Genetic Dissection of the AZF Regions of the Human Y Chromosome: Thriller or Filler for Male (In)fertility?

Paulo Navarro-Costa,1,2,3 Carlos E. Plancha,2 and João Gonçalves1

1 Departamento de Genética, Instituto Nacional de Saúde Dr. Ricardo Jorge, 1649-016 Lisboa, Portugal
2 Faculdade de Medicina de Lisboa, Instituto de Histologia e Biologia do Desenvolvimento, 1649-028 Lisboa, Portugal
3 Faculdade de Medicina de Lisboa, Instituto de Medicina Molecular, 1649-028 Lisboa, Portugal

Correspondence should be addressed to Paulo Navarro-Costa, navarro-costa@fm.ul.pt

Received 17 December 2009; Accepted 23 April 2010

Copyright © 2010 Paulo Navarro-Costa et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The azoospermia factor (AZF) regions consist of three genetic domains in the long arm of the human Y chromosome referred to as AZFa, AZFb and AZFc. These are of importance for male fertility since they are home to genes required for spermatogenesis. In this paper a comprehensive analysis of AZF structure and gene content will be undertaken. Particular care will be given to the molecular mechanisms underlying the spermatogenic impairment phenotypes associated to AZF deletions. Analysis of the 14 different AZF genes or gene families argues for the existence of functional asymmetries between the determinants; while some are prominent players in spermatogenesis, others seem to modulate more subtly the program. In this regard, evidence supporting the notion that DDX3Y, KDM5D, RBMY1A1, DAZ, and CDY represent key AZF spermatogenic determinants will be discussed.

1. Introduction

The notion that functional determinants of spermatogenesis map to the Y chromosome (Y) was established in the 1970s [1]. Ever since the pioneering observation that deletions in the long arm of the Y chromosome (Yq) could be associated to defects in sperm production, researchers have tried to precisely map and identify such factors. In the course of this paper, a thorough genetic and functional analysis of the Y regions involved in spermatogenesis will be undertaken. These are designated as azoospermia factor (AZF) regions and they represent an area of significant interest in the field of human reproduction. In order to give added insight to this topic, the present manuscript will start with a brief overview of the major developments in the mapping of the AZF domains.

2. Historical Perspective on the Mapping of AZF

2.1. The Early Years. Following initial reports tentatively linking the loss of genetic material in Yq to azoospermia and hypogonadism, Tiepolo and Zuffardi established in 1976 the first solid association between Y chromosome deletions and abnormal spermatogenesis [1–4]. The authors screened 1170 infertile men for karyotypic abnormalities and observed deletions removing both fluorescent (heterochromatic) and nonfluorescent (euchromatic) Yq segments in 6 azoospermic men. This association signalled the advent of a new era in the study of the Y chromosome: the identification and characterization of the Yq genetic determinants involved in spermatogenesis (Figure 1). In this regard, significant efforts were invested into mapping the azoospermia factor to a specific Yq region. These initial studies were based on the development of linear deletion interval maps using Y-specific DNA probes in samples from infertile men with cytogenetically visible Y chromosome abnormalities (illustrative examples: [5–7]). Although Bardoni and colleagues did manage to map the human spermatogenesis locus to a more precise Yq interval (Yq11.23), the early deletion mapping strategies were met with meagre success. This stemmed both from the lack of suitable DNA markers, a consequence of the highly repetitive organization of Yq; and a dependence on the relatively rare occurrence of cytogenetically visible Yq abnormalities.
2.2. Beyond the Microscope: The Concept of Y Microdeletions. Nevertheless, these efforts were of importance in establishing “rough drafts” of the Yq genomic map, particularly in identifying markers to be used in subsequent projects. Appropriately, Ma and colleagues in 1992 mapped a marker panel consisting of 28 DNA probes using a collection of patients with Yq structural abnormalities [8]. The real breakthrough associated to the study was the screening of “chromosomally normal” azoospermic men, leading to the identification of two deletion patterns not detectable by karyotype visualization (and therefore dubbed microdeletions). The implications of this result were paramount. Firstly, it suggested that the AZF region might in fact have a multipartite organization, with the authors referring to them as AZFa and AZFb in a subsequent report (Figure 1) [9]. Secondly, it established the notion that small Yq interstitial deletions not visualized in a standard karyotype analysis might be a causative agent of spermatogenic failure. Therefore, it became evident that the mapping of AZF could benefit from microdeletion screening programs in infertile men with apparently normal karyotypes.

Advances in molecular biology techniques, more specifically the use of PCR-based analyses of Yq genomic markers heralded a new stage in the quest for the AZF domains [10–12]. Despite the failure of some early studies in confirming the existence of two AZF regions, both the notions of multiple AZF loci and of the advantages of screening karyotypically normal infertile men (not necessarily azoospermic) gradually became entrenched in the scientific community [13–15]. The corollary of this strategy was the screening of 370 idiopathic infertile men (either azoospermic or severe oligozoospermic) with a marker panel consisting of 76 Yq sequence-tagged sites (STs), most of them previously mapped by Vollrath and colleagues in 1992 [16]. The use of a large cohort of infertile men was crucial for the identification of less frequent microdeletion patterns that would otherwise pass undetected. This study revealed the existence of not two but three AZF regions (AZFa, AZFb, and AZFc) corresponding to three deletion intervals, each associated to a specific infertility phenotype (Figure 1). Therefore, the criterion for defining the AZF regions was above all functional since it was based on particular spermatogenic disruption phenotypes as means to delineate genomic regions. AZFa deletions were associated to complete absence of germ cells in the testis tubules (Sertoli cell-only syndrome; SCOS) and AZFb deletions to maturation arrest at the spermatocyte stage. Contrary to the azoospermia phenotype recorded in AZFa and AZFb deletions, AZFc deletions were shown to be compatible with sperm production (albeit at reduced levels) and could be transmitted to the progeny. More specifically, AZFc deletions were associated to hypospermatogenesis (abnormally decreased sperm production) that stemmed from a mixed degree of germ cell atrophy in the testis tubules [15, 17]. Although Vogt and colleagues proposed estimates of AZF sequence length and a series of gene candidates responsible for the AZFb and AZFc deletion phenotypes, the exact length, structure and gene content of the three AZF intervals would only be fully characterized in subsequent studies. These would show that the functional partition of AZF into three individual regions was not reflected in structural terms, since the AZFb and AZFc sequences overlap (Figure 1).

2.3. From Microdeletion Screening Programs to Sequencing the Y. The concept of AZF microdeletion screening adopted by Vogt and colleagues was taken one step further in 1997 with the analysis of infertile men irrespectively of their spermatogenic phenotype [18]. This revealed not only that Yq microdeletions were present in ~7% of the infertile population, but also, more significantly, that a considerable variability in sperm counts was associated to such microdeletions. In fact, some microdeletion types were even detected in infertile men with normal sperm concentrations. Although it was later shown that only some partial AZF deletions might be compatible with normozoospermia, this study signalled the importance of a systematic screening of these molecular defects in the infertile population. Accordingly, AZF microdeletions are, alongside karyotype abnormalities, the most common known genetic cause of spermatogenic failure [19]. Pryor and colleagues also screened fertile men, detecting microdeletions in 2% of the individuals. This led the authors to conclude that some microdeletion patterns correspond to Y variants devoid of any obvious phenotypical consequences for male fertility. Thus, an adequate deletion screening protocol should require a validated selection of genetic markers, as well as a precise understanding of the AZF sequence in order to rule out functionally meaningless polymorphisms. Such degree of knowledge was dependent on the availability of a reference sequence for the male-specific region of the Y, which only materialized in the early 2000s [20].

After this brief overview of the historical landmarks on the identification of AZF, the following paragraphs contain a thorough genetic and functional characterization of the three intervals. For an abridged analysis of the mapping and functional properties of the AZF genes please consult Table 1 and Figure 2.

3. The AZFa Region of the Y Chromosome

The AZFa region totals 792 kb and was fully sequenced in 1999 [21]. AZFa maps to proximal Yq (chromosome location: ∼12.9–13.7 Mb) and unlike either AZFb or AZFc, is exclusively constituted by single-copy DNA (Figure 2). The region is flanked by two human endogenous retrovirus (HERV) elements, spanning approximately 10 kb each and displaying considerable levels of sequence identity. Although the degree of similarity varies along the elements (with the distal HERV copy having an additional insertion of ∼1.5 kb of transposon material—the L1 insertion), an overall sequence identity of 94% potentiates the occurrence of HERV-mediated rearrangements [22]. Accordingly, the complete AZFa deletion is the result of non-allelic homologous recombination (NAHR) between the two HERV elements [22–24]. This deletion is always associated with SCOS and is a fairly rare event, representing less than 5% of the reported
Table 1: Functional and genomic characterization of the AZF genes.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>AZF location (amplicon unit)</th>
<th>mRNA expression/protein evidence (germ cell type)</th>
<th>Protein type</th>
<th>Role in the male germline</th>
<th>Homologous genes (location)</th>
</tr>
</thead>
<tbody>
<tr>
<td>USP9Y</td>
<td>Ubiquitin specific peptidase 9, Y-linked</td>
<td>AZFa</td>
<td>Ubiquitous/Yes (spermatid)</td>
<td>Ubiquitin-specific protease</td>
<td>Involved in protein turnover in spermatogenesis (?)</td>
<td>USP9X (X chr.)</td>
</tr>
<tr>
<td>DDX3Y</td>
<td>DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked</td>
<td>AZFa</td>
<td>Ubiquitous/Yes (spermatogonia only)</td>
<td>ATP-dependent RNA helicase (deduced by similarity)</td>
<td>RNA metabolism in pre-meiotic germ cells (deduced by similarity)</td>
<td>DDX3X (X chr.)</td>
</tr>
<tr>
<td>C1orf15A</td>
<td>Chromosome Y open reading frame 15A/B</td>
<td>AZFb</td>
<td>Ubiquitous/No (n/a)</td>
<td>n/a</td>
<td>n/a</td>
<td>C1orf15 (X chr.)</td>
</tr>
<tr>
<td>C1orf15B</td>
<td></td>
<td>AZFb</td>
<td>Ubiquitous/No (n/a)</td>
<td>n/a</td>
<td>n/a</td>
<td>C1orf15 (X chr.)</td>
</tr>
<tr>
<td>RPS4Y2</td>
<td>Ribosomal protein S4, Y-linked 2</td>
<td>AZFb</td>
<td>Testis-specific/No</td>
<td>Ribosomal protein subunit (deduced by similarity)</td>
<td>Regulation of mRNA binding to the ribosome (deduced by similarity)</td>
<td>RPS4Y1 (Y chr.) RPS4X (X chr.)</td>
</tr>
<tr>
<td>EIF1AY</td>
<td>Eukaryotic translation initiation factor 1A, Y-linked</td>
<td>AZFb</td>
<td>Ubiquitous/Yes (n/a)</td>
<td>Demethylase of di- and tri-methylated H3K4</td>
<td>Chromatin remodelling in meiosis</td>
<td>KDM5C (X chr.)</td>
</tr>
<tr>
<td>KDM5D</td>
<td>Lysine (K)-specific demethylase 5D</td>
<td>AZFb</td>
<td>Ubiquitous/Yes (primary spermatocytes)</td>
<td>Transcription factor</td>
<td>Gene expression regulation (?)</td>
<td>HSFX1 (X chr.) HSFX2 (X chr.)</td>
</tr>
<tr>
<td>XKRY</td>
<td>XK, Kell blood group complex subunit-related, Y-linked</td>
<td>XKRY: AZFb (yel3) XKRY2: AZFb (yel4)</td>
<td>Testis-specific/No</td>
<td>Multipass transmembrane protein (?)</td>
<td>Gamete interaction (?)</td>
<td>—</td>
</tr>
<tr>
<td>XKRY2</td>
<td></td>
<td>HSFY1: AZFb (b5) HSFY2: AZFb (b6)</td>
<td>Testis-predominant/Yes (up to the spermatid stage)</td>
<td>Signalling molecule (deduced by similarity)</td>
<td>Germ cell apoptosis (?)</td>
<td>—</td>
</tr>
<tr>
<td>PRY</td>
<td>PTPN13-like, Y-linked/2</td>
<td>PRY: AZFb (b1) PRY2: AZFb/c (b2)</td>
<td>Testis-specific/Yes (in some post-meiotic cells)</td>
<td>RNA-binding protein with protein interaction domain</td>
<td>RNA splicing and metabolism, signal transduction and meiotic regulation</td>
<td>RBMX (X chr.) RBMXL1 (Chr. 1) RBMXL2 (Chr. 11) RBMXL9 (Chr.9)</td>
</tr>
<tr>
<td>PRY2</td>
<td></td>
<td>RBMY1A1</td>
<td>RNA binding motif protein, Y-linked, family 1, member A1</td>
<td>Testis-specific/Yes (mainly meiotic and post-meiotic cells)</td>
<td>Transcriptional co-repressor with histone acetyltransferase activity</td>
<td>CDYL (Chr. 6) CDYL2 (Chr. 16)</td>
</tr>
<tr>
<td>BPY2</td>
<td>Basic charge, Y-linked, 2 B/C</td>
<td>BPY2: AZFb/c (g1) BPY2B: AZFb/c (g2) BPY2C: AZFc (g3)</td>
<td>Testis-specific/Yes (all spermatogenic stages)</td>
<td>Highly charged protein</td>
<td>Regulation of the cytoskeletal network (?)</td>
<td>—</td>
</tr>
<tr>
<td>BPY2B</td>
<td></td>
<td>CDY1</td>
<td>Testis-specific/Yes (post-meiotic cells)</td>
<td>Transcriptional co-repressor with histone acetyltransferase activity</td>
<td>Gene expression regulation and and post-meiotic nuclear remodelling</td>
<td>CDYL (Chr. 6) CDYL2 (Chr. 16)</td>
</tr>
<tr>
<td>BPY2C</td>
<td></td>
<td>CDY2</td>
<td>Testis-specific/Yes (post-meiotic cells)</td>
<td>RNA-binding protein with protein interaction domain</td>
<td>RNA splicing and metabolism, signal transduction and meiotic regulation</td>
<td>RBMX (X chr.) RBMXL1 (Chr. 1) RBMXL2 (Chr. 11) RBMXL9 (Chr.9)</td>
</tr>
</tbody>
</table>
Table 1: Continued.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>AZF location (amplicon unit)</th>
<th>mRNA expression/protein evidence (germ cell type)</th>
<th>Protein type</th>
<th>Role in the male germline</th>
<th>Homologous genes (location)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAZ1</td>
<td></td>
<td>DAZ1: AZFb/c (r1)</td>
<td>Testis-specific/Yes (spermatogonia)</td>
<td>RNA-binding protein with protein interaction domain</td>
<td>Pre-meiotic regulation of transcript transport/storage, translation initiation and protein interaction</td>
<td>DAZ1 (Chr. 3) BOLL (Chr. 2)</td>
</tr>
<tr>
<td>DAZ2</td>
<td></td>
<td>DAZ2: AZFb/c (r2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAZ3</td>
<td></td>
<td>DAZ3: AZFc (r3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAZ4</td>
<td></td>
<td>DAZ4: AZFc (r4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For appropriate references please consult the manuscript. (?): inference lacking appropriate experimental validation, n/a: not available

Figure 1: Mapping the azoospermia factor (AZF) region in the human Y chromosome. Heterochromatin is depicted in grey, euchromatin in blue and the two pseudoautosomal regions in black. Starting from the initial identification in 1976 of a genomic domain in Yq euchromatin regulating spermatogenesis, several studies have progressively led to a more refined mapping of this determinant. In 1991 Bardoni and colleagues mapped AZF to the distal end of Yq euchromatin, while in the next year Vogt and colleagues provided evidence for a bipartite organization (AZFa and AZFb). A major landmark was recorded in 1996, with a large-scale deletion mapping project defining three AZF regions (AZFa, AZFb and AZFc) based on the association of specific spermatogenic disruption phenotypes to nonoverlapping deletion intervals. Yet, Repping and colleagues demonstrated in 2002 that the distal end of AZFb corresponded to the proximal domain of AZFc. For complete references please consult the manuscript text.

AZF deletions [25, 26]. The low prevalence most likely stems both from limitations of the deletion mechanism (such as the lack of multiple homology domains and a relatively short recombination target), and from its considerable deleterious effect on fertility. Fittingly, the corresponding NAHR product, the AZFa duplication, is detected at a four-fold higher frequency when compared to that of the deletion [27].

AZFa contains two ubiquitously expressed genes with X homologues that escape inactivation: ubiquitin specific peptidase 9, Y-linked (USP9Y) and DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked (DDX3Y). The precise roles these genes play in the spermatogenic process are relatively unknown, with most of the available data arising from the functional characterization of partial AZFa deletions. Interestingly, despite being ubiquitously expressed, the deletion of both genes only appears to have phenotypical consequences in the male germline, suggesting that their function is tissue-specific and/or that the X homologues can exert a rescue effect in somatic lineages.
3.1. AZFa Gene Content

3.1.1. USP9Y. The USP9Y protein is an ubiquitin-specific protease and member of the C19 cystein peptidase family. These enzymes promote the intracellular cleavage of ubiquitin molecules from ubiquitinated proteins [28, 29]. Appropriately, a role for USP9Y in the regulation of protein turnover during spermatogenesis has been proposed [30, 31]. USP9Y shares 91% identity with its X homologue (USP9X) suggesting that both target similar molecules and
may overlap functionally [32]. Studies in murine gametogenesis have shown that while USP9X expression starts as early as in the establishment of the primordial germ cell (PGC) population in both sexes, USP9Y only starts to be expressed in the male germline at the spermatid stage [32, 33]. This markedly distinct expression window hints at a temporal constrain in the regulation of USP9Y function, a probable consequence of its molecular targets only being present at later spermatogenic stages. Yet, available data are inconsistent with USP9Y being a key player in male gametogenesis. Although USP9Y deletions were initially thought to be exclusively associated to azoospermia, and the cause of the AZFa deletion phenotype, two more recent reports demonstrate otherwise. Indeed, USP9Y deletions that are compatible with sperm production and with natural conception have already been identified, the latter corresponding to the complete deletion of the gene [34, 35]. Thus, published data points to USP9Y not being essential for male fertility, as observed in other primate lineages where the gene became inactive [36, 37]. Nevertheless, it is still premature to discard the involvement of USP9Y in the epistatic regulation of male gametogenesis [38].

3.1.2. DDX3Y. The DDX3Y protein has the hallmarks of an ATP-dependent RNA helicase belonging to the DEAD box protein family (characterized by the conserved motif Asp-Glu-Ala-Asp). The exact molecular role of DDX3Y is unknown, although the DEAD box proteins have been implicated in several key processes of RNA metabolism such as secondary structure alteration, splicing, spliceosome assembly and translation initiation [39]. DDX3Y and its X homologue (DDX3X; 91.7% sequence identity) are ubiquitously expressed, with expression levels peaking in testis [40, 41]. However, the widespread presence of both transcripts in adult tissues does not directly correlate with actual protein expression since DDX3Y, unlike DDX3X, is testis-specific. An additional layer of regulation can be invoked as both genes encode for testis-specific transcripts characterized by an overall shorter length and the presence of extended untranslated regions (UTRs) [32, 42]. This specificity in transcriptional profiles most likely serves to ensure a precise expression window as DDX3Y is detected predominantly in the cytoplasm of spermatogonia whereas DDX3X is mainly detected in spermatids [40]. The divergent expression window of the two genes suggests that DDX3Y may represent a specialization of DDX3X functions for premeiotic developmental stages. Yet, despite differences in cell type expression, the molecular functions of DDX3Y and DDX3X are probably analogous, as evidenced by functional rescue studies in murine cell lines [41].

Taking into account data from USP9Y deletions as well as the regulation of DDX3Y function, it is tempting to consider that the absence of the latter is the main causative agent for the complete AZFa deletion phenotype. However, this hypothesis still requires the validation warranted by the unambiguous identification of DDX3Y-specific deletions.
4. The AZFb Region of the Y Chromosome

As previously stated, the three AZF regions were defined on a functional basis: specific spermatogenic impairment phenotypes associated to specific deletion patterns. It should be noted that phenotypic specificity does not necessarily translate into genotypical individuality, as emphatically illustrated by the AZFb and AZFc regions. Although the AZFb and AZFc deletion phenotypes are noticeably different (maturation arrest and hypospermatogenesis, respectively), both sequences overlap in Yq (Figure 2). Actually, despite some early sequencing efforts pointing to a non-overlapping AZFb domain spanning 3.2 Mb, the molecular breakpoint characterization of AZFb deletions revealed not only a far larger extension for this region but also that the distal portion of the AZFb interval is part of AZFc [43, 44]. While several AZFb deletion patterns have been reported [45, 46], in the present paper the extension of this domain will be considered defined by Repping and colleagues in 2002 [43]. According to this definition, AZFb spans a total of 6.23 Mb and maps to ~18.1–24.7 Mb of the Y. AZFb contains three single-copy regions (from the large proximal u1 domain to the more distal and considerably shorter u2 and u3 regions), a DYZ19 satellite repeat array (embedded in the u1 region) and 14 multicopy sequence units (Figure 2). These units are termed amplicons and are organized in sequence families, with intrafamily homology levels exceeding 99%. Amplicon families are defined by a specific colour code (yellow, blue, turquoise, green, red or grey), with each family member identified by a numeral. Therefore, amplicons are referred to in an ab notation, where a represents the family colour code and b the corresponding member number. The AZFb amplicons are divided in 6 families and of the 14 amplicon units, half of them (yel3, yel4, b5, b6, b1, t1, t2) are exclusive to AZFb, with the remaining being shared with AZFc (Figure 2). Amplicons can also be categorized by a higher-order structural organization based on symmetrical arrays of contiguous repeat units. Such arrays are designated as palindromes and are defined by a symmetry axis separating two largely identical arms constituted by single or multiple palindrome units. AZFb contains palindromes P2 to P5, as well as the proximal part of P1. Indeed, the first description of AZFb deletions used palindrome notations to identify the NAHR recombination targets giving rise to the deletion pattern [43]. According to such notation, the complete AZFb deletion (P5/proximal P1) corresponds to the interval encompassed between amplicons yel3 and yel1.

The presence of extensive ampliconic domains in AZFb makes for very peculiar rearrangement dynamics. The complete AZFb deletion seems to occur at a similar or slightly increased rate to that of AZFa (~3 to 10 % of all Yq microdeletions), despite a much larger recombination target [47–49]. This result is somewhat counter-intuitive if we consider recombination target length as the main factor driving NAHR frequency. Nevertheless, this figure may rise five-fold if AZFb+c deletions are included, suggesting that the propensity for rearrangements may vary between amplicon units. Regarding the spermatogenic impairment phenotype of the complete AZFb deletion, patients are azoospermic with testicular analysis revealing the presence of arrested germ cells. This maturation arrest is usually at the spermatocyte/spermatid stage, yet some very rare instances of complete spermatogenesis in a small number of testis tubules have been reported [47, 48]. Thus, the chances of finding sperm in the testis of these patients are extremely remote.

The AZFb gene content reflects the mesh of different sequence types constituting this region, with single-copy genes mapping alongside ampliconic gene families (Figure 3). A total of 5 different single-copy transcription units map to AZFb: KDM5D [lysine (K)-specific demethylase 5D], EIF1AY (eukaryotic translation initiation factor 1A, Y-linked), RPS4Y2 (ribosomal protein S4, Y-linked 2), CYorf15A (chromosome Y open reading frame 15A) and CYorf15B (chromosome Y open reading frame 15B) [20, 43, 44]. Despite several efforts to assess their functional and regulatory properties, as a whole they can be considered as still poorly characterized.

4.1. Single Copy AZFb Genes

4.1.1. CYorf15. The CYorf15A and CYorf15B sequences have an X homologue (CXorf15) that belongs to the taxilin family and has been linked to transcriptional regulation in osteoblasts [50]. Yet, the role of CYorf15 sequences for either general or reproductive functions is unknown. Although CYorf15A and CYorf15B apparently encode for proteins homologous to the amino and carboxy-terminal domains of CXorf15, respectively, evidence for their existence is restricted to the identification of ubiquitously expressed transcripts [20].

4.1.2. RPS4Y2. RPS4Y2 corresponds to a fairly recent duplication of the RPS4Y gene, the latter encoding for a ribosomal protein subunit required for mRNA binding to the ribosome [51]. Since RPS4Y2 expression is testis-specific, a putative role in the posttranscriptional regulation of the spermatogenic program can be postulated [52]. Indeed, evidence for positive selection in the RPS4Y2 coding sequence suggests a hypothetical acquisition of germline-specific functions. Yet, confirmation of both the existence and functional properties of the RPS4Y2 protein are prerequisites for any further developments.

4.1.3. EIF1AY. EIF1AY is a ubiquitously expressed Y-linked member of the EIF-1A family—a sequence family involved in translation initiation [53]. The EIF-1A proteins are required for a high rate of protein biosynthesis since they enhance ribosome dissociation into subunits and stabilize the binding of the 43S complex (a 40S subunit, elF2/GTP/Met-tRNAi and elF3) to the 5′ end of capped RNA [54]. EIF1AY has an X-homologue (EIF1AX) and although evidence at the protein level is available, its functions are largely deduced by similarity to EIF1AX. In this regard, the acquisition of male-specific regulatory features by EIF1AY and/or the existence of partial functional overlap with EIF1AX are valid hypotheses.
4.1.4. KDM5D. KDM5D encodes for a histone H3 lysine 4 (H3K4) demethylase that forms a protein complex with the MSH5 DNA repair factor during spermatogenesis [55, 56]. This complex locates to condensed DNA during the leptotene/zygotene stage, suggesting an involvement in male germ cell chromatin remodelling. In accordance, by demethylating di- and tri-methylated H3K4, KDM5D may be involved in chromosome condensation during meiosis. Such possibility fits with the instances of maturation arrest at the spermatocyte stage associated to AZFb deletions. Despite the apparently male germline-specific functions, this gene is ubiquitously expressed and is homologous to KDM5C, an X-borne gene associated to X-linked mental retardation [57–60].

4.2. Multicopy AZFb Genes. Due to the presence of ampiclonic sequences AZFb contains a set of 7 different multicopy gene families: XKRY (XK, Kell blood group complex subunit-related, Y-linked), HSFY (heat shock transcription factor, Y-linked), RBMY1A1 (RNA binding motif protein, Y-linked, family 1, member A1), PRY (PTPN13-like, Y-linked) CDY (chromodomain protein, Y-linked), BPY2 (basic charge, Y-linked, 2), and DAZ (deleted in azoospermia). The members of these gene families make for a total of 20 transcription units in the reference AZFb sequence. Since several of them also map to AZFc, only genes with functional copies exclusively located in AZFb will be analysed in this section.

4.2.1. XKRY. The XKRY gene is expressed specifically in testis and maps to the yellow-coded ampiclon family. Although protein evidence is still lacking, sequence analysis suggests that the two active copies of XKRY (mapping to yel3 and yel4) encode for a multipass transmembrane transport protein similar to the XK protein. The latter locates to neuromuscular and hematopoietic cell membranes, with XK mutations causing specific disruption phenotypes [61–63]. However, a role for XKRY in spermatogenesis has yet to be validated despite tentative links to the fertilization process [64].

4.2.2. HSFY. HSFY encodes for a member of the heat shock factor family of transcriptional activators and displays testis-predominant expression. This gene maps to the blue amplicon, with the two functional units being restricted to b1 and b2. HSFY is subjected to alternative splicing, generating 3 different protein-coding transcripts with varying expression patterns [66]. HSFY has been identified in spermatogenic cells up to the spermatid stage and in Sertoli cells [67]. The protein’s stage-dependent translocation from the cytoplasm to the nucleus is suggestive of a developmentally regulated functional window, consistent with its role as a transcription factor. Nevertheless, as previously argued, the HSFY polyclonal antiserum used in the former study might have detected epitopes from the more widely expressed X homologues (HSFX1/2) [32]. In this regard, a subsequent report by the same team established that the mouse orthologue of HSFY is predominantly expressed in round spermatids [68]. Regardless of several uncertainties on the exact function and regulation of HSFY, a role in spermatogenesis has already been proposed based on observations in animal models and in infertile males. More specifically, Vinci and colleagues detected a partial AZFb deletion purportedly only affecting the functional copies of HSFY in an azoospermic man [69]. Yet, it should be noted that the heritability of the deletion was unknown. More recently, HSFY protein levels in spermatogenic cells were shown to be decreased in samples with maturation arrest, associating once again this gene to the regulation of male gametogenesis [70].

4.2.3. PRY. The PRY gene copies map to the blue ampiclons, with the two functional units being restricted to b1 and b2. These are designated as PRY and PRY2, respectively, and encode for a gene product with a low degree of similarity to protein tyrosine phosphatase, nonreceptor type 13. The latter corresponds to a signalling molecule involved in the regulation of a myriad of cellular processes, particularly in programmed cell death (illustrative example: [71]). PRY and PRY2 display testis-specific expression and additional regulation via alternative splicing [72]. Nevertheless, the alternative transcript seems to correspond to a nonfunctional isoform since it contains a premature stop codon truncating the product at about half. The expression of PRY in germ cells is irregular, with the protein being detected only in a few sperm and spermatids [73]. Interestingly, both transcript and protein levels were shown to be higher in the defective germ cell fraction of the ejaculate. Furthermore, PRY levels are increased in ejaculated sperm obtained from men with abnormal semen parameters, suggesting a link between its expression and defective spermatogenesis. Appropriately, a role for PRY in male germ cell apoptosis has been suggested based on the observation that approximately 40% of PRY-positive cells show DNA fragmentation. Yet, as acknowledged by the authors of the paper, results were insufficient to fully back the claim. Regardless of such considerations, available evidence points to a postmeiotic expression of PRY in restricted subsets of developing germ cells.

4.2.4. RBMY1A1. The RBMY1A1 gene family was identified in the early 1990s [13]. At the time, the functional properties of the then dubbed Y chromosome RNA recognition motif gene (YRRM) made it the first candidate azoospermia factor. Although RBMY1A1 is present in multiple copies along the Y chromosome, the six functional units cluster to the AZFb ampiclons [74, 75]. This complex arrangement, characterized by an extensive array of RBMY1A1 pseudogenes and sub-families, had thwarted initial attempts to precisely map this determinant [76, 77]. RBMY1A1 is part of the RBM gene family that also includes an X homologue (RBMX) and a set of autosomal retrogene-derived copies of RBMX (of these only RBMXL1, RBMXL2 and RBMXL9 are expressed, and protein evidence is only available for RBMXL2) (for review: [78]). Unlike its ubiquitously expressed X homologue, RBMY1A1 is expressed solely in male germ cells, with the protein displaying a nuclear location [75]. The main feature of the RBM family is the presence of a N-terminal RNA
recognition motif (RRM) responsible for the interaction with target RNA molecules [74, 79]. In this regard, RBM family members display characteristics of canonical RNA-binding proteins involved in nuclear RNA processing. In fact, this gene has been linked to the storage and transport of mRNA from the nucleus during spermatogenesis [80]. Contrasting with the other RBM genes, RBMY1A1 also contains a C-terminal protein interaction repeat domain enriched in serine, arginine, glycine, and tyrosine (SRYG) [78]. This serves as a probable regulatory region for the modulation of RBMY1A1 function.

The nuclear localization of RBMY1A1 is pinpointed to domains enriched in pre-mRNA splicing components, as evidenced in prophase I nuclei [81]. In accordance, efforts to identify RBMY1A1-interacting proteins have shown that pre-mRNA splicing regulators, particularly the SR and the SR-related proteins, are bona fide partners [82, 83]. These ubiquitously expressed factors also contain RRM domains, therefore their functional modulation via RBMY1A1 interaction emerges as a distinct possibility [84]. Additionally, RBMY1A1 may modulate cellular processes other than splicing regulation and mRNA metabolism since it has been shown to interact with the STAR and T-STAR proteins [85]. These act not only as splicing regulators but also as members of signal transduction pathways involved in cell cycle control. In this regard, RBMY1A1 can be involved in several aspects of meiotic and premeiotic regulation via the establishment of multiple protein complexes. Interestingly, the male germ cell-specific expression of RBMY1A1 is also mimicked by the autosomal RBMXL2 gene. In this case the nuclear localization of the protein during and immediately after meiosis is suggestive of meiotic specialization [86].

In accordance, haploinsufficiency of the murine RBMXL2 orthologue results in abnormal spermatogenesis in animal models [87].

The identification of the RNA targets of RBMY1A1 has been partially successful. It is believed that the RRM domain can bind RNA at both high and low affinity, making the characterization of target molecules complex [78]. Furthermore, the protein has an unique two-step mechanism for RNA recognition that starts with a sequence-specific interaction with the target molecule before eliciting a conformational modification [88]. This complex mechanism warrants RBMY1A1 a significant plasticity in terms of RNA partners. Studies in murine models have identified 12 different potential mRNA targets for RBMY1A1, most of them expressed in testis starting from the neonatal period [89]. Interestingly, the protein seems to be able to bind to its own alternative transcript, suggesting a complex regulatory network. The existence of alternative RBMY1A1 transcripts has also been detected in humans [90].

All the aforementioned properties seem to indicate that the disruption of RBMY1A1 plays a significant role in the AZFb deletion phenotype. In reality, both its expression pattern and putative role in male germ cell development support the notion that RBMY1A1 deletions perturb the meiotic program. Similarly, the disruption of KDM5D may also contribute to the deletion phenotype. In fact, RBMY1A1 and KDM5D are located in the germ cell nucleus during prophase I, suggesting involvement in meiotic orchestration. While this regulation may be directly exerted by KDM5D (via changes in chromatin structure), the role of RBMY1A1 might be mediated by effector proteins or by transcriptional regulation of mRNA targets.

Regardless of the actual contribution of the AZFb genes for the maturation arrest phenotype, a predominantly structural effect of the AZFb deletion on meiotic progression cannot be discarded. The removal of such a large stretch of Yq chromatin (~6.23 Mb) may result in X-Y pairing impairment during meiosis and lead to meiotic breakdown. This effect has already been identified in patients with AZFb+c deletions [91]. In such cases, a significant decrease in spermatocyte X-Y bivalent formation was recorded, with only 29% of the cells having juxtaposed telomere signals. It can be argued that the lower rate of sex chromosome pairing arises from DNA conformational changes that undermine meiotic efficiency. However, it is impossible to dissociate the effect of gene loss from the observed pairing impairment, particularly since we are dealing with genes involved in cell cycle progression. Therefore, and in light of all evidence, the maturation arrest phenotype associated to AZFb deletions most probably stems from a combination of genetic disruption with structural defects in the chromosome.

5. The AZFc Region of the Y Chromosome

The AZFc region is one of the most remarkable domains of the human genome, displaying a structural and functional intricacy only paralleled by the major histocompatibility complex in chromosome 6. The sequencing of AZFc represented a monumental effort based on laborious data compilation and inventive analytical tools [65]. Paradoxically, the effort put into sequencing AZFc revealed that the obtained sequence corresponds to just one of the plethora of expected genomic variants in the Y chromosome population. This observation arises from the fact that AZFc is almost exclusively constituted by amplicons. Indeed, the extensive homology between intra-family ampliconic units is a fertile substrate for large-scale AZFc structural rearrangements (deletions, duplications and inversions) as well as more subtle sequence modifications. Yet, both the molecular drivers and phenotypical consequences of AZFc variability fall outside the scope of the present review, having been thoroughly discussed elsewhere [92].

5.1. Genomic Assembly of AZFc. Approximately 95% of the reference AZFc sequence is constituted by ampliconic units belonging to five different colour-coded families (blue, green, red, grey and yellow). The remainder corresponds to a duplicated spacer for the red amplicons (present in the two red amplicon clusters) and a single copy domain (u3) similar to other Y regions (Figure 2). Structurally, the region contains one large (P1) and one smaller (P2) palindrome, as well as the b2-u3-g1 segment. Different models for the genomic assembly of the reference AZFc structure have been proposed. A simple, two-step model states that the palindromes arose from supragenic tandem duplication.
followed by inversion [65]. Traces of such events have already been detected in the P1 palindrome, where Alu elements were probably involved in a large-scale duplication prior to an IR (inverted repeat)-mediated inversion. Recently, a more complex model for the progressive assembly of the AZFc ampliconic organization was proposed. The model states that 3 major waves of amplicon acquisition were required for the establishment of AZFcs, starting from a basal structure constituted by the blue, turquoise and the distal part of the yellow amplicons [93]. In the first wave, the green and red amplicons were transposed; in the second the middle segment of the yellow amplicon was acquired, and in the third both the proximal yellow amplicon segment and the grey amplicon were transposed. In parallel with these acquisition waves, other molecular processes such as deletions, duplications and inversions shaped AZFc by operating on the progressively acquired blocks. A limitation to this model is the assumption that the ampliconic families of the ancestral AZFc state are identical to those of the reference sequence. This limitation is also evident when calculating the minimum-mutation history of AZFc architectures [94]. According to the most parsimonious model, the ancestral AZFc architecture was already multicopy, with the majority of the observed diversity arising from sequence inversions (and to a lesser degree from deletions and duplications) in the ancestral sequence.

In clinical terms, men with complete AZFc deletions have variable seminal and testicular phenotypes, with sperm production levels ranging from azoospermia to severe oligozoospermia (but rarely exceeding 1 million sperm/ml) [16, 47, 48]. Although in these patients the presence of sperm in semen is a frequent event (in ~50 to 60% of the cases), natural conception is extremely rare due to low sperm counts [16, 95–101]. The variable phenotype associated to these deletions suggests an intricate regulation of the AZFc genetic determinants, making this region particularly prone to a genetic background effect. Complete AZFc deletions total 3.5 Mb (mapping from ~23 to 26.8 Mb of the chromosome) and are the product of NAHR between the b2 and b4 amplicons. They account for approximately 60% of all recorded AZF deletions, occurring in one out of every ~4000 males [47, 48, 65]. The ampliconic organization of AZFc is also responsible for partial deletions that arise from NAHR between the more internal units. These partial AZFc deletions are associated to extremely variable spermatogenic disruption phenotypes (if any), leading to a debate on whether such rearrangements represent a male infertility risk or not (for selected reading: [92, 102, 103]).

5.2. AZFc Gene Content. As previously stated, 3 protein-coding gene families map to the AZFc interval: BPY2, DAZ and CDY (Figure 3). AZFc is also enriched in other transcription units, mainly for spliced but apparently non-coding transcripts of the TTTY family (TTTY3 and TTTY4). Additionally, it contains an extensive array of pseudogenes. These correspond to inactive copies of AZFb and AZFc genes (RBM1A1, PRY, CDY and BPY2), as well as AZFc-exclusive sequence families [GOLGA2LY1 (golgi autoantigen, golgin subfamily a, 2-like, Y-linked 1) and CSPG4LYP1 (chondroitin sulfate proteoglycan 4-like, Y-linked pseudogene 1)].

5.2.1. BPY2. The BPY2 gene family maps to the green AZFc amplicons (one active copy per amplicon), encoding for a testis-specific highly charged protein tentatively linked to cytoskeletal regulation in spermatogenesis [64, 104]. This gene family is further expanded by a set of pseudogene sequences also mapping to the green amplicons [65]. Despite the existence of a region of homology with chromosome 8, no autosomal homologues of BPY2 have been identified [105]. The genomic organization of the gene is quite unique since it is constituted by nine exons but only five of which are translated into amino-acids [106]. The BPY2 protein displays a nuclear localization throughout all male germ cell developmental stages, persisting even in ejaculated sperm [107]. The exact role played by BPY2 in spermatogenesis is unclear, with most of the available knowledge being inferred from its protein partners. Using the yeast two-hybrid assay BPY2 has been shown to interact with ubiquitin protein ligase E3A (UBE3A), a widely-expressed member of the ubiquitin protein degradation system [108]. This interaction is mediated by the HECT domain of UBE3A. Since UBE3A corresponds to a testis-expressed E3 ubiquitin protein ligase (responsible for the transfer of the ubiquitin group to the targeted substrates), BPY2 may modulate its target specificity. Additionally, the two-hybrid assays have also identified microtubule-associated protein 1S (MAP1S) as an interacting protein [104]. MAP1S is a member of the microtubule-associated proteins (MAPs) family and is involved in microtubule binding, bundling and stabilization, as well as in the crosslinking of microtubules with microfilaments [109]. Since MAP1S is predominantly expressed in testis, a putative role of BPY2/MAP1S in the control of the male germ cell cytoskeletal network has been proposed [104]. The functional properties of the MAP1S complex are regulated by changes to its heavy chain, making this molecule a suitable target for posttranslational regulation [109]. In this context, BPY2 emerges as a very strong candidate regulator, possibly through an UBE3A-mediated ubiquitinization event. Protein structure prediction models also suggest the existence of a DNA/RNA binding domain (a HTH-like motif) in BPY2, yet experimental validation is still lacking [110].

The screening for BPY2 mutations in infertile males has thus far been inconclusive, with no identifiable exon mutations in a cohort of 106 SCOS patients [111]. Therefore, despite suggestions that a specific promoter genotype might be associated to spermatogenic defects [111], both BPY2 function and the phenotypical consequences associated to its disruption remain to be elucidated.

5.2.2. CDY. The chromodomain protein family (CDY) consists of two Y-encoded genes (CDY1 and CDY2) and two autosomal copies (CDYL in chromosome 6 and CDYL2 in chromosome 16) [112]. These genes are involved in postmeiotic nuclear remodelling and transcriptional regulation. The Y family members map to the yellow amplicons, with
the CDY1 copies in AZFc (amplicons yel1 and yel2) and the CDY2 copies in AZFb (yel3 and yel4) [44, 65]. A fairly large number of pseudogene sequences are also scattered throughout AZFb and AZFc. As expected, the Y-linked copies have testis-specific expression whereas the autosomal units display a more general expression pattern (CDYL is even ubiquitously expressed) [64, 112]. CDY1 displays two additional transcript variants (minor and short CDY1) with the former showing evidence of the excision of a single intron [112, 113]. The expression of these alternative transcripts correlates significantly with complete spermatogenesis in testicular samples of azoospermic men [114].

The CDY proteins are characterized by two functional motifs: an N-terminal chromatin-binding domain (the chromodomain) and a C-terminal catalytic domain (responsible for the CoA-dependent acetyltransferase activity). The chromodomain is a typical signature of proteins involved in chromatin remodelling and gene expression regulation [115]. Accordingly, in vitro assays have demonstrated that recombinant CDY proteins can acetylate histone H4 (and, to a lesser degree, H2A) [116]. Furthermore, it was established that mouse Cdyl (mCdyl) transcript and protein levels peak at the elongating spermatid stage, a time frame coinciding with histone H4 hyperacetylation [116]. Given the nuclear localization of the protein and the post-meiotic expression window, the CDY family is considered a nuclear remodelling factor promoting histone H4 hyperacetylation [116]. The latter, by inducing a more relaxed chromatin configuration, may serve as trigger for the histone-to-proline transition and subsequent nuclear condensation.

The function of the CDY proteins is not restricted to histone acetylation. Studies have associated CDYL and its paralogues to transcriptional corepressor complexes consisting of multiple chromatin modifying proteins [117–119]. Accordingly, the primary function of CDYL may be that of a transcriptional co-repressor, as observed in murine models when histone deacetylases (HDACs) bind to its catalytic domain [117]. The protein acquires its role in chromatin remodelling only when HDACs are degraded (in the elongating spermatid stage) and the CoA-binding activity of mCdyl is activated. This fits with data obtained from protein structure analysis indicating that the CDY proteins do not show obvious similarities to canonical histone acetyltransferase motifs [120]. Recently, CDY1 has also been shown to interact with lysine 9-methylated histones (H3K9me2 and H3K9me3), although the exact functional role of this interaction is unknown [121]. The analysis of such binding properties further suggests that CDYL2, not CDYL, is the ancestor of the gene family [122].

The CDY1 and CDY2 proteins are isoforms with an amino-acid identity of 98% and a similar expression window [112]. This contradicts previous views that CDY2 was required at earlier spermatogenic stages [114]. On the other hand, the global identity score between the Y-linked CDY proteins and CDYL is just 63%. The accelerated protein evolution rate of the Y-borne CDY sequences seems to suggest that these copies have evolved under positive selection for germline specific functions [123]. Nevertheless, identity levels are slightly higher when comparing just the functional domains of the Y-derived and autosomal copies. In this regard, functional complementation between CDY genes may rescue, to some extent, the loss of the AZFb and/or AZFc variants. Fittingly, complete AZFc deletions do not alter H4 hyperacetylation levels in developing spermatids when compared to those recorded in nondeleted hypospermatogenic men [124].

5.2.3. DAZ. The members of the DAZ gene family are RNA binding proteins that play prominent roles in the establishment and maintenance of the male germ line (for selected reviews: [125, 126]). This gene family consists of three different genetic determinants: BOLL (bol, boule-like), DAZL and DAZ [127]. Of these, DAZ maps to AZFc (consequently being organized as a multi-copy gene family) with the remaining two being single-copy autosomal genes. Since the DAZ gene family contains the Y-borne DAZ copies, for the sake of disambiguation the italicized reference (DAZ genes) will refer solely to the Y determinants. The DAZ genes are present in one copy per red AZFc amplicon, for a total of four in the reference AZFc sequence (DAZ1 to DAZ4) [65, 128]. The particularities of the palindromic organization of the reference sequence results in the clustering of the DAZ copies to two red amplicon duplets, with the more proximal cluster containing DAZ1 and DAZ2, and the more distal cluster DAZ3 and DAZ4. Nevertheless, variations in gene number have been recorded among different Y chromosomes [129, 130]. The DAZ genes encode for four protein isoforms varying in the number of functional domains, with the most recent data suggesting that all four are expressed in human testis [131]. They are expressed in spermatogonia, with the protein displaying a cytoplasmic localization [131–134].

Historically, DAZ has been the focus of considerable attention both for its link to Yq microdeletion phenotypes and evolutionary origin. In fact, almost 10 years before the sequencing of the MSY, DAZ was considered to be the azoospermia factor [17]. This gene also corresponded to the first reported instance of an autosome to Y transposition, an observation that triggered the resurgence of the controversy on the evolutionary fate of the Y chromosome. DAZ was the result of the transposition of autosomal DAZL to Yq, with the newly acquired sequence being subjected to several bouts of intra-and intergenic amplification followed by degeneration of some of the amplified exonic units [128, 135]. In this regard, DAZ corresponds to the product of diverse evolutionary forces that have ensured that a RNA-binding protein could evolve in a male-specific genomic context. The fact that its reading frame emerged unsathed from all these intense rearrangements serves as testament of positive selection, in opposition to previous reports suggesting a lack of selective pressure [128, 136].

The DAZ family proteins are characterized by two functional domains: a N-terminal RRM and a C-terminal DAZ repeat. The number of these domains varies between the DAZ genes and may even be polymorphic in the DAZ copies [137]. The DAZ repeat consists of a unit of 24 amino-acids that is involved in protein–protein interactions [126]. While the number of DAZ repeat units varies in the DAZ
genes (8 to 24), both BOLL, and DAZL contain a single unit. Several DAZ-interacting proteins have been identified, most of them also displaying RNA-binding properties. In fact, DAZAP1 (DAZ associated protein 1), PUM2 (pumilio homolog 2), DZIP1 (DAZ interacting protein 1) and DZIP3 (DAZ interacting protein 3) are not only able to interact with DAZ family members but also have RNA binding activity (DZIP3 couples this function with that of an ubiquitin protein ligase) [138–142]. Additionally, DAZAP2 (DAZ interacting protein 2) although devoid of RNA binding properties is also a regulator of the spermatogenic program. An interesting property of the DAZ family is that the proteins can interact with other members in the form of homo- or heterodimers [143]. Therefore, multiple interaction patterns might modulate the functional status of the protein in a stage-dependent manner.

The RNA binding properties of the DAZ family are associated to the translational activation of developmentally regulated transcripts. Fittingly, studies in Drosophila have shown that bol (ortologue of BOLL) mutations diminished the protein level of a regulated gene (twine) but not of its mRNA [144]. These properties can be ascribed to the RRM, an RNP-type motif with a preference for poly(U) and poly(G) UTR sequences [145–148]. The DAZ family proteins have only one RRM, except the DAZ1 and DAZ4 copies that contain 3 and 2, respectively. Several model-based studies have tried to identify target mRNAs, yet the full range of these molecules is still open for debate [145–149]. This can be illustrated by the meagre overlap of identified candidates between the different studies. It should be noted that this repertoire of mRNA targets may be more extensive in humans since murine models lack the DAZ genes. Indeed, DAZ orthologues are absent in mammals lower than Old World monkeys. Nevertheless, the currently available list of purported mRNA targets shows some very interesting associations. The majority of the identified transcripts correspond to regulators of cell cycle progression and of general RNA metabolism. Transcripts for genes involved in spermiogenesis also seem to be targeted by murine Dazl, consistent with DAZL expression during cytodifferentiation [125]. Such data, although still awaiting more extensive validation, are indicative that the DAZ family plays an important role in the orchestration of spermatogenesis. Fittingly, overexpression of BOLL, DAZL and DAZ promotes the formation of haploid cells in human embryonic stem cell differentiation systems [150].

The molecular mechanism through which the DAZ family exerts its control over protein expression seems to involve the enhancement of translation initiation. A model has been proposed based on the binding of the DAZ proteins to the 3′UTR of target mRNAs followed by the recruitment of poly(A)-binding proteins to the transcripts [151]. These in turn enhance the recruitment of ribosomal subunits and consequently the onset of translational activation. Accordingly, a report has associated DAZL to the actively translating ribosome fraction of testis extracts, although the robustness of this observation has been questioned [125, 152]. Moreover, the DAZ family is also involved in the transport and storage of transcripts [153]. This function seems to be dependent on the dynein-dynactin complex and leads to the storage of the molecules as transcriptionally quiescent particles waiting for proper developmental cues to trigger their activation. This is particularly relevant in light of the transcriptional shutdown associated to chromatin remodelling in spermiogenesis. On a more general note, the DAZ family corresponds to an active regulator of the spermatogenic program, operating at multiple levels via mediation of transcript transport/storage, translation initiation and protein function.

Despite the significant roles in spermatogenesis attributed to the DAZ genes, their complete deletion is not incompatible either with sperm production (albeit at extremely low levels) or with rare instances of natural conception [47, 48, 95–97]. This can be explained in part by some functional overlap between DAZ family members. Indeed, DAZ and DAZL share 93% similarity in the RRM region and 80–90% in the DAZ repeat domain [125]. In this context, the loss of the Y-borne DAZ copies may be compensated by DAZL. This functional overlap is illustrated by the fact that a human DAZ transgene can partially rescue the spermatogenic impairment phenotype of Dazl-null mice [154, 155]. Nevertheless, despite an increase in the germ cell population and meiotic progression up to the pachyten stage, the rescue phenotype is insufficient to ensure mature sperm production. It should be noted that not even a DAZL transgene was able to extend the rescue effect to post-meiotic stages [155], a clear indicator that interspecies differences played a decisive role in determining the degree of rescue.

In summary, AZFc genes play important roles for male fertility. The majority of published studies focus on the functional properties of CDY and DAZ. Such reports indicate that both genes may be the main functional determinants of the interval. Yet, the effects of their deletion can presumably be minimized by some degree of phenotypical rescue ensured by the autosomal homologues. This might account for the less severe spermatogenic impairment phenotype associated to AZFc deletions when compared to those removing AZFa or AZFb. Evidence for functionally related copies being able to partially rescue deletion phenotypes is available in other genomic contexts. Specifically, the dynamics of the SMN1 and SMN2 genes (survival of motor neuron 1 and 2, respectively) in the context of spinal muscular atrophy is a prime example [156]. Nevertheless, the still largely uncharacterized properties of BPY2 advise some caution on relegating this gene to a secondary role in spermatogenesis. An additional consideration regards the possibility of the different copies inside each AZFc gene family varying slightly in function. In this scenario, each copy would contribute differently to the overall function of the gene family, with the effects of some determinants being more decisive than others. To test this hypothesis, several authors have suggested the study of evolutionary branches of the Y chromosome where partial AZFc deletions have become fixed without any apparent consequences for male fertility [37, 157]. Although such authors argue that the copies remaining in these chromosomes represent the key determinants of the AZFc gene families, advances have thus far been inconclusive. Regardless of such considerations, the
variable levels of spermatogenic impairment observed in AZFc deletions are a clear indicator that this region is prone to a more pronounced phenotypical modulation than the other two AZF intervals. In this regard, factors such as genetic background and other epigenetic/environmental cues may play a crucial role in defining the deletion’s outcome.

Since the complete AZFc deletion removes a still significant Yq stretch, a deleterious effect on cell cycle progression arising from pairing deficiencies may be considered. Fittingly, an association between AZFc deletions and minor impairment of telomere clustering has been reported [91]. AZFc deletions have equally been linked to instances of prolonged zygote stage and reduced XY condensation [158]. However, since the level of these disturbances is small, their effective impact on spermatogenesis is highly speculative.

6. Final Considerations and Future Perspectives

Despite the tremendous breakthroughs recorded in the past few years, our knowledge of AZF gene function is still considerably limited. The reasons for this can be ascribed to both technical issues and to the inherent complexity of this biological system. In technical terms, the lack of easily accessible animal models (AZF sequence architectures are only present in some primate lineages) and of in vitro spermatogenic cell lines introduce clear restrictions to a faster development of the field. Additionally, the biological properties of the AZF regions further complicate matters, as attested by the tremendous variability associated to the AZFb and AZFc sequences, as well as the intricate regulation of the corresponding genetic determinants. Clearly, the future research lines to be pursued in this field consist of the full sequencing of AZF diversity across the Y chromosome population and of a more in-depth functional characterization of AZF genes. Both represent considerable challenges that will ultimately yield benefits for a significant fraction of infertile patients.

Based on the state-of-the-art discussed in this paper, the AZF genes display considerable differences in their genomic organization and molecular role. The available evidence indicates that DDX3Y in AZFa, KDM5D and RBMY1A1 in AZFb, and DAZ and CDY in AZFb/c represent key determinants for spermatogenesis. Yet, the characterization of the remaining AZF genes is still quite incipient. Thus, advances in this area are of paramount importance for a more comprehensive outlook on the reproductive fitness of the Y chromosome. Appropriately, the screening of infertile males for specific deletions or other (epi)genetic alterations in AZF may reveal new clinically-relevant mutations. Coupling such knowledge with functional data on the affected biological processes would translate into significant conceptual advances in male reproductive genetics.

Abbreviation List

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZF</td>
<td>Azoospermia factor</td>
</tr>
<tr>
<td>HDAC</td>
<td>Histone deacetylase</td>
</tr>
<tr>
<td>HERV</td>
<td>Human endogenous retrovirus</td>
</tr>
<tr>
<td>IR</td>
<td>Inverted repeat</td>
</tr>
<tr>
<td>MAP</td>
<td>Microtubule-associated protein</td>
</tr>
<tr>
<td>MSY</td>
<td>Male specific region of the Y chromosome</td>
</tr>
<tr>
<td>NAHR</td>
<td>Non-allelic homologous recombination</td>
</tr>
<tr>
<td>PGC</td>
<td>Primordial germ cell</td>
</tr>
<tr>
<td>RRM</td>
<td>RNA recognition motif</td>
</tr>
<tr>
<td>SCOS</td>
<td>Sertoli cell-only syndrome</td>
</tr>
<tr>
<td>STS</td>
<td>Sequence-tagged site</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>X</td>
<td>X chromosome</td>
</tr>
<tr>
<td>Y</td>
<td>Y chromosome</td>
</tr>
<tr>
<td>Yq</td>
<td>Long arm of the Y chromosome</td>
</tr>
</tbody>
</table>

Acknowledgments

This work was partially funded by CIGMH (Centro de Investigação em Genética Molecular Humana). P. Navarro-Costa was supported by a Ph.D. fellowship from Fundação para a Ciência e a Tecnologia (no. SFRH/BD/16662/2004). All authors declare that they have no competing interests.

References


40. H. J. Ditton, J. Zimmer, C. Kamp, E. Rajpert-De Meyts, and P. H. Vogt, “The AZFa gene DBY (DDX3Y) is widely
transcribed but the protein is limited to the male germ cells by translation control," Human Molecular Genetics, vol. 13, no. 19, pp. 2333–2341, 2004.


[55] A. Roll-Mecak, M. Pratt, P. T. Mehman, and D. Goldman, "PTPN13/PTPL1 induces apoptosis through insulin receptor


S. E. Kleiman, L. Yogeov, R. Hauser et al., "Members of the CDY family have different expression patterns: CDY1 transcripts have the best correlation with complete spermatogenesis," *Human Genetics*, vol. 113, no. 6, pp. 486–492, 2003.


