ANDROLOGY

ISSN: 2047-2919

ANDROLOGY

REVIEW ARTICLE

Correspondence:
Kenneth I. Aston, PhD, HCLD, 675 Arapeen Dr Ste 201, Salt Lake City, UT 84108, USA.
E-mail: kiaston@utah.edu

Keywords:
copy number variant, gene mutation, gene polymorphism, gene sequencing, genetic disorders, genome studies, genome-wide association study, male infertility, microarray, single-nucleotide polymorphism, whole genome sequencing

Received: 20-Nov-2013
Revised: 6-Jan-2014
Accepted: 7-Jan-2014

doi: 10.1111/j.2047-2927.2014.00188.x

Genetic susceptibility to male infertility: news from genome-wide association studies

K. I. Aston

Department of Surgery, University of Utah School of Medicine, Salt Lake City, UT, USA

SUMMARY

A thorough understanding of the genetic basis of male infertility has eluded researchers in spite of significant efforts to identify novel genetic causes of the disease, particularly over the past decade. Approximately half of male factor infertility cases have no known cause; however, it is likely that the majority of idiopathic male factor infertility cases have some unidentified genetic basis. Well-established genetic causes of male infertility are limited to Y chromosome microdeletions and Klinefelter's syndrome, together accounting for 10–20% of cases of severe spermatogenic failure. In addition to these, several genetic polymorphisms have been demonstrated to be significantly associated with male infertility. The discovery of new genetic associations with male infertility has been hampered by two primary factors. First, most studies are underpowered because of insufficient sample size and ethnic and phenotypic heterogeneity. Second, most studies evaluate a single gene, an approach that is very inefficient in the context of male infertility, considering that many hundreds of genes are involved in the process of testicular development and spermatogenesis. Significant recent advances in microarray and next-generation sequencing technologies have enabled the application of whole-genome approaches to the study of male infertility. We recently performed a pilot genome-wide association study (GWAS) for severe spermatogenic failure, and several additional male infertility GWAS have since been published. More recently, genomic microarray tools have been applied to the association of copy number variants with male infertility. These studies are beginning to shed additional light on the genetic architecture of male infertility, and whole-genome studies have proven effective in identifying novel genetic causes of the disease. This review will discuss some of the recent findings of these whole-genome studies as well as future directions for this research that will likely be the most productive moving forward.

BACKGROUND

In spite of the prevalence of male infertility, which affects up to 7% of men of reproductive age, a cause for infertility cannot be identified in up to half of cases. Severe spermatogenic impairment, a condition in which the sperm-producing capacity of the testis is diminished or depleted resulting in non-obstructive azoospermia (NOA) or severe oligozoospermia (SO), is observed in approximately 5% of men presenting for infertility diagnostic testing. While severe spermatogenic impairment is likely a genetic condition in many cases, few genetic causes have been identified to date.

This deficit of understanding significantly limits our ability to counsel patients regarding prognosis for treatment or to optimize personalized treatment strategies for individual patients. In addition, without the identification of the cause for a man's infertility, it is impossible to tell patients the likelihood of infertility in offspring that might result from advanced reproductive techniques (ART). In large part, the significant limitations in current diagnostic and treatment strategies are because of our limited understanding of the underlying causes of male infertility.

Established genetic causes of spermatogenic impairment

Although our understanding of the genetic basis for male factor infertility is limited, several genetic lesions that result in spermatogenic impairment are well characterized. Two of the most important are azoospermia factor (AZF) deletions of the Y

chromosome that are present in approximately 10–15% of men with severe spermatogenic defects (Poongothai *et al.*, 2009), and Klinefelter's syndrome, which is the underlying aetiology for 3–4% of infertile men (Paduch *et al.*, 2009).

Although these two aetiologies account for a minority of men with severe spermatogenic impairment, their prognostic value is significant. If a man is diagnosed with azoospermia because of an AZF deletion, the magnitude and location of the deletion is very prognostic of the probability for successful sperm retrieval on biopsy, and some deletions (AZFa, AZFb/c) virtually guarantee that no spermatozoa will be retrieved by biopsy (Krausz *et al.*, 2011). Furthermore, patients with AZF deletions who do have spermatozoa and undergo ART can be counselled that male offspring will be infertile, while female offspring will be unaffected. Likewise, for men diagnosed with Klinefelter's syndrome, sperm retrieval is often successful, and by most indications, risk of Klinefelter's in offspring following ART is not higher than the general population (Greco *et al.*, 2013). The identification of additional genetic causes of male infertility will further improve our ability to appropriately diagnose and treat the disease.

GENETIC TOOLS FOR THE STUDY OF MALE INFERTILITY

AZF deletions and Klinefelter's syndrome were initially both identified by cytogenetic techniques decades ago. AZF deletions have since been fully characterized using advanced molecular techniques (Repping *et al.*, 2006). Since the identification of these two well-established causes of severe spermatogenic impairment, numerous researchers have expended considerable effort to identify additional clinically relevant genetic causes of male infertility.

Sanger sequencing

By far the most exploited method employed to identify novel genetic causes of male infertility is the targeted resequencing study, largely popularized by the broad accessibility and affordability of Sanger sequencing technology. The workflow for these studies involves first, identification of a gene that is involved in spermatogenesis, often based on a male infertility phenotype identified in a knockout mouse model. DNA, usually from blood is then collected from a group of infertile men and a group of fertile controls, and the gene of interest is polymerase chain reaction amplified and sequenced. Analysis involves identifying single-nucleotide polymorphisms (SNPs) that occur more frequently in one group than the other, or gene disruptions that are unique to the infertile group. This workflow has been repeated by a number of laboratories, and the result is an impressive list of well over 100 genes that have been screened with significantly associated SNPs identified in many of them (Carrell & Aston, 2011). Unfortunately, the majority of genes have been evaluated in a single study, and issues including small sample size, differences in ethnicities between cases and controls and inadequate phenotyping in patients often leads to spurious associations.

While individually these types of studies are relatively inexpensive, the cumulative resources devoted to these studies are enormous. Unfortunately, in spite of significant efforts and investment, the fruits of those efforts are modest at best. Most reported associations either exist as a single study, or they fail to be replicated in subsequent studies. In spite of the limitations, several robust associations have been identified across multiple studies representing many hundreds of individuals including SNPs in MTHFR, GSTM1 and FSHB (Tuttelmann *et al.*, 2012; Wei *et al.*, 2012; Song *et al.*, 2013). While polymorphisms in these genes show convincing associations with male infertility based on multiple studies and meta-analyses, the risk conferred by these polymorphisms is modest.

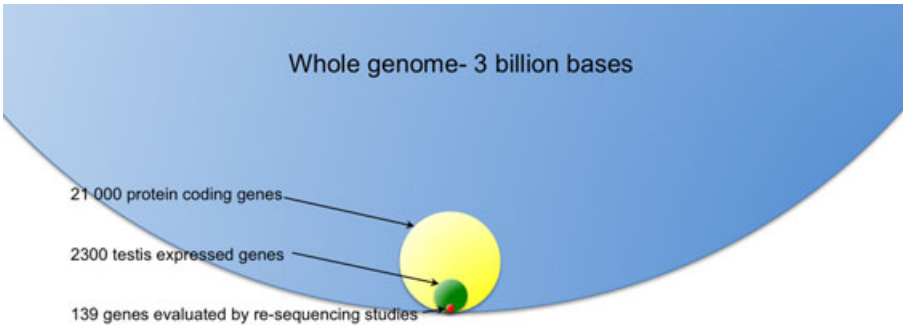
Clearly, the resequencing approach is inefficient owing to the complexity of the genome as it relates to the requirements for spermatogenesis and male fertility.

In all, at least 139 genes have been evaluated in resequencing studies. This number of genes, while an impressive effort, only represents a small fraction of the estimated 2300 genes involved in spermatogenesis (Schultz *et al.*, 2003), and less than 1% of the protein-coding genes in the genome. More importantly, the protein-coding region of the genome represents only about 0.7% of the entire genome. Recent data generated through the ENCODE project estimates that about 80% of the non-protein-coding portion of the genome is actually involved in regulatory functions. Hence, in spite of the tremendous efforts by many groups, we have really only evaluated less than 1/10 000th of the genome by resequencing in the context of male infertility (Fig. 1).

Whole-genome approaches

Clearly, the challenge to characterize the genetic basis for male infertility is largely a function of the complexity of the process of spermatogenesis, which requires the concerted action of many

Figure 1 Genes evaluated by male infertility resequencing studies in context. We have identified 139 genes evaluated by resequencing studies in the context of male infertility (red circle). This represents only about 6% of the 2300 genes that are likely required for spermatogenesis (green circle), and less than 0.7% of the approximately 21 000 protein-coding genes in the human genome. Moreover, the amount of genomic sequence evaluated by male infertility resequencing studies represents less than 1/10 000th of the entire human genome.



hundreds to several thousand genes under the control of mostly uncharacterized regulatory elements. So, herein lies the significant advantage of whole-genome approaches for discovery research.

While whole-genome approaches provide the coverage necessary to identify new genetic associations, a second very important requirement for the successful identification of novel genetic associations with infertility is a large number of carefully phenotyped samples. Sufficient sample sizes for well-powered studies can only be achieved through multicentre collaborations. This model has proven critical in the success of genetic association studies for other complex diseases. As costs for whole-genome studies have dropped considerably in the past decade, male infertility studies using these tools have begun to emerge. There are several available platforms that can be used for genome-wide studies, each with its advantages and disadvantages in terms of resolution, genomic coverage, cost and complexity of analysis (Table 1). Broadly classified, the two approaches are microarray and next-generation sequencing.

Genomic microarrays

Microarrays, including SNP arrays and comparative genomic hybridization (CGH) arrays provide a representative whole-genome snapshot. Arrays vary widely in their gnomic coverage and resolution. Although CGH arrays are designed primarily for the detection of copy number variants (CNVs), SNP arrays are capable of providing genotype information for several hundred thousand to several million SNPs as well as CNVs. Probe density of these arrays dictates the minimal size of detectable CNVs, and neither platform is capable of identifying point mutations or in most cases CNVs smaller than a few thousand bases. Typical costs for genomic microarrays range from approximately $200 to $500 per sample.

Next-generation sequencing

The costs for next-generation sequencing have dropped dramatically in the past few years, making exome and even whole-genome sequencing a viable option for basic research. Exome sequencing, the predominant whole-genome sequencing strategy for the study of complex disease involves selective enrichment of the coding regions across the genome, followed by sequencing. This strategy significantly reduces the cost of sequencing because it excludes the approximately 99% of the genome that is not protein coding. Current costs for exome sequencing range from approximately $500 to $1000 per sample. In addition to the cost advantage of this approach over whole-genome sequencing, an exome-centric approach significantly reduces analysis complexity, as annotation for the protein-coding portion of the genome makes assessment of functional impact of observed genomic changes much easier to predict.

Although exome sequencing offers certain advantages, the primary disadvantage of this approach is the significant potential for missing genetic events in regulatory regions that could be responsible for a particular male infertility phenotype. Whole-genome sequencing is the alternative approach, with disadvantages in cost and complexity of analysis. Current whole-genome sequencing costs are in the range of $5000–$10 000 per sample.

In an effort to overcome some of the disadvantages of limited coverage with exome sequencing strategies while striking a balance with cost, several companies including Agilent (Santa Clara, CA, USA), Illumina (San Diego, CA, USA), Life Technologies (Carlsbad, CA, USA) and NimbleGen (Madison, WI, USA) now offer expanded exome capture kits that extend beyond the exons to include untranslated regions (UTRs) and in some cases miRNAs. In addition, in most cases, user-specified custom content can be easily added to an exome capture kit allowing the addition of several mega bases of sequence not included in the off-the-shelf kits. These options can significantly improve the power of a study to detect important variants at minimal additional cost.

WHOLE-GENOME STUDIES IN MALE INFERTILITY

Over the past several years, microarray platforms have been used increasingly in the study of male infertility. Existing studies can be broadly classified into rare phenotype studies, whole-genome association studies and CNV studies. Table 2 summarizes the design and findings of the whole-genome studies that have been performed in infertile men to date.

Rare phenotype studies

To the authors' knowledge, the earliest study that utilized SNP microarrays in the study of male infertility involved the study of three brothers from a consanguineous family with complete globozoospermia (Dam *et al.*, 2007). In this study, the authors used a low-resolution microarray to localize a 17 Mb region of homozygosity shared by all three brothers. The causal mutation that was likewise shared by all three brothers was localized to the gene SPATA16 (Dam *et al.*, 2007). Several years later, two other articles were copublished that implicated DPY19L2 deletions in a significant proportion of men with complete globozoospermia. Both of these studies employed a 250 k SNP array to identify the causal 200 kb homozygous deletion (Harbuz *et al.*, 2011; Koscinski *et al.*, 2011). These studies illustrate the power of whole-genome studies to identify causal genetic variants without prior hypotheses of genomic location. These studies have been largely successful owing to the phenotype studied, which is very rare in the population and generally behaves very much like a simple Mendelian disorder. Identification and whole-genome analysis of additional familial cases of male infertility will likely be a fruitful area of future study.

Table 1 Advantages and disadvantages of the available 'whole genome' options for genetic analysis

Approach	Advantages	Disadvantages
CGH microarray	CNV information, less costly than sequencing	Limited resolution, no SNP/point mutation information, relatively complex analysis
SNP microarray	SNP/CNV information, less costly than sequencing	Limited resolution, no point mutation information, relatively complex analysis
Exome sequencing	High-resolution: SNP/mutation/CNV information	Limited to coding regions, complex analysis, more costly than microarrays
Whole-genome sequencing	High-resolution: SNP/mutation/CNV information, including non-coding regions	Complex analysis, about 10 × more costly than exome sequencing

CGH, comparative genomic hybridization; CNV, copy number variants; SNP, single-nucleotide polymorphism.

Table 2 Summary of whole-genome studies performed in infertile men

Publication date	Variants assessed	Manuscript title	Phenotypes (*n*)	Platform	Summary findings
August, 2007	Point mutation	Homozygous mutation in SPATA16 is associated with male infertility in human globozoospermia	Globozoospermic brothers (3)	Affy 10 k SNP array	Homozygous mutation in SPATA16
November, 2009	SNP	Genome-wide study of single-nucleotide polymorphisms	Control (80), SO (52), NOA (40)	Illumina 370 k SNP array	21 marginally significant SNPs
April, 2010	SNP	Evaluation of 172 candidate polymorphisms for association with oligo or azoospermia in a large cohort of men of European descent	Control (158), SO (141), NOA (80), Moderate Oligo (63)	Illumina BeadXpress	Nine marginally significant SNPs
March, 2011	CNV	A recurrent deletion of DPY19L2 causes infertility in man by blocking sperm head elongation and acrosome formation	Globozoospermia (20)	Affy 250 k SNP array	15/20 globo pts had 200 kb homozygous deletion in DPY19L2
March, 2011	CNV	A recurrent deletion of DPY19L2 causes infertility in man by blocking sperm head elongation and acrosome formation	Globozoospermic brothers (5)	Affy 250 k SNP array	4/5 globo brothers had 200 kb homozygous deletion in DPY19L2
April, 2011	CNV	CNV in patients with SO and sertoli-cell-only syndrome	SO (89), SCOS (37), Control (100)	Agilent 244A/ 400 k arrays	14 patient-specific CNVs, Two additional CNVs significantly associated with SCOS
December, 2011	SNP	A genome-wide association study of men with symptoms of testicular dysgenesis syndrome and its network biology interpretation	Infertile (107,) TGCC (212), Hypospadia (31), Cryptorchid (138), Control (439), Replication (671)	Affy 6.0 array (946 k probes)	Two marginally significant SNPs
December, 2011	SNP	A genome-wide association study in Chinese men identifies three risk loci for non-obstructive azoospermia	NOA (1000), Control (1703), Replication NOA (1747), Replication control (3652)	Affy 6.0 array (946 k probes)	Three strongly significant SNPs (OR 1.23–1.39)
January, 2012	CNV	Array CGH in male infertility	Maturation arrest (9), Controls (20), Replication controls (130)	Agilent 244A/ arrays	Eight patient-specific CNVs
May, 2012	SNP	A genome-wide association study reveals that variants within the HLA region are associated with risk for NOA	NOA (802), Control (1863), Replication NOA (1424), Replication control (2713)	Affy 6.0 array (946 k probes)	Four strongly significant SNPs in HLA region (OR 1.29–1.60)
June, 2012	CNV	Genome-wide screening of severe male factor infertile patients using BAC array CGH	NOA/SO (37), Control (10)	BAC array (1 Mb resolution)	Patient -specific CNVs
June, 2012	SNP	Genome-wide association study identifies candidate genes for male fertility traits in humans	Hutterite men (269), Replication men w/SA (123)	Affy 500 k, 5.0, and 6.0 arrays	Nine SNPs significantly associated with SA results
October, 2012	CNV	High-resolution X chromosome-specific array CGH detects new CNVs in infertile males	NOA (49), Crypto (25), Oligo (22), Control (103), Rep NOA (57), Replication crypto (53), Replication oligo (153)	Agilent custom 80k X chr array	Increased deletion burden in infertiles, 31 patient-specific CNVs (including several recurrent), 33 control-specific CNVs
March, 2013	CNV	Human spermatogenic failure purges deleterious mutation load from the autosomes and both sex chromosomes, including the gene DMRT1	NOA/SO (323), Control (1100), Replication NOA (1460), Replication control (1703)	Illumina 370 k, OmniExpress; Affy 6.0 SNP arrays	Increased burden of rare CNVs in infertiles, multiple patient-specific CNVs (including several recurrent, e.g. DMRT1 del), patient enriched CNVs (e.g. SPACA5)
April, 2013	SNP	SNP array analysis in men with idiopathic azoospermia or oligoasthenozoospermia syndrome	NOA (26), Crypto (1), OAT (12)	Illumina Omni1	Partial deletion of CLCA4 in azoo patient

CGH, comparative genomic hybridization; Crypto, cryptozoospermia; NOA, non-obstructive azoospermia; OAT, oligoasthenoteratozoospermia; oligo, oligozoospermia; SNP, single-nucleotide polymorphism; SO, severe oligozoospermia.

Genome-wide association studies

Another whole-genome approach that has been widely used to identify variants associated with a variety of complex diseases employs genotyping arrays to evaluate the relative incidence of SNPs in cases vs. controls.

We first performed a pilot GWAS in 2009 that included 80 controls and 92 azoospermic and severe oligozoospermic men using the Illumina 370 k array (Aston & Carrell, 2009). While we were able to identify several SNPs with marginal associations with spermatogenic impairment, the initial results indicated that common polymorphisms do not contribute appreciably to severe male infertility. As a follow-up we used a targeted medium throughput approach to evaluate 172 polymorphisms in a larger cohort of infertile men and controls (Aston *et al.*, 2010). These SNPs were selected based on the pilot GWAS as well as previously reported SNP associations. In this study, we identified a small number of SNPs that again showed moderate associations.

The next GWAS was a larger two-stage study to identify SNPs associated with testicular dysgenesis syndrome (TDS) phenotypes (Dalgaard *et al.*, 2012). The first phase of the study

involved Affymetrix 6.0 array analysis of 488 men with TDS symptoms and 439 healthy controls, with the most significant SNPs evaluated in a replication cohort of an additional 436 men with TDS symptoms and 235 additional controls. In this study, the authors identified SNPs in two TGF beta genes, TGFBR3 and BMP7 that were significantly associated with TDS (Dalgaard *et al.*, 2012).

Two Chinese groups recently performed two large GWAS in non-obstructive azoospermia (NOA) men (Hu *et al.*, 2012; Zhao *et al.*, 2012). Both studies utilized a two-stage design using genotyping microarrays to screen a discovery cohort, followed by targeted genotyping of a large number of cases and controls. In the first study, 1000 NOA men and 1703 controls were screened using Affymetrix 6.0 microarrays, and 32 marginally significant SNPs were carried forward for targeted genotyping in an additional 1747 NOA men and 3652 controls. The results were four SNPs that were significantly associated with NOA, with odds ratios (O.R.) ranging from 1.23 to 1.39 (Hu *et al.*, 2012). The second study genotyped 802 NOA men and 1863 controls by microarray followed by genotyping of 32 SNPs in cohorts of an additional 1424 NOA men and 2713 controls. In this study, four SNPs in the human leukocyte antigen (HLA) region were significantly associated with NOA (O.R. 1.29–1.60) (Zhao *et al.*, 2012). Importantly, none of SNPs identified by the two studies were concordant between studies. The discordant results may be the result of subtle population stratification in the latter study, which has been demonstrated to yield spurious associations in the hypervariable HLA region (Cardon & Palmer, 2003).

The most recent male infertility GWAS took a different approach. In this study, the authors performed microarray genotyping in 269 Hutterite men, a group derived from a founder population of European descent that desires large families and proscribes contraception (Kosova *et al.*, 2012). Forty-one SNPs with marginal associations with fertility traits including family size and birth interval were evaluated in 123 ethnically diverse men from Chicago with semen analysis data. In all, nine SNPs selected from the GWAS were significantly associated with semen parameters in the Chicago cohort (Kosova *et al.*, 2012).

Although the GWAS cited here was an important step in characterizing the genetic basis for idiopathic male infertility, the clinical relevance of the findings are minimal because of discordant results between studies as well as modest risks conferred by the identified SNPs. As is the case with the genetic architecture of most complex diseases, these studies have demonstrated conclusively that common polymorphisms do not individually contribute clinically significant risk to severe male infertility.

Copy number variants studies

A more recent, and arguably more fruitful strategy to identify genetic causes of male infertility has been to use genomic microarrays to identify CNVs. CNVs are a class of structural variation in which relatively large regions are either duplicated or deleted in the genome.

Tüttelmann and colleagues reported on the first whole-genome CNV study for male infertility in 2011 (Tuttelmann *et al.*, 2011). In this study, the authors performed CGH microarray analysis on 89 severely oligozoospermic men, 37 men with sertoli cell only syndrome (SCOS) and 100 controls. They identified 14 patient-specific CNVs as well as two CNVs that were found significantly more frequently in SCOS men that in controls. In addition, it was found that the number of CNVs in the controls was negatively correlated with sperm counts, and sex chromosome CNVs were significantly overrepresented in men with SCOS (Tuttelmann *et al.*, 2011).

A second small study used CGH arrays to screen for CNVs in nine men with maturation arrest and 20 controls (Stouffs *et al.*, 2012). The 10 most promising patient-specific CNVs were subsequently screened by qPCR in 130 additional controls, and eight of the 10 CNVs remained to be patient specific, indicating a potential involvement in the maturation arrest phenotype (Stouffs *et al.*, 2012).

Another small CGH-based study identified several patient-specific CNVs after screening 37 infertile men with NOA or SO and 10 control men (Song *et al.*, 2012). The impact of this study was diminished by the small sample size as well as the inclusion of 10 men with AZF deletions.

Recently, a larger CNV study used a high-resolution CGH array specific to the X chromosome to identify CNVs associated with infertility (Krausz *et al.*, 2012). A total of 96 infertile men ranging from NOA to oligozoospermic were screened against 103 controls. The most promising CNVs were further screened by qPCR in an additional group of 263 infertile men, and a total of 31 patient-specific CNVs were identified, including several recurrent CNVs. In addition, the authors found that infertile men harboured significantly more X chromosome deletions than controls (Krausz *et al.*, 2012).

We recently reported the results of the largest male infertility CNV study to date that included an international group of collaborators (Lopes *et al.*, 2013). In this study, we identified several important features of spermatogenic failure as well as some likely causal variants. This study included 323 infertile men evaluated by SNP microarrays. Several of the most relevant findings were evaluated in a targeted manner in replication cohorts that included nearly 1000 Han Chinese cases and 2000 controls, 400 additional Caucasian NOA men and more than 4500 unscreened Caucasian men with CNV data reported in the database for genomic variants. In this study, data were analysed to identify locus-specific associations as well as CNV burden (Lopes *et al.*, 2013).

In this study, in agreement with the Tüttelmann and Krausz studies, we found a CNV burden effect on male infertility. We reported an increased frequency of rare CNVs genome wide in infertile men compared with controls. In addition, as initially reported by Kraus *et al.* we found the X chromosome to be particularly enriched for CNVs in infertile men. Unexpectedly, we found a similarly strong increased burden of both deletions and duplications on the Y chromosome in infertile men that were previously screened for the canonical AZF deletions (Lopes *et al.*, 2013).

In addition to the finding of increased CNV burden, we identified several other likely important CNVs including a recurrent deletion on Xp11.23, apparently mediated by non-allelic homologous recombination (NAHR) that was present significantly more frequently in infertile men than controls. This deletion was also reported in a single infertile man and was absent in controls evaluated by Krausz and colleagues (Lopes *et al.*, 2013).

One of the most important findings of this study was a recurrent deletion that included the gene DMRT1, a gene orthologous to the putative sex determination locus of the avian ZW chromosome system. Larger deletions in this region have been

associated with severe phenotypes including sex reversal. We first identified the deletion in two apparently unrelated men from the Utah cohort and later identified three distinct, but overlapping deletions in the same region in the Han Chinese cohort, and this deletion was absent in more than 7700 controls (Lopes *et al.*, 2013).

While there are clear limitations in each of the CNV studies performed to date, the CNV burden effect that has been reported by three separate groups represents a potentially important finding, as does a handful of patient-specific CNVs that displayed some concordance between studies. Larger, carefully designed studies will be instrumental in identifying additional rare CNVs associated with male infertility.

DATA ANALYSIS

An extremely important component of whole-genome studies, and generally the most difficult component to deal with, is data analysis. A thorough discussion of data analysis is beyond the scope of the present review, but it is worth mentioning that careful consideration of data analysis workflow is critical from inception of the study.

Study design considerations should include power calculations to determine appropriate sample sizes and technical planning to avoid potentially confounding batch effects resulting from systematic differences in handling of cases and controls during DNA extraction, library preparation and microarray analysis or sequencing. In addition, careful phenotyping and avoidance or correction of ethnic admixture are critical for avoiding spurious associations.

Currently, the majority of in-depth analyses, particularly involving next-generation sequence data, are best handled by bioinformaticians and other analytical specialists. However, a number of software companies are developing tools to make whole-genome analyses accessible to non-specialists. While some of the available software packages are very powerful and user friendly, appropriate analysis including quality control, data normalization, variant calling and statistical testing are critical for proper interpretation of the vast amounts of data generated in whole-genome studies.

Collaborations that include clinicians, geneticists and bioinformaticians are critical for ensuring that huge investments in performing whole-genome studies are not wasted because of deficiencies in phenotyping, study design and data generation, analysis and interpretation.

CONCLUSIONS

The whole-genome studies that have been performed thus far have shed some light on the genetic architecture of male infertility, but we are in the very early stages of characterizing the aetiology of the disease. Given their modest sample size, existing studies have really only been powered to identify independent variants with strong effect. With the most severe phenotypes such as NOA and complete globozoospermia, it is reasonable to assume that the disorder may be the result of inactivation of a single gene or genomic region critical in the process of spermatogenesis. Targeting of these phenotypes is important to identify genes that are absolutely required for spermatogenesis and sperm function. However, future studies to investigate the spectrum of male infertility and subfertility will be important, as only a small percentage of infertile men display these severe phenotypes. The genomic architecture of subfertility is likely much more complex, and it is expected that variants in multiple genes and regulatory regions likely all contribute cumulatively to overall spermatogenic output and sperm function. Again, thorough and careful data analysis using the appropriate framework will be critical in proper interpretation of the data from future studies.

These early whole-genome studies illustrate the value of whole-genome approaches in identifying novel genetic causes of male infertility. Based on the results of the genome-wide male infertility studies that have been performed to date, we can conclude that common polymorphisms generally do not contribute appreciably to severe male infertility, but larger structural variations as well as individually rare events including point mutations are an important contributor to these phenotypes. We still have a long way to go to more completely characterize the genetic basis for spermatogenic failure and other idiopathic cases of male infertility. As demonstrated by other complex disease models, characterization of the genetic basis for male infertility will likely ultimately require the analysis of thousands of carefully phenotyped samples using tools capable of identifying rare variants including CNVs and point mutations. In addition, careful analysis using polygenic models and considering the involvement not only of spermatogenic genes but also of their regulatory elements are necessary to understand the genetic basis of male infertility. These aims will only be accomplished through the development of well-funded collaborations and with the concerted efforts of many researchers. With the increased understanding of the aetiologies of male infertility that will come through these collaborative research efforts, as well as continued advances in genomic tools, it is likely that the hope of personalized medicine in the Andrology laboratory will be realized in the coming decades. Personalized genomic approaches to the diagnosis and treatment of male infertility will significantly improve our ability to care for patients.

REFERENCES

Aston KI & Carrell DT. (2009) Genome-wide study of single-nucleotide polymorphisms associated with azoospermia and severe oligozoospermia. *J Androl* 30, 711–725.

Aston KI, Krausz C, Laface I, Ruiz-Castane E & Carrell DT. (2010) Evaluation of 172 candidate polymorphisms for association with oligozoospermia or azoospermia in a large cohort of men of European descent. *Hum Reprod* 25, 1383–1397.

Cardon LR & Palmer LJ. (2003) Population stratification and spurious allelic association. *Lancet* 361, 598–604.

Carrell DT & Aston KI. (2011) The search for SNPs, CNVs, and epigenetic variants associated with the complex disease of male infertility. *Syst Biol Reprod Med* 57, 17–26.

Dalgaard MD, Weinhold N, Edsgard D, Silver JD, Pers TH, Nielsen JE *et al.* (2012) A genome-wide association study of men with symptoms of testicular dysgenesis syndrome and its network biology interpretation. *J Med Genet* 49, 58–65.

Dam AH, Koscinski I, Kremer JA, Moutou C, Jaeger AS, Oudakker AR *et al.* (2007) Homozygous mutation in SPATA16 is associated with male infertility in human globozoospermia. *Am J Hum Genet* 81, 813–820.

Greco E, Scarselli F, Minasi MG, Casciani V, Zavaglia D, Dente D *et al.* (2013) Birth of 16 healthy children after ICSI in cases of nonmosaic Klinefelter syndrome. *Hum Reprod* 28, 1155–1160.

Harbuz R, Zouari R, Pierre V, Ben Khelifa M, Kharouf M, Coutton C *et al.* (2011) A recurrent deletion of DPY19L2 causes infertility in man by

blocking sperm head elongation and acrosome formation. *Am J Hum Genet* 88, 351–361.
Hu Z, Xia Y, Guo X, Dai J, Li H, Hu H *et al.* (2012) A genome-wide association study in Chinese men identifies three risk loci for non-obstructive azoospermia. *Nat Genet* 44, 183–186.
Koscinski I, Elinati E, Fossard C, Redin C, Muller J, Velez de la Calle J *et al.* (2011) DPY19L2 deletion as a major cause of globozoospermia. *Am J Hum Genet* 88, 344–350.
Kosova G, Scott NM, Niederberger C, Prins GS & Ober C. (2012) Genome-wide association study identifies candidate genes for male fertility traits in humans. *Am J Hum Genet* 90, 950–961.
Krausz C, Chianese C, Giachini C, Guarducci E, Laface I & Forti G. (2011) The Y chromosome-linked copy number variations and male fertility. *J Endocrinol Invest* 34, 376–382.
Krausz C, Giachini C, Lo Giacco D, Daguin F, Chianese C, Ars E *et al.* (2012) High resolution X chromosome-specific array-CGH detects new CNVs in infertile males. *PLoS ONE* 7, e44887.
Lopes AM, Aston KI, Thompson E, Carvalho F, Goncalves J, Huang N *et al.* (2013) Human spermatogenic failure purges deleterious mutation load from the autosomes and both sex chromosomes, including the gene DMRT1. *PLoS Genet* 9, e1003349.
Paduch DA, Bolyakov A, Cohen P & Travis A. (2009) Reproduction in men with Klinefelter syndrome: the past, the present, and the future. *Semin Reprod Med* 27, 137–148.
Poongothai J, Gopenath TS & Manonayaki S. (2009) Genetics of human male infertility. *Singapore Med J* 50, 336–347.
Repping S, van Daalen SK, Brown LG, Korver CM, Lange J, Marszalek JD *et al.* (2006) High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nat Genet* 38, 463–467.
Schultz N, Hamra FK & Garbers DL. (2003) A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. *Proc Natl Acad Sci U S A* 100, 12201–12206.
Song SH, Shim SH, Bang JK, Park JE, Sung SR & Cha DH. (2012) Genome-wide screening of severe male factor infertile patients using BAC-array comparative genomic hybridization (CGH). *Gene* 506, 248–252.
Song X, Zhao Y, Cai Q, Zhang Y & Niu Y. (2013) Association of the Glutathione S-transferases M1 and T1 polymorphism with male infertility: a meta-analysis. *J Assist Reprod Genet* 30, 131–141.
Stouffs K, Vandermaelen D, Massart A, Menten B, Vergult S, Tournaye H *et al.* (2012) Array comparative genomic hybridization in male infertility. *Hum Reprod* 27, 921–929.
Tuttelmann F, Simoni M, Kliesch S, Ledig S, Dworniczak B, Wieacker P *et al.* (2011) Copy number variants in patients with severe oligozoospermia and sertoli-cell-only syndrome. *PLoS ONE* 6, e19426.
Tuttelmann F, Laan M, Grigorova M, Punab M, Sober S & Gromoll J. (2012) Combined effects of the variants FSHB -211G>T and FSHR 2039A>G on male reproductive parameters. *J Clin Endocrinol Metab* 97, 3639–3647.
Wei B, Xu Z, Ruan J, Zhu M, Jin K, Zhou D *et al.* (2012) MTHFR 677C>T and 1298A>C polymorphisms and male infertility risk: a meta-analysis. *Mol Biol Rep* 39, 1997–2002.
Zhao H, Xu J, Zhang H, Sun J, Sun Y, Wang Z *et al.* (2012) A genome-wide association study reveals that variants within the HLA region are associated with risk for nonobstructive azoospermia. *Am J Hum Genet* 90, 900–906.

