Genomic and post-genomic approaches to polycystic ovary syndrome—progress so far: Mini Review

Catherine Hughes¹, Mustafa Elgasim¹, Robert Layfield²,³ and William Atiomo¹

¹Department of Obstetrics and Gynaecology, School of Human Development and ²School of Biomedical Sciences, University of Nottingham, Nottingham, UK
³To whom correspondence should be addressed at: School of Biomedical Sciences, University of Nottingham Medical School, Queen’s Medical Centre, Nottingham NG7 2UH, UK. E-mail: robert.layfield@nottingham.ac.uk

Genomic studies in polycystic ovary syndrome (PCOS) have focused on ovarian tissues and gene expression changes related to the gynaecological manifestations of PCOS. These studies have revealed a variety of altered genes that fall into many functional categories. Of these, the genes involved in steroidogenesis, including genes related to retinoic acid biosynthesis and LH-stimulated gene pathways, are generally up-regulated in PCOS samples. Genes involved in the Wnt signalling pathway appear down-regulated. Immune response genes and those involved in apoptosis are altered, but the net effect of these alterations is unclear at present. However, these altered gene expression patterns are yet to produce a defined aetiological basis or diagnostic biomarker for PCOS. The use of proteomic technologies for the study of the PCOS proteome is in its infancy; however, a few pilot studies have been published and the data are reviewed. Proteomics looks directly at the functional units within a cell, the proteins. This approach should thus serve to validate some of the gene expression changes identified and then build on the genomic results collected to date.

Key words: genomics/metabonomics/ovary/polycystic/proteomics

Introduction

Polycystic ovary syndrome (PCOS) is a heterogeneous disorder affecting 5–10% of reproductive aged women in the UK (Asuncion et al., 2000). The syndrome was first recognized by Stein and Leventhal (1935) in a small group of women presenting with amenorrhoea, hirsutism, obesity and histological evidence of polycystic ovaries, and yet 70 years on, the aetiological mechanisms remain unknown. Phenotypic manifestations of the syndrome vary from patient to patient, with many of the women suffering reproductive aberrations including anovulatory sub/infertility. The current diagnostic consensus for PCOS was established at a workshop (The Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group, 2004). It is a diagnosis of the exclusion of other androgenic-based diseases following clinical identification of at least two of three defined criteria: oligo/anovulation, clinical and/or biochemical evidence of hyperandrogenaemia and the ultrasonographic findings of polycystic ovaries. Subsequent treatment strategies vary depending on the prevalent symptom(s) and the patient’s desired outcome, tending towards symptom-based interventions for gynaecological and/or cosmetic improvements or simple lifestyle changes and weight management to increase fertility and reduce insulin resistance. It is clear that more definitive diagnostic testing and better understanding of the heterogeneity in the PCOS aetiology leading to improved therapeutic intervention would be invaluable to the medical community and women with PCOS alike.

Although the aetiology of PCOS remains unknown, it is thought that the hyperandrogenism underpins all other manifestations of the syndrome (Ehrmann, 2005). Current theories based on hypothesis-driven studies suggest that this stems from hypothalamic changes in GnRH secretion, creating aberrations in the pituitary release of LH, which in turn instigates the over-production of androgens by the ovaries in PCOS women. Nonetheless, the mechanisms and cause/effect nature of these alterations in the hypothalamic-pituitary-ovarian axis remain unclear (Doi et al., 2005). Today, the complexity of the syndrome has also widened to encompass metabolic and cardiovascular problems in addition to the original gynaecological and endocrinological ones. Insulin resistance, particularly in skeletal muscle and adipose tissue with sensitivity in ovarian tissue, affects up to 70% of women with PCOS and is a risk factor in these women for developing type 2 diabetes. Resulting high serum insulin levels can then act synergistically with LH to stimulate ovarian theca-cell androgen production and inhibition of sex hormone-binding globulin (SHBG) synthesis, intensifying the already high free androgen index in these women. An altered lipid profile and alterations in other cardiovascular disease markers such as C-reactive protein and homocysteine suggest PCOS could be a risk factor, independent of, although exacerbated by, PCOS-related obesity and diabetes, for developing cardiovascular disease (Carmina et al., 2005). Potential links between PCOS and endometrial (Hardiman et al., 2000). The syndrome was first recognized by Stein and Leventhal (1935) in a small group of women presenting with amenorrhoea, hirsutism, obesity and histological evidence of polycystic ovaries, and yet 70 years on, the aetiological mechanisms remain unknown. Phenotypic manifestations of the syndrome vary from patient to patient, with many of the women suffering reproductive aberrations including anovulatory sub/infertility. The current diagnostic consensus for PCOS was established at a workshop (The Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group, 2004). It is a diagnosis of the exclusion of other androgenic-based diseases following clinical identification of at least two of three defined criteria: oligo/anovulation, clinical and/or biochemical evidence of hyperandrogenaemia and the ultrasonographic findings of polycystic ovaries. Subsequent treatment strategies vary depending on the prevalent symptom(s) and the patient’s desired outcome, tending towards symptom-based interventions for gynaecological and/or cosmetic improvements or simple lifestyle changes and weight management to increase fertility and reduce insulin resistance. It is clear that more definitive diagnostic testing and better understanding of the heterogeneity in the PCOS aetiology leading to improved therapeutic intervention would be invaluable to the medical community and women with PCOS alike.

Although the aetiology of PCOS remains unknown, it is thought that the hyperandrogenism underpins all other manifestations of the syndrome (Ehrmann, 2005). Current theories based on hypothesis-driven studies suggest that this stems from hypothalamic changes in GnRH secretion, creating aberrations in the pituitary release of LH, which in turn instigates the over-production of androgens by the ovaries in PCOS women. Nonetheless, the mechanisms and cause/effect nature of these alterations in the hypothalamic-pituitary-ovarian axis remain unclear (Doi et al., 2005). Today, the complexity of the syndrome has also widened to encompass metabolic and cardiovascular problems in addition to the original gynaecological and endocrinological ones. Insulin resistance, particularly in skeletal muscle and adipose tissue with sensitivity in ovarian tissue, affects up to 70% of women with PCOS and is a risk factor in these women for developing type 2 diabetes. Resultant high serum insulin levels can then act synergistically with LH to stimulate ovarian theca-cell androgen production and inhibition of sex hormone-binding globulin (SHBG) synthesis, intensifying the already high free androgen index in these women. An altered lipid profile and alterations in other cardiovascular disease markers such as C-reactive protein and homocysteine suggest PCOS could be a risk factor, independent of, although exacerbated by, PCOS-related obesity and diabetes, for developing cardiovascular disease (Carmina et al., 2005). Potential links between PCOS and endometrial (Hardiman...
et al., 2003) and ovarian (Schildkraut et al., 1996) cancers have been made, but as with cardiovascular disease, the evidence for PCOS as an independent risk factor for these pathologies remains controversial (Gadducci et al., 2005).

Discovery-driven genomic techniques such as linkage analysis and DNA microarrays for global gene expression profiling have recently begun to be used in PCOS and other complex diseases such as multiple sclerosis (Martin, 2004; Robinson et al., 2005). The aim of these studies was to define potential candidate genes for more detailed hypothesis-driven research. This article will review, in turn, the various genomic as well as proteomic and metabonomic technologies currently being used in the quest to find definitive diagnostic and aetiological definitions for PCOS. We will outline the common molecular pathways identified in such studies, focusing on the potential clinical implications these may have in PCOS. Methodological challenges will be commented on, and the need to exploit the findings for future therapeutic advantage will also be addressed. In conclusion, this article argues for post-genomic research initiatives to support and build on genomic profiling to obtain a comprehensive representation of disease aetiology at a molecular level. Even though the proteomic and metabonomic studies of PCOS are in their infancy, they are expected to generate an extensive pool of novel biomarkers for the disease.

The genetic basis of PCOS

Evidence of familial clustering has also been shown in PCOS (Amato and Simpson, 2005), suggesting a genetic mode of inheritance. However, no single mode of inheritance has been revealed. Some PCOS-associated genes known to be involved in metabolic or androgen biosynthetic pathways have been identified. For example, the ghrelin gene, previously shown to be involved in appetite suppression, was found to be expressed at lower levels in women with PCOS compared with healthy control women (Schofl et al., 2005). Numerous genetic polymorphisms associated with PCOS have also been identified, including a 5′-UTR repeat polymorphism in the steroid synthesis gene, CYP11A (Gharani et al., 1997; Wu et al., 2005) that is positively associated with PCOS and serum testosterone levels, and an extended polymorphic trinucleotide repeat in exon 1 of the androgen receptor that is associated with infertility in PCOS women compared with fertile PCOS women and healthy controls (Hickey et al., 2002). The insulin signalling pathway is another area of intense genetic research. Linkage analysis identified PCOS associations with a variable number of tandem repeats (VNTRs) in the 5′ region of the insulin gene (Waterworth et al., 1997). Other studies revealed a microsatellite marker near the insulin receptor gene on chromosome 19p13.3 that was significantly associated with PCOS (Urbanek et al., 2000; Tucci et al., 2001) and PCOS associations with specific genetic variants of the insulin receptor substrate 1 (Dilek et al., 2005).

The spectrum of findings may be accounted for, in part, by the lack of uniform diagnoses between study groups, small study sizes and the lack of an identified male phenotype. However, it is clear that PCOS is not a single-gene disorder that would be inherited in a predictable Mendelian fashion. It is most likely a complex multigenic disease with epigenetic and environmental factors playing a significant role in the phenotypic expression of disease symptoms. It is hoped that the ability to profile gene and protein expression on a global scale in PCOS tissues will begin to unravel the molecular mechanisms underlying the syndrome and also generate useful diagnostic biomarkers for use in a clinical setting. This is where the research tools of genomic and proteomic technologies are most valuable, and already studies are identifying potential gene networks and molecular pathways that may be important contributors to the PCOS aetiology (see later). The findings from these profiling studies can then form the basis of more detailed analysis of the molecular mechanisms of PCOS, as shown by the recent publication of an investigation into the increased GATA-6 mRNA levels in PCOS theca cells (Ho et al., 2005) identified in earlier genomic profiling by this research group.

Genomic approaches to PCOS

The publication of the first draft of the human genome sequence (International Human Genome Sequencing Consortium, 2001) heralded the true beginnings of genome-wide genetic exploration. The field of PCOS research was no exception to this; many groups took advantage of the new gene sequence availability to undertake hypothesis-free analysis of the PCOS genome in the search for genes of interest to drive further research initiatives.

DNA microarrays

The genomic studies of PCOS to date have focused on ovarian tissues and gene expression changes relating to the gynaecological manifestations of the disease, with DNA microarrays using GeneChip technology being the most common method employed. This technique facilitates high-throughput screening of mRNA transcripts within a given tissue or cell type, based on designated gene sequences identified in the human genome project and earlier genetic studies.

The first study was published by Wood et al. (2003), looking at theca cells from size-matched follicles in PCOS women and healthy controls. The study was small, only five PCOS women and four controls were involved, but it produced reproducible and statistically significant results that were later validated by quantitative real-time RT–PCR (Q-PCR). Of the 244 genes with a statistically significant difference in gene expression between PCOS and normal theca-cell populations only four genes exhibited a greater than 5-fold difference, suggesting a relatively stable transcriptome, with only small fluctuations in gene expression levels in the PCOS patients compared with controls in this study. Functional classification of all the 244 differentially expressed genes (Figure 1) revealed the up-regulation of metabolic genes with the down-regulation of membrane receptors, cell adhesion genes and structural protein encoding genes. The expression of various genes encoding secreted factors, intracellular signalling proteins, gene expression/
protein processing factors and cell surface antigens showed both up- and down-regulations, indicative of altered cell fate and/or function. In particular, alterations in the retinoic acid biosynthesis and Wnt signalling pathways were identified. An increased conversion of retinol to all-trans-retinoic acid was shown to contribute to increased androgen production by PCOS theca cells. In addition, the up-regulation of Wnt antagonists and the down-regulation of the Wnt ligand indicated the inhibition of the Wnt signalling pathway, which is hypothesized to normally participate in folliculogenesis; thus, an inhibition may contribute to the PCOS phenotype. This study also noted an up-regulation of GATA-6 (Wood et al., 2004), a transcription factor involved in transcription of CYP11A, a finding that supports the earlier linkage analysis. The substantiation of this kind shows that the gene-specific technologies, such as linkage analysis, are not redundant and can be used subsequently for more in-depth analyses of microarray results. This has recently been substantiated in a follow-up study by Ho et al. (2005) looking at the mechanisms responsible for the increased levels of GATA-6 mRNA in PCOS theca cells.

Subsequent studies have analysed total ovarian tissue collected during surgical investigations. Diao et al. (2004) reported 119 incidences of altered gene expression in three PCOS ovaries compared with three normal controls; again, the data were validated by Q-PCR and by western blot to look for corresponding protein changes. The expression changes were related to changes in gene regulation, apoptosis and metabolism (Figure 2). The apoptosis data highlighted the overexpression of anti-apoptotic factors and the down-regulation of apoptosis in PCOS ovaries that may contribute to the ovarian polycystic appearance. Additionally, many of the altered metabolic genes were shown to be those with a role(s) in insulin signalling, suggesting that the metabolic features of PCOS extend to the ovaries as well as peripheral tissues. An alternative ovarian tissue study (Jansen et al., 2004) compared six PCOS women and 11 normal controls with the validation of the microarray data for five genes performed using Q-PCR. Extracellular matrix-related genes were reported as up-regulated in this study (Figure 3), a phenomenon that may be related to the morphological changes seen in PCOS ovaries. However, their analysis of the differentially expressed genes showed a general trend for gene down-regulation in PCOS samples, in particular genes known to be positive regulators of cell proliferation and gene expression, those involved in immune responses and various...
Genomics, proteomics, metabonomics and PCOS

Heat shock proteins. The lowered immunological activity was postulated to be the resultant of decreased periovulatory processes in the PCOS ovaries, whilst the decrease in cell proliferation and heat shock proteins were speculated to be indicative of the follicular atresia seen in PCOS. Concurrent with the study by Wood et al. (2003), differential expression of the Wnt signalling pathway components was reported although the net outcome of this dysregulation was unclear in this study with one pathway inhibitor (carboxypeptidase Z) up-regulated and another (dickkopf homologue 3) down-regulated. Conversely, the findings of Jansen et al. (2004) contradicted those of Diao et al. (2004), with a reported increase in the expression of apoptotic factors in PCOS ovarian tissue, supporting, they suggest, the follicular atresia hypothesis. This study also compared the PCOS ovarian profiles to that of long-term androgen-treated female-to-male transsexuals. The two profiles were very comparable, an observation that supports the case for hyperandrogenism as the underlying aetiological mechanism in PCOS.

The most recent studies of gene expression in PCOS ovaries looked specifically at follicular granulosa cells (Hu et al., 2004) and ovarian connective tissue (Oksjoki et al., 2005). The granulosa cell study revealed the up-regulation of 25 genes and the down-regulation of 21 genes in five PCOS cases compared with five controls. It is of note that all the women in this study were undergoing IVF treatment at the time of sample collection—the harsh hormone treatment regimen involved in this would have confounded the resultant ovarian gene expression profiles. Nonetheless, differential expression of genes involved in cellular metabolism of fatty acids, a process linked to steroidogenesis, and immune response genes were reported, supporting the earlier studies.

The ovarian connective tissue study revealed a great deal of heterogeneity between the eight PCOS samples when compared with two baseline control samples, illustrating on a genetic level the phenotypic heterogeneity seen in PCOS patients. The overall profiles (Figure 4) reported in this study showed the down-regulation in PCOS of genes involved in cell proliferation, cell adhesion and immune response. Up-regulated genes included secreted factors, such as growth factors and chemokines, and intracellular signalling genes including the retinoic acid receptor α and Wnt-5A/Wnt-13, in agreement with the earlier studies.

Figure 2. Functional classification of the 119 differentially expressed genes in ovarian cells from women with polycystic ovary syndrome (PCOS) compared with healthy control women, as identified by Diao et al. (2004). The classification is a modified version of the designated categories, based on biological function of the gene product, published in the original paper. The complementary table provides the examples of differentially expressed genes that are of potential relevance to PCOS aetiology or genes that are identified in other genomic studies (see comment column for details).

![Figure 2](https://academic.oup.com/humrep/article-abstract/21/11/2766/2939149)
Proteomic approaches to PCOS

From a clinical point of view, the most easily accessible patient sample is plasma or serum; however, these are devoid of genetic material and so unsuitable for genomic testing. On the contrary, plasma and serum have an extensive and rich proteome that is representative not only of events in the circulatory system but also of many organs within the body. Both pathogenic and disease response mechanisms can often be detected in patient plasma or serum samples owing to tissue leakage following cellular death or damage or alterations in the endocrine system during disease progression.

As the functional units within the cellular environment, proteins arguably provide what is currently the best representation of disease aetiology at a molecular level. Technologies for the analysis of proteomes for a given tissue or clinical sample have been developed and can now be targeted to look at specific subsets of the proteome based on generic protein properties without any pre-requisite knowledge of protein identity or biochemical characteristics. The comprehensive and complex nature of the plasma and serum proteomes is of great value for proteomic studies but can also bring analytical problems. Many disease-specific changes in the plasma proteomes will occur in low-abundance proteins that can be more than 10 orders of magnitude less abundant than albumin and some immunoglobulins (Anderson and Anderson, 2002), which account for >70% of the total protein in plasma. Although this presents a hurdle for proteomic analysis, it can be overcome, placing proteomic analysis ahead of genomic techniques in both functional representation of the molecular disease mechanisms and clinical application. The use of proteomic technologies for the study of PCOS proteome is in its infancy; however, a few pilot studies have recently been published, summarized in Table I.

Two-dimensional electrophoresis

Two-dimensional electrophoresis (2DE) is a commonly used research tool that facilitates high-throughput analysis of the expression of numerous protein species. Protein spots excised from the gels can then be directly identified by, for example, matrix-assisted laser-desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS). In addition, the 2DE techniques are amenable to pre-fractionation of protein samples and can be adapted to focus on a specific region of the proteome.

Figure 3. Functional classification of the 135 differentially expressed genes in ovarian cells from women with polycystic ovary syndrome (PCOS) compared with healthy control women, as identified by Jansen et al. (2004). The classification is a modified version of the tabulated categories, based on the biological function of the gene product, published in the original paper. The complementary table provides the examples of differentially expressed genes that are of potential relevance to PCOS aetiology or genes that are identified in other genomic studies (see comment column for details).
based on either protein hydrophobicity (during the first-dimension isoelectric focusing of the sample) or molecular weight (a second-dimension refinement). The technique itself has limited clinical application, owing to resource consumption and the technicalities of analysing multiple samples at once. However, it is a good foundation tool for identifying biomarkers to which conventional, protein-specific molecular techniques such as cell-based expression studies and interaction assays can be applied for defining aetiological disease mechanisms and improving therapeutics. The discovery of protein biomarkers can also initiate the commercial development of diagnostic and disease monitoring tools such as enzyme-linked immunosorbent assay (ELISA)-based protein assays.

The first 2DE analysis of PCOS tissues was published by Cortón et al. (2004). This study used intra-abdominal adipose samples from obese women with and without PCOS. A detailed optimization of the 2DE techniques was performed to accommodate the high levels of triglycerides in the samples.
and to include strong alkaline protein species in the analytical gels. The group was then able to demonstrate the successful identification of individual spots by MS; however, they did not look at differential expression in PCOS compared with control women for this publication.

Our group has also recently employed 2DE to analyse the PCOS proteome. Using serum samples, we optimized the 2DE protocol, adding a pre-focusing depletion step to remove the highly abundant albumin and immunoglobulin G proteins from the serum and produced well-resolved, reproducible gel images for analysis. In a study of five age- and BMI-matched PCOS and control sample pairs, we found a relatively stable PCOS proteome compared with controls with only one observed protein spot change when compared with the control proteome. However, this finding is yet to be validated. This stability is a phenomenon that was apparent to a lesser extent in the genomic studies, further corroborating the fact that post-translational protein processing can alter the relationship between the functional proteome and the initial transcriptomic data. The up-regulated protein spot in the PCOS samples was identified as an inactivating cleavage product of the complement component 4b, a finding that is consistent with the down-regulation of the immune response reported by the earlier genomic studies. We remain optimistic that future proteomic analysis of PCOS serum samples will reveal further biomarkers with clinical relevance.

Despite the limited number of 2DE analysis studies published in the field of PCOS, the future looks promising when one considers the fact that the same techniques have been used for biomarker identification in other complex diseases. A good example is that of sporadic Creutzfeldt–Jakob disease (sCJD). The 2DE analysis of cerebrospinal fluid (CSF) samples from sCJD cases (Harrington et al., 1986), previously only diagnosed post-mortem, revealed two protein species that were present only in sCJD and not in other neurodegenerative dementias or healthy controls. These proteins were identified as the 14-3-3 anti-apoptotic protein (Zerr et al., 1996) and cystatin C (Righetti et al., 2005). Simple western blot tests for the presence of 14-3-3 in CSF from living patients are now used to diagnose sCJD, although the validity of this test for use in the related transmissible spongiform encephalopathies of scrapie and bovine spongiform encephalopathy (BSE) is controversial.

**Liquid chromatography and MS**

One of the restrictions of 2DE is in the sensitivity of the technique: even with specific refinements to look at proteins in a specific pH range or by using a depletion step to remove the highly abundant albumin and immunoglobulin G from serum, some highly acidic, basic and low-abundance proteins will be missed. Very small peptides, especially many hormones, are also beyond the resolving power of 2D gels. This is exemplified by our study that did not pick up the well-documented alterations in many hormone levels such as LH and SHBG. Thus, complementary proteomic techniques are recommended to obtain a comprehensive overview of the proteome. As an alternative method to 2DE, some research groups have used liquid chromatography coupled with MS (LC-MS/MS) to obtain a proteomic profile for a given sample in other disease states, but this has not yet been applied in PCOS. For example, low-molecular-weight protein biomarkers have been identified (Yeo et al., 2004), using LC-MS/MS, in an animal model of asthma with implications for both diagnostic and prognostic application in the various types of allergic and non-allergic asthmatics. As with 2DE, LC-MS allows the multi-dimensional separation of the proteome. The 1D separation is a chromatographic separation of the sample based on biochemical properties, such as hydrophobicity, of the individual protein species through a liquid phase. The differing fractions are then subjected to TOF-MS separation of the component protein species based on their mass/charge (m/z) ratio. Resultant peaks of interest are identified by peptide mass sequencing using a second MS step, as is the case for 2DE protein identification. LC-MS/MS does have the capability to investigate protein species outside the normal range of 2DE analysis, including small peptides and highly basic or acidic protein species. It also maintains the proteins in solution; thus, MS characterization is of intact proteins, making it a valuable complementary technique.

A new system for high-throughput, 2D LC separation of protein species was recently launched (Simonian and Betgovarguez, 2003). The ProteomeLab PF 2D separates a biological sample, such as serum, by chromatofocusing (protein separation based on isoelectric point) followed by reversed-phase chromatography (protein separation based on hydrophobicity) to generate both qualitative and quantitative analysis of a proteome. The resultant liquid fractions again maintain the intact nature of the proteins for subsequent protein identification using MALDI-TOF-MS as with previous 2DE and LC-MS technologies. This new technology may, in the near future, come to serve as a welcome addition or alternative for research groups working on large proteomic analysis studies.

**Surface-enhanced laser desorption/ionization TOF-MS (SELDI-TOF-MS)**

The final tool in the proteomic collection is SELDI-TOF-MS. Again a 2D separation of the individual species within a proteome, this technique uses a first-dimension affinity-based fractionation of a sample on a ProteinChip. There are numerous chips available to capture different fractions of the proteome, based on either biochemical properties (e.g. hydrophobicity) or protein–protein interactions (e.g. antibody capture). The most commonly used chip in proteomic studies to date is the weak cation exchange (WCX2) surface. The 2D separation is then performed using laser irradiation and TOF-MS analysis to produce m/z peak profiles, which can be compared for two samples. Although this technique is expensive, it has been widely used in cancer research to generate tumour-specific, serum and plasma proteomic profiles for both diagnostic and disease progression monitoring (Seibert et al., 2005; Kong et al., 2006). The popularity of this technology was highlighted at the recent International Conference on Obstetrics and Gynaecology (ICOG, 2005); of the 10 abstracts covering proteomic studies of gynaecological conditions, seven had used SELDI-TOF-MS as their proteomic tool of choice. It is also encouraging that a SELDI study of sCJD (Sanchez et al., 2004), carried out independently of the 2DE
analysis discussed above, simultaneously identified cystatin C as a biomarker of sCJD.

The first full SELDI study of PCOS was published by Zhao et al. (2005a) and reported the identification of four differentially expressed mz/z peaks in the serum of 31 PCOS patients compared with healthy controls. Using the mz/z values for these peaks, researchers developed a model for PCOS diagnosis and showed that it had sensitivity, specificity and positive predictive values all >80% in subsequent testing. Nonetheless, the identification of the proteins corresponding to the differential peak values and the validation of these expression differences are needed to drive further studies into the aetiological mechanisms of PCOS. The ICOG (2005) also brought news of a second SELDI study in PCOS. The study by Zhao et al. (2005b) was an expansion of the first—this time looking at insulin-resistant PCOS women compared with non-insulin-resistant PCOS women and healthy control women. Distinct peaks profiles were identified for each group with specific changes seen not only when the two PCOS groups were individually compared with the control group but also between the two PCOS groups. This is the first experimental evidence of proteomic subsets existing in PCOS, a concept that will aid the development of targeted therapeutic interventions in PCOS women based on their individual molecular profile.

**Probing the metabonome**

Beyond proteomics, metabolic profiling may also help unravel the complexity of PCOS. Endogenous metabolites often exhibit disturbances in concentration or flux during disease progression in an attempt to maintain cellular homeostasis in the organism. These metabolites can be rapidly analysed in a quantitative manner using easily accessible biological samples such as blood serum. This up-and-coming metabolic research discipline is yet to be investigated in the field of PCOS. It is unlikely that the independent metabolomic studies of complex disorders will solve all the diagnostic problems or aetiological mechanisms involved; however, it may provide novel insights complementary to proteomic studies. A variety of analytical methods can be used for metabolomic studies, but the favoured technique to study biofluids and intact tissue samples is high-resolution 1H nuclear magnetic resonance (NMR) spectroscopy. With serum or plasma samples it is even possible to look at different metabolite spectra to cover both small-molecule and macromolecular profiles. Principal component analysis can then be used for disease-related pattern recognition. Metabonomic (and also SELDI) studies are highly sensitive to changes instigated in samples during the collection and storage/handling procedures (Banks et al., 2005). Thus, the standardization of a protocol is an important consideration when embarking on and analysing such a study. Despite this, the case for metabonomic studies in PCOS is strong, supported by the phenotypic evidence of changes in both metabolic and endocrine pathways. With this in mind, our group has recently initiated an NMR pilot study of the metabonome in serum samples from PCOS and control women. We hope to use any results from this investigation to support our proteomic data and to drive our future investigations.

**Conclusions**

The case for proteomic analysis as the true functional assessment of cellular activity and as the technology of choice for comprehensive and representative biomarker discovery in complex diseases, such as PCOS, is, even in these early days of investigations, persuasive. Owing to alternative RNA splicing, the differential rates of protein synthesis from individual mRNA species and post-translational protein modifications, the association between mRNA abundance analysed in genomic studies and the corresponding protein abundance is often weak. The rates of protein turnover and subcellular localization are also not reflected in a sample’s mRNA pool, further justifying the need for proteomic studies. Despite the current methodological challenges and the need for larger data sets to facilitate more robust data analysis, the next few years should potentially yield a range of validated protein biomarkers for definitive diagnosis of the syndrome in a clinical setting. There also exists the potential for proteomic classification of PCOS subsets to be established to facilitate targeted therapeutic intervention tailored to each individual’s molecular profile. Metabonomic profiling alongside peptidomic studies and the analysis of the microRNA-ome may prove useful, also allowing the incorporation of bioinformatics for statistical analysis of the data sets to look for trans-omic biomarker relationships. The identification of biomarkers will subsequently open the door to detailed studies for the elucidation of the molecular basis of PCOS and the improved therapeutics this will precede. We hope that this article will serve as a call for ‘omic’ integrations into a single systems biology discipline within biomedical research (Figure 5). Proteomics should not be taken as a stand-alone discipline but as the major player alongside genomic, metabolomic and other –omic disciplines in the bid to unravel the complexities of many multi-faceted diseases such as PCOS.

**Conflicts of interest**

The authors C. Hughes, R. Layfield and W. Atiomo have a GB patent application for a PCOS biomarker identified using proteomic techniques.

**References**


Submitted on March 4, 2006; resubmitted on May 8, 2006; accepted on May 15, 2006