Preface

Lessons from Genome-Wide Association Studies in Reproductive Medicine

Joop S. E. Laven, MD, PhD[1]

[1] Division of Reproductive Medicine, Department of Obstetrics and Gynecology, Erasmus MC, Rotterdam, The Netherlands

Semin Reprod Med 2016;34:193–195

Joop S. E. Laven, MD, PhD

Although the genetic origin of human disease has been recognized for quite some time, identification of the causative genes was limited, slow, and cumbersome. This changed after the introduction of genome-wide association studies (GWAS). Fueled by several human genome reference projects along with the development of novel high-throughput systems, these studies generated large datasets in several international consortia. In the latter GWAS, data were combined with existing, large epidemiological studies and provided many new insights in the genetic background of several generally occurring traits as well as in different communicable diseases. GWASs have over flooded many clinical and basic research areas with gene discoveries, including those in reproductive medicine. It also led to the establishment of large-scale collaborations within international consortia, leading to high quality as well as high-impact publications. This special issue of *Seminars in Reproductive Medicine* tries to summarize recent GWAS findings and tries to dissect out how these new genetic data might impact on our daily practice in reproductive medicine.

Menarche and Menopause

Younger age at menarche appears to be associated with increased risks for type 2 diabetes, obesity, and cardiovascular disease. Thus, the timing of puberty may provide insight into future health trajectories and outcomes. Despite the limitations of some studies and that age at menarche was assayed by recall, available GWAS data suggest that the genetic architecture of age at menarche may be shared among European, African, and Asian populations. The strongest single nucleotide polymorphism (SNP) identified was strongly related to the so-called tumor-related gene network. Other loci detected in the study include nuclear hormone receptors and genes known to be associated with disorders of puberty. Moreover, there is a strict relationship between some alleles associated with body mass index and age at menarche. The former were also associated with various pubertal growth parameters. Finally, there seems to be some connection between the onset of menarche with insulin and obesity-associated genes. Although some of the loci identified using GWAS do explain the genetic background of the onset of menarche, environmental factors might also play an important role. Moreover, dynamic epigenetic mechanisms can modify neuroendocrine responses to exogenous and endogenous indicators to adjust reproductive development and function in the context of real-time exposures and experiences. In addition to short-term effects, epigenetic marks can provide a collective memory of early life exposures leading to longer-term, even transgenerational, changes in gene expression. The timing of puberty, using menarche as marker, may be a harbinger of chronic disease risk and guide for future health and development. Characterization and cataloguing of genetic risk markers may allow for early interventions for our current patients and, perhaps, for their daughters.

Age at menopause is strongly associated with mother's age at menopause. In recent years, common genetic variants have been identified by GWAS that have led to the identification of 57 genetic loci associated with approximately 6% of common variation in age at menopause. GWAS has been very successful at identifying novel biological pathways involved in reproductive aging. Approximately two-thirds of the loci reported so far include genes involved in the DNA damage response, highlighting the importance of this pathway in determining oocyte reserve. In addition, GWAS demonstrates that the hypothalamic–pituitary axis is involved in menopause timing as well as puberty timing, showing the first genetic link between timing of the start and end of reproductive life. Finally, later age at menopause seems to be related with an increased risk of breast cancer. However, it seems that prolonged exposure to estrogens seems to be the key driver behind this relationship,

Address for correspondence Joop S. E. Laven, MD, PhD, Division of Reproductive Medicine, Department of Obstetrics and Gynecology, Erasmus MC, PO Box 2040, 3000 CA, Rotterdam, The Netherlands (e-mail: j.laven@erasmusmc.nl).

Issue Theme Lessons from Genome-Wide Association Studies in Reproduction; Guest Editor, Joop S. E. Laven, MD, PhD

DOI http://dx.doi.org/10.1055/s-0036-1585407.
ISSN 1526-8004.

indicating that there is only a minor genetic component to this increased cancer risk.

Polycystic Ovary Syndrome and Primary Ovarian Insufficiency

Polycystic ovary syndrome (PCOS) is the most common endocrinopathy in women, and several twin and family studies have confirmed its genetic background. Several GWASs have explored this genetic background and identified genes that are involved in gonadotropin and insulin signaling and type 2 diabetes mellitus. Several genes involved in organ size and cell proliferation were identified to be associated with PCOS. Moreover, some genes important for chromatin remodeling and transmembrane receptor processing seem to be associated with PCOS. Recently, along with the previously identified genes involved in gonadotropin signaling, loci associated with the follicle-stimulating hormone B polypeptide (*FSHB*) gene have been found to be involved in women with PCOS. Recently, it was also discovered that genes associated with late menopause are more frequently found in women with PCOS. PCOS is a heterogonous disease with different sub-phenotypes, which seem to have distinctly different genetic backgrounds. Exploring these differences might provide more and better insight into the diagnosis and/or specific treatment modalities of the different subtypes of PCOS.

Taken together, recent old as well as newer genetic approaches have revealed numerous genetic causes of primary ovarian insufficiency (POI). Up till now, however, they only explain the genetic background in approximately 20 to 25% of POI cases. By using modern genetic detection techniques, many genes have arisen that play some role in the pathophysiology of POI. Some of them have been replicated in other studies; however, the majority have not been proven yet to be unequivocally causative through functional validation studies. There are notable differences among populations as well as between different ethnicities. Some of the causative genes are not only expressed in the ovary but also play a role in other pathways. Moreover, many genes that currently appear isolated in function actually may be interrelated within yet to be defined pathways. Uncovering the remainder of the causative genes will be facilitated not only by whole genome approaches involving larger cohorts in multiple populations but also incorporating environmental exposures and exploring signaling pathways in intragenic and intergenic regions that point to perturbations in regulatory genes and networks.

Elucidating the genetic and molecular basis of POI as well as PCOS is of paramount importance not only in understanding ovarian physiology but also in providing appropriate genetic counseling and fertility guidance. Once additional variants are detected, it might become possible to predict the occurrence of PCOS or POI, aiding a more patient-tailored treatment and or prevention.

Endometriosis

Endometriosis is a heritable, complex chronic inflammatory disease, for which much of the causal pathogenic mechanism remains unknown. GWASs to date have identified 12 SNPs at 10 independent genetic loci associated with endometriosis. Most of these were more strongly associated with the more advanced stages of the disease. Unfortunately, due to the lack of other phenotypic information regarding the extent of the disease in the individual case datasets, further dissection of genetic heterogeneity between the surgically defined stages is not yet feasible. The loci are almost all located in intergenic regions that are known to play a role in the regulation of expression of target genes yet to be identified. Because of this, it has been difficult to highlight specific causal molecular mechanisms through which endometriosis-associated variants impact on disease, and the genes through which their effects are mediated. Moreover, although information on endometrial tissue and cells is very limited, major limitations for endometriosis research, many of the genomic annotation features are shared across different cell and tissue types. Pathway analysis indicated shared genetic origins between endometriosis and fat distribution mostly through the WNT/β catenin signaling pathway. Similarly, strong associations were found between endometriosis and ovarian cancer, showing strong genetic correlations between endometriosis and clear-cell, endometrioid, and low-grade serous ovarian cancer. To understand these observations, further phenotypic dissection requires the collection of much more detailed, standardized surgical and clinical data, integrated with genomic and other molecular profiling of endometrium and other relevant samples from the same woman. The WERF EPHect standardized data collection instruments and sample collection protocols now allow such data to be collected and compared across different endometriosis research centers, paving the way for studies focused on the translation of GWAS results into results that are meaningful for patients and practitioners: novel treatments that target subtypes of disease.

Uterine Leiomyomata

Uterine leiomyomata (UL), or more commonly referred to as fibroids, are the most common tumor of the female reproductive system. Epidemiological analyses, including familial aggregation, twin studies, and racial discrepancies in disease prevalence and morbidity, indicated that genetic factors influence risk for developing UL. Early genetic evidence came from studies identifying chromosomal rearrangements and point mutations as a cause of fibroids. Moreover, twin and family studies as well as early linkage studies in affected family members indicated that some germline variants also could contribute to a woman's genetic susceptibility for uterine leiomyomata. Moreover, there seems to be an ethnic predisposition which is probably also conveyed through germ-line variants. GWAS discovered several potent genome-wide significant variants. The exact biological role of these variants needs to be determined yet; however, some signals relate to growth and muscle development and as such provide important clues for the pathogenesis of fibroids.

Future Directions and Possibilities

A major advantage of pathway-based GWASs is their ability to correlate the wealth of information embedded in GWAS

data with our knowledge of functional biological pathways that is readily available in public databases. The pathway-based approach integrates GWAS results with genes in biological pathways or gene sets from predefined human databases, ranking all genes according to their statistical significance.

Since one genetic variant might exert a similar effect in different pathways, one variant might be involved in different diseases. One strategy for identifying such pleiotropic genes is to analyze potentially correlated disease phenotypes simultaneously via a multivariate GWAS approach. This new strategy will be more effective than univariate GWAS approach in identifying pleiotropic genes underlying human diseases of shared genetic susceptibility, thereby revealing interconnected pathophysiological networks for a spectrum of common diseases. Thus, bivariate analysis may offer new insights into the pathophysiology of PCOS, and potentially reveal key mechanistic links between PCOS and related disease such as T2D and obesity.

There are certainly ethnic variations between different traits or diseases that might be genetically determined. Future studies should therefore also explore these ethnic genetic differences to elucidate the evolutionary history of reproductive diseases as well as the impact the environment might have had on the evolution of the current phenotype.

In general, there are three areas of potential utility for GWAS results for reproductive diseases or traits like menopause and menarche. First, these genetic data might further improve the understanding of pathogenesis and the pathophysiology of reproductive traits. Second, dissection of phenotypic heterogeneity (discovery of "subtypes") might further facilitate a better diagnosis and/or treatment. Finally, based on the genetic determinants that have been associated with the reproductive traits and or diseases, one might predict the risk for having or developing that specific disease in the future more accurately. As such, there is still a lot to be done and future research will provide some definite answers to questions which have puzzled us for many years.