Molecular and Cellular Endocrinology xxx (2017) 1–11

Contents lists available at ScienceDirect

Molecular and Cellular Endocrinology

journal homepage: www.elsevier.com/locate/mce

The molecular complexity of primary ovarian insufficiency aetiology and the use of massively parallel sequencing

Paul Laissue*

Center For Research in Genetics and Genomics-CIGGUR, GENIUROS Research Group, School of Medicine and Health Sciences, Universidad del Rosario, Bogotá, Colombia

ARTICLE INFO

Article history:
Received 29 May 2017
Received in revised form
21 July 2017
Accepted 22 July 2017
Available online xxx

Keywords:
Primary ovarian insufficiency
Next generation sequencing
Female infertility
Molecular aetiology

ABSTRACT

Primary ovarian insufficiency (POI) is a frequently occurring pathology, leading to infertility. Genetic anomalies have been described in POI and mutations in numerous genes have been definitively related to the pathogenesis of the disease. Some studies based on next generation sequencing (NGS) have been successfully undertaken as they have led to identify new mutations associated with POI aetiology.

The purpose of this review is to present the most relevant molecules involved in diverse complex pathways, which may contribute towards POI. The main genes participating in bipotential gonad formation, sex determination, meiosis, folliculogenesis and ovulation are described to enable understanding how they may be considered putative candidates involved in POI. Considerations regarding NGS technical aspects such as design and data interpretation are mentioned. Successful NGS initiatives used for POI studying and future challenges are also discussed.

Contents

1. Introduction

The World Health Organization has defined human infertility as a disorder characterised by the failure to become pregnant after at least 12 months of regular and unprotected intercourse. It can be considered a public health concern since it affects more than 50 million couples worldwide (Datta et al., 2016; Mascarenhas et al., 2012). Although precise infertility prevalence has been difficult to establish, especially due to differences regarding its definition and because study populations vary depending on the pertinent inclusion/exclusion factors, 5%–20% has been reported (Datta et al., 2016; Mascarenhas et al., 2012).

Abnormalities explaining infertility in 70% of couples can be detected after standard clinical testing; however, despite considerable advances in human infertility diagnosis the aetiology remains unexplained in 30% of cases. Exclusively female factors

* Corresponding author. Center For Research in Genetics and Genomics-CIGGUR, GENIUROS Research Group, School of Medicine and Health Sciences, Universidad del Rosario, Carrera 24 N° 63C-69, Bogotá, Colombia.
E-mail address: paul.laissue@urosario.edu.co.

http://dx.doi.org/10.1016/j.mce.2017.07.021
0303-7207/

(accounting for a third of all causes of infertility) include dysfunction leading to ovulation disorders (e.g. oligomenorrhoea, amenorrhoea), tubal disease (e.g. obstruction) and endometrial pathology (endometriosis, miscarriage) (Smith et al., 2003). The correct functioning of each step (e.g. sex determination, meiosis, gametogenesis, folliculogenesis/oocyte maturation, ovulation) leading to healthy competent oocyte ovulation is certainly crucial for ensuring reproductive success in mammals. Dysregulation regarding any of these steps may lead to primary ovarian insufficiency (POI), currently affecting ~1.5% of women worldwide (Luborsky et al., 2003). POI is defined as the cessation of menses or spaniomenorrhea over 4 months and plasmatic levels of FSH over 25 IU/l, reflecting ovarian impairment thereby closing a feed-back loop in the pituitary gland (European Society for Human Reproduction and Embryology (ESHRE) Guideline Group on POI, 2016). It can result from disturbances during development (formation and recruitment) during follicles' stock establishment during embryo life or by its premature depletion due to abnormally high levels of atresia (Persani et al., 2011). Several POI aetiologies have been described, but most cases are still classified as idiopathic, thereby enhancing genetic factor research. Chromosomal abnormalities have been described, especially those involving the X chromosome, as deletions (Turner's syndrome) and translocations (Elsheikh et al., 2002; Lacombe et al., 2006). Point mutations have also been identified in some POI candidate genes, some of which have been formerly validated by functional tests as being aetiological (see below) (Laissue, 2015) (and references therein).

From a molecular view, it has been shown that hundreds of genes belonging to overlapping complex molecular cascades in ovarian tissue are subtly regulated during the numerous steps involved (i.e. from sex determination to ovulation) (Biason-Lauber and Chaboissier, 2015; Eggers et al., 2014; Hunter, 2015; Koopman, 2016; Matzuk and Burns, 2012; Matzuk and Lamb, 2008, 2002; Suzuki et al., 2015; Wood and Rajkovic, 2013) (and references therein). These genes are particularly interesting as their mutations might contribute towards POI aetiology.

Numerous genes have been analysed during the past two decades (especially by Sanger sequencing) to identify sequence variants related to POI. However, this task has been particularly challenging due to the high number of candidate genes and because of the limited genomic coverage of Sanger sequencing. Indeed, analyzing large genomic regions via direct sequencing involves numerous time-consuming PCR/sequencing experimental setup steps.

In an attempt to overcome this drawback, recent next generation sequencing (NGS) genomic screening approaches to family-based and non-related cases have identified new genes and mutations associated with specific female infertility disorders, such as POI (AlAsiri et al., 2015; Bouilly et al., 2016; Caburet et al., 2014, 2012; Carlosama et al., 2017; de Vries et al., 2014; Desai et al., 2017; Fauchereau et al., 2016; Fonseca et al., 2015; Guo et al., 2017; Patiño et al., 2017a; Qin et al., 2015a; Wood-Trageser et al., 2014; Bramble et al., 2016; França et al., 2017). NGS has certainly revolutionised medical genomics research as it has led to hundreds of monogenic and complex disease-related genes being described (McCarthy et al., 2013). NGS is currently being widely used for research and clinical purposes due to a concomitant decrease in its costs. However, data filtering and analysis is still challenging since it implies (especially for complex frequently-occurring pathologies) an in-depth knowledge of the molecular cascades regulating numerous physiological steps, as well as expertise in interpreting the potential pathogenicity conferred by particular mutations.

This review has thus been aimed at targeting the most relevant molecules involved in the diverse pathways playing key roles regarding ovarian tissue which may contribute towards POI. The main genes and proteins participating in bipotential gonad formation, sex determination, meiosis, folliculogenesis and ovulation are described to enable understanding how they may be considered putative candidates involved in POI. Considerations regarding NGS technical aspects such as design and data interpretation are mentioned. Successful NGS initiatives and future challenges are also discussed.

Taken together, the information presented here should enable a better understanding of the molecular origin of a frequently-occurring female disorder for clinicians and basic science researchers. It should also facilitate designing new diagnostic approaches and interpreting the molecular tests used in clinical environments.

2. NGS: design and data interpretation

During the past 10 years NGS has substantially contributed to the understanding of the molecular origin of rare and frequent diseases. This technique, which has evolved technically leading to improved levels of sensitivity, is available nowadays at affordable costs. All NGS platforms need DNA pre-processing to create a library for sequencing. Various steps are required: DNA fragmentation into specific sizes, generating blunt-ended DNA fragments, adapter ligation, PCR to increase library concentration and enrichment with specific capture probes (e.g. exons for exome sequencing). Varied NGS sequencing technologies are currently being offered by commercial brands, such as Illumina, Ion Torrent, Pacific Biosciences, Solid and Roche 454, each having advantages and pitfalls regarding genomic coverage of specific regions (repeated and GC rich, read sensitivity and costs) (Buermans and den Dunnen, 2014; Goodwin et al., 2016). Distinct NGS formats, mainly defined by the length of the genomic region which is to be sequenced, are available for research and diagnostic purposes. Three main formats are currently available: target sequencing microarrays (TSM) as panels of specifically selected regions, whole-exome sequencing (WES) which includes the entire genome's coding regions and whole-genome sequencing (WGS). TSM and WES sequencing are used for research and diagnosis while WES is reserved for investigative purposes due to the inclusion of extensive intron regions for which the functional nature remains largely unknown. One technically important NGS variable (named sequencing depth) is associated with the number of times that each nucleotide is read during a specific experiment. An appropriate read depth is essential for obtaining reliable results due to NGS' inherent nature which is based on resequencing DNA fragments susceptible to carrying unexpected variations caused by PCR amplification (Sims et al., 2014) (and references therein). For instance, it has been estimated that a 40X to 80X average depth is necessary to cover >90% of target genes for WES experiments and for high sensitivity genotyping of heterozygous SNPs (Clark et al., 2011; Meynert et al., 2014, 2013; Sims et al., 2014). Regarding diagnostic purposes, it has been proposed that a 120X depth in WES is suitable for identifying most non-synonymous variants causing disease (Kim et al., 2015). However, most research reports use PCR/direct sequencing to validate candidate sequence variants identified by NGS screenings as the average depth read ranges between 10X-60X. After confirmation by Sanger sequencing, variants are used to establish links with normal or pathogenic traits. Distinct pipelines and variables have been used to select variants underlying potential pathogenic effects, depending on a disease's genetic nature (mono/polygenic), type of inheritance (dominant, recessive) and disease presentation (familial/non-familial). Several bioinformatics methods, some of them involving automatic tools for directly analysing NGS data, have been used for predicting missense variants' potential pathogenic effect. Some of them (e.g.

Please cite this article in press as: Laissue, P., The molecular complexity of primary ovarian insufficiency aetiology and the use of massively parallel sequencing, Molecular and Cellular Endocrinology (2017), http://dx.doi.org/10.1016/j.mce.2017.07.021