Proteomic analysis of the proliferative and secretory phases of the human endometrium: Protein identification and differential protein expression

Leroi DeSouza\textsuperscript{1, 2}, Georg Diehl\textsuperscript{1}, Eric C.C. Yang\textsuperscript{1, 2}, Jingzhong Guo\textsuperscript{1}, Mary Joe Rodrigues\textsuperscript{3}, Alexander D. Romaschin\textsuperscript{4, 5}, Terence J. Colgan\textsuperscript{3, 5} and K.W. Michael Siu\textsuperscript{1}

\textsuperscript{1} Department of Chemistry and Centre for Research in Mass Spectrometry, and
\textsuperscript{2} Department of Biology, York University, Toronto, Ontario, Canada
\textsuperscript{3} Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada
\textsuperscript{4} Biochemistry, Toronto General Hospital, Toronto, Ontario, Canada
\textsuperscript{5} Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada

Proteomic analyses of the proliferative and secretory phases of the human endometrium were carried out to identify proteins and discover differentially expressed proteins using isotope-coded affinity tags, three stages of chromatographic separation and online tandem mass spectrometry (MS/MS). From an initial list of 346 proteins identified by ProICAT, manual inspection of MS/MS spectra and confirmatory searches pared the list down to 119 positively identified proteins. Only five of the proteins showed consistent differential expression. The utility of some of these proteins as indicators of true differential expression in the endometrium is open to discussion. The two proteins with unquestionable differential expressions in the secretory endometrium are: glutamate NMDA receptor subunit zeta 1 precursor and FRAT1. Some of the proteins that show no differential expression have previously been examined in gene-expression studies with similar conclusions.

Keywords: Differential protein expression / Isotope-coded affinity tag / Multidimensional liquid chromatography / Protein identification / Secretory and proliferative endometrium

1 Introduction

The endometrium is the mucosal lining of the uterus, the function of which is to provide a suitable site for implantation and development of a fertilized ovum. Underneath the endometrium is a thick muscular wall, the myometrium, which expands considerably to accommodate, protect and nourish the fetus, and is responsible for expulsion of the fetus during birth. The endometrium comprises three histologically distinctive layers: The deepest layer adjacent to the myometrium is the stratum basalis; this is the persisting germative layer that undergoes minimal changes during the menstrual cycle. The intermediate layer, stratum spongiosum, is characterized by the presence of a spongy stroma. The thinner, most superficial layer, is the stratum compactum. These last two layers undergo dramatic and characteristic changes during the menstrual cycle, culminating in menstrual sloughing, and are therefore often jointly referred to as the stratum functionalis.

Following menses, the endometrium undergoes major phenotypic changes characterized by two major phases. The proliferative phase begins after menses and terminates at
ovulation. During this phase, the simple tubular glands, stimulated by increasing levels of estrogen, increase in tortuosity as well as length, and the stromal amount increases. Shortly following ovulation, the glands and stroma move through an orderly process of maturation in a secretory phase, prompted and supported by the appearance of progesterone. The endometrium is receptive to embryonic implantation for approximately a period of five days in the midsecretory phase, coinciding with peaking of the progesterone level. Appropriate secretory development and glycogen-rich secretion of the endometrium are important to the successful implantation and nutrition of the fertilized ovum, and are critical to its survival. In the absence of any implanted fertilized ovum, menses occurs two weeks following ovulation [1].

Consistent failure of a fertilized ovum to implant and thrive in a post-ovulatory, secretory endometrium is recognized as one site-specific cause of female infertility. The precise reason(s) for such endometrial causes of infertility is incompletely understood, but may be due to intrinsic defects in endometrial development and receptivity, gamete factors, and/or immune defects. Protein expression profiling of physiologic proliferative and secretory endometria using quantitative proteomics offers a new opportunity to understand and measure the physiologic changes occurring within the endometrium. Such an understanding may provide new insights into secretory endometrial developmental defects that are responsible for infertility. Protein biomarkers of fertility may be useful clinically. Traditionally, differential expression studies rely on using gene expression profiling; one such study of endometrial tissue identified approximately 150 genes that are up-regulated and 380 genes that are down-regulated within the implantation window of the uterus [2]. Another common approach is the use of 2-DE, followed by densitometry for relative quantification, and subsequent identification of proteins by MS analysis. A study of breast cancer cell lines identified approximately 70 proteins; differential expression of a number of proteins was noted between a drug-resistant and a drug-sensitive line [3].

Comparing the two approaches, an obvious advantage of the latter is that any differential expression correlates directly with differences in protein levels. This circumvents the potential for misleading results, a factor inherent to gene expression-type studies where issues such as translational regulation are not addressed. This problem was clearly illustrated by a recent study which investigated the differential expression of proteins in two different human T-cell types [4]. The authors of that study used the 2-DE approach to detect 14 proteins which, when compared to a gene expression study performed by the same group on the same two cell types, showed only partial correlation between the results from the two methods.

The last few years have seen the development of a third approach in which the proteins from two samples to be compared are differentially labeled by tags differing in isotopic composition. The two samples are then combined and processed in a single batch, thereby allowing relative quantification to be performed through mass spectrometric analysis irrespective of protein losses in purification and separation. Differential labeling may be achieved by proteolytic labeling such as 18O-labeling of the C-terminus [5, 6] or derivatization with reagents such as α-C-1 N-acetoxy succinamide of the C-terminus [7] and 1-δ4-nicotinoyloxysuccinamide of the N-terminus [8]. A particularly effective labeling strategy in the form of an isotope-coded affinity tag (ICAT) targets the cysteine residue [9]. Cysteine is a relatively rare amino acid residue, but is present in the vast majority of proteins. Tryptic peptides that do not contain the cysteiny1 residue are eliminated in a subsequent clean-up step. Thus, in addition to enabling relative quantification, this tagging also reduces sample complexity. Other ICAT-type reagents that earmark different residues or functional groups have also been developed, including PhlAT [10] which is selective for phosphorylation. A drawback in some of the tagging reagents is that the tag may unduly perturb the peptide’s chemistry. Small differences in the retention times of labeled peptide pairs have been noted, which may confuse identification and quantification. In addition, the larger tags may dominate fragmentation of the labeled peptides in MS/MS and complicate peptide identification. These drawbacks have largely been overcome by the use of a second-generation ICAT reagent [11]. This reagent introduces weight differences by means of differing numbers of 12C and 13C, which minimizes differences in retention times, and eliminates potential MS/MS complications by means of incorporating an affinity tag that is cleaved prior to MS analysis. Herein we report results of a study that aims to discover proteins in the human endometrium, and especially proteins that are differentially expressed in the proliferative and secretory phases. Differences in protein expression levels will be highlighted and determined by the use of the cleavable ICAT reagent [11]. Proteins that are discovered will be compared with genes that have been examined for regulatory changes.

2 Materials and methods

2.1 Tissue samples

Endometrial tissues were retrieved from an in-house dedicated endometrial tissue bank. All tissues had been snap-frozen in liquid nitrogen within 15–20 min of devitalization at the time of hysterectomy. All tissue procurements were carried out after informed patient consents. The consent, procurement and banking procedures were approved by the Research Ethics Boards of York University, Mount Sinai Hospital, University Health Network, and North York General Hospital. In every case, the endometrium was classified as proliferative or secretory by a pathologist (Terence J. Cogan). The histological classification was verified by examination of a histopathologic section from the frozen tissue. The mirror-face of the residual block was used for proteomic
analysis. After addition of 1 mL of Hanks’ Balanced Salt Solution containing protease inhibitors (1 mM 4-(2-aminoethyl)benzenesulfonyl fluoride, 10 μM leupeptin, 1 μg/mL aprotinin and 1 μM pepstatin), the tissue sample was mechanically homogenized at 30,000 rpm using a Polytron PT 1300D handheld homogenizer (Brinkmann, Westbury, USA). The whole tissue homogenate contained not only endometrial epithelium, but supportive stroma, vessels as well as secretions. The homogenate was stored in aliquots at −80°C and/or submitted for proteomic analysis. Samples from six different individuals were selected for this study. Three of these tissues had been classified as proliferative endometria (randomly designated as PRO1, PRO2 and PRO3), and the other three as secretory endometria (randomly designated as SEC1, SEC2, SEC3). Figure 1 shows an example of the histologic appearances of (a) a proliferative (PRO2) and (b) a secretory (SEC2) endometrium. In both endometria, the stratum basalis is characterized by a denser stroma than the physiologic responsive stratum functionalis above. Across the top of the stratum functionalis is the surface epithelium, which lines the endometrial cavity. The proliferative endometrium shows small, coiled glands with lining columnar epithelium reaching to the surface. By contrast, the secretory endometrium is thicker, and contains more tortuous glands with intraluminal secretions. The endometria of both types of physiologic phases have abundant supportive stroma and vessels among the epithelial glands.

2.2 Chemicals

Solvents, chemicals (except otherwise noted), and Hanks’ Balanced Salt Solution were obtained from Sigma (Oakville, ON, Canada). All reagents and buffers for the cleavable ICAT sample preparation procedure were from Applied Biosystems (Foster City, CA, USA).

2.3 Sample preparation

After removal of cell debris by centrifugation, the total protein content for each of the six clarified homogenates was measured using a commercially available Bradford protein assay (Bio-Rad, Mississauga, ON, Canada). The ICAT sample preparation procedure was carried out according to the cleavable ICAT protocol (Applied Biosystems). One hundred micrograms of proteins was used per sample. The proteins were denatured with the denaturing buffer supplied with the ICAT reagent kit. Afterwards, disulfide bonds were cleaved by adding the reducing reagent supplied, which contained 50 mM tris-(carboxyethyl)phosphine, and boiling for 10 min. The proliferative samples were then labeled with the light ICAT reagent and the secretory samples with the heavy reagent by incubating with the respective ICAT label for 2 h at 37°C. The labeled PRO1 and SEC1 samples were combined to form ICAT sample A, PRO2 and SEC2 to form ICAT sample B, and PRO3 and SEC3 to form ICAT sample C. The final volume of each sample was 0.2 mL. Mixing of the labeled proliferative and secretory samples in pairs in this manner ensures that any protein or peptide losses during subsequent processing steps is the same for both samples in a pair. The sample pairs were then digested by incubating each pair with 0.2 mL of a 50 μg/mL trypsin solution at 37°C overnight. The resulting peptides were fractionated by means of strong cation exchange chromatography using an HP1050 LC system (Agilent, Palo Alto, CA, USA) with a 1.5 mL injection loop and a 2.1 mm internal diameter (ID) × 100 mm length PolyLC Polysulfoethyl A column packed with 5 μm beads with 300 Å pores (The Nest Group,

© 2005 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim www.proteomics-journal.de
Southborough, MA, USA). A 2.1 mm ID × 10 mm length guard column of the same material was plumbed upstream from the analytical column. Fractionation was effected by a binary mobile-phase gradient at a total flow rate of 0.2 mL/min. Eluent A consisted of a 10 mM KH$_2$PO$_4$ solution in 25% ACN and 75% deionized water acidified to a pH of 3.0 with phosphorus acid. Eluent B consisted of a 10 mM KH$_2$PO$_4$ and 350 mM KCl solution in 25% ACN and 75% deionized water acidified to a pH of 3.0 with phosphorus acid. The trypsinized ICAT sample pair (0.4 mL) was mixed with 2.0 mL of Eluent A. A 1.5 mL portion of the 2.4 mL sample was injected. Initially, the gradient comprised 100% Eluent A. At the 2nd min, the % Eluent B was changed linearly from 0 to 100% at the 58th min. The run was terminated after 60 min. A total of 30 fractions were collected, one every 2 min, using an SF-2120 Super Fraction Collector (Advancet MFS, Dublin, CA, USA). UV monitoring of the chromatographic eluent revealed abundant peptides in fractions 11–20. These 10 fractions were further purified by affinity chromatography according to the ICAT protocol recommended by Applied Biosystems. Eluted labeled peptides were treated with the Cleaving Reagent, which contains TFA in order to remove the biotin label, and resolved in a third stage of chromatographic separation using reverse-phase nanobore LC with online MS/MS.

2.4 Nanobore LC-MS/MS

The nanobore LC system was from LC Packings (Amsterdam, The Netherlands) and consisted of a Famos autosampler and an Ultimate Nano LC system. It was interfaced to an API QSTAR Pulsar QqTOF mass spectrometer (Applied Biosystems/MDS Scienx) using a Protana NanoES ion source (Protana Engineering, Odense, Denmark). The spray capillary was a PicoTip SilicaTip emitter with a 10 μm ID tip (New Objective, Woburn, MA, USA). The nanobore LC column was a 75 μm ID × 150 mm length reverse-phase PepMap C18 nano capillary column (LC Packings) packed with 3 μm beads with 100 Å pores. One microliter of sample was injected via the full-loop mode. Separation was performed using a binary mobile-phase gradient at a total flow rate of 200 nL/min. Eluent A consisted of 94.9% deionized water, 5.0% ACN and 0.1% formic acid (pH = 3). Eluent B consisted of 5.0% deionized water, 94.9% ACN and 0.1% formic acid. The following binary gradient was used:

Time (min) 0 5 125 135 157 160 163 190 % Eluent B 5 5 30 60 80 80 5 Stop

For nanospray, the source conditions were a curtain-gas setting of 20 and an ionspray voltage in the range of 1800–3800 V that was optimized daily. In the Q0 region, the instrument parameters were a declustering potential of 65 V and a focusing potential of 265 V. Nitrogen was used as the collision gas at a setting of CAD = 5 for both TOF-MS and MS/MS scans. All LC-MS/MS data were acquired in information-dependent acquisition (IDA) mode in Analyst QS SP5 with Bioanalyst Extension 1.1 (Applied Biosystems/MDS Scienx). A TOF-MS survey scan with an m/z range of 400–1500 and 1 s scan time was followed by two product ion scans with an m/z range of 70–2000 and 2 s scan time. The collision energy (CE) was automatically controlled by the IDA CE Parameters script. The switching criteria were set to ions greater than m/z = 400 and smaller than m/z = 1500 with a charge state of 2 to 5 and an abundance of ≥ 10 counts/s. Former target ions were excluded for 60 s and ions within a +4 Th window were ignored. In addition, the IDA Extensions II script was set to two repetitions before dynamic exclusion and to select a precursor ion nearest to a threshold of 15 count/s every 4 cycles. These settings ensured examination of not only high abundance ions, but low abundance ones as well. Each of the three sample pairs, A, B and C, was injected twice, thus yielding a total of six sets of runs. Data analysis was performed using ProICAT 1.0 SP3 software (Applied Biosystems/MDS Scienx). We used an initial setting of a confidence value of 75 and a score of 15. Relative quantification of proteins between the light and heavy labels was performed on the TOF-MS scans by calculating the relative areas under the series of peaks. Differential levels of expression were based upon measured protein expression ratios in proliferative versus secretory endometrium normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) expression and were considered significant when they exceeded three SDs from the mean expression of housekeeping proteins (actin, tubulin α-chain and triose phosphate isomerase). All expression ratios were based on a mean of two essays. The sample size calculation for determination of minimum numbers of tissue samples for ICAT analysis (N) was based upon the difference between two means, n = 8 (SD / precision)$^2$. Based on preliminary experiments, an SD value of 0.3 was observed with a target precision of 0.5, yielding a minimum sample size of three for each tissue group. The significance of differential protein expression between proliferative and secretory endometrium was evaluated using the Wilcoxon rank sum test.

3 Results

Figure 2 shows a typical nano LC/MS total ion chromatogram from fraction 16 of sample A. As a total of 10 fractions were analyzed per sample and each sample was injected twice, there were a total of 60 total ion chromatograms. The use of the IDA mode generated hundreds of MS/MS spectra per total ion chromatogram. Using the initial setting of a confidence value of 75 (ProICAT recommends a lower confidence value of 50) and a score of 15 in ProICAT 1.0 SP3 resulted in the identification of 346 nonredundant proteins. Manual inspection of the MS/MS spectra revealed that identification of some proteins that scored between 15 and 20 might be problematic. A random selection of 50 spectra from proteins scoring between 15 and 20 showed that 15 were an-
analyzed and sequenced correctly, 13 were probably correct, and 22 were incorrect. Manual inspection of the MS/MS spectra of the 145 proteins that scored 20 or higher also revealed occasional errors in assignment, but at a rate considerably lower than that between scores 15 and 20. Unfortunately, the use of ICAT-labeling results in a reduced opportunity for corroborating identification based on a second tryptic peptide from the same protein. Although the majority of human proteins do contain more than one cysteine residue per protein, some of these residues may be located within tryptic peptides that are too small or large and others may fall into peptides that have poor ionization efficiencies. Of the 346 proteins, only 23 were identified with more than one peptide; 15 of these 23 proteins were identified with two peptides.

Table 1 lists the 119 proteins that have been identified by ProICAT and verified by manual inspection of MS/MS data. Figure 3 shows the distributions of the proteins in the form of a pie chart. Of the 119 proteins, 15 were added after manual inspection of 50 randomly selected MS/MS spectra. If we had inspected the MS/MS spectra of the remaining 151 proteins, we would have expected to confirm the identifications of an additional ~45 proteins. This, however, was considered impractical, especially after it was known that none of the 15 additional proteins were observed in all three runs and would contribute towards knowledge of differential expression. Our results show that expression levels of the majority of proteins identified are not consistently different between the proliferative and secretory phase of the endometrial cycle. As expected, the majority of identified proteins fall under the metabolic/housekeeping and structural categories. The proteins classified under Others include antibodies. A small percentage of the proteins are viral or pathogenic in origin, reflecting the possibility of infection in the biopsied endometria. A few proteins have no known functions or were identified from cDNA matches; these are, therefore, classified as hypothetical.

To normalize small variations of protein amounts, relative quantification of the protein levels is normalized to the ratio of GAPDH. Gene expression studies involving different stages in the endometrial cycle have used GAPDH expression to normalize differential mRNA expressions [12–14]. Our assumption is that the constant mRNA level of GAPDH through the endometrial cycle translates to a constant protein level (no variation caused by translational controls). The GAPDH ratio of 1.5 ± 0.5 (one SD) before self-normalization in the six runs is in accordance with this assumption. The ratios of the proteins in the secretory/proliferative phase (heavy/light label) listed in Table 1 are the averages ± SDs of the peptides in all the identified runs. In the case of more abundant proteins, e.g. serum albumin and serotransferrin, individual peptides were detected in as many as 3–4 fractions. Less abundant proteins, however, were frequently detected only in 1–2 fractions (see Fig. 4). In cases of multiple fractions, multiple peptides, and/or multiple runs, the ratios listed in Table 1 are the averages and the SDs of all contributing elements. Twenty-four proteins show extreme differential expression in that only one labeled form was observable (these are noted as Singleton H or L). However, only five of these 24 proteins were observed in all three samples (and an additional one in two samples). All five proteins were identified by single tryptic peptides. The expression levels of three housekeeping proteins (actin, tubulin α-chain and triose phosphate isomerase) normalized against GAPDH were pooled to establish a mean expression level of 1.07–1.1 and an SD of 0.3. The SD of these expression levels
Table 1. Proteins identified and secretory/proliferative expression levels

<table>
<thead>
<tr>
<th>Accession no.</th>
<th>Description</th>
<th>Average H:L ratio</th>
<th>SD</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>143T_HUMAN (P27348)</td>
<td>14–3-3 protein tau (14–3-3 protein theta)</td>
<td>0.81</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>RS12_ORENI (O13019)</td>
<td>40S ribosomal protein S12.</td>
<td>0.52</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>RS21_HUMAN (P35265)</td>
<td>40S ribosomal protein S21.</td>
<td>NQ</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>HPPD_MYCGR (O42764)</td>
<td>4-hydroxyphenylpyruvate dioxygenase (EC 1.13.11.27)</td>
<td>Singleton L</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>RL36_VIBCH (P78001)</td>
<td>50S ribosomal protein L36.</td>
<td>0.84</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ACTA_HUMAN (P03996)</td>
<td>Actin, aortic smooth muscle or Actin alpha skeletal muscle</td>
<td>1.86</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ACTB_HUMAN (P02570)</td>
<td>Actin, cytoplasmic</td>
<td>0.93</td>
<td>0.04</td>
<td>3</td>
</tr>
<tr>
<td>SAH1_XENLA (P1893)</td>
<td>Adenosylhomocysteine synthase (EC 3.3.1.1)</td>
<td>0.70</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>AFAH_HUMAN (P43652)</td>
<td>Afamin precursor (Alpha-albumin) (Alpha-Alb).</td>
<td>NQ</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ADHL_HUMAN (P11766)</td>
<td>Alcohol dehydrogenase class III L chain</td>
<td>0.67</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>DHA1_HUMAN (P00352)</td>
<td>Alpha-2-HS-glycoprotein precursor (Fetuin-A)</td>
<td>NQ</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>RS21_HUMAN (P35265)</td>
<td>40S ribosomal protein S21.</td>
<td>0.52</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ACTA_HUMAN (P03996)</td>
<td>Actin, aortic smooth muscle or Actin alpha skeletal muscle</td>
<td>1.86</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ACTB_HUMAN (P02570)</td>
<td>Actin, cytoplasmic</td>
<td>0.93</td>
<td>0.04</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 1. Continued

<table>
<thead>
<tr>
<th>Accession no.</th>
<th>Description</th>
<th>Average H:L ratio</th>
<th>SD</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS70_HUMAN (08107)</td>
<td>Heat shock cognate 71 kDa protein.</td>
<td>1.3</td>
<td>1.1</td>
<td>3</td>
</tr>
<tr>
<td>HS9A_HUMAN (P07900)</td>
<td>Heat Shock Protein HSP90-alpha</td>
<td>NQ</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>HBB_HUMAN (P02023)</td>
<td>Hemoglobin beta chain.</td>
<td>0.60</td>
<td>0.45</td>
<td>3</td>
</tr>
<tr>
<td>HEMO_HUMAN (P02790)</td>
<td>Hemopexin precursor (Beta-1B-glycoprotein).</td>
<td>1.1</td>
<td>0.3</td>
<td>3</td>
</tr>
<tr>
<td>ROD_HUMAN (Q14103)</td>
<td>Heterogeneous nuclear ribonucleoprotein D0 (hnRNPD0)</td>
<td>1.0</td>
<td>0.2</td>
<td>3</td>
</tr>
<tr>
<td>ROK_HUMAN (Q07244)</td>
<td>Heterogeneous nuclear ribonucleoprotein K (hnRNPK)</td>
<td>0.68</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>HNT1_HUMAN (P49773)</td>
<td>Histidine triad nucleotide binding protein 1</td>
<td>1.4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>YF48_HUMAN (Q9HC77)</td>
<td>Hypothetical protein KIAA1548</td>
<td>0.85</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>YB85_MYCTU (Q10546)</td>
<td>Hypothetical protein Rv0885.</td>
<td>NQ</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GC1_HUMAN (P01857)</td>
<td>Ig gamma-1 chain C region.</td>
<td>0.76</td>
<td>0.23</td>
<td>3</td>
</tr>
<tr>
<td>GC2_HUMAN (P01859) or GC4_HUMAN (P01861)</td>
<td>Ig gamma-2 chain C region.</td>
<td>0.85</td>
<td>0.22</td>
<td>3</td>
</tr>
<tr>
<td>HNT1_HUMAN (P49773)</td>
<td>Histidine triad nucleotide binding protein 1</td>
<td>1.4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>YF48_HUMAN (Q9HC77)</td>
<td>Hypothetical protein KIAA1548</td>
<td>0.85</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>YB85_MYCTU (Q10546)</td>
<td>Hypothetical protein Rv0885.</td>
<td>NQ</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GC1_HUMAN (P01857)</td>
<td>Ig gamma-1 chain C region.</td>
<td>0.76</td>
<td>0.23</td>
<td>3</td>
</tr>
<tr>
<td>GC2_HUMAN (P01859) or GC4_HUMAN (P01861)</td>
<td>Ig gamma-2 chain C region.</td>
<td>0.85</td>
<td>0.22</td>
<td>3</td>
</tr>
<tr>
<td>KAC_HUMAN (P01834)</td>
<td>Ig kappa chain C region.</td>
<td>0.88</td>
<td>0.37</td>
<td>3</td>
</tr>
<tr>
<td>HUMAN (P01876) or (P01877)</td>
<td>Ig alpha-1 chain C region or Ig alpha-2 chain C region</td>
<td>1.4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IDI2_RICCNc) (Q92HM7)</td>
<td>Isopentenyl-diphosphate delta-isomerase (EC 5.3.3.22).</td>
<td>1.0</td>
<td>0.3</td>
<td>3</td>
</tr>
<tr>
<td>LIT_HUMAN (P00338)</td>
<td>Lactate dehydrogenase (EC 1.1.1.27) (LDH).</td>
<td>1.4</td>
<td>0.3</td>
<td>3</td>
</tr>
<tr>
<td>MIF_HUMAN (P14174)</td>
<td>Macrophage migration inhibitory factor (MIF)</td>
<td>1.2</td>
<td>0.7</td>
<td>3</td>
</tr>
<tr>
<td>MSRE_MOUSE (P30204)</td>
<td>Macrophage scavenger receptor types I and II</td>
<td>1.1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MDHM_HUMAN (P04926)</td>
<td>Malate dehydrogenase, mitochondrial precursor</td>
<td>0.88</td>
<td>0.37</td>
<td>3</td>
</tr>
<tr>
<td>MSP1_PLA F3 (P19598)</td>
<td>Merozoite surface protein 1 precursor</td>
<td>1.0</td>
<td>0.2</td>
<td>2</td>
</tr>
<tr>
<td>M51A_HUMAN (P59088)</td>
<td>Mucin 5AC (Mucin 5 subtype AC, tracheobronchial)</td>
<td>0.85</td>
<td>0.22</td>
<td>3</td>
</tr>
<tr>
<td>MYH9_HUMAN (P35579) or MYH1A_HUMAN (P35580)</td>
<td>Myosin heavy chain non muscle type A or B</td>
<td>0.29</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>KML2_RABIT (P07313)</td>
<td>Myosin light chain kinase 2</td>
<td>Singleton H</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>NIR_NEUCR (P08681)</td>
<td>Nitrite reductase [NAD(P)H] (EC 1.7.1.4).</td>
<td>1.3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>NOG3_BRARE (Q9YHV3)</td>
<td>Noggin 3 precursor.</td>
<td>1.3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>NDHA_HUMAN (P15531) or NDKB_HUMAN (P22392)</td>
<td>Nucleotide diphosphate kinase A or B</td>
<td>1.2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CYP_HUMAN (P00589)</td>
<td>Peptidyl-prolyl cis-trans isomerase A (EC 5.2.1.8)</td>
<td>1.2</td>
<td>0.2</td>
<td>3</td>
</tr>
<tr>
<td>PEBB_BOVIN (P13696)</td>
<td>Phosphatidylethanolamine-binding protein (PEBP)</td>
<td>0.99</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>PGK_HUMAN (P00559)</td>
<td>Phosphoglycerate kinase (EC 2.7.1.1)</td>
<td>0.88</td>
<td>0.37</td>
<td>3</td>
</tr>
<tr>
<td>PMG1_HUMAN (P18669)</td>
<td>Phosphoglycerate mutase 1 (EC 5.4.2.1)</td>
<td>0.73</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>POL_MSG Charge 1 and 2/3 complex 34 kDa subunit (P34-ARC).</td>
<td>3.5</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>PIRG_HUMAN (P01833)</td>
<td>Probable ARP2/3 complex 34 kDa subunit (P34-ARC).</td>
<td>3.5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AR34_DROME (Q9YIM5)</td>
<td>Probable ARP2/3 complex 34 kDa subunit (P34-ARC).</td>
<td>3.5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AMPA_CAMJ E (P99P04)</td>
<td>Probable cytosol aminopeptidase (EC 3.4.11.1)</td>
<td>0.93</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>PRO1_HUMAN (P07737)</td>
<td>Profilin I</td>
<td>1.1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>PROP_HUMAN (P27918)</td>
<td>Properdin precursor (Factor P).</td>
<td>1.2</td>
<td>0.2</td>
<td>3</td>
</tr>
<tr>
<td>PRTH_PORGI (P46071)</td>
<td>Prothrombin (EC 3.4.22.-).</td>
<td>2.1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>PSA5_HUMAN (P28066)</td>
<td>Proteosome subunit alpha type 5 (EC 3.4.25.1)</td>
<td>0.93</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>VN02_VACC (P20641)</td>
<td>Protein N2.</td>
<td>1.1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FRT1_HUMAN (Q92387)</td>
<td>Proto-oncogene FRAT1</td>
<td>1.1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>O2C_DROME (P81911)</td>
<td>Putative odorant receptor 22c.</td>
<td>0.90</td>
<td>0.31</td>
<td>3</td>
</tr>
<tr>
<td>KPY1_HUMAN (P14618)</td>
<td>Pyruvate kinase, M1 isozyme (EC 2.7.1.40)</td>
<td>1.2</td>
<td>0.4</td>
<td>3</td>
</tr>
<tr>
<td>CUT1_SCHPO (P18296)</td>
<td>Separin (EC 3.4.22.-).</td>
<td>NQ</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>TRFE_HUMAN (P02787)</td>
<td>Serotransferrin precursor (Transferrin)</td>
<td>1.3</td>
<td>0.4</td>
<td>3</td>
</tr>
<tr>
<td>ALBU_HUMAN (P02768)</td>
<td>Serum albumin precursor.</td>
<td>1.1</td>
<td>0.3</td>
<td>3</td>
</tr>
<tr>
<td>STC_DROME (P40798)</td>
<td>Shuttle craft protein.</td>
<td>1.1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>SL1_HUMAN (Q13642)</td>
<td>Skeletal muscle LIM protein 1</td>
<td>2.0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>PCP1_SCHPO (P92351)</td>
<td>Spindle pole body protein pcp1.</td>
<td>Singleton H</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>SODC_HUMAN (P00441)</td>
<td>Superoxide dismutase [Cu-Zn]</td>
<td>0.32</td>
<td>0.33</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 1. Continued

<table>
<thead>
<tr>
<th>Accession no.</th>
<th>Description</th>
<th>Average H:L ratio</th>
<th>SD</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>THIO_HUMAN (P10599)</td>
<td>Thioredoxin.</td>
<td>0.87</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>TIPS_HUMAN (P00938)</td>
<td>Triosephosphate isomerase (EC 5.3.1.1) (TIM).</td>
<td>1.1</td>
<td>0.5</td>
<td>3</td>
</tr>
<tr>
<td>TPM1_HUMAN (P09493)</td>
<td>Tropomyosin 1 alpha chain (Alpha-tropomyosin).</td>
<td>2.3</td>
<td>0.7</td>
<td>2</td>
</tr>
<tr>
<td>TPM4_HUMAN (P07226) or TPM2_HUMAN (P07951)</td>
<td>Tropomyosin alpha 4 chain or beta 2 chain</td>
<td>0.75</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>TBA_HUMAN (P05209)</td>
<td>Tubulin alpha chain.</td>
<td>1.2</td>
<td>0.5</td>
<td>3</td>
</tr>
<tr>
<td>TBB_HUMAN (P07437 or P05217)</td>
<td>Tubulin beta chain (Fragment).</td>
<td>0.77</td>
<td>0.25</td>
<td>3</td>
</tr>
<tr>
<td>PTK7_HUMAN (Q13308)</td>
<td>Tyrosine-protein kinase like 7 precursor</td>
<td>0.35</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>SPK1_DUGTI (P42687)</td>
<td>Tyrosine-protein kinase SPK-1 (EC 2.7.1.112).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V5I1_TRYBB (P26326)</td>
<td>Variant surface glycoprotein ILTAT 1.21 precursor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCGV_MACNE (Q28858)</td>
<td>Versican core protein</td>
<td>4.3</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>VIME_HUMAN (P08670)</td>
<td>Vimentin (Fragment).</td>
<td>1.6</td>
<td>2.2</td>
<td>3</td>
</tr>
<tr>
<td>VNC_HUMAN (P18206)</td>
<td>Vinculin (Metavinculin).</td>
<td>3.1</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>VTB_HUMAN (P02774)</td>
<td>Vitamin D-binding protein precursor (DBP)</td>
<td>1.3</td>
<td>0.4</td>
<td>2</td>
</tr>
<tr>
<td>WDR1_HUMAN (Q75083)</td>
<td>WD-repeat protein 1 (Actin interacting protein 1)</td>
<td>1.6</td>
<td>1.0</td>
<td>2</td>
</tr>
<tr>
<td>Z363_HUMAN (Q96PM5)</td>
<td>Zinc finger protein 363</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a) Subsequently identified by MASCOT as clostridium putative phosphatidylserine decarboxylase proenzyme (gi: 28209853)
b) Internal standard to which other proteins’ peak areas are normalized
c) Rickettsia conorii protein
d) Subsequently identified by MASCOT as unlabeled serum albumin

Frequency is the number of sample pairs in which the protein is observed; NQ = not quantified; Singleton L = only observed in proliferative phase; Singleton H = only observed in secretory phase

Figure 4. Appearance of selected proteins in the strong cation exchange fractions.

Figure 4 was calculated from expression data normalized to a value of 1 for any expression ratios less than 1 to generate linear data. (This was done by taking the reciprocal of expression values less than 1 in the original data set; for example, an expression ratio of 0.7 was converted to its reciprocal 1.42.) This data was then used to calculate the SD of housekeeping protein expression. A significant differential expression value was, therefore, defined as a normalized expression ratio that was three SDs above the ratio of 1.0. On the basis of the observed data, this ratio was set as 2.0 (approximately 1.0 + 3(0.3)). Consequently, only 2-fold or larger changes in expression ratios were considered to be of significance, based on the analysis of the variation in measured expression of housekeeping proteins. An additional stringency requirement of this level of expression in all three sample pairs was imposed due to the small initial sample set. No additional proteins in Table 1 meet the differential expression criteria of abundance ratios larger than 2.0 or smaller than 0.5 in all three sample pairs. As it will be shown in section 4, further inspection and additional searches will question the utility of some of these proteins as indicators of differential expression.

An example of extreme differential expression observable in all three sample pairs is shown in Fig. 5, which shows very significant expression enhancement of glutamate NMDA receptor subunit zeta 1 precursor in the secretory endometrium. The absence of the light-labeled analogue, which would have manifested as a cluster of peaks beginning at 578.6 Th, demonstrates the significantly lower expression of this protein in the proliferative endometrium. The triply charged tryptic peptide at 581.6 Th was identified as LLTLALLFSCSVAR, which maps to the N-terminal region of the protein. A second example is shown in Fig. 6, again showing extreme differential expression in the secretory phase. This protein was identified as FRAT1. For all five proteins that met the differential expression criteria stipu-
Figure 5. (a) Mass spectral window of sample A, fraction 19, showing enhanced expression of glutamate receptor subunit zeta 1 precursor in the secretory endometrium. The line to the left of the isotope envelope of the heavy-labeled version of the protein denotes where the light-labeled version is expected; (b) resulting MS/MS spectrum showing a partial sequence for the peptide LLTLALLFSCSVAR.

Figure 6. Mass spectral windows of the three sample pairs (fraction 15), showing enhanced expression of FRAT1.

lated above, a minimum p value of less than 0.04 was obtained on the basis of the Wilcoxon rank sum test, suggesting that the differences observed were of statistical significance.

4 Discussion

The observation of similar expression levels in the majority of proteins shown in Table 1 is not surprising, as many proteins are structural and housekeeping in nature. In addition, structural and housekeeping proteins, e.g., tubulin and actin, are expressed in many tissue types. This is of significance as the biopsies that we examined comprise not only epithelial cells, but also supportive stroma cells, blood and vessels that are not expected to be affected by the endometrial cycle. In this initial examination of three sample pairs, we are adopting the conservative strategy of recognizing proteins as differentially expressed only when their expression ratios are larger than 2.0 or smaller than 0.5 in all three sample pairs. Of the 119 proteins listed in Table 1, five proteins are in this category. In fact, differential expression is in the extreme form that only single labels are apparent. There is one protein (POL polypeptide) that was seen only in the secretory endometrium, but was observed in only two sample pairs and which therefore does not meet our criteria of differential expression in this study (see later).

One of the five differentially expressed proteins is glutamate NMDA receptor subunit zeta 1 precursor, which was only observed in the secretory endometrium. Literature searches have not yielded any particular evidence of this protein being associated with the endometrium. This protein is a subunit of a ligand gated ion channel and is known to be involved with synaptic plasticity in neurons. A recent paper suggests that it may also play a role in glutamate-mediated toxicity to mitochondria, eventually leading to apoptosis [15].

As stated earlier, the peptide maps to the N-terminal region of the precursor, which is normally cleaved to form the mature protein. This suggests a need for a closer investigation into the localization of the mature protein as well as the peptide itself by immunohistochemistry. As far as we can tell, this is the first evidence of this peptide being a marker for the secretory endometrium. Similarly, FRAT1 is detected only in the secretory endometrial samples under our experimental conditions. There is no direct connection between FRAT1 and the endometrium reported in the literature; however, FRAT1 is expressed in a wide range of human tissues [16]. The gene expressing FRAT1 is a known protooncogene that activates the WNT pathway [17, 18]. This pathway is important in the endometrial cycle; FRAT1 is known to inhibit c-Jun activity, thereby inhibiting subsequent apoptosis [19]. c-Jun, in turn, has been shown to be expressed in the endometrium and this expression is at a higher level during the proliferative phase than the secretory phase [20]. Thus, there was indirect evidence for a role of FRAT1 in endometrium biochemistry.

Myosin light chain kinase 2 is another protein that was observed exclusively in the secretory endometrium in all three sample pairs. While we have found no studies on the
relative amounts of this enzyme during the endometrial cycle, previous studies showed that it is present in the human myometrium at a level that is 4- to 5-fold higher than that in endometrial stromal cells [21]. It was also demonstrated that the specific activity of myosin light chain kinase, which is believed to be essential for smooth muscle contraction, was 20-fold higher in the myometrium than in non-muscle cells, e.g., skin fibroblasts [21]. Additional work will be needed before it can be ascertained that the enhanced expression in the secretory phase observed here is truly a reflection of the endometrium and not of other tissue components that were also present in the tissue homogenates. Isopentenyl diphosphate delta-isomerase, which is present only in the secretory endometrium is an interesting case. This protein was identified via the peptide ITMCTGSRTLK. In addition to this identification with ProICAT, a subsequent search with MASCOT (Matrix Science) verified this assignment. This peptide, and therefore the protein, is unique and found only in the bacterium *Rickettsia conorii*. A BLAST search comparing human and *R. conorii* isopentenyl diphosphate delta-isomerase proteins showed no region of homology. In addition, an MS BLAST of the peptide ITMCTGSRTLK against nonredundant protein databases returned no other matches.

A protein that appeared in all three sample pairs was initially identified by ProICAT as a simian immunodeficient viral protein, envelope polyprotein GP160 precursor. However, searching MASCOT with the same MS/MS spectra returned a different protein from clostridium, putative phosphatidylserine decarboxylase proenzyme (gi 28209853). Endometrial receptors for a clostridium toxin have been documented [22]; thus the identification of a clostridium protein is not entirely surprising. Another protein initially identified by ProICAT as POL polyprotein from Rous sarcoma virus was observed only in the secretory endometrium in two of the three sample pairs. A subsequent MASCOT search, however, returned a positive hit for an unlabeled peptide from serum albumin, and strongly suggested a wrong initial identification. These last results help to illustrate the importance in validating results returned by automated searches. Tropomyosin 1 alpha chain was identified in two out of three sample pairs to be selectively expressed in the secretory phase at a ratio of 2.3 ± 0.7. An enhanced expression of this protein in the secretory endometrium is in accordance with the result of a previous gene expression study that reported an up-regulation of 3.7 times [2].

A protein that showed no difference in expression levels (1.2 ± 0.7 in three samples) is macrophage migration inhibitory factor (MIF). This observation is consistent with those in previous studies that demonstrated MIF as being expressed by the human endometrium throughout the menstrual cycle and, in particular, predominantly in the glandular epithelial cells [23]. MIF was found to localize throughout the glandular epithelial cytoplasm in the proliferative phase; however, this distribution changes in the secretory phase, when MIF localizes to the apical portion of the glandular epithelial cells and is also present in glandular secretions. Macrophages are common in female reproductive tissues. In the endometrium, they play an important role in defense. In addition, macrophage degradation of cellular debris and foreign material may play an important role in endometrial shedding and repair [23]. Another protein observed in this study and for which there was precedence in the literature is glycodelin. On the basis of gene expression studies [2], glycodelin was believed to be up-regulated during the secretory phase. In this study, we identified the precursor of this protein in only one sample pair, but it was identified in both runs in only the secretory endometrium (Fig. 7).

Cathepsin B which, according to a previous study, does not appear to be differentially regulated during the menstrual cycle [24], was seen in two of our sample pairs with opposite trends (Fig. 8). Glutathione-S-transferase is known to show large variability which is believed to result from in-
dividual differences, rather than endometrial cycling [25]. The relative expression level observed in this study is 1.3 ± 0.4. Lactate dehydrogenase is another protein not believed to have any cyclic trend [26]; the relative level measured in this study, 1.4 ± 0.3, is in accordance with this expectation. Peptidyl-propyl cis-trans isomerase A or cyclophilin A has been observed in placental and decidual tissues, but there are no studies on relative amounts between the phases of the endometrial cycle [27]. Our relative expression results of 1.2 ± 0.2 show a constant level of expression. Haptoglobin is thought to be up-regulated during the secretory phase [28]. Our average relative expression level of 2.2 ± 1.6 is based on two enhanced secretory expression level measurements in two sample pairs and one reduced secretory expression level measurement in another sample pair that account for the very large SD. Haptoglobin is a liver protein secreted into blood. As the amounts of blood in the biopsied tissues could not be controlled prior to homogenizing, the large SD might simply reflect that the samples contain different amounts of blood as opposed to truly representing increased or decreased haptoglobin expression.

It should perhaps be emphasized that protein identifications using ICAT-labeled peptides are reliable as the identification of one labeled peptide in a pair corroborates the identification of the other in the pair. This offsets somewhat the reduced opportunities in seeing multiple peptides because ICAT targets only cysteine-containing peptides. The chance of a missed identification increases when only one member of the labeled peptide pair is observable in the form of an extreme differential expression and there is uncertainty as to whether the observable peptide is labeled with the light or heavy ICAT reagent. In addition, the presence of an unlabeled peptide may be mistakenly interpreted as a labeled peptide that is differentially expressed in the extreme. An unlabeled peptide will need to have a combination of “right” chemical properties and very high initial concentrations (e.g. a serum albumin peptide) to “survive” the affinity chromatography stage intended to select only labeled peptides. The above is intended to be a cautionary note in pointing out potential shortcomings in a powerful technique that readily highlights and quantifies differential protein expressions, and not as a negative opinion on the suitability and utility of ICAT in protein discovery or quantification.

Of the 46 studies published to date that centered on ICAT, the vast majority were on method development that compared differential protein expressions in two different cell states. Most of these studies were on yeast cultured under different medium conditions and they compared protein levels of one sample pair. Only one recent study [29] targeted clinical samples, and even in this study only one sample pair was analyzed for differential protein expression. Our present study is the only work in which multiple sample pairs have been examined. It is, of course, possible that a given observed deviation in protein ratio from 1, i.e. a heavy/light label ratio ≠ 1, is due to individual differences, and that different pairing of the proliferative and secretory samples may produce a different ratio. It is because of this possibility that the criterion for differential expression was stringently set to a ratio of larger than 2.0 and smaller than 0.5 in all three sample pairs. It turns out that for the five proteins that met the above criterion of differential expression, the differences were in the extreme in that the proteins were observed in only one endometrial phase in a pair. Expression differences at this extreme level in all three sample pairs are highly unlikely to have originated from individual differences.

5 Concluding remarks

From an initial list of 346 proteins identified by ProICAT 1.0 SP3, 119 proteins were identified after manual inspection of MS/MS data and additional searches using MASCOT. The expression levels of the majority of proteins do not vary significantly between the proliferative and the secretory phases. Only five proteins satisfy our criteria on differential expression in having an expression ratio larger than 2.0 or smaller than 0.5 in all three sample pairs. In fact, the differential expressions observed are extreme in that the proteins are present in only one of the two endometrial phases. Further inspection and additional searches question the utility of some of these proteins as indicators of differential expression. The two proteins with unquestionable differential expressions in the secretory endometrium are: glutamate NMDA receptor subunit zeta 1 precursor and FRAT1. Some of the proteins that show no differential expression have previously been examined in gene-expression studies with similar conclusions. Notwithstanding the limitations of the current-generation software, quantitative proteomic analysis as practiced with ICAT is viable and will lead to discovery of marker proteins that exhibit differential expression.

We gratefully acknowledge the support and cooperation of our surgical and clinical laboratory collaborators, including Drs. K. Joan Murphy (University Health Network), Wusun Paek (Mount Sinai Hospital), and Titus Owalabi, Ali Qizilbash and Dennis MacDonald (North York General Hospital). This study was financed primarily by the National Cancer Institute of Canada (NCIC). Instrumental and additional funding was supplied by the Natural Sciences and Engineering Research Council (NSERC) of Canada, MDS Scix, the Ontario Research and Development Challenge Fund (ORDCF), and Genome Canada, G. D. thanks the German Academic Exchange Service (DAAD) for funding of a postdoctoral fellowship.

6 References