and could be predicted to originate from unknown bovine transcripts. Expression analysis was performed to validate this hypothesis. Examination of a subset of the trapped putative transcripts showed exemplarily that in a lactating cow numerous ETS displayed a divergent, tissue-specific expression pattern (Figure 2D, Figure 3). Most expression signals were observed in liver, thyroid gland, small intestine, kidney, and pituitary gland. Tissue-dependent expression pattern of the ETS may indicate to potentially specific functions in the corresponding tissues of the lactating cow. Some of the ETS were found to be expressed in all tissues examined, indicating a ubiquitous expression pattern and suggesting them being probably part of housekeeping genes or conserved structural genes, if similarity to repetitive sequences could be excluded. As shown in Figure 3, 16% of the analyzed ETS did not display expression signals in the multi-tissue panel. This could likely be due to the fact, that these transcripts were not expressed in the tissues contained in the panel of a lactating cow analyzed here. In this context it should be mentioned that the advantage of the exon trapping approach is that the method is independent of spatio-temporal expression patterns due to the identification of transcripts based only on intrinsic characteristics of the genuine genomic sequence. But non-expressed ETS, possibly, could also represent false-positive sequences isolated from regions of the genomic DNA due to existing sequence similarities to splice site consensus sequences (e.g., splice donor/acceptor, branch point region), which the exon trapping technique is based on.

From the results of the expression analysis it could be inferred that the ETS revealing expression in the bovine multi-tissue panel may represent *bona fide* transcribed sequences. However, the ETS have to be characterized in further studies with regard to their functional significance. Based on the presented data, we can not exclude that a part of the identified ETS may be attributed to the class of pseudogenes or to non-functional RNA. Pseudogene transcription has been observed in small-scale gene-centred studies and genome-scale unbiased mapping of transcriptionally active regions in the human and mouse genomes. Surveys of Gerstein and Zheng [36,37] have revealed that for example, 5–20% of human pseudogenes can be transcriptionally active. However, considering the relatively high percentage of ETS (84%), for which expression has been demonstrated in our study, it could be assumed that

a number of them might attain to another category of transcribed sequence elements as for example noncoding RNA.

Continued submission of ESTs and other sequence information in a variety of species points to the existence of transcripts that do not map to currently annotated genes [12,38-40]. These transcripts may possibly correspond to novel protein coding genes, genes encoding small unknown peptides, pseudogenes or noncoding RNA. Evidence of transcription had increasingly been found in unannotated intergenic genome regions of the human genome, which were thought to be transcriptionally silent (e.g., [41-46]). The ENCODE consortium reported that a vast amount of DNA, not annotated as known genes, is transcribed into RNA. While the majority of the genome appears to be transcribed at the level of primary transcripts, only about the half of the processed transcripts is mapped as currently annotated genes [36,42,46]. Particularly, a high number of new transcriptionally active regions (more than 50%) were detected in non-annotated intergenic regions. These studies indicated that genomic regions previously considered as "junk" encode for multiple polyadenylated and non-polyadenylated transcripts of unknown function. According to Gerstein et al. [36] the ENCODE project provided evidence that there is much activity between annotated genes and intergenic space in the human genome contributed by transcribed non-protein-coding RNAs and transcribed pseudogenes. The authors highlighted that a number of these transcribed pseudogenes and noncoding RNA genes are located even within introns of protein-coding genes and assumed that these components may possibly influence the expression of their host genes. It is also possible that these transcripts themselves do not have a direct function, but rather are important for a particular process (e.g., chromatin accessibility for transcription factor binding). Continuously, numerous noncoding RNA sequences are recognized in the transcriptomes of different eukaryotes as having important regulatory functions in controlling various levels of gene expression in physiological and developmental processes and diseases of complex organisms (e.g., [44,47-49]). The detailed investigation of the functional relevance of the numerous unknown transcripts was postulated as a prospective task in the post ENCODE era.

The findings of our study provide experimental support for transcripts lacking ESTs or other cDNA evidence in the targeted regions of BTA6. The majority of unknown ESTs presumably identified novel noncoding transcripts located in intergenic regions of the chromosome. However, prospective studies should be performed to further characterize the transcripts with regard to their putative functional significance. In this respect it has to be proven, if these transcripts belong to non-functional RNA or if

they have any specific regulatory function in the bovine genome. Currently, there is scarce information on the function of bovine noncoding RNA genes compared to the state in mouse and human. The prevalence of bovine noncoding RNAs, their regulatory impact on gene expression and their physiological effects are not yet examined in detail.

While the mammalian genomes contain nearly similar repertoires of protein-coding gene sequences comprising only a fraction of about 1.5% of the whole genome, the majority of the mammalian genome is obviously transcribed. Our results support the increasingly accepted concept suggesting that the physiological complexity and the unique phenotypes of species-specific or individual genomes might evolve from combinatorial features contributed by the entire genome sequence including previously neglected genome regions [33,42,45,46,50]. Consequently, variation in noncoding sequences might be important effectors of phenotypic variation in complex traits and diseases (see reviews [51,52]) in livestock.

The results of our study on BTA6 demonstrate that the exon trapping method based on region-specific BAC clones is applicable to targeted screening for novel transcripts located within a defined chromosomal region sparsely covered with annotated genes. The novel transcript sequences obtained will contribute to establish a detailed transcription map for targeted specific subchromosomal BTA6 regions. Our results show that the computational prediction and identification of genes and transcripts and manual inspection solely are not sufficient to annotate the final bovine genome in the absence of experimentally derived data. Experiences from genome studies in other species revealed that genome annotation is never complete or final (e.g., [39,40,42,45,53,54]). Therefore, correcting and refining the genome annotation is a reiterative task, which is continuously being done and depends on experimental data for final validation, especially for the identification of rare transcripts and alternative splice variants. Compared to high-throughput sequencing technologies like transcriptome sequencing initiated currently, the method of exon trapping has some advantages. Detection of transcripts by high-throughput sequencing requires knowledge about the temporal-spatial expression pattern of the targeted group of unknown transcripts. Frequently, the time point of expression is difficult to predict, e.g., for transcripts of high relevance for developmental regulation. Furthermore, high-throughput technologies have their limits regarding detection of rare transcripts. Therefore, a targeted approach independent of amount, time and locus of expression using a method like exon trapping will complement high throughput technologies for the analysis of defined chromosomal intervals,



for example to trace transcripts in fine-mapped QTL regions.

### Conclusion

In this context, the identification of novel transcripts for BTA6 can contribute to improve the accuracy of gene annotation of the bovine genome and provide new, additional sequence information to help to complete the current genome sequence assembly. Hence, the data obtained in this study was submitted to the database of the annotation community of the bovine genome [55] for implementation in the current international efforts of annotation of the bovine genome. The major objective of these activities is to generate a set of manually annotated bovine genes, connect them to biological function and incorporate them into the Bovine Official Gene Set, which will be made available at NCBI.

### Methods

#### BAC library screening and identification

DNA superpools of bovine BAC libraries [BBI\_B750, available at ImaGenes, the former German Resource Center for Genome Research (RZPD) Berlin, [56] were screened by PCR as described previously [9]. Primers of 12 loci including nine microsatellite markers (FBN12, BM143, BMS382, BL1099, TGLA37, BMS518, FBN13, FBN14, BM4528) and three targeted sequence tagged sites (FBNS4, FBNS10, FBNS13) isolated from a microdissection library specific for the BTA6 region [57] were used for the identification of specific BAC clones. These loci, which were previously mapped on BTA6 by radiation hybrid (RH) mapping [8], were selected according to their position in the QTL interval and the poor gene coverage within the chromosomal region they were mapped to. Information concerning primer sequences and references of these loci is given by Weikard et al. [8]. The specificity of the identified BAC clones was determined by direct BAC sequencing with primers used for BAC library screening and/or Fluorescence *in situ* hybridization (FISH) [58]. BAC DNA extractions were performed using the Qiagen Large-Construct Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

The genome coverage of the of BAC clones was estimated from *in silico* similarity search of BAC end sequences [9] on the bovine genome sequence assembly of BTA6. Calculating insert sizes of BAC clones by *in silico* mapping on the basis of the recent sequence assembly (NCBI, Btau4.0) revealed experimentally unachievable insert sizes for two clones (BBI\_750K15348: of 865 Kb and BBI\_750M0258: 198 Kb, see Additional file 2), which could possibly be due to errors in the sequence assembly. Therefore, medium inserts sizes were applied for these two clones in our estimation of genome coverage.

#### Exon trapping on BAC clones

Exon trapping relies on the conservation of cis-acting sequences at intron-exon boundaries required for splicing in all eukaryotic species. By subcloning a genomic fragment into the intron of an expression vector, internal exons encoded in the genomic fragment will be spliced into the transcript encoded on the expression vector. Successful capture of exons by internal exon trapping relies on the prerequisite, that the cloned genomic DNA fragment focused on, contains at least one exon flanked by intronic sequences [22,23,59]. Reverse transcriptase polymerase chain reaction (RT-PCR) using primers specific for the transcript on the expression vector will provide a product for analysis by electrophoresis and sequencing.

Essentially, exon trapping was performed using the exon trapping system formerly provided by Gibco/Life Technologies (now Invitrogen) according to the manufacturer's instructions and published protocols [22,23] with minor modifications. We replaced the splicing vector pSPL3 with pSPL3b provided by Genzyme Corp., Framington, MA, in order to increase exon trapping efficiency as has been reported by Burn et al. [24,60].

Individual BAC clones (500 ng) were completely double digested with *Bam*HI and *Bgl*II and partially digested with *Sau*3AI, respectively, gel-purified and shotgun cloned into the dephosphorylated *Bam*HI site of the splicing vector pSPL3b. Recombinant colonies were pooled and plasmid DNA was prepared with a plasmid purification kit (Macherey & Nagel). The DNA was used to transfect COS-7 cells for transient expression of the reporter system by lipofection with lipofectamine plus (Invitrogen). Exon trapping reporter transcripts were examined by extracting total cytoplasmic RNA from COS-7 cells 24 h posttransfection using the RNeasy kit (Qiagen). The RNA was reverse-transcribed with oligo (dT) primer and Superscript II RNase H- reverse transcriptase (Invitrogen) followed by PCR amplification using vector-specific primers (SD6 and SA2 provided by the supplier). To minimize the recovery of products due to cryptic or vector-self splicing, BSTXI digestion of the PCR mixture was performed. Exon trapping libraries were established by subcloning the BSTXI digested PCR products of the transcribed sequences into pAMP10 plasmid vector using the UDG cloning system (Gibco/Life Technologies, now Invitrogen) as recommended by the manufacturer.

Plasmid DNA from recombinant clones was isolated from each exon trapping library and sequenced with vector-cassette-specific primers SD2 and SA4 using ABI-Prism BigDye terminator cycle sequencing kit (Applied Biosystems). Prior to sequencing we included amplification of the inserts of recombinant clones directly from the colo-

nies with primers SD2 and SA4 as a preselection step to minimize sequencing of small exon trapping products. PCR products were analyzed by agarose gel electrophoresis. PCR products exceeding the amplicon size of the empty plasmid vector (>230 bp) were subjected to sequencing. Sequences were compared among themselves and with the vector sequence to identify non-redundant genuine trapped exons.

Primers were designed for unique exon trapped sequences (ETS) exceeding a size of 60 bp (see Additional file 1) and tested for successful PCR amplification on a panel of genomic DNA from their parent BAC clones to verify the genomic origin of the trapped sequences.

Unique ETS were deposited in the Genbank nucleotide database. Accession numbers are given in Additional file 1.

#### **Analysis and characterization of identified exon trapped sequences**

##### *Sequence similarity search and in silico mapping on the bovine genome sequence*

ETS from BACs specific to BTA6 were used for sequence similarity screening in the currently available 7.15× coverage sequence of the bovine genome at NCBI [61]. The similarity search was performed in the NCBI database [genome (reference only)] using the Basic Local Alignment Search Tool (MegaBLAST) with default parameters. Map positions were retrieved from the current bovine genome sequence assembly map of BTA6 at NCBI (Btau build 4.0 [31]). Generally, each match found in the sequence databases was manually curated. Sequence matches were excluded when revealing less than 95% identity for blast searches.

##### *Identification of comparative similarity to human and mouse genomic sequences*

In addition to homology search in the bovine genome sequence, ETS were applied to similarity search against human and mouse genomic sequences of the NCBI database [genome (all assemblies)] [62,63] using the cross-species MegaBLAST tool. Comparative sequence similarity screening was performed with low complexity filter and default parameters. Sequence matches were accepted with at least 80% identity across species.

##### *Identification of similarity to expressed and coding sequences*

Analysis of ETS for protein coding sequences was performed by using the ORF Finder program [64]. Protein and transcript sequence similarity searches were performed at NCBI [65] using BLASTX against the non-redundant protein sequence database (nr) and TBLASTX against the nucleotide collection (nr/nt) and the expressed sequence tags (est) databases with default

parameters. Furthermore, transcribed sequences inferred from the ORF Finder analysis were applied to screen the nucleotide collection (nr/nt) and the expressed sequence tags (est) databases using the TBLASTN algorithm. Identity to known bovine sequences was accepted at a threshold of 95% and with more than 70% in the case of non-bovine sequences.

##### *Sequence similarity search for non-protein coding RNA*

The ETS were analyzed for similarity to known noncoding RNA by screening the comprehensive mammalian non-coding RNA database (RNAdb 2.0, all ncRNA Datasets) using the BLAST algorithm [32].

##### *Expression analysis*

Expression patterns of identified ETS in cattle were analyzed by RT-PCR on a multi-tissue panel. Total RNA was extracted from nine bovine tissues (liver, small intestine, kidney, thyroid gland, mammary gland, pituitary gland, skeletal muscle, subcutaneous and intestinal fat) isolated from a lactating German Holstein cow. Principally, total RNA was prepared using the Nucleospin II RNA kit (Macherey & Nagel). In case of fat tissues, the RNeasy lipid kit (Qiagen) was applied. Genomic DNA was carefully eliminated by repeated on-column digestion using twice the amount of RNase-free DNaseI solution the manufacturers recommended in the protocols. The cDNAs were synthesized using the Superscript II RNase H- reverse transcriptase system for first strand cDNA synthesis (Invitrogen,) with oligo (dT) primer according to the manufacturer's instructions.

Quality check-up of the prepared cDNA for contamination with genomic DNA was performed by PCR using primers spanning intron 4 of the bovine PPARGC1A gene. Primers, forward 5'-AAGAAGCTCTTACTGGCACC-3' and reverse 5'-ATGTTGTGTCTGCGATTGTG-3', generated fragments of 318 bp in cDNA and 1177 bp in genomic DNA [66]. Expression of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as a control gene was analyzed in cDNA of all tissues (forward primer: 5'-TACATGGTCTACATGTTCCAGTATG-3', and reverse primer: 5'-CAGTCTTCTGGGTGGCAGTGATG-3', amplicon: 440 bp). Tissue-specific expression profiling of the identified putative transcripts was performed with primers given in Additional file 1. Bovine genomic DNA was tested to assess amplification on genome level.

#### **Authors' contributions**

RW performed BAC library screening, exon trapping, analyses on genome sequence databases, and expression analyses. TG performed FISH mapping of BAC clones. AE supported database and annotation analyses. RW and CK wrote the manuscript. All authors read and approved the final manuscript.

## Additional material

### Additional file 1

*Unique putative exon trapped sequences (ETS) from BAC clones mapped to bovine chromosome 6 (BTA6). The data provided represent BTA6 markers used for BAC library screening, BAC clones subjected to exon trapping and details of the isolated ETS (e.g., accession numbers, primers). Locus: region-specific sequence (markers) used for BAC library screening.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-186-S1.pdf>]

### Additional file 2

*Sequence similarity of identified unique exon trapping sequences (ETS) specific to bovine chromosome 6 (BTA6). Genomic positions on BTA6 and insert sizes of BAC clones were determined according to the bovine genome assembly Btau4.0 at NCBI by in silico screening of BAC end sequences as described previously [9]. Chromosomal positions of ETS were determined by in silico sequence similarity screening to the bovine genome sequence database and chromosome alignment to the bovine genome assembly Btau4.0 at NCBI [31,61]. Comparative sequence similarity of ETS was determined by screening across sequence resources of the human and mouse genomes [62,63]. Similarity to ESTs from Ovis aries and Sus scrofa and to predicted Bos taurus gene models were inferred from the NCBI MapViewer [31] including the additional map options rnaOar, rnaSsc and ab initio. Column 5: UN: unassigned. Column 7: Altern. assembly: human genome assembly based on Celera assembly.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-186-S2.pdf>]

## Acknowledgements

We thank Astrid Kühl, Anne Berndt, and Marlies Fuchs for their expert technical assistance. Genzyme Corp., Framington, MA is thankfully acknowledged for providing the exon trapping vector pSPL3b. This work was supported by the grants We1786/2-2 and We1786/2-3 from the German Research Council (Deutsche Forschungsgemeinschaft, DFG).

## References

1. **Reprogen QTL Map of Dairy Cattle Traits** [[http://www.vet.sci.usyd.edu.au/reprogen/QTL\\_Map/](http://www.vet.sci.usyd.edu.au/reprogen/QTL_Map/)]
2. **Bovine QTL Viewer at Texas A&M University** [<http://genomes.sapac.edu.au/bovineqtl/>]
3. **Cattle QTL database** [<http://www.animalgenome.org/QTLdb/cattle.html>]
4. Itoh T, Watanabe T, Ihara N, Mariani P, Beattie CW, Sugimoto Y, Takasuga A: **A comprehensive radiation hybrid map of the bovine genome comprising 5593 loci.** *Genomics* 2005, **85**:413-424.
5. Jann OC, Aerts J, Jones M, Hastings N, Law A, McKay S, Marques E, Prasad A, Yu J, Moore SS, Floriot S, Mahe MF, Eggen A, Silveri L, Negrini R, Milanese E, Jimone-Marsan P, Valentini A, Marchitelli C, Savarese MC, Janitz M, Herwig R, Hennig S, Gorni C, Connor EE, Sonstegard TS, Smith T, Drogemuller C, Williams JL: **A second generation radiation hybrid map to aid the assembly of the bovine genome sequence.** *BMC Genomics* 2006, **7**:283.
6. McKay SD, Schnabel RD, Murdoch BM, Aerts J, Gill CA, Gao C, Li C, Matukumalli LK, Stothard P, Wang Z, Van Tassell CP, Williams JL, Taylor JF, Moore SS: **Construction of bovine whole-genome radiation hybrid and linkage maps using high-throughput genotyping.** *Anim Genet* 2007, **38**:120-125.
7. Weikard R, Kühn C, Goldammer T, Laurent P, Womack JE, Schwerin M: **Targeted construction of a high-resolution, integrated, comprehensive, and comparative map for a region specific to bovine chromosome 6 based on radiation hybrid mapping.** *Genomics* 2002, **79**:768-776.
8. Weikard R, Goldammer T, Laurent P, Womack JE, Kuehn C: **A gene-based high-resolution comparative radiation hybrid map as a framework for genome sequence assembly of a bovine chromosome 6 region associated with QTL for growth, body composition, and milk performance traits.** *BMC Genomics* 2006, **7**:53.
9. Weikard R, Goldammer T, Womack JE, Kuehn C: **Forward to a detailed sequence map of bovine chromosome 6.** *Cytogenet Genome Res* 2007, **116**:72-79.
10. Everts-van der Wind A, Larkin DM, Green CA, Elliott JS, Olmstead CA, Chiu R, Schein JE, Marra MA, Womack JE, Lewin HA: **A high-resolution whole-genome cattle-human comparative map reveals details of mammalian chromosome evolution.** *Proc Natl Acad Sci USA* 2005, **102**:18526-18531.
11. Zhang Z, Pang AWC, Gerstein M: **Comparative analysis of genome tiling array data reveals many novel primate-specific functional RNAs in human.** *BMC Evol Biol* 2007, **7**(suppl 1):S14.
12. Kumar CG, Larson JH, Band MR, Lewin HA: **Discovery and characterization of 91 novel transcripts expressed in cattle placenta.** *BMC Genomics* 2007, **8**:113.
13. Larson JH, Kumar CG, Everts RE, Green CA, Everts-van der Wind A, Band MR, Lewin HA: **Discovery of eight novel divergent homologs expressed in cattle placenta.** *Physiol Genomics* 2006, **25**:405-413.
14. Degrelle SA, Champion E, Cabau C, Piumi F, Renaud P, Richard C, Renard JP, Hue I: **Molecular evidence for a critical period in mural trophoblast development in bovine blastocysts.** *Dev Biol* 2005, **288**:448-460.
15. Ghanem N, Holker M, Rings F, Jennen D, Tholen E, Sirard MA, Torner H, Kanitz W, Schellander K, Tesfaye D: **Alterations in transcript abundance of bovine oocytes recovered at growth and dominance phases of the first follicular wave.** *BMC Dev Biol* 2007, **7**:90.
16. Roberts RM, Ealy AD, Alexenko AP, Han CS, Ezashi T: **Trophoblast interferons.** *Placenta* 1999, **20**:259-264.
17. Holtenius P, Holtenius K: **New aspects of ketone bodies in energy metabolism of dairy cows: A review.** *J Vet Med A Physiol Pathol Clin Med* 1996, **43**:579-587.
18. Bell AW, Bauman DE: **Adaptations of glucose metabolism during pregnancy and lactation.** *J Mammary Gland Biol Neoplasia* 1997, **2**:265-278.
19. Vernon RG, Pond CM: **Adaptations of maternal adipose tissue to lactation.** *J Mammary Gland Biol Neoplasia* 1997, **2**:231-241.
20. Vernon RG: **Lipid metabolism during lactation: a review of adipose tissue-liver interactions and the development of fatty liver.** *J Dairy Res* 2005, **72**:460-469.
21. Chagas LM, Bass JJ, Blache D, Burke CR, Kay JK, Lindsay DR, Lucy MC, Martin GB, Meier S, Rhodes FM, Roche JR, Thatcher WW, Webb R: **New perspectives on the roles of nutrition and metabolic priorities in the subfertility of high-producing dairy cows.** *J Dairy Sci* 2007, **90**:4022-4032.
22. Buckler AJ, Chang DD, Graw SL, Brook JD, Haber DA, Sharp PA, Housman DE: **Exon Amplification – A Strategy to Isolate Mammalian Genes Based on RNA Splicing.** *Proc Natl Acad Sci USA* 1991, **88**:4005-4009.
23. Church DM, Stotler CJ, Rutter JL, Murrell JR, Trofatter JA, Buckler AJ: **Isolation of Genes from Complex Sources of Mammalian Genomic DNA Using Exon Amplification.** *Nature Genet* 1994, **6**:98-105.
24. Burn TC, Connors TD, Klinger KW, Landes GM: **Increased Exon-Trapping Efficiency Through Modifications to the pSPL3 Splicing Vector.** *Gene* 1995, **161**:183-187.
25. Korn B, Sedlacek Z, Kioschis P, Konecki D, Poustka A: **Establishment of An Xq27.3-Xqter Specific cDNA Library Using cDNA Selection.** *Cytogenet Cell Genet* 1994, **67**:356.
26. Sedlacek Z, Korn B, Konecki DS, Siebenhaar R, Coy JF, Kioschis P, Poustka A: **Construction of A Transcription Map of A 300 Kb Region Around the Human G6PD Locus by Direct cDNA Selection.** *Hum Mol Genet* 1993, **2**:1865-1869.
27. Yaspo ML, Gellen L, Mott R, Korn B, Nizetic D, Poustka AM, Lehrach H: **Model for A Transcription Map of Human-Chromosome-21 –**

- Isolation of New Coding Sequences from Exon and Enriched cDNA Libraries.** *Hum Mol Genet* 1995, **4**:1291-1304.
28. Lovett M, Kere J, Hinton LM: **Direct Selection – A Method for the Isolation of cDNAs Encoded by Large Genomic Regions.** *Proc Natl Acad Sci USA* 1991, **88**:9628-9632.
  29. Delmastro RG, Wang LP, Simmons AD, Gallardo TD, Clines GA, Ashley JA, Hilliard CJ, Wasmuth JJ, McPherson JD, Lovett M: **Human Chromosome-Specific cDNA Libraries – New Tools for Gene Identification and Genome Annotation.** *Genome Res* 1995, **5**:185-194.
  30. Parimoo S, Patanjali SR, Shukla H, Chaplin DD, Weissman SM: **cDNA Selection – Efficient PCR Approach for the Selection of cDNAs Encoded in Large Chromosomal DNA Fragments.** *Proc Natl Acad Sci USA* 1991, **88**:9623-9627.
  31. **NCBI Bos taurus Map Viewer of BTA6** [<http://www.ncbi.nlm.nih.gov/projects/mapview/maps.cgi?taxid=9913&chr=6>]
  32. **RNAdb 2.0, BLAST Search** [<http://jsm-research.imb.uq.edu.au/rnadblseqsearch/default.aspx>]
  33. Tellam RL: **Capturing benefits from the bovine genome sequence.** *Aust J Exp Agr* 2007, **47**:1039-1050.
  34. Hillier LV, Graves TA, Fulton RS, Fulton LA, Pepin KH, Minx P, Wagner-McPherson C, Layman D, Wylie K, Sekhon M, Becker MC, Fewell GA, Delehaunty KD, Miner TL, Nash WE, Krenitzki C, Oddy L, Du H, Sun H, Bradshaw-Cordum H, Ali J, Carter J, Cordes M, Harris A, Isak A, van Brunt A, Nguyen C, Du FY, Courtney L, Kalicki J, Ozersky P, Abbott S, Armstrong J, Belter EA, Caruso L, Cedroni M, Cotton M, Davidson T, Desai A, Elliott G, Erb T, Fronick C, Gaige T, Haakenson W, Haglund K, Holmes A, Harkins R, Kim K, Kruchowski SS, Strong CM, Grewal N, Goyea E, Hou S, Levy A, Martinka S, Mead K, McLellan MD, Meyer R, Maher JR, Tomlinson C, Kohlberg SD, Kozlowski-Reilly A, Shah N, Swearingen-Shahid S, Snider J, Strong JT, Thompson J, Yoakum M, Leonard S, Pearman C, Trani L, Radionenko M, Waligorski JE, Wang CY, Rock SM, Tin-Wollam AM, Maupin R, Latreille P, Wendl MC, Yang SP, Pohl C, Wallis JW, Spieth J, Bieri TA, Berkowicz N, Nelson JO, Osborne J, Ding L, Meyer R, Sabo A, Shotland Y, Sinha P, Wohldmann PE, Cook LL, Hickenbotham MT, Eldred J, Williams D, Jones TA, She XW, Ciccarelli FD, Izaurralde E, Taylor J, Schmutz J, Myers RM, Cox DR, Huang XQ, McPherson JD, Mardis ER, Clifton SW, Warren WC, Chinwalla AT, Eddy SR, Marra MA, Ovcharenko I, Furey TS, Miller W, Eichler EE, Bork P, Suyama M, Torrents D, Waterston RH, Wilson RK: **Generation and annotation of the DNA sequences of human chromosomes 2 and 4.** *Nature* 2005, **434**:724-731.
  35. Ovcharenko I, Loots GG, Nobrega MA, Hardison RC, Miller W, Stubbs L: **Evolution and functional classification of vertebrate gene deserts.** *Genome Res* 2005, **15**:137-145.
  36. Gerstein MB, Bruce C, Rozowsky JS, Zheng DY, Du J, Korbel JO, Emanuelsson O, Zhang ZDD, Weissman S, Snyder M: **What is a gene, post-ENCODE? History and updated definition.** *Genome Res* 2007, **17**:669-681.
  37. Zheng DY, Gerstein MB: **The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they?** *Trends Genet* 2007, **23**:219-224.
  38. Seemann SE, Gilchrist MJ, Hofacker IL, Stadler PF, Gorodkin J: **Detection of RNA structures in porcine EST data and related mammals.** *BMC Genomics* 2007, **8**:316.
  39. Choy KW, Wang CC, Ogura A, Lau TK, Rogers MS, Ikeo K, Gojobori T, Lam DSC, Pang CP: **Genomic annotation of 15,809 ESTs identified from pooled early gestation human eyes.** *Physiol Genomics* 2006, **25**:9-15.
  40. Peters BA, Croix BS, Sjoblom T, Cummins JM, Silliman N, Ptak J, Saha S, Kinzler KW, Hatzis C, Velculescu VE: **Large-scale identification of novel transcripts in the human genome.** *Genome Res* 2007, **17**:287-292.
  41. Carninci P, Hayashizaki Y: **Noncoding RNA transcription beyond annotated genes.** *Current Opinion in Genetics & Development* 2007, **17**:139-144.
  42. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng ZP, Snyder M, Dermitzakis ET, Stamatoyannopoulos JA, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SCJ, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Dutta A, Guigo R, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng DY, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Flicek P, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermuller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu JQ, Lian Z, Lian J, Newburger P, Zhang XQ, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Dermitzakis ET, Margulies EH, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan YJ, Snyder M, Birney E, Struhl K, Gerstein M, Antonarakis SE, Gingeras TR, Brown JB, Flicek P, Fu YT, Keefe D, Birney E, Denoeud F, Gerstein M, Green ED, Kapranov P, Karaoz U, Myers RM, Noble WS, Reymond A, Rozowsky J, Struhl K, Siepel A, Stamatoyannopoulos JA, Taylor CM, Taylor J, Thurman RE, Tullius TD, Washietl S, Zheng DY, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Collins FS, Margulies EH, Cooper GM, Asimenos G, Thomas DJ, Dewey CN, Siepel A, Birney E, Keefe D, Hou MM, Taylor J, Nikolaev S, Montoya-Burgos JI, Loytynoja A, Whelan S, Pardi F, Massingham T, Brown JB, Huang HY, Zhang NR, Bickel P, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Sringhaus M, Church D, Rosenbloom K, Kent WJ, Stone EA, Gerstein M, Antonarakis SE, Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Pachter L, Green ED, Sidow A, Weng ZP, Trinklein ND, Fu YL, Zhang ZDD, Karaoz U, Barrera L, Stuart R, Zheng DY, Ghosh S, Flicek P, King DC, Taylor J, Ameur A, Enroth S, Bieda MC, Koch CM, Hirsch HA, Wei CL, Cheng J, Kim J, Bhing AA, Giresi PG, Jiang N, Liu J, Yao F, Sung WK, Chiu KP, Vega VB, Lee CWH: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.** *Nature* 2007, **447**:799-816.
  43. Pheasant M, Mattick JS: **Raising the estimate of functional human sequences.** *Genome Res* 2007, **17**:1245-1253.
  44. Soares LMM, Valcarcel J: **The expanding transcriptome: the genome as the 'Book of Sand'.** *EMBO J* 2006, **25**:923-931.
  45. Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G, Sementchenko V, Piccolboni A, Bekiranov S, Bailey DK, Ganesh M, Ghosh S, Bell I, Gerhard DS, Gingeras TR: **Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution.** *Science* 2005, **308**:1149-1154.
  46. Gingeras TR: **Origin of phenotypes: Genes and transcripts.** *Genome Res* 2007, **17**:682-690.
  47. Mattick JS: **Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms.** *Bioessays* 2003, **25**:930-939.
  48. Mattick JS, Makunin IV: **Non-coding RNA.** *Hum Mol Genet* 2006, **15**:R17-R29.
  49. Carninci P: **Constructing the landscape of the mammalian transcriptome.** *J Exp Biol* 2007, **210**:1497-1506.
  50. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, Kodzius R, Shimokawa K, Bajic VB, Brenner SE, Batalov S, Forrest ARR, Zavolan M, Davis MJ, Wilming LG, Aidinis V, Allen JE, mbesi-Impombato X, Apweiler R, Aturaliya RN, Bailey TL, Bansal M, Baxter L, Beisel KW, Bersano T, Bono H, Chalk AM, Chiu KP, Choudhary V, Christoffels A, Clutterbuck DR, Crowe ML, Dalla E, Dalrymple BP, de Bono B, la Gatta G, di Bernardo D, Down T, Engstrom P, Fagiolini M, Faulkner G, Fletcher CF, Fukushima T, Furuno M, Futaki S, Gariboldi M, Georgii-Hemming P, Gingeras TR, Gojobori T, Green RE, Gustincich S, Harbers M, Hayashi Y, Hensch TK, Hirokawa N, Hill D, Huminicki L, Iacono M, Ikeo K, Iwama A, Ishikawa T, Jakt M, Kanapin A, Katoh M, Kawasawa Y, Kelso J, Kitamura H, Kitano H, Kollias G, Krishnan SPT, Kruger A, Kummerfeld SK, Kurochkin IV, Lareau LF, Lazarevic D, Lipovich L, Liu J, Liuni S, McWilliam S, Babu MM, Madera M, Marchionni L, Matsuda H, Matsuzawa S, Miki H, Mignone F, Miyake S, Morris K, Mottagui-Tabar S, Mulder N, Nakano N, Nakauchi H, Ng P, Nilsson R, Nishiguchi S, Nishikawa S, Nori F, Ohara O, Okazaki Y, Orlando V, Pang KC, Pavan WJ, Pavesi G, Pesole G, Petrovsky N, Piazza S, Reed J, Reid JF, Ring BZ, Ringwald M, Rosol B, Ruan Y, Salzberg SL, Sandelin

- A, Schneider C, Schonbach C, Sekiguchi K, Semple CAM, Seno S, Sessa L, Sheng Y, Shibata Y, Shimada H, Shimada K, Silva D, Sinclair B, Sperling S, Stupka E, Sugiura K, Sultana R, Takenaka Y, Taki K, Tammoja K, Tan SL, Tang S, Taylor MS, Tegner J, Teichmann SA, Ueda HR, van Nimwegen E, Verardo R, Wei CL, Yagi K, Yamanishi H, Zabarovsky E, Zhu S, Zimmer A, Hide W, Bult C, Grimmond SM, Teasdale RD, Liu ET, Brusica V, Quackenbush J, Wahlestedt C, Mattick JS, Hume DA, Kai C, Sasaki D, Tomaru Y, Fukuda S, Kanamori-Katayama M, Suzuki M, Aoki J, Arakawa T, Iida J, Imamura K, Itoh M, Kato T, Kawaji H, Kawagashira N, Kawashima T, Kojima M, Kondo S, Konno H, Nakano K, Ninomiya N, Nishio T, Okada M, Plessy C, Shibata K, Shiraki T, Suzuki S, Tagami M, Waki K, Watahiki A, Okamura-Oho Y, Suzuki H, Kawai J, Hayashizaki Y, FANTOM Consortium, RIKEN Genome Explorat Res Grp: **The transcriptional landscape of the mammalian genome.** *Science* 2005, **309**:1559-1563.
51. Glazier AM, Nadeau JH, Aitman TJ: **Finding genes that underlie complex traits.** *Science* 2002, **298**:2345-2349.
  52. Georges M: **Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals.** *Annu Rev Genomics Hum Genet* 2007, **8**:131-62.
  53. Stark A, Lin MF, Kheradpour P, Pedersen JS, Parts L, Carlson JW, Crosby MA, Rasmussen MD, Roy S, Deoras AN, Ruby JG, Brennecke J, Hodges E, Hinrichs AS, Caspi A, Park SW, Han MV, Maeder ML, Polansky BJ, Robson BE, Aerts S, van Helden J, Hassan B, Gilbert DG, Eastman DA, Rice M, Weir M, Hahn MW, Park Y, Dewey CN, Pachter L, Kent WJ, Haussler D, Lai EC, Bartel DP, Hannon GJ, Kaufman TC, Eisen MB, Clark AG, Smith D, Celniker SE, Gelbart WM, Kellis M: **Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures.** *Nature* 2007, **450**:219-232.
  54. Siepel A, Diekhans M, Brejova B, Langton L, Stevens M, Comstock CLG, Davis C, Ewing B, Oommen S, Lau C, Yu HC, Li JF, Roe BA, Green P, Gerhard DS, Temple G, Haussler D, Brent MR: **Targeted discovery of novel human exons by comparative genomics.** *Genome Res* 2007, **17**:1763-1773.
  55. **BovineGenome.Org** [[http://genomes.arc.georgetown.edu/bovine/annotation\\_information.html](http://genomes.arc.georgetown.edu/bovine/annotation_information.html)]
  56. **ImaGenes** [<http://www.imagenes-bio.de/>]
  57. Weikard R, Goldammer T, Kühn C, Barendse W, Schwerin M: **Targeted development of microsatellite markers from the defined region of bovine chromosome 6q21-31.** *Mamm Genome* 1997, **8**:836-840.
  58. Brunner RM, Sanftleben H, Goldammer T, Kühn C, Weikard R, Kata SR, Womack JE, Schwerin M: **The telomeric region of BTA18 containing a potential QTL region for health in cattle exhibits high similarity to the HSA19q region in humans.** *Genomics* 2003, **81**:270-278.
  59. Brennan MB, Hochgeschwender U: **So Many Needles, So Much Hay.** *Hum Mol Genet* 1995, **4**:153-156.
  60. Burn TC, Connors TD, VanRaay TJ, Dackowski WR, Millholland JM, Klinger KW, Landes GM: **Generation of a transcriptional map for a 700-kb region surrounding the polycystic kidney disease type 1 (PKD1) and tuberous sclerosis type 2 (TSC2) disease genes on human chromosome 16p13.3.** *Genome Res* 1996, **6**:525-537.
  61. **NCBI Bos taurus BLAST Search Home Page** [<http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9913>]
  62. **NCBI Human BLAST Search Home Page** [<http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9606>]
  63. **NCBI Mouse BLAST Search Home Page** [<http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=10090>]
  64. **ORF Finder** [<http://www.ncbi.nlm.nih.gov/gorf/>]
  65. **NCBI BLASTN Search Site** [<http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi>]
  66. Weikard R, Kuhn C, Goldammer T, Freyer G, Schwerin M: **The bovine PPARGCIA gene: molecular characterization and association of an SNP with variation of milk fat synthesis.** *Physiol Genomics* 2005, **21**:1-13.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

