Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Andrology. Author manuscript; available in PMC 2016 September 01.

Published in final edited form as:

Andrology. 2015 September ; 3(5): 825–833. doi:10.1111/andr.12063.

Rare *DMRT1* regulatory variants in severe spermatogenic failure

Ana Cristina Lima[1,2,3,4], **Filipa Carvalho**[5], **João Gonçalves**[6], **Susana Fernandes**[5], **Patrícia Isabel Marques**[2,3,7], **Mário Sousa**[8], **Alberto Barros**[5], **Susana Seixas**[2,3], **António Amorim**[2,3,9], **Donald Franklin Conrad**[4,10], and **Alexandra Manuel Lopes**[2,3]

[1]Graduate Program in Areas of Basic and Applied Biology (GABBA), Abel Salazar Institute of Biomedical Sciences, University of Porto, 4050-313 Porto, Portugal

[2]Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal – I3S

[3]Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal – IPATIMUP, 4200-465 Porto, Portugal

[4]Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110, USA

[5]Department of Genetics, Faculty of Medicine of the University of Porto, 4200-319 Porto, Portugal

[6]Department of Human Genetics - National Institute of Health Dr. Ricardo Jorge, 1649-016 Lisboa, Portugal

[7]Abel Salazar Institute of Biomedical Sciences, University of Porto, 4050-313 Porto, Portugal

[8]Laboratory of Cell Biology, UMIB, ICBAS, University of Porto, 4050-313 Porto, Portugal

[9]Faculty of Sciences of the University of Porto, 4169 - 007 Porto, Portugal

[10]Department of Pathology & Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA

Abstract

The *DMRT1* (doublesex and mab-3 related transcription factor 1) gene has long been linked to sex-determining pathways across vertebrates and is known to play an essential role in gonadal development and maintenance of spermatogenesis in mice. In humans, the genomic region harboring the *DMRT* gene cluster has been implicated in disorders of sex development and recently *DMRT1* deletions were shown to be associated with non-obstructive azoospermia (NOA). In this work we have employed different methods to screen a cohort of Portuguese NOA patients for *DMRT1* exonic insertions and deletions (by MLPA; n=68) and point mutations (by Sanger sequencing; n=155). We have found three novel patient-specific non-coding variants in

Corresponding author: Alexandra Manuel Lopes, Address: Ipatimup, Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal, alopes@ipatimup.pt.

Disclosure: The authors declare no competing interests.

Authors' Contributions: ACL performed the *DMRT1* sequence analysis in Portuguese patients and controls, conducted the data analysis, extracted DNA of fertile control samples, interpreted the results and drafted the manuscript. FC, JG, SF, MS and AB performed the clinical and genetic characterization of the patients' samples for the routine workup of male infertility. PIM performed DNA extraction of the normozoospermic patients and contributed to the genotyping of the controls. AA, SS and DFC contributed with a critical review of the manuscript. AML obtained financial support, designed the study and supervised the data analysis, interpretation of results and manuscript preparation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

heterozygosity that were absent from 357 geographically matched controls. One of these is a complex variant with a putative regulatory role (c.-223_-219CGAAA>T), located in the promoter region within a conserved sequence involved in *Dmrt1* repression. Moreover, while DMRT1 domains are highly conserved across vertebrates and show reduced levels of diversity in human populations, 2 rare synonymous substitutions (rs376518776 and rs34946058) and 2 rare non-coding variants that potentially affect *DMRT1* expression and splicing (rs144122237 and rs200423545) were overrepresented in patients when compared to 376 Portuguese controls (301 fertile and 75 normozoospermic). Overall our previous and present results suggest a role of changes in *DMRT1* dosage in NOA potentially also through a process of gene misregulation, even though *DMRT1* deleterious variants seem to be rare.

Keywords

alternative splicing; cis-regulatory variants; DMRT1; DMRT1 promoter; non-obstructive azoospermia (NOA)

Introduction

Male infertility is a complex phenotype often arising from the inability of men to produce viable spermatozoa capable of fertilization. Non-obstructive azoospermia (NOA) is a severe form of male infertility characterized by the lack of sperm in the ejaculate, which results from conditions other than hypothalamic-pituitary disease or obstruction of the male genital tract (Jungwirth *et al.* 2013). It is estimated that NOA accounts for 10-15% of the cases of male infertility, affecting 1% of male individuals worldwide (Practice Committee of American Society for Reproductive Medicine in collaboration with Society for Male Reproduction and Urology 2008). Importantly, over the past years several studies have been showing that a significant portion of the genes involved in severe spermatogenic failure (SSF) are spread throughout the genome rather than restricted to the sex chromosomes [Reviewed in (Lima & Lopes 2014)]. One particular genomic region (9p), harboring the doublesex and mab-3 related transcription factor (DMRT) gene cluster, has been linked with disorders of sex development (Õunap *et al.* 2004, Raymond *et al.* 1999, Tannour-Louet *et al.* 2010, Veitia *et al.* 1997). It appears that isolated genetic defects in *DMRT1*, one of the genes in the cluster, may be the cause of alterations in gonadal development in some cases (Ledig *et al.* 2010, 2012) but not all, as no inactivating mutations or deletions were found in this gene in a study of patients with gonadal dysgenesis (Machado *et al.* 2012). However, we have recently identified recurrent 9p24.3 deletions spanning *DMRT1* in patients with idiopathic azoospermia, suggesting that loss of function of this gene is a risk factor and potential genetic cause of human SSF (Lopes *et al.* 2013).

DMRT proteins are transcription factors that bind to the promoters of target genes through their zinc-finger DM (doublesex and mab-3)-domain. This DNA-binding motif is highly conserved and requires dimerization for recognition of a target consensus palindromic sequence (Erdman *et al.* 1996). Also, in association with the DM-domain, the DMRT protein family shares a less conserved DMRT1-like domain of yet uncharacterized function. Particularly, DMRT1 has been extensively studied in vertebrates and is known to be crucial

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

in the pathway of sex determination and gonadal development in most, if not all, species of this group [Reviewed in (Zarkower 2013)]. Indeed DMRT1 has acquired specialized functions in different clades. It is essential for male gonad differentiation but not sex determination in mice, showing testis specific expression restricted to Sertoli and premeiotic germ cells. A similar pattern is observed for human *DMRT1* with the exception that it is also part of the network of sex determination [Reviewed in (Bratuś 2012)], being co-expressed with SRY (sex determining region) in the genital ridge of the male embryo (Moniot *et al.* 2000). Interestingly, a total of three cDNA isoforms have been detected in the human testis - DMRT1a, DMRT1b, DMRT1c - and result from the inclusion of intronic regions and *Alu* elements (Cheng *et al.* 2006), suggesting that the transcriptional diversity of this gene might be involved in the regulation of spermatogenesis.

Bearing in mind the expression pattern of *DMRT1*, its crucial role in the progression of spermatogenesis and previous findings of an association of *DMRT1* defects with SFF, we sought to evaluate the contribution of sequence variants in the promoter and coding regions of the *DMRT1* gene to NOA phenotypes. Here, we describe novel non-coding variants and rare non-coding and synonymous variants overrepresented in NOA patients when compared to Portuguese fertile and normozoospermic controls, which are predicted to affect the regulation of the canonical *DMRT1* transcript (DMRT1a) and alternative splicing.

Materials and Methods

Patient and Control Populations

DNA samples of 155 male individuals with idiopathic non-obstructive azoospermia (NOA) were collected at the Genetics Department from INSA-IP and at the Genetics Department from FMUP. Patients with known causes of infertility, including chromosome anomalies and Yq microdeletions were excluded from this study. The majority of these samples (n=110) had been analyzed for CNVs by SNP microarrays (Lopes et al., 2013) but none presented CNVs within *DMRT1*.

As controls, we obtained DNA of 376 Portuguese men: 75 normozoospermic (normal sperm parameters) and 301 males who fathered at least one child.

Ancestry of one NOA patient harboring two potential regulatory variants was determined by genotyping of 46 ancestry informative markers (insertions and deletions). Biogeographical ancestry was assigned using Snipper 2.0 (http://mathgene.usc.es/snipper/; See supplementary data and supplementary Fig.1)

The study was included in the project ‘Copy number variation in infertile men genomic regions: screening in the Portuguese population’ (PTDC/SAU-GMG/101229/2008), approved by the INSA Ethics Committee (Lisbon, Portugal on 6 November 2007). Molecular studies were performed after informed consent and all DNA samples were anonymized.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Analysis of DMRT1 Sequence

Multiplex Ligation Probe Assay (MLPA)—To detect deletions/duplications in the coding region, MLPA was performed in 68 randomly picked NOA patients. SALSA MLPA P334 Gonadal probemix (MRC-Holland, Amsterdam, the Netherlands), which contains two probes for each exon of the gene, was employed according to the manufacturer's instructions. Typically 150-200 ng of genomic DNA were used for amplification.

Sanger Sequencing—Primers for amplification and sequencing (Supplementary Table 1) of the *DMRT1* genomic reference sequence (NM_021951.2) were designed in Primer3 v. 0.4.0 (Koressaar & Remm 2007, Untergasser *et al.* 2012). DNA fragments were amplified by standard PCR, sequenced with the BigDye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems®, Life Technologies, Carlsbad, California, USA) and run on an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems®, Life Technologies, Carlsbad, California, USA). Due to low DNA quality, in 23 patients only a partial gene sequence was obtained, which was taken into account in the variant frequency calculations, performed independently for each exon (See supplementary data). All individuals showing novel/rare variants had the full *DMRT1* sequence analyzed. Sequences were assembled and analyzed using Geneious version 5.5.8 (Biomatters, Auckland, New Zealand) and all putative polymorphisms were manually inspected and individually confirmed.

Genotyping of Portuguese fertile and normozoospermic controls—Rare variants (<1%) in European populations which were overrepresented in patients and novel mutations identified exclusively in NOA cases were genotyped in our Portuguese controls (Tables 1 and 2). The presence of two indels (c.354+38_insG with a juxtaposed poly-T and rs59834456) was detected by fragment size separation in polyacrylamide and agarose gels respectively; variants in the promoter, 5′UTR and CDS of Exon 1 were genotyped by Sanger sequencing; 6 other variants were analyzed in a single SNaPshot Multiplex reaction. All variants detected by SNaPshot (SupplementaryFig.2) were confirmed by Sanger sequencing. See extended materials and methods section and Tables 1-3 from Supplementary material for more details on experiments and primers used.

In silico and Statistical Analysis

Sequence data were retrieved from the 1000 genomes Project (Genomes Project Consortium *et al.* 2012), NHLBI GO Exome Sequencing Project (ESP) and CLINSEQ projects to check for all the reported variants in the regions screened and used to set a group of unphenotyped controls of European ancestry (Supplementary material and Supplementary Table 4).

Allele counts of rare variants that were overrepresented in patients were compared to the Portuguese control population applying Fisher's exact test, using RStudio v.0.98.1080 (R Development Core Team 2012).

Also, we predicted *in silico* the effect of novel non-coding variants in mRNA splicing [Human Splicing Finder v.2.4.1 (Desmet *et al.* 2009); BGDP: Splice Site Prediction by Neural Network] and the transcription factor binding sites (TFBSs) within the promoter [MatInspector, included in the Genomatix software suite (Genomatix Software GmbH,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Munich, Germany); TFSEARCH 1.3, available at http://www.cbrc.jp/research/db/TFSEARCH.html.

We have retrieved the protein sequences of 12 vertebrates and *C. elegans* from Ensembl database of *DMRT1* orthologs. Manually curated protein sequences were then aligned using ClustalW.

Results

In order to characterize the mutational spectrum of potentially deleterious *DMRT1* variants in men with idiopathic SSF, we performed a survey of intragenic deletions/duplications by MLPA and point mutations by Sanger sequencing. By MLPA we aimed to find intragenic *DMRT1* rearrangements that might be more frequent than the large CNVs previously described in American and Chinese NOA patients (Lopes *et al.* 2013). No alterations were detected by MLPA after screening approximately half of our patient cohort (n=68) and therefore we conclude that, if present, intragenic *DMRT1* rearrangements are also rare in our population.

A New *DMRT1* Promoter Variant Restricted to NOA

Given the potential deleterious effect of base substitutions and smaller deletions/duplications, we have screened the *DMRT1* sequence of 155 NOA patients by Sanger sequencing. In total we detected 4 novel non-coding variants in heterozygosity in patients (Supplementary Table 2), three of which were absent from Portuguese controls (Table 1). All novel variants were further analyzed *in silico* for a regulatory role (see Methods and Supplementary material) and those with a predicted effect are described.

Interestingly, one of the variants detected only in a single NOA patient presents a complex genotypic configuration and is located within a region of the *DMRT1* known to be involved in the transcriptional repression of the gene in mice (Lei & Heckert 2002). We identified, by allele-specific amplification followed by Sanger sequencing, a single variant (c.-223_-219CGAAA>T; Table 1) in one haplotype (i.e. inherited from the same parent, Fig. 1A) that most likely resulted from a complex mutation event. Importantly, this variant was not present in a sample of 357 geographically matched control individuals (301 fertile and 56 normozoospermic men). To assess its potential impact, we performed a more refined *in silico* analysis of the human *DMRT1* promoter (HG_KWN:62427), located within a CpG-island (chr9:840690-842192; NCBI37/hg19), a feature often associated with TATA-less promoters. Typically the core promoter activity and basal gene transcription of TATA-less genes is regulated by an initiator element (Inr), with a consensus sequence Py Py A+ 1N T/A Py Py (Smale & Baltimore 1989), which we have found in the promoter region of *DMRT1* (Fig. 1A) within the canonical transcription start site.

In the alignment of the *DMRT1* promoter sequences from several species representative of different mammalian clades (Fig. 1B), a high degree of conservation is apparent, in accordance with previous findings (Lei & Heckert 2002). Moreover, no variants have been described in this region in human populations (chr9:841602-841626; GRCh37/hg19 assembly). In addition, we found that the G nucleotide (chr9:841617; hg19 assembly),

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

deleted in the patient presenting the c.-223_-219CGAAA>T variant, is only found in the human sequence, whereas the ancestral allele A is present in all the other species. A screening for TFBSs within this conserved region of the *DMRT1* promoter was performed using two bioinformatics tools: MatInspector and TFSEARCH 1.3 (see Methods and Supplementary Table 5). In the human reference sequence, binding of Heat-shock factors 1 (HSF1; + strand) and 2 (HSF2; - strand) was predicted by both programs. When the human-specific G is replaced by the ancestral A at this position or by the sequence variant found in the patient, a disruption of the HSF1 binding site (Fig. 1A) is predicted, while the score for HSF2 binding remains unchanged. Moreover, in the proximal region of the *DMRT1* transcription start site, the human reference sequence contains a second binding site for HSF1 (HSF and TFSEARCH 1.3 predictions) that is conserved between human (chr9:841,690-969,090; GRCh37/hg19 assembly) and mouse (chr19:25,505,706-25,604,328; GRCm38/mm10 assembly).

Another patient was heterozygous for the insertion of a G nucleotide (c.354+38_insG) and for repeat length in a juxtaposed poly-T (>10 T) in intron 1. While this poly-T region seems to have small length variability in both Portuguese patients and controls, it is ≥ 10 bp longer than the reference sequence on both chromosomes of this patient. Length variability in this region alters the sequence of a *DMRT1* alternative transcript (DMRT1c) that has been previously detected in human testis (Cheng *et al.* 2006) and may lead to a frameshift in the corresponding protein isoform depending on the number of repeats.

Rare *DMRT1* Variants May Play a Role in Splicing Regulation

A subset of variants identified in the NOA cohort (n=18) had been previously found in unphenotyped individuals from large genome sequencing projects (Table 2). We obtained the frequency of all *DMRT1* variants detected and set a control group of populations of European ancestry (See Supplementary material). Since any genetic variant leading to NOA should be rare in the population, we have selected those at ≤1% in European populations and overrepresented in patients for genotyping in the Portuguese controls (Table 2; rs3739584 and rs3739583 were genotyped in the same assay as rs144122237 and rs376518776). While only the rs59834456 deletion was present in Portuguese normozoospermic controls, two variants (one synonymous substitution and a 5′UTR variant) were also absent from fertile individuals with unknown sperm counts (Supplementary Table 2). The synonymous substitution, rs376518776 (c.213G>A), juxtaposed to the DNA binding DM domain, was found in heterozygosity in a single NOA patient. It showed a 6-fold increase in frequency in the NOA Portuguese cohort when compared to unphenotyped database controls of European ancestry and was absent from fertile and normozoospermic Portuguese controls. *In silico* predictions using the Human Splicing Finder software [HSF; (Desmet *et al.* 2009)] suggest that this substitution creates a new branch point site, a cis-element required for spliceosome assembly, with a score (89.44) above the average of naturally occurring branch point sequences [83.4±8.6; (Desmet *et al.* 2009)]. This new branch point site may interfere with the interplay of splicing elements in its vicinity and potentiate the occurrence of new alternative transcripts.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The single rare 5′-UTR variant (rs144122237; c. -54C>T) resides within a putative binding site for the transcription factor RFX3 (Regulatory factor X, 3; Supplementary Table 6) and was found in one NOA patient (0.34%) but was absent from Portuguese fertile and normozoospermic controls and from European populations. Given the presence of the same allele in Africans at low frequency (1.8%) we genotyped a panel of 46 ancestry informative markers (Pereira *et al.* 2012), and assigned an African ancestry for this patient (LR African versus European= 7.712×10^{22}; see supplementary data). This was the same individual who harbored two longer poly-T alleles (>10bp) within intron 1, which may reflect his African genetic background.

Other coding variants overrepresented in patients are rs146975077 (c.774G>C) and rs34946058 (c.783C>G), the latter possibly interfering with splicing. In fact, using the HSF software two novel splicing enhancer motifs were predicted for the binding of SC35 (score 82.18) and SRp40 (score 79.28), both with scores above the respective thresholds (75.05 and 78.08). These proteins are members of the SR protein family known to be involved in the regulation of alternative splicing (Liu *et al.* 1998). Nonetheless, this substitution was present in 2 Portuguese fertile controls, one of them homozygous for the minor allele.

Three rare (≤1%) intronic variants found in patients were also detected in unphenotyped individuals of large genome sequencing projects and are also predicted to affect splicing: rs55905583 (c.355-8G>C) increases the likelihood of expression of the canonical isoform, rs59834456 (c.822+141_822+144delATAT) creates a new enhancer site for SRp40 (score 91.86) and rs200423545 (c.968-26T>A) creates a new branch point in the vicinity of exon 5 (HSF score of 95.11). While rs200423545 is present at nearly twice the frequency in patients compared to Portuguese controls, rs55905583 and rs59834456 are found at similar or higher frequencies in the Portuguese controls (Table 2).

DMRT1 Domains are Highly Conserved in Mammals and Show Lower Genetic Variation in Controls

Even though DMRT1 is known to be involved in the sex determination pathway from *C. elegans* to humans, there is much uncertainty regarding the role of each functional domain. In order to grasp the potential functional impact of *DMRT1* variants, we analyzed the spectrum of coding variants found in unphenotyped control populations (Supplementary Table 7). In Figure 2 we show the alignment of the available vertebrate DMRT1 sequences, and the invertebrate *C. elegans*, with variants found in human populations marked along the sequence. The zinc finger-like DM domain is involved in binding to the promoter of its target genes and therefore missense variants in this region are expected to have a functional impact. Indeed, only one missense variant (rs201947617) is described in the DM domain, in a less conserved region of the sequence. The DMRT1-like domain, typically found in association with the DM domain but of unknown function, is also conserved in mammals, even though some sequence variability can be found in the central region. Missense mutations within the conserved region are predicted to be deleterious, with an average frequency in the populations surveyed (0.0307%) similar to those without predicted functional impact on the protein (missense tolerated, 0.0266%; and synonymous variants, 0.0225%). Outside the domains, in regions not predicted to affect protein function, missense

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

as well as synonymous substitutions are rare but more frequent (1.75% and 1.23% respectively) than potentially damaging missense variants (0.026%), which present a similar frequency within the domains (0.028%).

Compared to the rest of the *DMRT1* coding sequence (67/729 = 0.0919) domains also show less variants per base pair (DM: 3/141 = 0.0212; DMRT1-like: 17/252 = 0.0675), likely reflecting functional constraints. Also, the high conservation of a stretch of 83 amino acids comprising the DM and part of the DMRT1-like domains within vertebrates supports their importance for protein function (Fig. 2A). Another region (19 amino acids) in the vicinity of DMRT1-like domain stands out in the alignment as highly conserved in mammals, also exhibiting shared homology with avian DMRT1 (Supplementary Fig. 3).

Discussion

The role of DMRT1 orthologs in sex determination as well as in gonadal development and maintenance throughout evolution is well established in the literature. In fact, DMRT1 is crucial in the maintenance of the mammalian gonad and required for normal progression of spermatogenesis. Expression of *Dmrt1* is necessary in Sertoli cells to prevent female reprogramming of the adult testis (Matson *et al.* 2011) and in spermatogonia to maintain their proliferative state by avoiding meiotic entry (Zarkower 2013). Recently, Murphy and co-workers have shown that DMRT1 is a bifunctional transcriptional regulator, activating genes required for proper development of the male phenotype while repressing others that are specific of the female gonad (Murphy *et al.* 2010). These observations, together with the existence of human phenotypes associated with *DMRT1* alterations make this gene a strong candidate for a causal role in some of the cases of idiopathic NOA.

Following our initial discovery of recurrent *DMRT1* whole gene deletions in azoospermic patients from Utah (USA) and China (Lopes *et al.* 2013), we now describe novel and rare potentially regulatory variants in Portuguese patients, reinforcing that rare variants affecting *DMRT1* function may underlie some cases of severe spermatogenic impairment. In one NOA patient we uncovered a non-coding variant in the promoter (c.-223_-219CGAAA>T) that resides within a region shown to be involved in the repression of *Dmrt1* expression in rat and mouse Sertoli cells (Lei & Heckert 2002). Our *in silico* analysis not only supports the high conservation of this region across mammals (Fig. 1B) but also predicts that the human sequence is recognized by two heat-shock transcription factors (Fig. 1A and Supplementary Table 5), HSF1 and HSF2. Heat-shock transcription factors are known to activate heat-shock response genes, under stress conditions. Interestingly, these two transcription factors (HSF1 and HSF2) are highly expressed in testis, showing germ cell type specificity and are required for the progression of spermatogenesis. In adult mice testis, HSF2 expression is limited to spermatocytes and spermatogonia (Kallio *et al.* 2002) while HSF1 is highly and specifically expressed in the nuclei of spermatocytes and round spermatids (Akerfelt *et al.* 2010). Apparently, *Hsf1* is specifically expressed in those stages of spermatogenesis where *Dmrt1* is not, suggesting that HSF1 might block the transcription of *Dmrt1*, therefore allowing spermatogenesis to progress.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Our analysis indicates that the base substitution within the repression motif of the human sequence has resulted in the acquisition of a second HSF1 binding site (Fig. 1 and Supplementary Table 5) absent in the ancestral sequence (Fig. 1). Remarkably, this human-specific HSF1 binding site is disrupted by the c.-223_-219CGAAA>T variant, suggesting that it may result in *DMRT1* misregulation and the development of NOA. Indeed, the histology of this patient revealed maturation arrest, resembling the phenotype of *Hsf1-/-* male mice which display regions of the seminiferous tubules containing only spermatogonia (Akerfelt *et al.* 2010). It is noteworthy that a mutation in *HSF2*, also predicted to bind the *DMRT1* promoter, has been associated with azoospermia (Mou *et al.* 2013). It would be interesting to further explore the role for HSF1 in *DMRT1* regulation during spermatogenesis in order to understand how its dynamics can contribute to the phenotype.

Three alternatively spliced *DMRT1* transcripts have been described in the human testis: DMRT1a, DMRT1b and DMRT1c (Cheng *et al.* 2006). DMRT1a is the canonical DMRT1 isoform, showing the highest expression, and translates into a 375 amino acid protein. Both DMRT1b and DMRT1c result from the incorporation of intronic sequences in the mature mRNA (intron 3 and intron 1 respectively) and are less expressed. Canonical splice signals are known to be insufficient for proper splicing, requiring the regulation provided by cis-elements such as enhancers or repressors, especially important when alternative splicing is involved (Cartegni *et al.* 2003). Indeed, the creation of a new consensus exonic splicing enhancer (ESE) motif by a deep intronic mutation has already been shown to result in the production of an aberrant mRNA involved in afibrinogenemia (Davis *et al.* 2009).

We have identified a novel indel that is predicted to alter splicing either by interfering with the reading frame of DMRT1c (c.354+38_insG and linked poly-T in intron 1, Supplementary Table 2) However, the length variation in the poly T in intron 1, was also observed in our controls makes them less likely candidates for a strong deleterious effect on DMRT1 function. Also, an African ancestry was determined for this NOA patient suggesting that this locus harbors longer alleles in African populations when compared to Europeans.

One rare (≤1%) non-coding variants more frequent in patients than in Portuguese fertile controls were also predicted to affect splicing (rs200423545, Table 2).

Furthermore, the single rare variant detected in the 5′-UTR (rs144122237) of one patient potentially disrupts a binding site for the transcription factor RFX3, which is highly expressed in the testis. This variant has not been detected in the European control populations surveyed (our own and those included in databases) and was only present in the patient with African ancestry. The relatively low frequency of this variant in Africans (1.8%; 6 female and 3 male carriers) does not preclude an effect on fertility and overall our results highlight the need to consider the combinatorial effect of regulatory variants, given the interplay between factors with different roles (expression regulation, splicing, etc), as well as the genetic background of each individual.

Increasing data has been showing that synonymous variants might have an effect on transcription, splicing, mRNA transport or translation [Reviewed in (Goymer 2007)]. In

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

accordance, we have found two synonymous variants overrepresented in patients that were predicted to potentially alter splicing by the creation of either new branch point sequences (rs376518776; c.213G>A) or new motifs for ESEs (rs34946058; c.783C>G). At the rs34946058 triallelic locus the T allele, and not the G allele found in our patients, has been recently reported in two men of German ancestry with cryptozoospermia (sperm concentration <0.1 million/mL), but also in one normozoospermic control (Tewes *et al.* 2014). Interestingly, the Portuguese population seems to harbor only two alleles at this locus, the reference C and the minor allele G that reaches a frequency of approximately 6% in African populations. Three patients carried the G allele, all of which had been previously analyzed for genome-wide SNPs (Lopes *et al.* 2013) and show no evidence for African ancestry/admixture. One fertile control also carried the G allele and another fertile control was homozygous for the same allele. Thus, even though the sperm parameters for the fertile G homozygous individual are unknown, the relatively high frequency of this allele in Africans does not support a strong impact on fertility.

Overall, our *in silico* functional analysis suggests that some of the of *DMRT1* variants found in patients, which are not predicted to affect the protein (synonymous and non-coding) may however be involved in alternative splicing and other regulatory mechanisms. Per se, these variants are likely to have a mild impact on fertility, probably affecting the regulatory network controlling spermatogenesis but may also be in LD with intronic cis-acting SNPs and therefore a combinatorial effect of regulatory variants could be implicated in NOA. Moreover, the lack of deleterious coding mutations and intragenic rearrangements (screened by MLPA) in our Portuguese cohort supports the view that *DMRT1* genetic defects are rare (Lopes *et al.* 2013, Tewes *et al.* 2014) and strong association for individual variants may only be detected in multicenter studies with larger cohorts.

Concordantly, our analysis of the DMRT1 protein amongst vertebrates (Fig. 2) reinforces the previous observations. The highly conserved DM domain as well as some regions of DMRT1-like domain seem to be under stronger functional constraints displaying less variation. Together with our previous results for CNVs (Lopes *et al.* 2013), and a recent screening for *DMRT1* variants in German patients (Tewes *et al.* 2014), the finding of variants potentially affecting *DMRT1* expression supports a model where changes in dosage in this gene may underlie some cases of NOA. The association of *DMRT1* hemizygosity with human disorders of sex development (Barbaro *et al.* 2009, Õunap *et al.* 2004, Raymond *et al.* 1999, Tannour-Louet *et al.* 2010, Veitia *et al.* 1997) but also NOA (Lopes *et al.* 2013, Tewes *et al.* 2014), further points to other cis- and trans-acting factors contributing to these more complex phenotypes. It seems therefore that both *DMRT1* copies must be functional for proper gonadal development and spermatogenesis maintenance and that variation affecting one or both alleles would result in phenotypes of increasing severity.

In summary, we provide evidence suggesting that rare regulatory variants in coding and non-coding regions of the *DMRT1* gene and its promoter might be involved in idiopathic non-obstructive azoospermia (NOA). Altogether our results draw attention to the potential contribution of regulatory variants to male infertility, namely NOA, reinforcing the complexity and significance of regulatory networks involved in the control of spermatogenesis.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Rui Pereira for the analysis of ancestry performed for one patient. This work was partially funded by the Portuguese Foundation for Science and Technology FCT/MCTES (PIDDAC) and co-financed by European funds (FEDER) through the COMPETE program [PTDC/SAU-GMG/101229/2008 to AML], CIGMH and Pest-OE/SAU/UI00009/2011. IPATIMUP is an Associate Laboratory of the Portuguese Ministry of Science, Technology, and Higher Education and is partially supported by FCT. AML and ACL are funded by FCT fellowships SFRH/BPD/73366/2010 and SFRH/BD/51695/2011, respectively.

References

Akerfelt M, Vihervaara A, Laiho A, Conter A, Christians ES, Sistonen L, Henriksson E. Heat shock transcription factor 1 localizes to sex chromatin during meiotic repression. The Journal of biological chemistry. 2010; 285:34469–76. [PubMed: 20802198]

Barbaro M, Balsamo A, Anderlid BM, Myhre AG, Gennari M, Nicoletti A, Pittalis MC, Oscarson M, Wedell A. Characterization of deletions at 9p affecting the candidate regions for sex reversal and deletion 9p syndrome by MLPA. European journal of human genetics : EJHG. 2009; 17:1439–47. [PubMed: 19417767]

Bratuś A. DM Domain Genes : Sexual and Somatic Development During Vertebrate Embryogenesis. 2012

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic acids research. 2003; 31:3568–71. [PubMed: 12824367]

Cheng HH, Ying M, Tian YH, Guo Y, McElreavey K, Zhou RJ. Transcriptional diversity of DMRT1 (dsx- and mab3-related transcription factor 1) in human testis. Cell research. 2006; 16:389–93. [PubMed: 16617334]

Davis RL, Homer VM, George PM, Brennan SO. A deep intronic mutation in FGB creates a consensus exonic splicing enhancer motif that results in afibrinogenemia caused by aberrant mRNA splicing, which can be corrected in vitro with antisense oligonucleotide treatment. Human mutation. 2009; 30:221–7. [PubMed: 18853456]

Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic acids research. 2009; 37:e67. [PubMed: 19339519]

Erdman SE, Chen HJ, Burtis KC. Functional and genetic characterization of the oligomerization and DNA binding properties of the Drosophila doublesex proteins. Genetics. 1996; 144:1639–52. [PubMed: 8978051]

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]

Goymer P. Synonymous mutations break their silence. Nat Rev Genet. 2007; 8:92.

Jungwirth A, Diemer T, Dohle G, Giwercman A, Kopa Z, Krausz C, Tournaye H. Guidelines on Male Infertility. Uroweb. 2013

Kallio M, Chang Y, Manuel M, Alastalo TP, Rallu M, Gitton Y, Pirkkala L, Loones MT, Paslaru L, Larney S, Hiard S, Morange M, Sistonen L, Mezger V. Brain abnormalities, defective meiotic chromosome synapsis and female subfertility in HSF2 null mice. The EMBO journal. 2002; 21:2591–601. [PubMed: 12032072]

Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. Bioinformatics. 2007; 23:1289–91. [PubMed: 17379693]

Ledig S, Hiort O, Scherer G, Hoffmann M, Wolff G, Morlot S, Kuechler A, Wieacker P. Array-CGH analysis in patients with syndromic and non-syndromic XY gonadal dysgenesis: evaluation of array CGH as diagnostic tool and search for new candidate loci. Hum Reprod. 2010; 25:2637–46. [PubMed: 20685758]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Ledig S, Hiort O, Wunsch L, Wieacker P. Partial deletion of DMRT1 causes 46,XY ovotesticular disorder of sexual development. Eur J Endocrinol. 2012; 167:119–24. [PubMed: 22573722]

Lei N, Heckert LL. Sp1 and Egr1 regulate transcription of the Dmrt1 gene in Sertoli cells. Biol Reprod. 2002; 66:675–84. [PubMed: 11870074]

Lima, AC.; Lopes, AM. Autosomal Mutations and Spermatogenic Failure. In: Encyclopedia of Life Sciences. , editor. eLS. John Wiley & Sons, Ltd; Chichester: 2014.

Liu HX, Zhang M, Krainer AR. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. Gene Dev. 1998; 12:1998–2012. [PubMed: 9649504]

Lopes AM, Aston KI, Thompson E, Carvalho F, Goncalves J, Huang N, et al. Human Spermatogenic Failure Purges Deleterious Mutation Load from the Autosomes and Both Sex Chromosomes, including the Gene DMRT1. Plos Genet. 2013; 9

Machado AZ, da Silva TE, Costa EMF, dos Santos MG, Nishi MY, Brito VN, Mendonca BB, Domenice S. Absence of inactivating mutations and deletions in the DMRT1 and FGF9 genes in a large cohort of 46,XY patients with gonadal dysgenesis. Eur J Med Genet. 2012; 55:690–4. [PubMed: 22939835]

Matson CK, Murphy MW, Sarver AL, Griswold MD, Bardwell VJ, Zarkower D. DMRT1 prevents female reprogramming in the postnatal mammalian testis. Nature. 2011; 476:101. + [PubMed: 21775990]

Moniot B, Berta P, Scherer G, Sudbeck P, Poulat F. Male specific expression suggests role of DMRT1 in human sex determination. Mechanisms of development. 2000; 91:323–5. [PubMed: 10704857]

Mou LS, Wang YD, Li HG, Huang Y, Jiang T, Huang WR, Li ZS, Chen J, Xie J, Liu YC, Jiang ZM, Li XX, Ye JX, Cai ZM, Gui YT. A dominant-negative mutation of HSF2 associated with idiopathic azoospermia. Hum Genet. 2013; 132:159–65. [PubMed: 23064888]

Murphy MW, Sarver AL, Rice D, Hatzi K, Ye K, Melnick A, Heckert LL, Zarkower D, Bardwell VJ. Genome-wide analysis of DNA binding and transcriptional regulation by the mammalian Doublesex homolog DMRT1 in the juvenile testis. P Natl Acad Sci USA. 2010; 107:13360–5.

Õunap K, Uibo O, Zordania R, Kiho L, Ilus T, Oiglane-Shlik E, Bartsch O. Three patients with 9p deletions including DMRT1 and DMRT2: A girl with XY complement, bilateral ovotestes, and extreme growth retardation, and two XX females with normal pubertal development. Am J Med Genet A. 2004; 130A:415–23. [PubMed: 15481033]

Pereira R, Phillips C, Pinto N, Santos C, dos Santos SE, Amorim A, Carracedo A, Gusmao L. Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. Plos One. 2012; 7:e29684. [PubMed: 22272242]

Practice Committee of American Society for Reproductive Medicine in collaboration with Society for Male Reproduction and Urology. Evaluation of the azoospermic male. Fertility and sterility. 2008; 90:S74–7. [PubMed: 19007652]

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2012.

Raymond CS, Parker ED, Kettlewell JR, Brown LG, Page DC, Kusz K, Jaruzelska J, Reinberg Y, Flejter WL, Bardwell VJ, Hirsch B, Zarkower D. A region of human chromosome 9p required for testis development contains two genes related to known sexual regulators. Human molecular genetics. 1999; 8:989–96. [PubMed: 10332030]

Smale ST, Baltimore D. The Initiator as a Transcription Control Element. Cell. 1989; 57:103–13. [PubMed: 2467742]

Tannour-Louet M, Han S, Corbett ST, Louet JF, Yatsenko S, Meyers L, Shaw CA, Kang SHL, Cheung SW, Lamb DJ. Identification of De Novo Copy Number Variants Associated with Human Disorders of Sexual Development. Plos One. 2010; 5:e15392. [PubMed: 21048976]

Tewes AC, Ledig S, Tuttelmann F, Kliesch S, Wieacker P. DMRT1 mutations are rarely associated with male infertility. Fertility and sterility. 2014; 102:816–20e3. [PubMed: 24934491]

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3--new capabilities and interfaces. Nucleic acids research. 2012; 40:e115. [PubMed: 22730293]

Veitia R, Nunes M, Brauner R, DocoFenzy M, JoannyFlinois O, Jaubert F, LortatJacob S, Fellous M, McElreavey K. Deletions of distal 9p associated with 46,XY male to female sex reversal: Definition of the breakpoints at 9p23.3-p24.1. Genomics. 1997; 41:271–4. [PubMed: 9143505]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Zarkower D. DMRT Genes in Vertebrate Gametogenesis. Curr Top Dev Biol. 2013; 102:327–56. [PubMed: 23287039]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

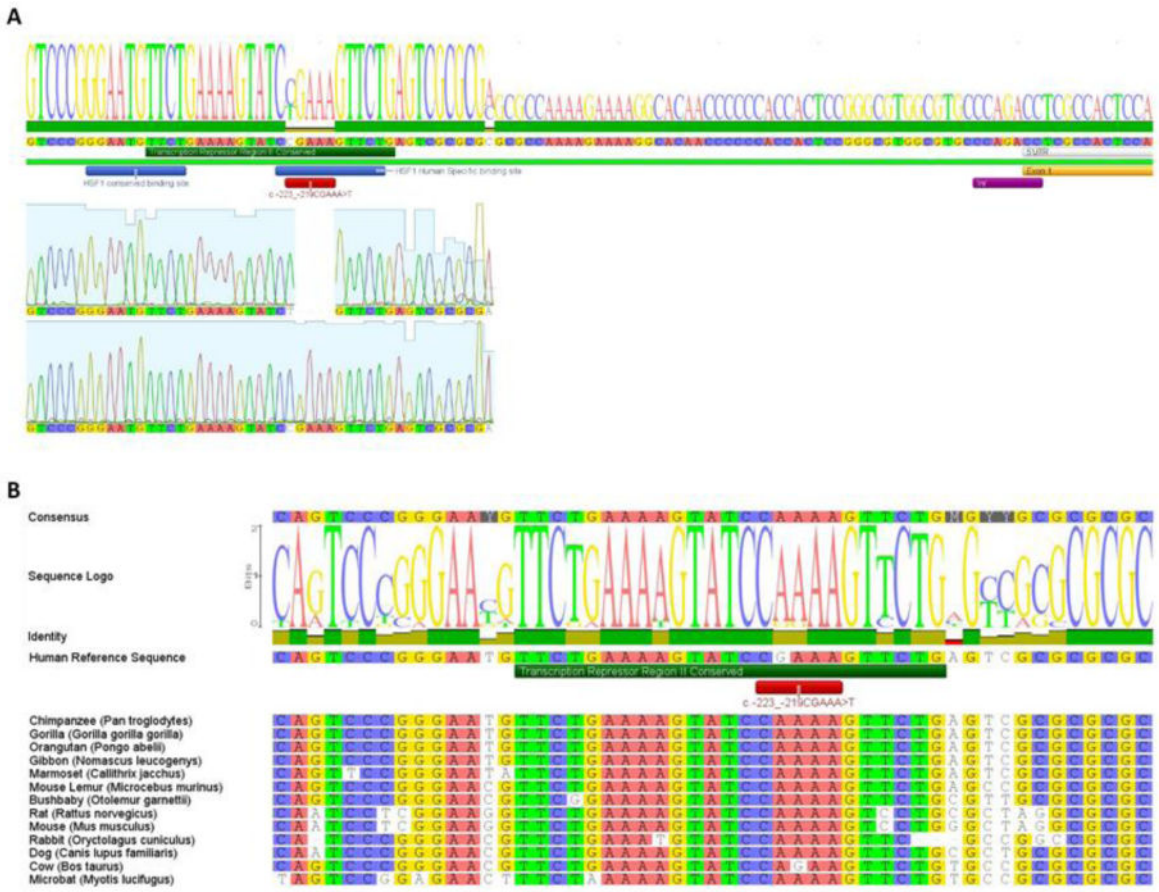

Figure 1. *DMRT1* promoter sequence in the proximal region of the transcription start site (TSS)

The *DMRT1* promoter region in humans (A) harbors an Inr element (purple filled box), which is characteristic of TATA-less genes, and also one region associated with DMRT1 repression from -88 to -63 bp upstream the TSS (green filled boxes). We have found by allele specific amplification a complex variant in the promoter of one azoospermic patient (A; red filled boxes). Analysis of the proximal region of TSS has predicted two binding sites for HSF1 within the conserved motif (A; blue boxes), one of which is disrupted by this novel patient-specific variant. Aligning the human sequence of this motif with the sequences from several species representative of mammalian clades (B) shows a high degree of conservation for this region (green filled boxes). Sequence annotations were performed using Geneious v.5.5.8 software and alignments were obtained applying the ClustalW algorithm available in the Geneious v.5.5.8 software.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Figure 2. Alignment of DMRT1 protein sequences from several species representative of vertebrates clades, including human

A high degree of identity is notably shared by the DM domain within vertebrates. Actually, a stretch of 83 amino acids containing the DM and part of the DMRT1-like domains seems to be fairly maintained amongst these clades. Interestingly, the DMRT1-like domain is greatly conserved in its extremities and more variability can be found at its center. This alignment was obtaining applying the ClustalW algorithm available with the Geneious v5.5.8 software. Amino acids are colored when they match at least 75% of the sequences. Purple filled boxes mark the domains; yellow and red filled boxes represent missense variants annotated in the Ensembl database, respectively predicted by at least one software (SIFT or PolyPhen) as not affecting the protein or as deleterious.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Novel DMRT1 non-coding variants restricted to patients

Name	Genomic location	Genomic context	Variant Type	Minor allele	MAF NOA	Global MAF	MAF PT Controls
c.-223_-219CGAAA>T	chr9:841616-841620	Promoter Region	Upstream gene variant	T	1/278	0	0/714
c.355-6T>C	chr9:846954	Intron 1	Intronic (Splice region)	C	1/280	0	0/700
c.823-64_823-62delATT	chr9:916699-916701	Intron 3	Intronic	-ATT	1/290	0	0/660

MAF: Minor allele frequency

Andrology. Author manuscript; available in PMC 2016 September 01.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Andrology. Author manuscript; available in PMC 2016 September 01.

Table 2

Annotated variants detected in the cohort of Portuguese NOA patients

Reference name	Genomic location	Genomic context	Variant Type	Minor allele detected	Patients MAF [a]	Global MAF [b]	EUR MAF [c]	PT Controls MAF [d]	p-value [*]
rs144122237	chr9:841785	5′ UTR	5′ UTR variant	T	0.0036	0.004	0.0000	0.0000	0.2875
rs3739584	chr9:841825	5′ UTR	5′ UTR variant	T	0.1475	0.141	0.1378	0.1362	.
rs3739583	chr9:841971	Exon 1	Missense	A	0.1475	0.132	0.1227	0.1344	.
rs376518776	chr9:842051	Exon 1	Synonymous	A	0.0036	0.000	0.0006	0.0000	0.3440
rs55905583	chr9:846952	Intron 1	Intronic (Splice Region)	C	0.0071	0.011	0.0006	0.0057	1.0000
rs2273929	chr9:847208	Intron 2	Intronic	G	0.4607	0.417	0.5409	.	.
rs16925431	chr9:894036	Exon 3	Synonymous	C	0.0109	0.032	0.0222	.	.
rs146975077	chr9:894147	Exon 3	Synonymous	C	0.0036	0.009	0.0002	0.0014	0.4866
rs34946058	chr9:894156	Exon 3	Synonymous	G	0.0109	0.042	0.0007	0.0043	0.3596
rs59168737	chr9:894224	Intron 3	Intronic	A	0.0109	0.039	0.0220	.	.
rs1033836	chr9:894297	Intron 3	Intronic	A	0.0471	0.101	0.0409	.	.
rs112866575	chr9:967939-967940	Intron 4	Intronic	+CTCCCTTT	0.0229	0.149	0.0985	.	.
rs200423545	chr9:967959	Intron 4	Intronic	A	0.0099	0.002	0.0024	0.0057	0.4291
rs376062302	chr9:967953-967954	Intron 4	Intronic	+TTCTCTCT	0.0915	0.1236 [e]	n.a	.	.
rs279895	chr9:967981	Intron 4	Intronic (Splice Region)	G	0.1765	0.111	0.1658	.	.
rs79358387	chr9:968150-968151	3′ UTR	3′ UTR variant	+T	0.1144	0.223	0.1135	.	.
rs279894	chr9:968334	3′ UTR	3′ UTR variant	G	0.1800	0.077	0.1541	.	.

MAF: Minor allele frequency detected in:

[a] the NOA Portuguese cohort

[b] all populations screened in large genome sequencing projects

[c] European populations of the 1000 Genomes (EUR and CEU), ESP (European-American) and CLINSEQ (CSAgilent)

[d] the Portuguese (PT) control population

[e] control population used by Machado et al. 2012

[*] Fisher's exact test using only the PT population as control

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Note: rs59834456 was detected in two Bantu individuals but no population data is available in the Ensembl release of March 2015