The use of genomics, proteomics, and metabolomics in identifying biomarkers of male infertility

Jason R. Kovac, M.D., Ph.D., Alexander W. Pastuszak, M.D., Ph.D., and Dolores J. Lamb, Ph.D.

Scott Department of Urology, Department of Molecular and Cellular Biology, The Center for Reproductive Medicine, Baylor College of Medicine, Houston, Texas

Although male factors account for approximately 50% of all infertility, the mechanisms underlying their origin are unknown. Currently, clinicians rely primarily on semen analyses to predict male reproductive potential and chart treatment success. Even when invasive procedures are performed, the causes of male factor infertility frequently remain elusive. Recently, the advent of new technologies has spurred the search for novel male infertility biomarkers, and the detection of genes, proteins, or metabolites unique to the infertile male holds much promise. The concept that a cost-effective, noninvasive, and accurate set of biomarkers can be identified to diagnose male factor infertility is tantalizing. This review focuses on the various methodologies used in the discovery of novel biomarkers along with their findings. Specific attention is paid to recent advances in the fields of genetics, proteomics, and metabolomics.

Key Words: Male infertility, proteomics, genomics, biomarker

Discuss: You can discuss this article with its authors and with other ASRM members at http://fertstertforum.com/kovacjr-genomics-proteomics-metabolomics-biomarkers/

Infertility affects ~15% of all couples with a male factor involved in ~50% (1–3). The complex nature of the condition is underscored by the fact that each individual contributes a complex array of genetic, proteomic, and metabolic differences that interact in unpredictable ways. The frustration that couples face is compounded by difficulties that clinicians have in diagnosing and treating infertility, particularly in men. Many causes of male factor infertility are still defined as idiopathic (1, 3) and as such, most diagnoses tend to be descriptive rather than specific. For example, the simple classification of azoospermia into obstructive or nonobstructive (NOA) simply delineates a physical blockage from testicular failure. However, a more accurate taxonomy is not possible when the very origins of the condition are unknown.

Infertility is defined as the inability to conceive after 12 months of regular, unprotected intercourse (4). Clinical investigations should not begin until this amount of time has elapsed—with several exceptions including advanced maternal age. Once inquiry begins, medical history and physical examination are performed and specific male factors addressed by semen analysis, karyotyping, Y-microdeletion analysis, and endocrine profiling. Unfortunately, for many men, these tests are normal or inconclusive, leading to a diagnosis of idiopathic infertility. A definitive diagnosis must then be pursued with surgical intervention, in the form of a testicular biopsy, which carries with it inherent complications. As a result, the identification of alternative, noninvasive methods to diagnose specific etiologies of male factor infertility are essential.

The use of novel genetic, proteomic, and metabolomic techniques may hold the key to more accurately diagnosing and treating male factor infertility. It is within these categories that the search for potential biomarkers can begin (5). A biomarker is a distinctive biological or biologically derived indicator of a process, event, or condition that can be objectively measured, evaluated, and compared (6). The ideal biomarker should identify disease at an early stage, be easily detectable, cost-effective, and accurate,
as well as having minimal side effects. The discovery of such noninvasive, highly sensitive and specific biomarkers would be helpful in eliminating the need for invasive testing in the infertile man and allowing an expanded and more specific classification of male factor infertility (7, 8).

For the purposes of this review, the field of male fertility-related biomarkers is classified into several areas. Basic laboratory analyses are discussed, including semen analysis, endocrine laboratory investigations, and antisperm antibodies, which are currently being used to assist clinicians in categorizing, diagnosing, and treating male factor infertility. This is followed by a summary of the current knowledge underlying state-of-the-art genomics, proteomics, and metabolomics. The difficulty involved with bringing these biomarkers from the bench to the bedside is also discussed. Although these areas are evolving and currently investigational, they have the potential to reinvent how male factor infertility is diagnosed and treated.

BASIC BIOMARKERS

A biomarker is any biological molecule that is measured and evaluated objectively and functions as an indicator of a physiological pathway (6). In the context of male factor infertility, biomarkers have the intent of evaluating, in an accurate and minimally invasive manner, a man’s potential for fathering children. At present the most widely used biomarker to predict male fertility potential is the semen analysis.

Semen is a complex fluid comprised of a cellular element (i.e., spermatozoa) as well as a plasma element that functions as a nourishing and protective medium for spermatozoa (9). Semen is primarily derived from the seminal vesicles (65%–75%) and prostate (25%–30%), and its multifocal origin makes the search for biomarkers more difficult (9). Although semen analysis yields basic, yet critical information, it is subject to significant modulation and is an overall poor predictor of male fertility (10). Several physiological factors (i.e., duration between ejaculates) as well as pathologies (i.e., diabetes, sarcoidosis [11]), systemic illnesses (i.e., flu), and environmental factors (i.e., smoking, alcohol) result in significant variability among semen samples (10). This leads to difficulties in the interpretation and management of men with abnormal findings (10).

To provide further utility to semen analysis, the components are broken down into macroscopic and microscopic factors (12). Coagulation, color, viscosity, pH, and volume are classified as macroscopic components, whereas agglutination, sperm counts and concentration, motility, morphology, and viability are microscopic (12). Sperm counts and concentration are primarily useful in the classification of male factor infertility into either azoospermic (absence of sperm) or oligospermic (<15 million sperm/mL) categories (12). Morphology is also an important component of the semen analysis; however, its utility in determining male fertility potential is unclear.

Although semen analysis is useful as a biomarker (13, 14), it is altered by a variety of environmental factors (15) including air pollution (16) and ozone levels (17). Organochlorines, such as polychlorinated biphenyls, used in motor oils, and dichlorodiphenyltrichloroethane, used in pesticides, have long half-lives that affect sperm counts and motility for years after exposure (18, 19). Phthalates present in industrial chemicals have also shown an inverse correlation with sperm motility, morphology, and concentration (20–22). Further complicating the interpretation of a semen analysis are lifestyle factors, such as cigarette smoking, alcohol consumption, psychological stress, and caffeine, all of which negatively affect semen parameters (14). As such, although semen analysis is at present the main biomarker used for male factor infertility, it is also the most unreliable (23, 24).

Antisperm antibodies were one of the first biomarkers used in widespread clinical practice. These antibodies develop after accidental or iatrogenic breach of the blood–testis barrier (i.e., due to torsion or trauma), accurately predict obstructive azoospermia (25) and affect fertility by numerous mechanisms, including altered sperm agglutination and cervical mucus penetration (26). No association has been found between antisperm antibodies and semen quality (27), although infectious diseases, like the human papillomavirus, may alter both (28). Unfortunately, antisperm antibodies do not correlate with spontaneous, IVF, or intracytoplasmic sperm injection (ICSI) pregnancy rates (PRs) (29, 30), casting doubts on their importance as a biomarker.

The acrosome reaction has also been explored in the search for a biomarker of male factor infertility. Contained within the distal end of the sperm head, the acrosome contains the enzymes required for oocyte penetration (12). The optimized sperm penetration assay (SPA; zona–free hamster oocyte penetration assay) (31, 32) collectively assesses the acrosome reaction, sperm capacitation, and the ability of sperm to penetrate ova and undergo the initial steps of sperm head decondensation and fertilization. The SPA has a high positive predictive value. If positive, there is a high likelihood the sperm will fertilize an ova in IVF (assuming the sperm can penetrate the human zona pellucida [ZP]) (31–33). For many years, the SPA was used to predict the fertilizing potential of sperm during IVF, but with the advent of ICSI, this was no longer a concern as the steps of acrosome reaction, capacitation, sperm zona and ova membrane binding, and penetration were bypassed (5, 34, 35). Nevertheless, the use of ICSI–SPA can accurately predict those patients whose sperm will fail to undergo decondensation in the ovum after ICSI, resulting in fertilization failure (34, 35).

GENOMIC BIOMARKERS

Because technology has advanced, the approach to the treatment and diagnosis of male factor infertility has rapidly evolved. Moving beyond semen analysis to genetic testing is a part of this evolution. Recent estimates indicate that genetic abnormalities cause 15%–30% of male factor infertility (36) and it is likely that this number will further increase as more genetic causes of infertility are identified. Although genetic abnormalities were previously detectable only in the form of large structural chromosomal aberrations defined with karyotype analysis, much smaller genomic regions have been found to be responsible for infertility (37–40). In the near future it is likely that single nucleotide changes...
leading to infertility will be identified, resulting in more male infertility biomarkers (37).

At present one of the most common genetic tests used to evaluate the status of the infertile male is a karyotype analysis. Karyotyping was first used in 1959 to determine the genetic composition of Turner syndrome (a single X and an absent Y chromosome; X0) as well Klinefelter syndrome (two X and one Y chromosome; XY) (41, 42). With the use of light microscopy to evaluate the appearance of chromosomes, karyotyping resolves variations in the DNA complement of ≥ 4 Mb.

With the use of the karyotype, initial reports suggested that ~2% of infertile men had chromosomal abnormalities, a rate five times higher than the normal population (43). Karyotypic anomalies vary with the severity of oligospermia or azoospermia and a higher incidence of aberrations have been identified in azoospermic (18.71%) versus severely (14.55%) or moderately (2.37%) oligozoospermic men (44). With regard to specific karyotypic defects, Robertsonian translocation occurs in 1.5% of oligozoospermic and 0.2% of azoospermic men (45, 46), whereas reciprocal translocations occur in 0.7% of oligospermic or azoospermic men (47). Other abnormalities that may occur in the infertile man include inversions, which are typically benign, and a 46,XX genotype, which does not produce sperm (48).

The concept that the Y chromosome determines maleness was initially put forth in 1976 by Tiepolo and Zuffardi (49). The Y chromosome was cloned in 1992 (50) and the first genetic marker of infertility, the azoospermia factor (AZF), was identified within its confines shortly thereafter (39, 40). These initial studies revealed that microdeletions in the Y chromosome (AZFa, AZFb, and AZFc) could act as biomarkers for male factor infertility. The presence of sperm was dependent on which specific AZF subregion was altered. For example, sperm were unlikely to be found in men with AZFa or AZFb deletions, whereas men with AZFc deletions had a 75% chance of having sperm identified on testicular biopsy (51).

Since the identification of the Y chromosome’s role in fertility, numerous other genes involved in male factor infertility have been identified (5, 52). Most are located on autosomes and the cystic fibrosis transmembrane regulator gene has become important in the evaluation of men with infertility, in general, and the congenital absence of the vas deferens, in particular (53). Although the full genetic complexity of this biomarker has yet to be elucidated, there are currently >1,000 known cystic fibrosis transmembrane regulator mutations that have yet to be explored (54).

Higher resolution assessment of a sperm's genetic complement can be achieved using fluorescence in situ hybridization, a technique that uses fluorescently labeled DNA to identify specific genetic sequences (55). Although not a first-line investigation in the United States, fluorescence in situ hybridization of sperm is used in cases of severe teratospermia or oligoasthenoteratospermia. In addition, in situations of recurrent pregnancy loss, it defines meiotic defects in the form of aneuploid sperm (56). Its primary disadvantage as an investigative tool is that the sequence of interest must be known before determining whether a region of interest exists in a specific patient.

Microarray technologies, which evaluate men for copy number variations, gene expression levels, and single nucleotide polymorphisms (SNPs) hold great promise for identifying highly sensitive and specific genetic biomarkers. Comparative genomic hybridization, a technique used to assess the relative quantities of DNA between samples, permits determination of gene copy number as a function of chromosomal location and can be applied to the entire genome using a microarray-based approach (array comparative genomic hybridization). This approach was used in children with mental retardation, demonstrating that 10%–20% have chromosomal abnormalities—a stark difference to the 3%–5% detected using standard karyotype analysis (57). In the setting of male factor infertility, array comparative genomic hybridization has identified Y-chromosome microdeletions as well as additional copy number variations outside of the known AZF regions including copy number variations in the pseudoautosomal regions of the Y chromosome (specifically in SHort stature Homeobox [SHOX] gene in pseudoautosomal region1) (58, 59). Additional candidate infertility genes have been identified using array comparative genomic hybridization, although their roles remain to be elucidated (52, 60).

Microarray-based SNP analysis has also yielded numerous male fertility gene candidates and potential biomarkers. A recent genomewide association study of family size and birth rates among 269 men evaluating ~250,000 SNPs revealed 41 SNPs with strong association to infertility and 9 SNPs related to reduced sperm quality and function (61). The study implicated the genes USP8, UBD, EPSTI1, and LRRC32 in male factor infertility, highlighting potential biomarkers (61). Other genes with SNPs associated in male fertility include: PDE3A, EFCA4B, COBL, ATP8A1, MASP1, PROK2 (62), AHRR (46, 47), MTHFR (58–60), andUBE2B, a homologue of the DNA repair gene rad6 in mice, which results in male infertility when knocked out (61). The genes associated with these changes are listed in Table 1.

Variations in the sperm epigenome may also contribute to male fertility by DNA protamination and methylation. Improper post-translational processing of protamine 2 (PRM2), one of the two proteins that replace histones in sperm, along with altered PRM expression, resulted in decreased sperm counts and function (63–66). In addition, lower levels of PRM1 and PRM2 messenger RNA were observed in asthenozoospermic men (67). In men with altered protamination, abnormal methylation of the CREM gene has also been observed (68). Methylation of genomic DNA, which regulates the accessibility of the DNA to histone binding, has also been shown to affect fertility. DNA methylation studies in oligozoospermic men demonstrated altered methylation of imprinted genes, in general (69–71), and IGF2/H19 and MEST, in particular (72, 73).

Sperm have been traditionally thought of as DNA storage and transport vehicles. However, the RNA complement of sperm, including both coding and noncoding RNAs, may also play a role in male factor infertility (74, 75). Full-length messenger RNAs and noncoding RNAs have been recently identified in sperm cytoplasm and play functional roles in fertilization and early zygote development (74, 75).
TSPY1, CLU, PRM2, PSG1, HLA-E, and PLCD1 transcripts were identified and theorized to be potential biomarkers (74, 76), as highlighted in Table 1. Microarray technologies have enabled rapid evaluation of spermatozoal RNAs and studies using these techniques have identified messenger RNA patterns that correlate with spermatogenesis, sperm motility, germ cell antiapoptotic processes, DNA repair, oxidative stress reduction, and histone modification (77–79).

Small noncoding RNAs produced during the process of spermatogenesis and localized in spermatozoa are theorized to contribute to early embryonic function; however, their exact nature and classification remain controversial (75, 80, 81). Approximately 24,000 small noncoding RNAs are believed to exist, with the best described being within the microRNA family that was identified in 2005 (82). A recent study by Krawetz et al. (75) found multiple classes of small noncoding RNAs to be present in human spermatozoa. By examining small RNAs (<200 bases) isolated from three fertile donors, bioinformatics revealed human spermatozoa to contain microRNAs (~11%), piwi-interacting piRNAs (~17%), and repeat-associated small RNAs (~65%) (75). A small subset, which framed the histone promoter-associated regions, comprised ~11% and was termed quiescent RNAs (~11%) (75). When mouse oocyte and one-cell zygotes were compared, several paternally localized microRNAs were identified including hsa-mir-34c, hsa-mir-375, hsa-mir-252, and hsa-mir-25 (75). The investigators proposed that hsa-mir-34c was of particular interest owing to its place in a group of highly conserved microRNAs recently described to play a critical role during male spermatogenesis (75, 83).

Interestingly, examination of the messenger RNA profiles of sperm after ICSI has found significant differences in the messenger RNA profiles from sperm that produced a viable pregnancy compared to those that did not (78). Specifically, sperm that resulted in pregnancies had increased levels of 44 transcripts, yielding a panel of potential biomarkers including several cathepsins and metallothioneins, ADD1, ACVRL1, AR, and ARNT (Table 1) (78). It is important to note, however, that as not all RNA is translated, protein expression and its complement (see Proteomic biomarkers section) also need to be examined.

Of the various genetic markers described detail, only a fraction of the genes, copy number variations, SNPs, and RNAs that may become important as biomarkers in the future.

### Table 1

<table>
<thead>
<tr>
<th>Gene</th>
<th>Name</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZF</td>
<td>Azoosperma factor</td>
<td>Yq11</td>
</tr>
<tr>
<td>CFTR</td>
<td>Cystic fibrosis transmembrane regulator</td>
<td>7q31.2</td>
</tr>
<tr>
<td>SHOX</td>
<td>Short stature homeobox</td>
<td>Xp22.33, Yp11.3</td>
</tr>
<tr>
<td>USP8</td>
<td>Ubiquitin-specific peptidase 8</td>
<td>15q21.2</td>
</tr>
<tr>
<td>UBD</td>
<td>Ubiquitin D</td>
<td>6p21.3</td>
</tr>
<tr>
<td>EPST1</td>
<td>Epithelial-Stromal interaction 1</td>
<td>13q13.3</td>
</tr>
<tr>
<td>LRC32</td>
<td>Leucine-rich repeat containing 32</td>
<td>11q13.5–q14</td>
</tr>
<tr>
<td>PDE3A</td>
<td>Phosphodiesterase 3A</td>
<td>12p12</td>
</tr>
<tr>
<td>EFCAB4B</td>
<td>EF-hand calcium-binding domain 4B</td>
<td>12p13.32</td>
</tr>
<tr>
<td>COBL</td>
<td>Cordon-bleu WH2 repeat protein</td>
<td>7p12.1</td>
</tr>
<tr>
<td>ATP8A1</td>
<td>ATPase, aminophospholipid transporter (APLT), class I, type 8A, member 1</td>
<td>4p13</td>
</tr>
<tr>
<td>MASP1</td>
<td>Mannan-binding lectin serum peptidase 1</td>
<td>3q27–28</td>
</tr>
<tr>
<td>PROK2</td>
<td>Prokineticin 2</td>
<td>3p13</td>
</tr>
<tr>
<td>AHRR</td>
<td>Aryl-hydrocarbon receptor repressor</td>
<td>5p15.3</td>
</tr>
<tr>
<td>MTHFR</td>
<td>Methylene-tetrahydrofolate reductase</td>
<td>1p36.3</td>
</tr>
<tr>
<td>UBE2B</td>
<td>Ubiquitin-conjugating enzyme E2B</td>
<td>5q31.1</td>
</tr>
<tr>
<td>CREM</td>
<td>cAMP responsive element modulator</td>
<td>10p11.2.1</td>
</tr>
<tr>
<td>TSPY1</td>
<td>Testis-specific protein, Y-linked 1</td>
<td>Yp11.2</td>
</tr>
<tr>
<td>CLU</td>
<td>Clustatin</td>
<td>8p21–p12</td>
</tr>
<tr>
<td>PRM2</td>
<td>Protamine 2</td>
<td>16p13.2</td>
</tr>
<tr>
<td>PSG1</td>
<td>Pregnancy-specific beta-1-glycoprotein 1</td>
<td>19q13.3</td>
</tr>
<tr>
<td>HLA-E</td>
<td>Major histocompatibility complex, class I, E</td>
<td>6p21.3</td>
</tr>
<tr>
<td>PLCD1</td>
<td>Phospholipase C, delta 1</td>
<td>3p22–p21.3</td>
</tr>
<tr>
<td>ADD1</td>
<td>Adducin 1 (alpha)</td>
<td>4p16.3</td>
</tr>
<tr>
<td>ACVR1L</td>
<td>Activin a receptor type II-like 1</td>
<td>12q13.13</td>
</tr>
<tr>
<td>AR</td>
<td>Androgen receptor</td>
<td>Xq12</td>
</tr>
<tr>
<td>ARNT</td>
<td>Aryl hydrocarbon receptor nuclear translocator</td>
<td>1q21</td>
</tr>
<tr>
<td>hCAP18</td>
<td>cAMP cathelicidin antimicrobial peptide</td>
<td>3p21.3</td>
</tr>
<tr>
<td>SPIN1</td>
<td>Spindlin 1</td>
<td>9q22.1</td>
</tr>
<tr>
<td>TEX101</td>
<td>Testis expressed 101</td>
<td>19q13.31</td>
</tr>
<tr>
<td>PGK2</td>
<td>Phosphoglycerate kinase 2</td>
<td>6p12.3</td>
</tr>
<tr>
<td>HIST1H2BA</td>
<td>Histone cluster 1, H2ba</td>
<td>6p22.2</td>
</tr>
<tr>
<td>SLC2A14</td>
<td>Solute carrier family 2 (facilitated glucose transporter), member 14</td>
<td>12p13.31</td>
</tr>
<tr>
<td>SPAC4A3</td>
<td>Sperm acrosome associated 3</td>
<td>17q11.2</td>
</tr>
<tr>
<td>GAPDH5</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase, spermatogenic</td>
<td>19q13.12</td>
</tr>
<tr>
<td>AKAP4</td>
<td>A kinase (PKA) anchor protein 4</td>
<td>19p11.2</td>
</tr>
<tr>
<td>SPAG11B</td>
<td>Sperm-associated antigen 11B</td>
<td>8p23.1</td>
</tr>
<tr>
<td>SAM32/SPACA1</td>
<td>Sperm acrosome associated 1</td>
<td>6q15</td>
</tr>
</tbody>
</table>

Thus, although the studies summarized provide insight into potential genetic biomarkers, many require validation and further experimentation.

**PROTEOMIC BIOMARKERS**

Proteomics joins the words “protein” and “genome” and is the study of both the structure and function of proteins with the goal of examining the total expressed protein complement of the genome, including modifications (5, 84, 85). This concept is critical, as not all RNA is translated and gene expression can differ from protein expression, particularly under different physiological states (i.e., obstructive azoospermia versus NOA) (86–88). Proteomics is arguably more complex when compared to genomics, given that although a genome is relatively constant between cell types and organisms, the proteome differs from cell to cell as well as temporally.

Because distinct genes are expressed in a cell-dependent manner, and messenger RNA is not always translated to protein, some suggest that proteomic analyses have the potential to yield highly accurate biomarkers (9, 13, 86, 88). At present studies focusing on proteomics in infertile men are in their infancy (5), but hold much promise. Analysis of semen has already produced several novel potential biomarkers for prostate cancer (89, 90). With respect to male factor infertility, recent work by Batruch et al. (91) compared the proteomes of pooled seminal plasma in men after vasectomy versus men (control) without vasectomy. Because after vasectomy men lack testicular and epididymal effluent, the 32 unique proteins identified in control men were theorized to originate from these regions, serving as candidate biomarkers (91).

Clinicians currently rely on semen analysis for the diagnosis of male factor infertility, although semen analysis alone is usually insufficient for diagnosis. Proteomic analysis can confirm the presence of a given protein, measure its quantity, and detect activated, phosphorylated, ubiquitinated, methylated, acetylated, glycosylated, oxidated, and/or nitrosylated protein forms (9, 13, 86, 88). Given these variables, it is easy to see how proteomic technology can be a powerful tool in identifying potential biomarkers for diagnosis and treatment (88). Using proteomics to measure changes in global protein levels, while monitoring specific protein interactions, heralds a new approach to the diagnosis of male factor infertility (84, 85). The challenge in identifying ideal biomarkers will be to distinguish relevant proteins from their chemically modified forms within the setting of such a complex milieu.

The use of semen for proteomic analysis is complicated because it contains sperm as well as seminal plasma. The seminal plasma is composed of products from the prostate, seminal vesicles, and bulbourethral glands (9) and offers sperm a nutritional environment that assists in capacitation, sperm–zona pellucida (ZP) interaction, and sperm–egg fusion (88, 92, 93). Unfortunately, proteomic analysis of semen is complex given the variations between and within individuals, as well as seasonal and age-related fluctuations (94, 95). Sperm also undergo distinct physiological changes after ejaculation that further complicate proteomic analysis.

For example, different proteins are activated at distinct phases of coagulation and liquefaction during fertilization (9). After ejaculation, the semen coagulates and approximately 30 minutes later, liquefies with the release of proteolytic enzymes (9). Several factors regulate this process including semenogelins 1 and 2 (9, 96), which undergo proteolytic digestion by prostate-specific antigen (PSA), causing liquefaction of the semen coagulum and release of motile spermatozoa (9). All of these complicating aspects serve to make the identification of a common set of proteins a challenge.

At present most proteomic investigations targeted the seminal plasma, which comprises ~90% of the semen and is the supernatant remaining after centrifugation of semen (9). The search for biomarkers in seminal plasma began in the 1940s (97, 98) with advances being limited by the technology available. In the 1980s, one of the first proteins found to be secreted by human Sertoli cells, transferrin, was also the first protein with binding sites identified on germ cells (99). The discovery that transferrin concentrations within the semen of NOA and oligospermic patients were significantly lower than in controls pointed to a critical role in male fertility (99–101). Although small gains were made in the intervening decades (102, 103), one of the first modern characterizations of pooled seminal plasma from healthy individuals was made in 2004 (104). In that study (104), 42 distinct proteins were identified including PSA, prostatic acid phosphatase, and semenogelin. This list of potential biomarkers was expanded in 2006 by Pilch and Mann (90) who identified 923 unique seminal plasma proteins in a single individual. The most abundant proteins identified included fibronectin, lactoferrin, laminin, albumin, as well as semenogelin 1 and 2 (90). The most prevalent molecular function of the identified proteins (~60%) was in the realm of catalytic enzymatic activity (90).

Another subset of proteins, which have been theorized to act as male infertility biomarkers, is the heparin-binding proteins (HBPs) (105). These glycosaminoglycans are potent enhancers of sperm capacitation in other animals (106). Within the seminal plasma, HBPs attach to the sperm surface, particularly to phosphoryl-choline-containing lipid groups, facilitating progress and capacitation in the female reproductive tract (105). Initial work reported purification of seven HBPs from human seminal plasma with the significant HBPs being lactoferrin, semenogelin 1 and 2, PSA, bovine seminal plasma proteins, zinc finger protein, and fibronectin (107). A more recent study using two-dimensional gel electrophoresis and mass spectrometry identified 40 proteins, including those previously mentioned (105, 107). Based on functional clustering, these proteins were suggested to be mostly involved with protein metabolism (38%), RNA processing/transcription (20%), cell transport/structure (18%), and signal transduction (16%) (105). Further work remains to better elucidate the value of HBPs as biomarkers of male infertility.

Prolactin inducible protein (PIP) is another component of seminal plasma that has been theorized to act as a biomarker for male factor infertility (108, 109). The PIP was found to be one of eight proteins with increased levels of expression in
azoospermic men (108). The others were fibronectin, prostatic acid phosphatase, proteasome subunit alpha type-3, beta-2-microglobulin, galectin-3-binding protein, and cytosolic nonspecific dipeptidase (108). Prolactin inducible protein is a 17-kDa glycoprotein known to be a marker of breast carcinoma (110), which has recently been found to interact and form complexes with human serum albumin (HSA) (109). The HSA is an abundant seminal plasma protein found to have a role in capacitation, also functioning to improve sperm motility. As such, it is tempting to speculate that PIP, HAS, and/or the PIP-HSA complex could serve as biomarkers. Although previous studies have failed to show a correlation between PIP and fertility status (111), the ability of these proteins to affect immunoglobulins and immunoreactivity makes them tempting targets (108, 109, 111, 112).

More recent work by Milardi et al. (88) examined the seminal proteins of five fertile men who had provided semen samples within 3 months, which resulted in conception. The investigators found between 919 to 1,487 unique proteins within each individual with 83 present in all five samples (88). Two proteins, human cationic antimicrobial protein (hCAP18) and spindlin1, were previously implicated in human reproduction (88). Present in the human epididymis and seminal plasma, hCAP18 has a key role in innate immunity and is closely associated with sperm (113, 114). It is thus tempting to speculate that hCAP18 may play a role in protecting against infection during fertilization (88, 113, 114). Spindlin1 is also involved with spermatogenesis and localizes to the tails of murine sperm, suggesting a role in sperm motility (88, 115). As such, both hCAP18 and spindlin1 are potential proteome biomarkers.

To further classify proteins contained in semen, Batruch et al. (91) evaluated the seminal plasma of controls and men after undergoing vasectomies. Those proteins most likely to be derived from the testes/epididymides were TEX101, after undergoing vasectomies. Those proteins most likely to be derived from the testes/epididymides were TEX101, PGK2, HIST1H2BA, SLC2A14, SPACA3, GAPDH, and AKAP4 (91). As such, these proteins are leading candidates for male infertility biomarkers (Table 1). Further work from the same group of investigators compared the seminal plasma proteomes of controls to men with NOA and identified two proteins, SPAG11B and TEX101, as key male fertility biomarkers (86). Although both protein concentrations were elevated in healthy controls and only SPAG11B was increased in men with NOA (86), this sets the course for further investigations into this protein as a biomarker.

Although the seminal plasma has been examined in the search for protein-based fertility biomarkers, the proteomes of spermatozoa have also been evaluated toward the goal of identifying contraceptive targets for drug development (116–118). These experiments are typically challenging given the limited amount of specimen involved (87). Recent work in the field of biomarkers has focused on the failure of sperm–egg recognition during fertilization. Redgrove et al. (119) identified a key role for heat shock protein 2 in this process. Heat shock protein 2 was found to have significantly reduced expression when the proteomic profiles of sperm from fertile men were compared with those of men whose sperm exhibited impaired fertilization capacity (119). Heat shock protein 2, present in the sperm acrosome, was also noted to exist in close interaction with sperm adhesion molecule 1 and arylsulfatase A. As such, the investigators (119) proposed that sperm adhesion molecule 1, arylsulfatase A, and heat shock protein 2 interacted together in a multimeric complex to mediate sperm–egg fusion. Any of these molecules could act as potential biomarkers of male fertility.

In addition recent reviews have highlighted four to six sperm surface antigens that may act as biomarkers of male fertility (116, 117). In addition, a study by Hao et al. (118) on sperm extracts and surface proteins identified several protein products with common peptide sequences. Interestingly, these investigators found that these proteins were the product of a single gene—SAMP32. In addition, this gene was found to have a testis-specific expression pattern and acrosomal localization, making it a very enticing potential biomarker for male factor infertility (118). In an excellent example of the use of proteomics in the identification of fertility-related proteins, SAMP32 (currently named SPACA1 or sperm acrosome membrane-associated protein), was recently found to code for a sperm membrane protein that functions in sperm–egg fusion (120) (Table 1). Fujihara and colleagues (120) furthered this work with the creation of a SPACA1 murine knockout in which males have both abnormally shaped sperm heads (similar to human globozoospermia) as well as absent acrosomes. The pathway of discovery for the SPACA1 biomarker is prototypical for how proteomics can lead the way to novel innovations.

**METABOLOMIC BIOMARKERS**

The study of cellular metabolic products (metabolomics) as potential male fertility biomarkers is currently in its infancy and is the least advanced of the three topics covered in this review. Metabolomics is the study of small, low molecular weight molecular metabolites that are the products of metabolism (121, 122). The physiological functions of these ~3,000 molecules span a diverse range and include growth, development, and reproduction (121). Metabolites can be either intrinsic, resulting from normal cellular physiology, or extrinsic, resulting from the influences of exogenously administered pharmaceuticals (121). Metabolomics reflects events downstream of gene expression and is considered to be closer to the actual phenotype than either proteomics or genomics (123–125).

In the search for male fertility biomarkers, metabolomics was initially focused on the changes seen in cases of oxidative stress. Due to excessive production and formation of reactive oxygen species (ROS) along with impaired antioxidant defence mechanism, oxidative stress results in spermatic abnormalities (122, 126). Previous studies have shown that oxidative stress markers (–CH, –NH, –OH, –SH) affect sperm and oocyte quality as well as embryo viability (123, 127). The ROS are elevated in a significant portion of semen samples from infertile men (122). Although few studies examining the role of ROS as biomarkers exist, their importance is well recognized. One recent study (128) identified the spontaneous generation of superoxide in spermatozoa and observed a negative correlation with
spontaneous motility. Furthermore, seminal plasma levels of citrate, lactate, and glycerolphosphorylcholine are altered in men with azoospermia compared to healthy controls, suggesting a possible involvement of ROS with infertility (129).

Reactive oxygen species are also generated by advanced glycosylation end products that accumulate in a variety of environments including the male reproductive tract (130–134). Advanced glycosylation end products may cause a wide range of cellular effects and are key pathogenic initiators of diabetic complications (130). Sperm DNA is susceptible to attack from ROS due to its high unsaturated fatty acid composition (131). Sperm from men with type 1 diabetes have a 1.6-fold increase in fragmented sperm nuclear DNA (131). Although most components of the semen analysis are similar between persons with and without diabetes, the most prominent advanced glycosylation end product, carboxymethyl-lysine, has been shown to be significantly higher in the sperm of persons without diabetes (133), suggesting a possible role in male factor infertility.

Other work examining the associations between small metabolite molecules in the testes has identified potential metabolic alterations between persons with type 1 diabetes and controls (135). For example, Mallidis et al. (135) examined the effects of diabetes on small molecular metabolites in a streptozotocin-induced diabetic mouse model. Diabetic mice exhibited decreased creatine, choline, and carnitine levels, along with increased lactate, alanine, and myoinositol (135). These studies provided support to the notion that diabetics affected changes in the metabolome of the testicle itself. Furthermore, considering that the decreases observed were in metabolites with antioxidant qualities, the investigators (135) speculated a link between ROS and the metabolic changes seen in the testes of diabetics—intriguing concepts that require further investigation.

Other work on metabolomics in the realm of infertility has focused on assisted reproductive technology (ART) by forging an association between ART outcomes and levels of ROS in follicular fluid (FF) and embryo culture media (122, 123). A recent study examining FF samples and ROS has suggested a link between the metabolic markers of oxidative stress (–CH, –NH, –OH, and ROH) and pregnancy outcomes (122, 123). Previous in vitro studies examining 35 embryos and their spent media noted different hydroxyl modifications in embryos that implanted successfully compared to those that did not (122). Although further work is required, the field of metabolomics and its potential to contribute to identification of male fertility-related biomarkers is great.

**BIOMARKER DEVELOPMENT**

The identification of novel biomarkers is a laborious and demanding process. Given that a biomarker may be a gene, protein, or messenger RNA (136), the possibilities are nearly endless. Once a biomarker has been identified and shown to be relevant, it must be brought into the clinical realm. However, in spite of the multitude of biomarkers that have been recently discovered, progress moving from the bench to the bedside has been slow. The most significant advances in the field have come from the field of oncology (137). The ideal cancer biomarker would be detectable only in malignant tissue while eliciting a biological signal that is distinct from the surrounding normal tissues (137). One example of a biomarker used in oncology is PSA (137). The difficulties surrounding PSA are well known and significant given its inability to differentiate cancer from other conditions such as urinary tract infection and prostatitis. It is the existence of these types of issues that makes the search for an ideal biomarker so difficult.

Furthermore, gene and protein expression data have identified thousands of potential markers that are differentially expressed between benign and malignant tissues (137). This “information overload,” however, is only the start of the problem. The pathways of several candidate biomarkers are intrinsic to both normal and malignant cells, whereas some biomarkers exhibit such low levels they are clinically useless. Because a majority of cancers exhibit significant mutagenic heterogeneity, this has currently precluded a single biomarker from being identified in any cancer.

Challenges also exist in developing and bringing biomarkers to market once they have been scientifically validated (137). For example, the use of PCA3 in prostate cancer requires special handling, thus limiting dissemination of the assay (137). From a business point of view, most of the research in biomarkers is currently reported in the academic literature, which exists in the public domain (137). The difficulty of establishing, and protecting, intellectual property is significant. If a test is eventually developed for a biomarker, it would require government approval, which is expensive and time consuming. All of these factors, and many more, currently limit the ability of a biomarker to move from discovery to application.

**CONCLUSIONS**

Many causes of male factor infertility are unknown, related to a large part to the lack of understanding of the molecular and genetic mechanisms responsible for fertility defects. However, many advances are currently being made at a rapidly increasing pace. Although reliance on semen analysis for diagnosis is a significant limiting factor, the dawn of novel genomic, proteomic, and metabolomic advances holds great promise. It is important to remember, however, that although the molecular technologies continue to advance, the clinical segment of the research must not be forgotten. Clinicians should be vigilant when seeing patients with male factor infertility and accurately document all previous medical and family history as well as phenotype. Awareness of the genetic underpinnings of male factor infertility will hopefully soon develop a combined molecular and clinical picture of the infertile male. The concepts, ideas, and studies detailed in this review highlight the most meaningful advances in the fields of genetics, proteomics, and metabolomics as they relate to male factor infertility.

**REFERENCES**

Munuce MJ, Berta CL, Pauluzzi F, Caille AM. Relationship between anti- 

Kovac JR, Flood D, Mullen JB, Fischer MA. Diagnosis and treatment of azo-


Keel BA. Within- and between-subject variation in semen parameters in 


Hauser R, Meeker JD, Duty S, Silva MJ, Calafot AM. Altered semen quality 


WHO. World Health Organization (WHO) manual for the standard investi-

Lamb DJ. A look towards the future: advances in andrology expected to 

VOL. 99 NO. 4 / MARCH 15, 2013


