Gene 542 (2014) 89–97

Contents lists available at ScienceDirect

Gene

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

Computational identification and analysis of functional polymorphisms involved in the activation and detoxification genes implicated in endometriosis

Roshni Panda, P.K. Suresh *

School of Biosciences and Technology, VIT University, Vellore, Tamil Nadu 632014, India

ARTICLE INFO

Article history:
Received 20 December 2013
Received in revised form 28 February 2014
Accepted 29 March 2014
Available online 1 April 2014

Keywords:
Endometriosis
Correlation
Epidemiology
SNP
Bioinformatics

ABSTRACT

Endometriosis is a complex disorder of the female reproductive system where endometrial tissue embeds and grows at extrauterine location leading to inflammation and pain. Hundreds of polymorphisms in several genes have been studied as probable risk factors of this debilitating disease. Bioinformatics tools have come a long way in augmenting the search for putative functional polymorphisms in human diseases. In this study we have explored 16 genes involved in the detoxification of xenobiotic chemicals that are implicated in endometriosis by utilising publically available programs like SIFT, Polyphen, Panther, FastSNP, SNPeffect and PhosSNP. The variations among different ethnic populations of the SNPs were studied. We then calculated the extent to which bioinformatics based predictions are concurrent with real world epidemiological, genotyping studies using a set of SNPs that have been studied in endometriosis case–control studies. Our study shows that there is a significant positive correlation ($r = 0.569$, $p < 0.005$) between the summary of the predicted scores taken from 4 different servers and the odds ratio found from epidemiological studies. This report has identified and catalogued various deleterious SNPs that could be important in endometriosis and could aid in further analysis by in vitro and in vivo methods for the better understanding of the disease pathophysiology.

1. Introduction

Single nucleotide polymorphisms (SNPs) are the most commonly occurring form of DNA variation in the genome that account for interindividual differences (Suh and Vijg, 2005). SNPs are single base pair positions in the genomic DNA at which different sequence alternatives (alleles) exist in normal individuals in some populations, wherein the least frequent allele has an abundance of 1% or greater (Brookes, 1999). Despite varying estimates, it is believed that the world population contains around 15 million SNP (H.S. Wang et al., 2012; M. Wang et al., 2012). SNP alleles are created either by transition (purine–purine or pyrimidine–pyrimidine substitution) or by transversion (purine–pyrimidine or pyrimidine–purine substitution). All of these occurrences have an almost equal likelihood of taking place at any given loci. However it is observed that, 70% of all SNPs found in human genome involve a C to T transition. Among these SNPs, nonsynonymous, missense single nucleotide polymorphism that leads to an amino acid change in their protein product are of particular interest with relevance to human diseases (Stenson et al., 2003). nsSNP resulting in amino acid substitution affects one or more roles of the concerned amino acid residue, such as protein stability and folding, protein expression, cellular localisation and post translational modification (Teng et al., 2008). SNPs can also act as associative bio-markers of diagnosis of complex diseases and serve as biomarkers of prognosis to determine drug response (Bader, 2001).

Endometriosis is characterised by the presence of endometrial tissue outside the endometrial cavity. It affects around 5–15% of women of child bearing age (Bellelis et al., 2011). Most common sites of endometriosis are the ovaries, peritoneum, utero sacral ligaments and Pouch of Douglas. Endometriosis remains an enigmatic disease with several theories regarding its etiopathogenesis, the most common being the reflux of menstrual debris and its subsequent attachment in the pelvic cavity.

Abbreviations: SNP, single nucleotide polymorphism; nsSNP, non-synonymous SNP; EDC, endocrine-disrupting chemical; PCDDs, polychlorinated dibenzo-p-dioxins; PCDFs, polychlorinated dibenzofurans; PCBs, polychlorinated biphenyls; PE, phthalate esters; BRCA1, breast cancer type 1 susceptibility protein; BRCA2, breast cancer type 2 susceptibility protein; *AHR*, aryl hydrocarbon receptor; *AHRR*, aryl-hydrocarbon receptor repressor; *ARNT*, aryl hydrocarbon receptor nuclear translocator; *CYP1A1*, cytochrome P450, family 1, subfamily A, polypeptide 1; *CYP1B1*, cytochrome P450, family 1, subfamily B, polypeptide 1; *CYP 2E1*, cytochrome P450, family 2, subfamily E, polypeptide 1; *EPHX1*, epoxide hydrolase 1, microsomal (xenobiotic); *GSTA1*, glutathione S-transferase alpha 1; *GSTM1*, glutathione S-transferase mu 1; *GSTP1*, glutathione S-transferase pi 1; *GSTT1*, glutathione S-transferase theta 1; *NAT1*, N-acetyltransferase 1; *NAT2*, N-acetyltransferase 2; *ESR1*, oestrogen receptor alpha; *ESR2*, oestrogen receptor beta; SIFT, sorting intolerant from tolerant; Polyphen, polymorphism phenotyping v2; Panther, Protein ANalysis THrough Evolutionary Relationships; Pmut, pathological mutation software SNPs3D; OR, odds ratio; dbSNP, databaseSNP.

* Corresponding author at: Division of Biomolecules and Genetics, School of Biosciences and Technology, VIT University, Vellore, Tamil Nadu 632014, India.
E-mail addresses: p.k.suresh@vit.ac.in, indian.ethos@gmail.com (S. P.K.).

http://dx.doi.org/10.1016/j.gene.2014.03.058
0378-1119/

Pelvic pain is the most common presenting symptom; other symptoms include dysmenorrhea (painful menstruation), back pain, dyspareunia (pain during intercourse), dyschezia (pain on defecation), and pain with micturition (Nnoaham et al., 2012).

Scientific evidence has shown that several environmentally occurring endocrine disrupting chemicals (EDCs) act to mimic the effect of oestrogen and disrupt the endocrine functions in human beings. Many of the EDCs are organochlorine compounds, including polychlorinated dibenzo-p-dioxins (PCDDs), polychlorinated dibenzofurans (PCDFs), polychlorinated biphenyls (PCBs) and phthalate esters (PE). Exposure to these chemicals might alter the tissue response to hormones and decrease the sensitivity of the endometrium to progesterone. Although the exact mechanism of action of these pollutants is not known, it is speculated that these endocrine disrupting chemicals, in conjunction with AHR complexes, can lead to the recruitment of pro inflammatory cytokines and help in the infiltration and adhesion of endometriotic cells in the extrauterine sites.

Over the years, several genetic polymorphisms in various candidate genes have been studied to be associated with endometriosis (Falconer et al., 2007). It has been postulated that gene encoding the detoxification of the xenobiotic chemicals like the EDCs could be associated as risk factors in endometriosis. The gene encoding for detoxification protects the body by metabolising and excreting the harmful pollutants. Therefore, it might be plausible that polymorphisms in the detoxification proteins could lead to impaired detoxification due to the absence of or diminished enzyme production (Guo, 2006). Currently, there appears to be a lack of consensus in the role of polymorphisms in the aetiology of endometriosis with several studies refuting and supporting the same set of polymorphisms in different patient sets, in part due to the paucity of experimental data and/or flawed experimental design. However, it is significant to point out that of all the candidate gene polymorphism studies conducted so far, the strongest evidence linking polymorphisms and endometriosis has arisen out of studies investigating phase II detoxification enzymes (Tempfer et al., 2009).

Currently, laparoscopy is the gold standard of diagnosis and management in endometriosis (Umaria and Olliff, 2001). A non-invasive or minimally-invasive test could detect the likelihood of a woman to suffer from endometriosis and this approach could prove invaluable in diagnosis, much like the BRCA1 and BRCA2 mutation testing in breast cancer patients.

The study focuses on understanding the effect of polymorphic variation in detoxification genes from an in silico standpoint. To the best of our knowledge, this study is the first of its kind, to explore the functional significance of all the major polymorphisms in 16 genes that associated with endometriosis and detoxification. Major genetic polymorphisms were studied in order to dissect the complex, multifactorial disease that is endometriosis. Differences in various ethnic populations, in terms of the chosen genetic polymorphisms, were also studied to identify geographical variations. Development of biomedical tools to aid in SNP functionality analysis has grown exponentially in the past few years. Several servers and tools now exist for the prioritisation of SNP functionality. Hence, it will be an interesting exercise to observe how far the bioinformatics predictions are concurrent with the real world epidemiological findings (Zhu et al., 2008). With this in mind, we wanted to test the hypothesis that SNPs predicted to have a functional significance by in silico methods might be responsible in increasing the risk of endometriosis, as observed by the odds ratio in clinical case–control studies. Hence, this study was undertaken to systematically analyse the effect of nsSNPs in genes implicated in endometriosis and to investigate the efficacy of the prediction servers. Further, this work may provide a systematic and logical approach for the evaluation of the utility of the cost-effective in silico prediction methods used.

2. Materials and methods

2.1. Information acquisition

The genes selected for this study were *AHR*, *AHRR*, *ARNT*, *CYP1A1*, *CYP1B1*, *CYP2C19*, *CYP2E1*, *EPHX1*, *GSTA1*, *GSTM1*, *GSTT1*, *GSTP1*, *ESR1*, *ESR2*, *NAT1* and *NAT2*. The aforementioned genes were selected based on the strength of evidence generated from studies involving detoxification genes and endometriosis. The only consistently reported evidence linking endometriosis with SNPs has come from genetic polymorphism studies involved in the detoxification machinery. The SNPs for all the genes were retrieved from the dbSNP database available at the National Centre for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/SNP). SNP analysis was performed for the nonsynonymous SNPs (nsSNPs) from the coding region and the SNPs from the 5′ and 3′ regulatory regions. The nonsynonymous SNPs were analysed using SIFT (sorting intolerant from tolerant) (Kumar et al., 2009), found at http://sift.jcvi.org/www/SIFT_dbSNP.html. SIFT is based on a position specific scoring matrix (PSSM) which calculates the probability of a mutation based on DNA sequence homology-based methods. SIFT predicts the nature of the substitution as "deleterious" if it affects protein structure and "tolerated" if it does not. The amino acid substitution is predicted as being damaging if the score is ≤ 0.05, and tolerated if the score is >0.05. Polyphen (Adzhubei et al., 2013; http://genetics.bwh.harvard.edu/pph2) was also used to predict the functional consequence of the nsSNP. Polyphen uses several features that are sequence, evolutionary and structure-based to predict the functional significance of an SNP. It classifies the SNP as "benign", "possibly damaging" and "probably damaging".

Polyphen differs from SIFT in that it predicts the functional effects of mutations based on the posterior probability that a given mutation be

Table 1
Distribution of SNPs in the selected genes.

Gene name	Number of missense variation	Number of SNP in 5′ UTR region	Number of SNP in 3′ UTR region
AHR (aryl hydrocarbon receptor)	192	18	113
AHRR (aryl-hydrocarbon receptor repressor)	364	60	114
ARNT (aryl hydrocarbon receptor nuclear translocator)	342	34	193
CYP1A1 (cytochrome P450, family 1, subfamily A, polypeptide 1)	643	46	155
CYP1B1 (cytochrome P450, family 1, subfamily B, polypeptide 1)	202	19	78
CYP2C19 (cytochrome P450, family 2, subfamily C, polypeptide 19)	137	02	08
CYP2E1 (cytochrome P450, Family 2, subfamily E, polypeptide 1)	315	19	65
EPHX1 (epoxide hydrolase 1, microsomal (xenobiotic))	190	20	16
GSTA1 (glutathione S-transferase alpha 1)	46	07	08
GSTM1 (glutathione S-transferase mu 1)	102	38	17
GSTP1 (glutathione S-transferase Pi 1)	73	10	17
GSTT1 (glutathione S-transferase theta 1)	90	24	121
NAT1 (N-acetyltransferase 1)	394	47	273
NAT2 (N-acetyltransferase 2)	103	17	08
ESR1 (oestrogen receptor alpha)	561	88	477
ESR2 (oestrogen receptor beta)	1093	83	231