Genetics: Early Online, published on January 22, 2015 as 10.1534/genetics.114.169680

Title: **The mammalian cervical vertebrae blueprint depends on the T (*brachyury*) gene**

Andreas Kromik[*], Reiner Ulrich[§], Marian Kusenda[†], Andrea Tipold[‡], Veronika M. Stein[‡], Maren Hellige[**], Peter Dziallas[‡], Frieder Hadlich[*], Philipp Widmann[*], Tom Goldammer[*], Wolfgang Baumgärtner[§], Jürgen Rehage[†], Dierck Segelke[§§], Rosemarie Weikard[*], Christa Kühn[*,††]

**Affiliations**:

[*] Leibniz-Institute for Farm Animal Biology (FBN), Institute for Genome Biology, 18196 Dummerstorf, Germany

[§] Department of Pathology, University of Veterinary Medicine Hannover, 30559 Hannover, Germany

[†] Clinic for Cattle, University of Veterinary Medicine Hannover, 30173 Hannover, Germany

[‡] Department of Small Animal Medicine and Surgery, University of Veterinary Medicine Hannover, 30559 Hannover, Germany

[**] Clinic for Horses, University of Veterinary Medicine Hannover, 30559 Hannover, Germany

[§§] Vereinigte Informationssysteme Tierhaltung w.V. (vit), 27283 Verden, Germany

[‡‡] Faculty of Agricultural and Environmental Sciences, University Rostock, 18059 Rostock, Germany

**Running title:** *T* gene effects on basic mammalian blueprint

**Key words:** homeotic transformation, genetic defect, brachyury

**Corresponding author:**

Christa Kühn

Leibniz-Institute for Farm Animal Biology (FBN)

Institute for Genome Biology

Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany

Phone: +49 38208-68709

E-mail: kuehn@fbn-dummerstorf.de

## ABSTRACT

A key common feature of all but three known mammalian genera is the strict seven cervical vertebrae blueprint suggesting the involvement of strong conserving selection forces during mammalian radiation. This is further supported by reports indicating that children with cervical ribs die before they reach reproductive age. Hypotheses had been put up associating cervical ribs (homeotic transformations) to embryonal cancer (e.g., neuroblastoma) or ascribing the constraint in cervical vertebral count to the development of the mammalian diaphragm. Here, we describe a spontaneous mutation *c.196A>G* in the *Bos taurus T* gene (also known as *brachyury*) associated with a cervical vertebral homeotic transformation that violates the fundamental mammalian cervical blueprint, but does not preclude reproduction of the affected individual. Genome-wide mapping, haplotype tracking within a large pedigree, resequencing of target genome regions, and bioinformatic analyses unambiguously confirmed the mutant *c.196G* allele as causal for this previously unknown defect termed vertebral and spinal dysplasia (VSD) by providing evidence for the mutation event. The non-synonymous VSD mutation is located within the highly conserved T-box of the *T* gene, which plays a fundamental role in eumetazoan body organization and vertebral development. To our knowledge, VSD is the first unequivocally approved spontaneous mutation decreasing cervical vertebrae number in a large mammal. The spontaneous VSD mutation in the bovine *T* gene is the first *in vivo* evidence for the hypothesis that the T protein is directly involved in the maintenance of the mammalian seven-cervical vertebra blueprint. It therefore furthers our knowledge of the T protein function and early mammalian notochord development.

## INTRODUCTION

High evolutionary diversification of the vertebral column exists in vertebrates, but the number of cervical vertebrae within mammals has been fixed at seven for more than 200 million years of evolution since the beginning of the long and wide mammalian radiation (Hautier *et al.* 2010). The reason why all mammals share this fundamental blueprint of cervical vertebrae, compared with a more relaxed rule for the number of posterior vertebrae analogous to other non-mammalian vertebrates, remains unknown. Nevertheless, evolutionary and clinical data indicate that the cervical vertebral development of mammals is under high selection pressure. For example, in human pediatrics, 83% of children with a deviating number of cervical vertebrae die in their first year, while the surviving individuals do not reach reproductive age (Galis *et al.* 2006). A detailed knowledge of the key factors involved in the spatial regulation of vertebral development will help to understand these forces.

Mutation models, either spontaneous or artificially induced, can reveal the complex processes that occur during vertebral development. Vertebral and accompanied spinal defects are described for many species including cattle [e.g., Complex Vertebral Malformation; (Agerholm *et al.* 2001)] and are often associated with urogenital and intestinal malformations (van de Ven *et al.* 2011). This association is conclusive due to the coordinated processes of notochord and cloaca formation during embryonic development. Mutations associated with spinal and vertebral cord defects are large in number and are located in coding but also in regulatory regions of many transcription factors (e.g., Ptf1a (Vlangos *et al.* 2013)). The murine *brachyury* (*T*) gene with its mutant alleles was the first gene that was identified and positionally cloned based on a genetic defect only, the long-known *brachyury* resulting in vertebral and spinal defects (Dobrovolskaia-Zavadskaia 1927; Herrmann *et al.* 1990). Numerous subsequent studies confirmed that the coordinated expression of the *T* gene during gastrulation is essential for appropriate notochord, neural tube, and mesoderm development

(Chesley 1935; Pennimpede *et al.* 2012; Satoh *et al.* 2012). Recently, the *T* gene has gained interest because of its association with the human chordoma, a sporadic and hereditary tumor originating from relicts of the notochord (Nibu *et al.* 2013; Pillay *et al.* 2012; Yang *et al.* 2009). Thus, the *T* gene is a prime candidate for investigating phenotypic alterations of the vertebral column and spinal cord.

In 2010, early data emerged about newborn calves with short, crooked tails in the Holstein cattle breed, the most widespread dairy cattle breed worldwide (FAO 2007). For this innate defect subsequently termed "vertebral and spinal dysplasia" (VSD), initial clinical data had indicated tail malformations and genealogical analyses a dominant mode of inheritance (Kromik et al. submitted, see Supplemental file). The aim of this study was to provide the detailed VSD-associated phenotype, to confirm its genetic background and to decipher the causal mutation for the VSD defect. In our study, we comprehensively i) disclose the malformations and neurological dysfunctions accompanied with VSD, ii) confirm a genetic origin and the mode of inheritance for VSD, iii) reveal the causal mutation in the *T* gene and the founder individual for the defect, and iv) indicate the functional relevance of the mutated nucleotide. Our study is the first report on a spontaneous mutation inducing a deviation from the fundamental seven-cervical-vertebrae blueprint in mammals and extends our knowledge on the functional relevance of the *T* gene regarding neuro-skeletal development.

## MATERIALS AND METHODS

**Animals:** The study included registered herdbook individuals with documented ancestry from the German Holstein dairy cattle population. From an initial on-farm screening for VSD-affected individuals (Kromik et al. submitted), we selected six calves of different ages and with different degrees of the congenital VSD associated tail defects (Table S1) for specific, detailed examinations by specifically trained experts in several specialized units of the University of Veterinary Medicine Hannover (Germany). This included i) an in-depth clinical/physical and neurological investigation (including electromyography (EMG) and motor nerve conduction velocity (mNCV)), ii) a radiological documentation involving X-rays, and CT and MRI scans with a focus on the spinal cord and vertebral column, iii) a post-mortem examination and iv) comprehensive laboratory diagnostic analyses of blood, cerebrospinal fluid (CSF) and serum (Table S2).

In addition, sire FBF0666 aging four years at the time of our study was included in phenotypic analyses, because although he had not shown any signs of a VSD phenotype at one year of age, but showed increasing locomotion problems with age, analogous to other reports from farmers of affected calves. For genetic analyses, from the initial on-farm monitoring (Kromik et al. submitted) individuals from 39 farms were included comprising 85 offspring of the VSD carrier sire FBF0666 (41 classified as VSD affected, 34 classified as non-VSD affected and 10 with ambiguous VSD classification) and 41 control individuals. (Table S3). Control calves were all classified as non-VSD affected and matched to target calves with respect to age, sex, housing conditions, and farm. Furthermore, we included the dams of the target calves, the carrier sire of the VSD defect (FBF0666), its ancestors and relatives covering eight generations, as well as 402 randomly selected Holstein and 126 Holstein x Charolais VSD-unaffected calves originating from 110 different sires.

**Ethics Statement:** All experimental procedures were carried out according to the German animal care guidelines and were supervised by the relevant authorities of the States Mecklenburg-Vorpommern and Niedersachsen, Germany.

**Characterization of the VSD Phenotype:** In addition to the standard bovine necropsy protocol, specific attention was given to those body compartments reported to be associated with vertebral defects and gait alterations in the literature (including the number and shape of vertebrae, the skull, peripheral nerves, limb bones, and muscular samples). The complete vertebral cord was meticulously examined, sampled, and partly macerated for final documentation.

To exclude an effect of epizootic virus diseases that might be involved in the observed congenital defects, tissue samples were investigated for virus antigens of Bovine virus diarrhea virus, Bovine herpes virus 1, and Bluetongue virus at the State Laboratory of the Department of Consumer and Food Safety of Lower-Saxony, Hannover, Germany.

For histopathological examination, samples taken during necropsy included the thymus, heart, lung, pancreas, kidney, bladder, genital apparatus, rumen, abomasum, small and large intestine, liver, spleen, lymphatic organs, muscles, bones, the central and peripheral nervous system, and endocrine organs. All samples were examined by light microscopy after hematoxylin-eosin staining. Furthermore, the spinal cord was investigated by additional histochemical assays: i) Luxol Fast Blue-Cresyl Echt Violet (myelin), ii) Azan and Masson-Goldner (collagenous and reticular fibers), and iii) Bielschowsky (neurofilaments). Additionally, the expression pattern of selected antigens was monitored by immunohistochemistry including i) glial fibrillary acidic protein (GFAP), ii) myelin basic protein (MBP), iii) amyloid precursor protein (APP), iv) factor VII related antigen, and v) vimentin. Histochemistry and immunohistochemistry were performed according to Ulrich and colleagues (Ulrich *et al.* 2010) .

**Karyotyping:** The karyotypes of the carrier sire and one severely affected offspring were investigated to identify chromosomal aneuploidy or translocation. Blood samples were taken and metaphase chromosomes were prepared according to standard procedures (Popescu *et al.* 2000). Chromosome morphology was visualized after Giemsa staining by light microscopy.

**Genetic Mapping of the VSD Locus:** For genotyping, blood/sperm samples from sire FBF0666, its dam FBF0266, its sire FBF0667, maternal grandsire FBF0669 and from all 126 calves included in the clinical and epidemiological survey and 73 dams were included. All DNA samples were genotyped with the BovineSNP50 v2 BeadChip (Illumina, San Diego, CA, USA) and analyzed using Genome Studio (Illumina, San Diego, CA, USA) software. SNPs were filtered for call frequency >0.97. All SNPs with heterozygote excess (deviation from Hardy-Weinberg equilibrium identified by p($\chi$2HWE) <0.05), gene train score <0.6, or minor allele frequency <0.01 were manually checked. Only those samples with a call rate >0.98 without pedigree conflicts were included in subsequent analyses.

Initial twopoint linkage mapping between each of the SNPs and the VSD locus was performed in the half-sibship originating from sire FBF0666. An autosomal dominant inheritance was assumed as indicated by the initial epidemiological analysis (Kromik et al., submitted) and an equal distribution of VSD cases across both sexes in the FBF0666 sibship. Consequently, the VSD locus was coded as heterozygous "1/2" in sire FBF0666 and all affected offspring, whereas all dams (assumed to be non-affected) and non-affected offspring were coded as homozygous "1/1". Mapping was carried out along the entire autosomal genome (BTA1 to BTA29) with the TWOPOINT option of CRIMAP version 2.50 (Green *et al.* 1990) incorporating modifications by Ian Evans and Jill Maddox (University of Melbourne, Australia).

After obtaining a strong indication of the genomic position of the VSD locus on BTA9, a multipoint mapping approach was conducted using MERLIN version 1.1.2 (Abecasis *et al.*

2002) with the affected code assigned to all VSD-affected offspring and sire FBF0666, and the non-affected status assigned to all dams and those offspring categorized as non-affected. For this purpose, a BTA9 marker map required for multipoint mapping was established with CRIMAP CHROMPIC options from the genotypes in the half-sib family. Markers with identical genetic positions were artificially separated by 0.001 cM to enable the running of the multipoint algorithm implemented in MERLIN. To account for potential incomplete penetrance of the defect, a 0.2, 0.6, and 1.0 penetrance of an autosomal dominant defect was modeled.

**Haplotyping:** All available offspring of sire FBF0666 were haplotyped for BTA9 using CRIMAP CHROMPIC options. After extracting the paternally inherited haplotype of each FBF0666 offspring, these haplotypes were aligned to identify the chromosomal segment shared by all VSD-affected individuals. All physical positions of SNPs and haplotype borders were indicated according to the bovine genome assembly UMD3.1 (Zimin *et al.* 2009).

To further trace the origin of the haplotype associated with VSD, we subsequently haplotyped all available dams and the FBF0666 ancestors in the German Holstein population using BEAGLE (Browning & Browning 2009). Haplotyping included a total of 55,384 individuals from the Holstein population with BovineSNP50Illumina SNP-Chip genotype information provided by VIT Verden (http://www.vit.de/index.php?id=milchrinder-zws-online&L=1), the central database for genomic evaluation in German Holstein cattle.

**Resequencing of the Candidate Locus:** The *T* gene was resequenced for a potentially causal mutation in VSD-affected and non-affected calves, in sire FBF0666, in the parents of sire FBF0666, and also in the maternal grandsire of sire FBF0666. All primers used for sequencing the *T* gene are indicated in Table S4. The obtained sequences were aligned to the mRNA reference sequence (http://www.ncbi.nlm.nih.gov/nuccore/NM_001192985) and the respective genomic sequence (http://www.ncbi.nlm.nih.gov/nuccore/AC_000166.1).

**Population Screening for the Causal Mutation:** We genotyped 94 sons of FBF0669, the sire FBF0666's maternal grandsire, at the *T c.196A>G* polymorphism to further confirm its causal characteristics and to validate the founder individual of the VSD mutation. All 94 offspring were sires themselves with at least 200 offspring each and with no report suggesting VSD cases in the first-generation descendants of these bulls. In addition, 39 of the VSD-unaffected control calves, 402 randomly selected purebred Holstein and 126 Holstein x Charolais crossbred calves were genotyped. All calves had shown no indication of VSD upon physical examination. For genotyping, a KASP assay addressing the mutation *T c.196A>G* was developed (LGC Genomics, KBiosience, Hoddesdon, UK). Genotyping was performed in a 10 µl reaction solution using 20 ng DNA on a Lightcycler 480 (Roche Applied Science, Mannheim, Germany) according to the manufacturer's recommendation for KASP assays but with the exception of an increase in $MgCl_2$ concentration by 0.3 mM (for primers see Table S4).

**Bioinformatic Analyses:** The wild-type and mutated (VSD) amino acid sequences of the bovine T protein were submitted for 3D protein structure prediction using Phyre2 (http://www.sbg.bio.ic.ac.uk/~phyre2/html/ page.cgi?id=index, (Kelley & Sternberg 2009)). To further predict the functional effects of the non-synonymous c.196A>G transition, wild-type and mutated (VSD) amino acid sequences of the bovine brachyury T were also submitted to Polyphen2 analysis (http://genetics.bwh.harvard.edu/pph2/, (Adzhubei *et al.* 2010)). Analysis of sequence homology across species was performed by Homologene (http://www.ncbi.nlm.nih.gov/homologene).

## RESULTS

### VSD is Characterized by a Variable Number of Vertebrae and Neurological Deficits

Radiological examination (X-ray, computed tomography (CT) and magnetic resonance imaging (MRI)) and necropsy of calves with divergent degrees of clinical VSD-associated tail malformations confirmed that the calves shared vertebral defects, including dysplasia and numerical aberrations in all parts of the spine except the sacrum (Table S1). The most striking feature was the cervical homeotic transformation (Figure 1) resulting in reduction of the cervical vertebrae number in four of the six necropsied calves. In addition to malformations of the vertebral column, variably expressed defects of the spinal cord restricted to the lumbosacral segment were found including syringomyelia (mostly accompanied with hydromyelia), diplomyelia, a duplicated central canal, and segmental hypoplasia (Figure 1, Figure S1, Table S1). The double central channel and the diplomyelia were exclusively observed in the sacral segment of the spinal cord and suggest a duplication event during neural development. Histochemistry and immunohistochemistry showed that in calves with prominent syringomyelia/hydromyelia a reduced number of axons in the lumbar white matter were detected that might be interpreted as hypoplasia. Furthermore, reactive astrogliosis was


**Figure 1.** Clinical, Radiological, Pathological and Histological Features of the VSD Phenotype in Affected Calves

[a] Macerated cervical vertebral column of a calf affected by VSD showing homeotic thoracic transformation of the seventh cervical vertebra (see red asterisk: the seventh vertebra articulating with the tuberculum costae of the first rib. [b, c] Transversal (b) and sagittal (c) MRI scans of a one day old calf with severe non-ambulatory paraparesis: prominent hyperintense fluid-filled central canal cavity (syringo-hydromyelia) in the lumbar spinal cord at the segment L1 to L2 and a massively reduced transverse diameter of the spinal cord at L3 and L4. [d]

Stepwise transverse sections of the lumbar spinal cord segments L1 to L4 (shown in b, c) displaying communicating hydromyelia and syringomyelia followed by segmental dysplasia and hypoplasia. [e] Calf with VSD phenotype showing a non-physiological forward positioning of the hind legs with straightened hocks. [f] Diplomyelia of the sacral segment of the spinal cord, scale bar 25 mm. [g] Hypo- and dysplasia of the middle lumbar segment of the spinal cord including missing ventral median fissure, scale bar 25 mm. [h] Duplication of the central canal in the sacral segment of the spinal cord, scale bar 500 µm. [i] Seven day old calf with slightly shortened and kinked tail defect combined with slightly hyperextended flexor tendons and external rotation of the hind limbs (left < right). [j] Seven months old calf with distinct kinked tail defect and slight rotation of the hind limbs (left < right). [k, l] Rear and dorsal view of an eight month old calf with a severe crooked tail defect and external rotation of the hind limbs. [m - n] Separation in coccygeal vertebral column as a part of a tail defect.

detected, shown as a small zone with strong accumulated GFAP positive cell dendrites around the syringomyelia. Further immunohistochemistry analyses of the spinal cord did not reveal additional abnormalities. All other tested neuroproteins were expressed regularly. Results from the neurological investigation matched the impaired posterior spinal structures and revealed multiple functional deficits associated with VSD. Specifically, VSD-affected calves displayed spasticity, paraparesis, impaired spinal reflexes, and ataxia which were predominantly expressed in the hind limbs (Table S5, File S1). However, VSD was not associated with intestinal, urogenital, cerebral, or skull defects in contrast to many other mammalian vertebral malformation defects (Vlangos *et al.* 2013). Biochemical and hematological tests monitoring enzyme activities, metabolites, electrolytes in serum as well as protein value and blood cell count in cerebrospinal fluid did not reveal any significantly increased incidence of deviation from norm values in VSD-affected calves. Furthermore, there was no evidence of Bovine Herpes Virus 1, Bluetongue or Bovine Virus Diarrhea virus in any of the necropsied, affected VSD calves.

## VSD is an Autosomal Dominantly Inherited Defect with Incomplete Penetrance

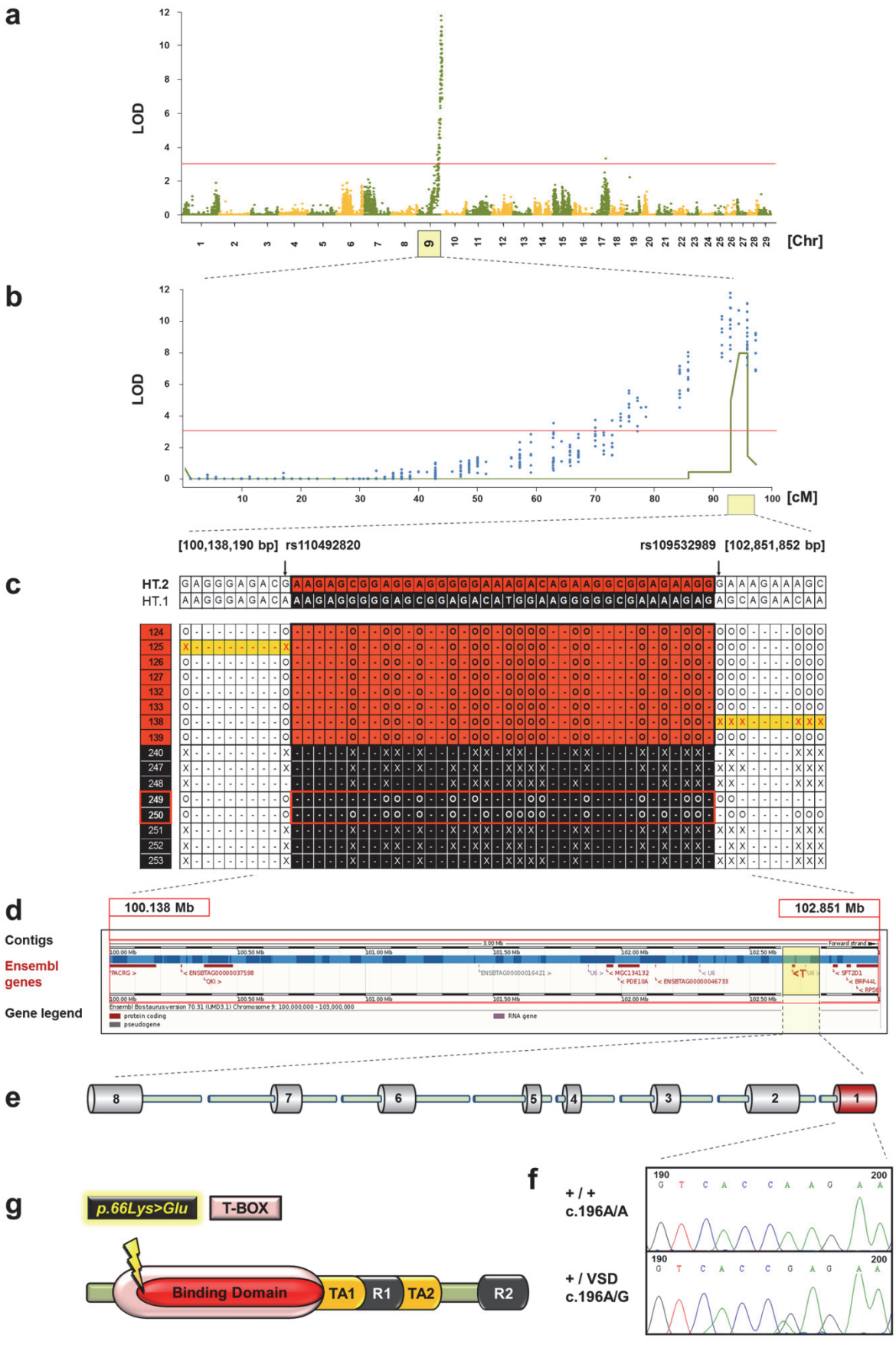
VSD cases showed substantial variation regarding the degree of physical and neurological alterations associated with the defect (severe cases with non-ambulatory paraparesis to mild cases displaying only minor tail defects, Tables S1 and S5). The hypothesis of a dominant VSD allele effect previously indicated by an almost equal proportion of VSD affected and

non-affected offspring of sire FBF0666 is further supported by sire FBF0666, which itself clearly expressed the VSD phenotype as determined by pathological examination (Table S1).

**VSD is Localized on Bovine Chromosome 9**

Initial karyotyping of sire FBF0666 and a severely affected offspring did not reveal any numerical abnormalities or large structural chromosomal aberrations. The equal distribution of VSD cases across both sexes in the FBF0666 sibship (Table S1) indicated an autosomal localization of the defect. The Crooked Tail Syndrome (CTS), a well-described bovine defect affecting tail morphology (Fasquelle *et al.* 2009), had been excluded as causal background for VSD due to a homozygous wild type genotype of sire FBF0666 at the causal mutation locus for CTS (Kromik et al. submitted).

A whole-genome scan in the *Bos taurus* genome yielded SNPs on two chromosomes with Logarithm of the Odds (LOD) scores > 3 for linkage to VSD: 99 SNPs on bovine chromosome (BTA) 9 and a single SNP on BTA17 (Figure 2, Table S6). On BTA9, exclusively SNPs located between 85,175,167 bp (rs41604518) and 105,074,182 bp (rs41619164) showed a significant LOD score > 3.0 in the twopoint analyses. The subsequent multipoint test statistic obtained by parametric linkage analysis placed the VSD locus in a LOD drop 3 confidence interval between rs110768165 (102,711,446 bp) and rs109233157 (104,196,469 bp). Alignment (Figure 2, Figure S2) of the paternally inherited BTA9 haplotypes of all FBF0666 offspring with VSD phenotype showed that all these individuals shared a common haplotype spanning from rs110492820 (100,138,190 bp) to rs109532989 (102,851,852 bp). This narrowed down the target interval for the VSD mutation to 2.7 Mb in the telomeric region of BTA9.


a
LOD
[Chr]
b
LOD
[cM]
[100,138,190 bp] rs110492820
rs109532989 [102,851,852 bp]
c
HT.2
HT.1
d
100.138 Mb
102.851 Mb
Contigs
Ensembl genes
Gene legend
protein coding
pseudogene
RNA gene
e
8
7
6
5
4
3
2
1
f
+ / +
c.196A/A
+ / VSD
c.196A/G
g
p.66Lys>Glu
T-BOX
Binding Domain
TA1
R1
TA2
R2

299

**Figure 2.** Mapping and Identification of the VSD Mutant Allele
[a] Manhattan plot showing the results (LOD scores) of the genome-wide twopoint linkage analysis between all tested SNPs and the VSD locus. LOD score threshold 3.0 is indicated by the red horizontal line. [b] LOD scores from twopoint linkage analysis (blue dots) and multipoint linkage analysis (green line) on BTA9. The light yellow box shows the LOD drop 3 confidence interval in the telomeric region on BTA9. 'x' and 'o' denote alternative paternal alleles inherited by the respective offspring, '-' indicates non-informative allele regarding paternal origin. [c] Selection of aligned paternally inherited BTA9 haplotypes (for all data see Figure S2) in the telomeric region of BTA9. The VSD-affected offspring of sire FBF0666 shared a common haplotype (HT2) spanning from rs110492820 (100,138,190 bp) to rs109532989 (102,851,852 bp). The phenotypically unaffected offspring of sire FBF0666 showed the alternative paternal haplotype (HT1) (black) except two individuals that had inherited the VSD-associated haplotype (red-boxed black). Yellow boxes indicate recombination events that set the limits of the VSD haplotype. [d] All annotated genes (Ensembl annotation release 75) in the chromosomal region 100-103 Mb including the prime candidate bovine *T* gene (light yellow box).
[e] Exon-intron structure of the bovine *T* gene according to Refseq sequence NM_001192985.1. Exon 1 containing the mutation causal for VSD is indicated in red. [f] Electropherogram showing a part of the exon 1 nucleotide sequence of the bovine *T* gene in a VSD-unaffected calf with the wild type genotype A/A at position c.196 and in a VSD-affected calf with the heterozygous genotype A/G at position c.196. [g] Domain composition of the bovine T protein with position 66 of the amino acid sequence affected by the polymorphism c.196A>G causal for VSD. The T-box is indicated as well as both transcription activation domains (TA1 and TA2) and both repression domains (R1 and R2). Domain annotation according to NCBI Conserved Domain Database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?seqinput= NP_001179914.1) and (Satoh *et al.* 2012).

## Tracing the Haplotype Associated with VSD in the Affected Pedigree

Haplotype tracking in an eight generation pedigree clearly demonstrated that sire FBF0666 had inherited the VSD-associated haplotype (position 100,138,190–102,851,852 bp) from its dam (FBF0266, Figure 3, Figure S3). Further tracing back of the inheritance of this haplotype showed that the dam had been inbred to its sire (FBF0669) and carried identical by state (IBS) chromosomal segments to both sire FBF0669's haplotypes in the VSD target region. However, analysis of the haplotypes for the entire BTA9 revealed that sire FBF0669 had forwarded to FBF0266 the respective haplotype (position 100,138,190–102,851,852 bp), which was shared by all VSD affected FBF0666 offspring (Figure S3, red haplotypes). The alternative haplotype of sire FBF0669 (Figure S3, blue haplotypes; Figure 3) was obviously not associated with VSD. This is supported by population data: In our eight generation pedigree, no previous reports on VSD-like defects were obtained in the first-generation offspring of confirmed carriers of the alternative non-VSD FBF0669 haplotype (sires FBF0670, FBF0671, FBF0672, and FBF0673; Figure 3), although these bulls had sired several hundred thousand offspring worldwide.

## VSD Is Caused by a *de-novo* Mutation in the T Gene

In the current bovine genome assemblies, the target interval for the causal mutation (BTA9: 100,138,190–102,851,852 bp) harbors 23 annotated or putative genes (Figure 2, NCBI annotation release 103: accession date 2014/03/18, http://www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid=9913&build=103.0; Ensembl: http://www.ensembl.org/Bos_taurus/Location/View?g=ENSBTAG00000018681;r=9:102662033-102680686;t=ENSBTAT00000024865, accession date 2014/03/18). Of these, the *T* gene stood out as the single prime functional candidate gene responsible for the vertebral and spinal malformations of VSD because of the previously reported effects of *T* gene mutations on embryonic notochord development and on tail length (Haworth *et al.* 2001; Herrmann *et al.* 1990). Resequencing of the *T* locus in cow FBF0266, in sires FBF0666, FBF0667, FBF0669, in VSD-affected and non-affected FBF0666 offspring as well as in unrelated individuals revealed an A>G transition polymorphism at position c.196 of the *T* gene (according to NM_001192985.1, Figure 2). This non-synonymous mutation is located in exon 1 of the *T* gene (according to NM_001192985.1) and results in a substitution of the amino acid lysine by glutamic acid at position 66 of the T protein sequence (p.66Lys>Glu). Only sire FBF0666, VSD-affected calves, five calves phenotypically unaffected but carrying the VSD-associated haplotype (e.g., FBF249 and FBF250, Figure 2) and dam FBF0266 carried the mutated allele (Figure 3). The observation of *T c.196G* carriers without clinical phenotype underlines the hypothesis of incomplete penetrance for VSD. However, sire FBF0669, from which cow FBF0266 had inherited the VSD-associated haplotype, was homozygous for the wild-type nucleotide at position c.196 (Figure 3).

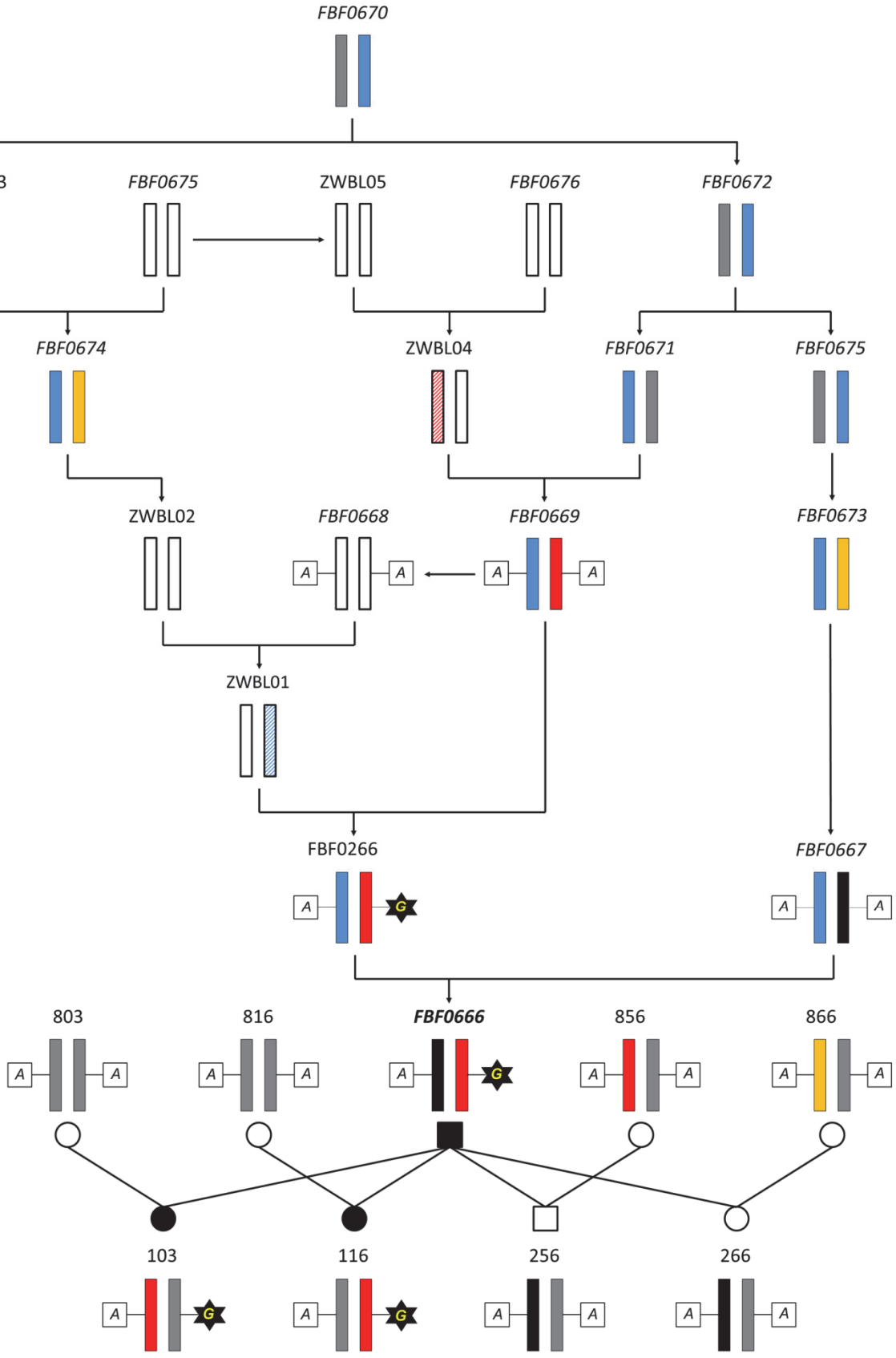



**Figure 3.** Tracing the VSD-associated Haplotype and the Origin of the VSD Mutation
Haplotypes in the target area of BTA9 (100,138,190 bp to 102,851,852 bp) are indicated by long rectangles within an eight generation Holstein pedigree segregating for the VSD. Red rectangle: maternally inherited haplotype of sire FBF0666; black rectangle: alternative haplotype of sire FBF0666; blue: non-VSD-associated

haplotype in the dam FBF0266 of sire FBF0666; fawn: haplotype identical by state to the VSD-associated sire FBF0666 haplotype except for the SNP rs29023535 (102,690,968 bp) at the telomeric end; grey: further haplotypes. Striped colored haplotypes were concluded from the haplotypes of the offspring according to Mendelian rules of inheritance; blank haplotypes are unknown. VSD-affected animals according to clinical, neurological and/or pathological examination are indicated by black boxes/circles. Individuals with confirmed non-affected phenotype are indicated by open boxes/circles. For confirmation of inherited haplotypes for dam FBF0266 see Figure S3. Letters in boxes or stars, respectively, indicate haplotype-associated alleles at position c.196A>G in the bovine *T* gene determined by sequencing.

Although sire FBF0669 has more than 140,000 registered daughters born in two decades, there are no reports of VSD within this large sibship suggesting that it is extremely unlikely that the sire carries the dominant causal VSD mutation. The homozygous wild-type genotype of 94 male offspring from sire FBF0669, as determined by genotyping of the VSD locus *T c.196A>G*, also supported the homozygous wild-type status of sire FBF0669 at this chromosomal position. These 94 offspring are themselves widely-used sires with at least 200 offspring born to each. The absence of VSD incidence reports in the first-generation descendants of these 94 bulls corresponds to their wild-type genotype at the VSD locus.

Thus, haplotype tracking and mutation analysis clearly demonstrate that *T c.196A>G* is a *de novo* mutation in cow FBF0266 not previously seen on the respective haplotype. Consequently, only the direct progeny of cow FBF0266 could possibly carry the mutated allele associated with VSD. Indeed, genotyping of 39 VSD-unaffected control calves (matched controls to FBF0666 offspring) and a further 528 randomly selected Holstein and Holstein x Charolais calves did not identify any carrier of the mutant *T c.196G* allele. In addition, seven VSD unaffected calves' dams in our dataset, which are no direct offspring to dam FBF0266, but which carried the VSD haplotype in a IBS homo- or heterozygous state (determined according to 50k SNP haplotyping), were all homozygous for the wild-type allele *T* c.196A.

## DISCUSSION

Our study is the first report of the inherited *Bos taurus* defect VSD that is associated with a reduced number of cervical vertebrae, a unique, striking feature that had not yet been described for a spontaneous mutation in any mammalian species before. The *T* gene belongs to the family of T-box genes that encode transcription factors consisting of transcriptional activator and/or repressor domains and a DNA binding T-box domain in many eukaryotic species including vertebrates and invertebrates (Satoh *et al.* 2012). The T protein is essential for development of the notochord and mesoderm formation in the primitive streak during early embryonic vertebrate development (Kispert & Herrmann 1994). Experimental crystallographic data for the T protein (Müller & Herrmann 1997) demonstrated that the amino acid position equivalent to the variant amino acid position p.66Lys>Glu in the bovine ortholog is located at a critical site in the DNA binding T-box domain of the T protein (Figure 2). Specifically, the amino acid position p.66 forms polar interactions with the DNA target and is directly involved in the DNA binding of the T-box domain and dimerization of the T protein during DNA binding. It is conclusive that replacing the wild-type basic amino acid lysine by the mutant acidic amino acid glutamic acid at p.66 in the bovine T protein will substantially disturb those T protein binding properties. This is supported by bioinformatic analyses predicting considerable changes in the three-dimensional peptide conformation of the bovine T protein as a result of the missense mutation (Figure 4) and also by estimating mutation effects ("probably damaging" score: 0.977, according to Polyphen2 (Adzhubei *et al.* 2010). Finally, HomoloGene analysis showed that the position homologous to bovine T p.66 is highly conserved from *Homo sapiens* down to *Drosophila melanogaster* and *Anopheles gambiae* (Figure 4). This strong conservation further confirms a fundamental relevance of the protein, particularly at the position affected by the mutation. Because classical gene rescue experiments to prove causality of a mutation are extremely difficult in cattle, we further followed the guidelines for investigating causality of sequence variants in human disease

(MacArthur *et al.* 2014). In this line, the conclusion of a causal role for the *T c.196A>G* mutation in VSD is further supported by comparative data. Already Chesley (Chesley 1935) reported that mice heterozygous for a mutant T allele showed effects on the notochord at the early stage of development (day 8) and also on the neural tube. Mutations in several parts of the *T* gene often show a similar mode of inheritance and a variable penetrance (e.g., the Manx syndrome in cats (Buckingham *et al.* 2013)). Furthermore, the mutations in the *T* gene are associated with tail defects or malformation of posterior parts of the body in many other species from drosophila to mice, cats and dogs (Buckingham *et al.* 2013; Haworth *et al.* 2001; Herrmann *et al.* 1990; Kispert *et al.* 1994; Odenthal *et al.* 1996). In human, a recessively acting mutation in the *T* gene has been identified to be associated with fusion of cervical vertebrae, with sacral agenesis and/or abnormal notochord features (Ghebranious *et al.* 2008; Postma *et al.* 2014). Furthermore, for the mouse *T* curtailed ($T^c$) allele there is one study reporting effects on the cervical vertebrae (Searle 1966), whereas *T* gene mutant alleles mostly affected the posterior part of the vertebral column. However, the specific effects observed in murine $T^c$ heterozygotes and human patients heterozygous for the *T c.1013C>T* mutation are different to those of VSD heterozygotes, because there is no lack, but a fusion of two or more vertebrae. Also in contrast to $T^c$, in the VSD-affected animals the sacrum is the only part of the bony vertebral column without malformation. To our knowledge, none of the known *T* mutations in other species showed effects of cervical vertebral deletions/homeotic transformations, not even for homozygous individuals. In *Bos taurus*, other lethal genetic defects associated with vertebral malformations (Complex cervical malformation, Brachyspina) could be excluded as background for the VSD defect, because both defects had been localized on BTA3 or BTA21, respectively (Charlier *et al.* 2012; Thomsen *et al.* 2006).

Our results suggest that the VSD mutation affects the primitive streak as well as the tail bud because vertebrae originating from both precursors are affected by the mutation: cervical vertebrae originating from the primitive streak and coccygeal vertebra originating from the

tail bud. This fits the observation that murine *T* +/- heterozygous embryos showed a 50% reduction of *T* gene expression in the tail bud and notochord compared with wild-type mice (Pennimpede *et al.* 2012).

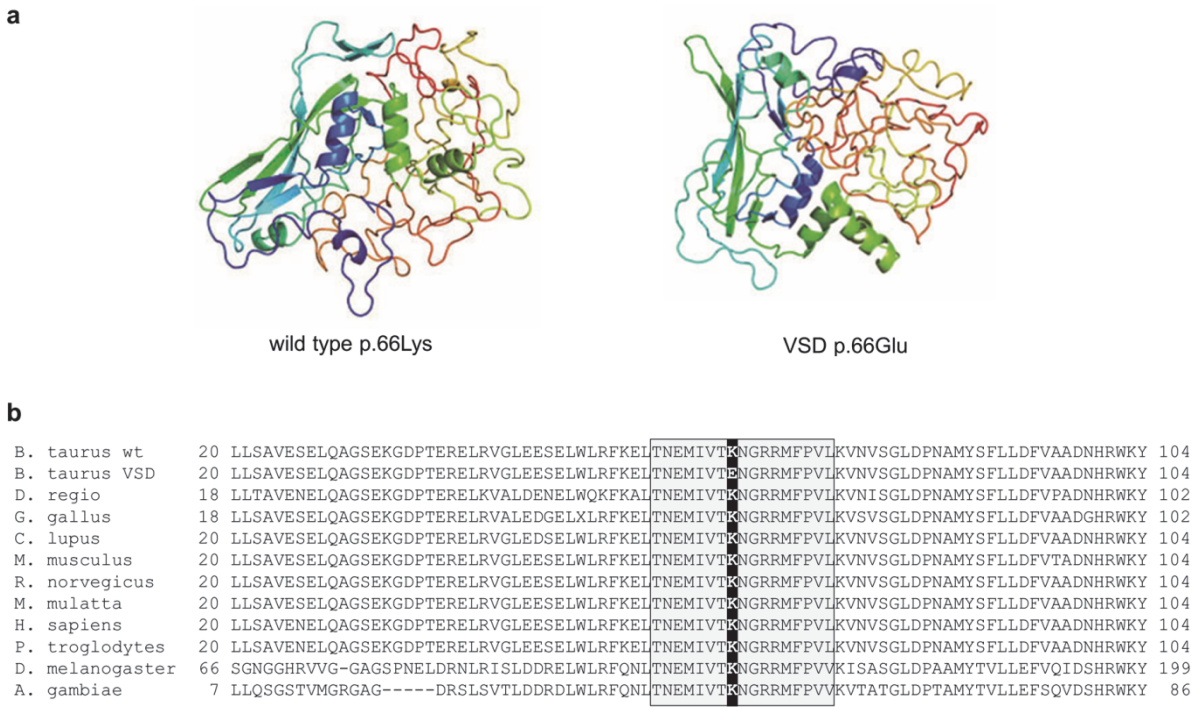



**Figure 4**. Predicted Conformation Change of the Wild-Type and VSD Bovine T Protein and T Protein Interspecies Amino Acid Sequence Comparison
[a] Predicted 3D-structure of the wild-type (p.66Lys) and mutated (VSD, p.66Glu) bovine T protein as determined by the bioinformatic prediction tool Phryre2 (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index).
[b] HomoloGene (http://www.ncbi.nlm.nih.gov/homologene) analysis of the T protein/homologue encompassing the variant bovine amino acid position 66 (indicated by black background) across vertebrates and insects (*Danio regio*: XP_001343633.3, *Gallus gallus* NP_990271.1, *Bos taurus* NP_001179914.1, *Canis lupus* NP_001003092.1, *Mus musculus* NP_033335.1, *Rattus norvegicus* NP_001099679.1, *Macaca mulatta* XP_001101514.1, *Homo sapiens* NP_003172.1, *Pan troglodytes* XP_527563.3, *Drosophila melanogaster* NP_524031.2, *Anopheles gambiae* XP_320606.4). Boxed and marked with gray background is the longest fully conserved segment within the entire T protein/homologue. wt: wild-type allele, VSD: VSD-associated allele.

Pennimpede and colleagues (Pennimpede *et al.* 2012) previously suggested that the T protein is directly involved in the maintenance of the mammalian seven-cervical vertebra blueprint because of the homeotic C7 > T1 transformation of cervical vertebrae in 30% of mice from *in vivo T* gene knockdown experiments. The spontaneous VSD mutation in the bovine *T* gene is the first *in vivo* evidence for this hypothesis from a mutation model. Our data also highlight a distinct amino acid position (p.66) that might be relevant for a coordinated Wnt–brachyury–

HOX signaling cascade, which is important for cervical vertebral and spinal cord development (Galis 1999; Yamaguchi *et al.* 1999). Remarkably, the heterozygous VSD genotype causes substantial phenotypic impairments, whereas even murine *T* null-alleles, in which the *T* locus is completely absent, only cause mild phenotypic defects in heterozygotes (Smith 1997). This expression pattern of the VSD phenotype suggests a dominant negative effect of the VSD allele. A similar mechanism was also suggested for some alleles at the murine *brachyury* locus ($T^c$, $T^{wis}$), although theses alleles alter the carboxy-terminus of the T protein (Herrmann & Kispert 1994), which potentially acts as activating domain and in contrast to the T-box domain shares little sequence similarity between species (Smith 1997). Although there are many similarities of the VSD mutation to tail defects in other species, to our knowledge no other spontaneous mutation in the *T* gene or other mammalian genes causes a homeotic transformation of cervical vertebrae similar to VSD. In addition, VSD is also unique, because in spite of congenital homeotic transformation of cervical vertebrae, affected individuals survive to reproductive age and show no primary defects outside vertebral spine and spinal cord.


## ACKNOWLEDGEMENTS

The project was funded by the Förderverein Biotechnologieforschung (FBF), Bonn, Germany. We thank Jill Maddox (University of Melbourne, Australia) for providing the modified CRIMAP Version 2.50. We are indebted to the Masterrind GmbH, Verden and its associated farmers for bringing the congenital effect to knowledge and assisting in data collection. Specifically, Dr. D. Frese and Dr. H. Osmers contributed valuable input during fruitful discussions. Technical assistance of Simone Wöhl, Antje Lehmann and Marlies Fuchs is thankfully acknowledged. Important help was provided by colleagues in the animal experimental units of the Leibniz Institute for Farm Animal Biology (FBN).

## LITERATURE CITED


Abecasis, G. R., S. S. Cherny, W. O. Cookson, and L. R. Cardon, 2002 Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. Nat. Genet. **30**: 97-101.

Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova *et al.* 2010 A method and server for predicting damaging missense mutations. Nat. Meth. **7**: 248-249.

Agerholm, J. S., C. Bendixen, O. Andersen, and J. Arnbjerg, 2001 Complex vertebral malformation in Holstein calves. J. Vet. Diagn. Invest. **13**: 283-289.

Browning, B. L., and S. R. Browning, 2009 A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. **84**: 210-223.

Buckingham, K. J., M. J. McMillin, M. M. Brassil, K. M. Shively, K. M. Magnaye *et al.* 2013 Multiple mutant T alleles cause haploinsufficiency of Brachyury and short tails in Manx cats. Mamm. Genome **24**: 400-408.

Charlier, C., J. S. Agerholm, W. Coppieters, P. Karlskov-Mortensen, W. Li *et al.* 2012 A Deletion in the Bovine FANCI Gene Compromises Fertility by Causing Fetal Death and Brachyspina. Plos One **7**: e43085.

Chesley, P., 1935 Development of the short-tailed mutant in the house mouse. J. Exp. Zool. **70**: 429-459.

Dobrovolskaia-Zavadskaia, N., 1927 Sur la mortification spontanee de la queue chez la souris nouveau-nee et sur l'existence d'un caractere hereditaire <<non viable>>. CR Soc Biol **97**: 114-116.

517 FAO, 2007 *The state of the World's Animal Genetic Resources for Food and Agriculture*.
518 Rome, Italy.

519 Fasquelle, C., A. Sartelet, W. B. Li, M. Dive, N. Tamma *et al.* 2009 Balancing Selection of a
520 Frame-Shift Mutation in the MRC2 Gene Accounts for the Outbreak of the Crooked
521 Tail Syndrome in Belgian Blue Cattle. Plos Genet. **5**: e1000666.

522 Galis, F., 1999 Why do almost all mammals have seven cervical vertebrae? Developmental
523 constraints, Hox genes, and cancer. J. Exp. Zool. **285**: 19-26.

524 Galis, F., T. J. Van Dooren, J. D. Feuth, J. A. Metz, A. Witkam *et al.* 2006 Extreme selection
525 in humans against homeotic transformations of cervical vertebrae. Evolution **60**: 2643-
526 2654.

527 Ghebranious, N., R. D. Blank, C. L. Raggio, J. Staubli, E. Mcpherson *et al.* 2008 A Missense
528 T(Brachyury) Mutation Contibutes to Vertebral Malformations. J. Bone Miner. Res.
529 **23**: 1576-1583.

530 Green, P., K. Falls, and S. Crooks, 1990 Documentation of CRI-MAP version 2.4.
531 Unpublished mimeo (http://www. animalgenome. org/hu/CRIMAPwkshp/crimap-doc.
532 html) .

533 Hautier, L., V. Weisbecker, M. R. Sanchez-Villagra, A. Goswami, and R. J. Asher, 2010
534 Skeletal development in sloths and the evolution of mammalian vertebral patterning.
535 Proc. Natl. Acad. Sci. **107**: 18903-18908.

536 Haworth, K., W. Putt, B. Cattanach, M. Breen, M. Binns *et al.* 2001 Canine homolog of the
537 T-box transcription factor T; failure of the protein to bind to its DNA target leads to a
538 short-tail phenotype. Mamm. Genome **12**: 212-218.

539 Herrmann, B. G., S. Labeit, A. Poustka, T. R. King, and H. Lehrach, 1990 Cloning of the T-
540 Gene Required in Mesoderm Formation in the Mouse. Nature **343**: 617-622.

541 Herrmann, B. G., and A. Kispert, 1994 The T genes in embryogenesis. Trends Genet. **10**: 280-
542 286.

543 Kelley, L., and M. J. E. Sternberg, 2009 Protein structure prediction on the Web: a case study
544 using the Phyre server. Nat. Prot. **4**: 363-371.

545 Kispert, A., B. G. Herrmann, M. Leptin, and R. Reuter, 1994 Homologs of the Mouse
546 Brachyury Gene Are Involved in the Specification of Posterior Terminal Structures in
547 Drosophila, Tribolium, and Locusta. Gene Dev. **8**: 2137-2150.

548 Kispert, A., and B. G. Herrmann, 1994 Immunohistochemical Analysis of the Brachyury
549 Protein in Wild-Type and Mutant Mouse Embryos. Developmental Biologie **161**: 179-
550 193.

551 MacArthur, D., T. Manolio, D. Dimmock, H. Rehm, J. Shendure *et al.* 2014 Guidelines for
552 investigating causality of sequence variants in human disease. Nature **508**: 469-476.

553 Müller, C. W., and B. G. Herrmann, 1997 Crystallographic structure of the T domain-DNA
554 complex of the Brachyury transcription factor. Nature **389**: 884-888.

555 Nibu, Y., D. S. Jose-Edwards, and A. Di Gregorio, 2013 From Notochord Formation to
556 Hereditary Chordoma: The Many Roles of Brachyury. Biomed Res. Internat. DOI
557 10.1155/2013/826435.

558 Odenthal, J., P. Haffter, E. Vogelsang, M. Brand, F. J. M. vanEeden *et al.* 1996 Mutations
559 affecting the formation of the notochord in the zebrafish, Danio rerio. Development
560 **123**: 103-115.

561 Pennimpede, T., J. Proske, A. Koenig, J. A. Vidigal, M. Morkel *et al.* 2012 In vivo
562 knockdown of Brachyury results in skeletal defects and urorectal malformations
563 resembling caudal regression syndrome. Dev. Biol. **372**: 55-67.

564 Pillay, N., V. Plagnol, P. S. Tarpey, S. B. Lobo, N. Presneau *et al.* 2012 A common single-
565 nucleotide variant in T is strongly associated with chordoma. Nat. Genet. **44**: 1185-
566 1187.

567 Popescu, P., H. Hayes, and B. Dutrillaux, 2000 Chromosome preparation, pp. 1-24 in
568 *Techniques in Animal Cytogenetics (Principles and Practice)*, Springer Berlin
569 Heidelberg.

570 Postma, A. V., M. Alders, M. Sylva, C. M. Bilardo, E. Pajkrt *et al.* 2014 Mutations in the T
571 (brachyury) gene cause a novel syndrome consisting of sacral agenesis, abnormal
572 ossification of the vertebral bodies and a persistent notochordal canal. J. Med. Genet.
573 **51**: 90-97.

574 Satoh, N., K. Tagawa, and H. Takahashi, 2012 How was the notochord born? Evol. Dev. **14**:
575 56-75.

576 Searle, A. G., 1966 Curtailed, a new dominant T-allele in the house mouse. Genet. Res. **7**: 86-
577 95.

578 Smith, J., 1997 Brachyury and the T-box genes. Curr. Opinion Genet. Developm. **7**: 474-480.

579 Thomsen, B., P. Horn, F. Panitz, E. Bendixen, A. H. Petersen *et al.* 2006 A missense
580 mutation in the bovine SLC35A3 gene, encoding a UDP-N-acetylglucosamine
581 transporter, causes complex vertebral malformation. Genome Res. **16**: 97-105.

582 Ulrich, R., A. C. Stan, M. Fehr, C. Mallig, and C. Puff, 2010 Desmoplastic ganglioglioma of
583 the spinal cord in a western European hedgehog (Erinaceus europaeus). J. Vet. Diagn.
584 Invest. **22**: 978-983.

585 van de Ven, C., M. Bialecka, R. Neijts, T. Young, J. E. Rowland *et al.* 2011 Concerted
586 involvement of Cdx/Hox genes and Wnt signaling in morphogenesis of the caudal
587 neural tube and cloacal derivatives from the posterior growth zone. Development **138**:
588 3859.

589 Vlangos, C. N., A. N. Siuniak, D. Robinson, A. M. Chinnaiyan, R. H. Lyons *et al.* 2013
590 Next-generation sequencing identifies the Danforth's short tail mouse mutation as a
591 retrotransposon insertion affecting Ptf1a expression. Plos Genet. **9**: e1003205.

592 Yamaguchi, T. P., S. Takada, Y. Yoshikawa, N. Y. Wu, and A. P. McMahon, 1999 T
593 (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. Gene
594 Dev. **13**: 3185-3190.

595 Yang, X. R., D. Ng, D. A. Alcorta, N. J. Liebsch, E. Sheridan *et al.* 2009 T (brachyury) gene
596 duplication confers major susceptibility to familial chordoma. Nat. Genet. **41**: 1176-
597 1178.

598 Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz *et al.* 2009 A whole-
599 genome assembly of the domestic cow, Bos taurus. Genome Biol. **10**: R42.

600
601

602

603 **Supporting Information**

604 **Kromik_Genetics_Supporting_information_2014_12_12.docx**

605

606 **File S1.** Video documentation of neurological deficit in a VSD-affected calf: non-ambulatory

607 paraparesis

608
609
610