

Sophia Li-Ming Wong
Michael G. Walker

Incyte Genomics, Inc.,
Palo Alto, CA, USA

A bioinformatics approach to identifying fetal development genes

Gene regulation of fetal development is not well understood. In part, insulin and insulin-like growth factors (IGF) modulate placental steroid synthesis (PSS), which in turn modulates fetal growth. However, many of the genes that participate in this function remain to be identified. To find such genes, we examined the expression patterns of known IGF and placental steroid synthesis (IGF/PSS) genes in 1176 human cDNA libraries. We found a set of eight known IGF/PSS genes (PL-4, hCG, PAPP-A, EMBP, PLAP, P450 aromatase, P450scc, and 3-beta-HSD) that shared a highly similar expression profile across these libraries. We used these eight as bait in a search for other genes that showed very similar expression, and that might thus be related in function. We found ten genes closely co-expressed with the eight bait genes, but not previously reported as linked to IGF/PSS. Of these ten, six were previously reported as associated with cell growth in fetal and/or cancer tissues (malignant melanoma metastasis suppressor, PLAC-1, PSG10, PSG-beta1, serine palmitoyl transferase, and TONDU). Four are EST sequences, here named PLAC2, PLAC3, PLAC4, and PLAC5. Co-expression provides a method to identify which human genes are promising candidates for further experiments to determine their roles in fetal development.

Keywords: expression analysis, fetal development genes

1 Introduction

Large databases containing the measured mRNA expression of tens of thousands of genes in thousands of tissue samples are now available from cDNA sequencing and microarray facilities. Such expression databases provide a potential source for identifying candidate genes that may be associated with fetal growth. To find such genes, we examined the expression patterns of insulin-like growth factors (IGF) and placental steroid synthesis (PSS) genes that are known to be expressed during fetal growth, and searched for previously uncharacterized genes that showed very similar expression patterns. The known IGF and PSS genes are expressed in placental and fetal tissue, but are also expressed in other adult tissues. An assumption of the method is that the genes that show the most similar expression to the IGF/PSS genes across a diverse set of tissues are the genes that are likely to be related in function. Hence, we do not restrict the search to placental tissue. Following identification of co-expressed genes, we examined the tissue-specific distribution of the genes in placental versus non-placental libraries and in tumor versus non-tumor libraries.

Correspondence: Michael G. Walker, Incyte Genomics, Inc., Palo Alto, CA 94304, USA. Phone: +1-650-845 5771; Fax: +1-650-621-8514; e-mail: mwalker@incyte.com

2 Materials and Methods

2.1 cDNA library preparation

cDNA libraries were prepared from a set of 1176 human tissue samples. These 1176 tissue samples were from diverse anatomic and pathologic states, mainly from surgery, biopsy, or post-mortem samples, or were prepared from cell lines and include all the libraries that were in our database at the time of the analysis. Approximately 5000 cDNAs from each library were sequenced by gel electrophoresis, assembled, and aligned against known genes. Sequences that were significantly different from known genes were assigned new identification numbers. All genes that were detected in at least five of the 1176 libraries were included in the analysis.

2.2 Computational analysis

For the purpose of this analysis, we encode each gene as present in a library if at least one cDNA fragment from the gene was detected in the sample, and absent if no cDNA fragment from the gene was detected. We summarize the co-occurrences of two genes in a contingency table, such as Table 1, which shows the co-occurrence of two genes, aromatase P450 and Cholesterol Side-Chain Cleavage Enzyme (P450scc), in this set of 1176 cDNA libraries. We observe that aromatase P450 was detected in a total of 27

Table 1. Co-expression of aromatase P450 and P450scc in 1176 libraries.

Columns indicate the number of libraries in which P450scc is present or absent. Rows indicate the number of libraries in which aromatase P450 is present or absent. Aromatase P450 was detected in a total of 27 libraries, and P450scc was detected in 13 of those 27 libraries. For these data, the probability that aromatase P450 and P450scc co-occur by chance is $8.65e-12$ (Fisher Exact test).

Number of libraries	P450scc present	P450scc absent	Total
Aromatase present	13	14	27
Aromatase absent	40	1109	1149
Total	53	1123	1176

libraries, and P450scc was detected in 13 of those 27 libraries. From the co-occurrence contingency table, we calculate the probability that the co-occurrences arose by chance using a Fisher Exact test. For the data in Table 1, the probability that aromatase P450 and P450scc co-occur by chance is $8.65e-12$. Genes known to be unrelated typically have p -values of $1.0e-2$ or higher in this data set. (The Fisher Exact test is preferable to the chi-square test for this application, because the chi-squared test is an approximation to the Fisher Exact. The chi-square approximation is only appropriate when the expected number of occurrences in each entry in the contingency table is greater than 10, which is not the case for many pairs of genes.)

Because we perform multiple statistical tests on each gene, we must deal with the question of statistical significance and interpretation of p -values. We may, for example, apply a Bonferroni correction (dividing the desired alpha, say, $P = 0.01$, by the number of comparisons performed) to determine a suitable p -value. For n genes, we perform $n(n-1)/2$ pairwise comparisons; thus 40,000 genes yield 8×10^8 pairwise comparisons, requiring a Bonferroni-corrected p -value of $0.01/(8 \times 10^8)$ or $\sim 10^{-11}$. However, the Bonferroni correction is extremely conservative, and yields almost no false positives at the price of failing to detect many real associations. For practical interpretation, we can only claim that, of the most widely-expressed genes, the genes identified here are the most closely co-expressed with these known placental steroid-associated genes.

3 Results

We found a set of eight known IGF/PSS genes that shared a highly similar expression profile across these libraries (PL-4, hCG, PAPP-A, EMBP, PLAP, P450 aromatase, P450scc, and 3-beta-HSD) (Table 2). We briefly describe here the IGF/PSS functions of these eight genes. Placental Lactogen Hor-

mone (PL-4) and the human growth hormones (hGH) regulate maternal and fetal metabolism and the growth and development of the fetus. PL and hGH stimulate maternal production of insulin-like growth factor (IGF). In the fetus, PL stimulates the production of IGFs, insulin, and adrenocortical hormones [1,2]. Human chorionic gonadotropin (hCG) is a differentiating agent that modulates fetal testicular, ovarian, and adrenal function, and that modulates expression of IGF-1, IGF-1 receptor, the IGF binding proteins, and collagenase (MMP9) [3,4]. Pregnancy-Associated Plasma Protein-A (PAPP-A) is a metalloprotease expressed in the placenta that cleaves IGFBP to release bound IGF [5,6]. Eosinophil Major Basic Protein (EMBP) irreversibly binds PAPP-A and prevents PAPP-A from cleaving IGFBP. Both EMBP and PAPP-A are expressed at significantly higher levels in term placenta than in other tissues that have been examined [7,8]. Cholesterol Side-Chain Cleavage Enzyme (P450scc) catalyzes the first regulatory reaction controlling steroid hormone synthesis. It is expressed in the placenta in early and midgestation. Insulin and the insulin-like growth factors regulate placental steroidogenesis in part by modulating the activity of P450scc [9–11]. 3-beta-hydroxysteroid dehydrogenase (3-beta HSD) catalyzes oxidation of beta-hydroxysteroid precursors, leading to the synthesis of all classes of steroid hormones. Several isoenzymes are known, including a placentally-expressed isoenzyme. Insulin and the insulin-like growth factors regulate placental steroidogenesis by modulating the activity of 3 beta-HSD, among other placental steroid synthesis enzymes [12–14]. P-450 aromatase catalyzes estrogen 2-hydroxylase activity in human placenta and catalyzes the conversion of testosterone to estradiol. Insulin and the insulin-like growth factors regulate placental steroidogenesis by modulating the activity of placental aromatase P450, among other placental steroid synthesis enzymes [11,15]. Placental Alkaline Phosphatase (PLAP) is a placental enzyme whose expression is regulated by estradiol. PLAP appears to participate in placental differentiation [16,17]

We chose these eight genes as our bait to look for other genes that showed very similar expression. We found ten genes that were not previously linked to IGF/PSS that had expression patterns similar to the eight bait genes. Of these ten, six were previously reported as associated with cell growth in fetal and/or cancer tissues (malignant melanoma metastasis suppressor, placenta specific 1 (PLAC1), pregnancy-specific glycoprotein 10 (PSG10), pregnancy-specific beta 1 glycoprotein (PSG-beta1), serine palmitoyl transferase (SPT), and TONDU). We briefly describe the reported functions of these six genes.

Malignant melanoma metastasis suppressor (alternate name KISS1) is expressed predominantly in the placenta and in tumor cells. Like hCG, it regulates expression of collagenase (MMP9). Experimental evidence indicates that KISS1 reduces metastatic potential, possibly by regulating

Table 2. Co-expression of known and putative IGF/PSS genes.

Entries in this table are the probability that the co-expression of each pair of genes is due to chance (expressed as negative log of the p-value from the Fisher Exact test. ns = not significant).

	3betaHSD	aromP450	EPIL	MBP	PAPP-A	PL-4	PLAP	scc P450	hCG
3betaHSD	33	12	8	11	12	11	16	11	11
aromatase P450	12	62	10	17	11	10	10	12	10
EPIL	8	10	27	6	ns	4	6	9	7
MBP	11	17	6	106	11	12	10	7	13
PAPP-A	12	11	ns	11	154	6	10	6	6
PL-4	11	10	4	12	6	37	6	6	21
PLAP	16	10	6	10	10	6	37	11	9
scc P450	11	12	9	7	6	6	11	103	7
hCG	11	10	7	13	6	21	9	7	75
mmms	13	12	12	9	7	7	11	11	9
PAPP-E	13	15	9	14	18	7	14	11	8
PLAC1	15	16	11	10	10	9	11	13	12
PSG beta1	11	12	6	12	10	10	10	8	13
PSG10	20	13	9	13	11	10	16	13	12
SPT	13	18	8	19	11	10	12	9	9
TONDU	11	12	8	10	10	5	10	8	7
PLAC2	14	7	5	5	ns	7	9	7	6
PLAC3	10	14	8	11	13	6	10	8	5
PLAC4	12	11	11	9	6	7	9	9	10
PLAC5	12	10	7	8	6	9	9	8	7

events downstream of cell-matrix adhesion [18–20]. Placenta-specific 1(PLAC1) is a recently-identified X-linked gene with placenta-specific expression. Human PLAC1 has 60% identity to mouse PLAC1. During mouse embryogenesis, PLAC1 is expressed in the ectoplacental cone, giant cells, and labyrinthine trophoblasts [21]. Pregnancy-Specific Beta-1-Glycoprotein (PSBG) (alternate name Sp1) is a major product of the placenta. Sequence similarity between carcinoembryonic antigen (CEA) and PSBG suggest a possible common function in the control of cell growth [1, 22]. Serine palmitoyl transferase (SPT) is the enzyme that initiates the biosynthetic pathway for ceramide and sphingosine, leading to formation or repair of the epidermal barrier, and thus participates in genesis of fetal skin. Sphingolipids are expressed in placenta and appear to mediate trophoblast apoptosis [23–25]. TONDU is the human homologue of the *Drosophila* morphogenesis gene *vestigial*, and substitutes for *vestigial* in wing formation when that gene is knocked out [26].

Four of the ten genes are previously uncharacterized EST sequences, here named PLAC2, PLAC3, PLAC4, and PLAC5. PLAC5 shows extensive regions of similarity to the Human endogenous retrovirus strain XA34 pol gene sequence (Accession HEU29659), and to several other human endogenous retrovirus strain genes. It may be a splice variant or a close homologue of one of these genes. Human endogenous retrovirus genes, including XA34, are ex-

Table 3. Tissue distribution of PLAC2 through PLAC5.

Column 1 is the gene name. Columns 2, 3, and 4 are the number of libraries of each tissue type in which the gene was detected.

Gene ID	placental/fetal	tumor	non-tumor
PLAC2	8	9	5
PLAC3	9	1	1
PLAC4	9	2	1
PLAC5	5	2	0

pressed in placental and fetal tissues and appear to participate in morphogenesis [27,28].

Following identification of the co-expressed genes, we examined the tissue-specific distribution of PLAC2 through PLAC5. For non-placental tissues, we also examined the distribution of the genes in tumor versus non-tumor libraries. Table 3 shows the tissue distribution of the PLAC genes. The four genes occur predominantly in placental/fetal tissue or in tumors suggesting that, like the IGF/PSS genes, the PLAC genes may participate in tissue growth.

Other known IGF/PSS genes, such as the metalloprotease ADAM12, early placenta insulin-like peptide (EPIL), IGF binding proteins, placental growth factor, and other placental genes were co-expressed with these genes, but less consistently or at less significant *p*-values.

4 Discussion

Several aspects of the library selection and preparation are likely to affect the results of the co-expression analysis, because they violate one or more assumptions of the p -value calculation. Specifically, because more than one library may be obtained from a single patient (for example, multiple organs, or matched tumor and non-tumor tissue, or normalized and non-normalized), libraries are not completely independent. Libraries were, in many cases, normalized or subtracted to enrich the proportion of genes expressed at low levels; this alteration will make the detection of associations between genes expressed at different levels more difficult. Some libraries in which many novel genes were found were sampled to greater depth, leading to an inconsistent measure of presence or absence. Because of the random sampling, genes that are expressed at low levels may not be detected. However, the consequence of all these effects is most likely to obscure a true relationship. It is unlikely that a spurious relationship would be introduced consistently across 1176 libraries by chance. Co-expression analysis may fail to detect known relationships for a variety of reasons, including effects of sampling, temporal or tissue differences in expression, and errors in sequence assembly or annotation.

Traditional approaches to understanding genes focus on in-depth analysis, in which we learn a great deal about a small set of genes. The bioinformatics approach described here is a breadth-first search, in which we learn a small amount about each of a large number of genes. Expression analysis using a database of cDNA libraries provides hypotheses about the likely function of genes, but these hypotheses need confirmation in direct experiments. The primary utility of an expression database analysis is to suggest experiments that are most likely to be fruitful, thereby saving research time and expense. The genes identified here by co-expression analysis are useful candidates for further experiments to determine their possible roles in fetal development.

Accession numbers

Gene name	Genbank accession number
PLAC2	BG354568
PLAC3	BG354569
PLAC4	BG354570
PLAC5	BG354571

Acknowledgments

We gratefully acknowledge the assistance and support of our colleagues. LifeSeq is a registered trademark of Incyte Genomics, Inc.

References

- [1] Gardner, M. O., Goldenberg, R. L., Cliver, S. P., Boots, L. R., Hoffman, H. J. (1997) Maternal serum concentrations of human placental lactogen, estradiol and pregnancy specific beta 1-glycoprotein and fetal growth retardation. *Acta Obstet. Gynecol. Scand. Suppl.* **165**: 56–58.
- [2] Handwerker, S., Freemerk, M. (2000) The roles of placental growth hormone and placental lactogen in the regulation of human fetal growth and development. *J. Pediatr. Endocrinol. Metab.* **13**: 343–356.
- [3] Huynh, H. (1998) In vivo regulation of the insulin-like growth factor system of mitogens by human chorionic gonadotropin. *Int. J. Oncol.* **13**: 571–575.
- [4] Stadtmauer, L., Vidali, A., Lindheim, S. R., Sauer, M. V. (1998) Follicular fluid insulin-like growth factor-I and insulin-like growth factor-binding protein-1 and -3 vary as a function of ovarian reserve and ovarian stimulation. *J. Assist. Reprod. Genet.* **15**: 587–593.
- [5] Kristensen, T., Oxvig, C., Sand, O., Moller, N. P., Sottrup-Jensen, L. (1994) Amino acid sequence of human pregnancy-associated plasma protein-A derived from cloned cDNA. *Biochemistry* **33**: 1592–1598.
- [6] Bersinger, N. A., Leporrier, N., Herrou, M., Leymarie, P. (1999) Maternal serum pregnancy-associated plasma protein A (PAPP-A) but not pregnancy-specific beta-1-glycoprotein (SP1) is a useful second-trimester marker for fetal trisomy 18. *Prenat. Diagn.* **19**: 537–541.
- [7] Wasmoen, T. L., Coulam, C. B., Benirschke, K., Gleich, G. J. (1991) Association of immunoreactive eosinophil major basic protein with placental septa and cysts. *Am. J. Obstet. Gynecol.* **165**: 416–420.
- [8] Overgaard, M. T., Oxvig, C., Christiansen, M., Lawrence, J. B., Conover, C. A., Gleich, G. J., Sottrup-Jensen, L., Haaning, J. (1999) Messenger ribonucleic acid levels of pregnancy-associated plasma protein-A and the proform of eosinophil major basic protein expression in human reproductive and nonreproductive tissues. *Biol. Reprod.* **61**: 1083–1089.
- [9] Albrecht, B. A., Daels, P. F. (1997) Immunolocalization of 3 beta-hydroxysteroid dehydrogenase, cytochrome P450 17 alpha-hydroxylase/17,20-lyase and cytochrome P450 aromatase in the equine corpus luteum of dioestrus and early pregnancy. *J. Reprod. Fertil.* **111**: 127–133.
- [10] Schiff, R., Arensburg, J., Itin, A., Keshet, E., Orly, J. (1993) Expression and cellular localization of uterine side-chain cleavage cytochrome P450 messenger ribonucleic acid during early pregnancy in mice. *Endocrinology* **133**: 529–537.
- [11] Nestler, J. E. (1993) Regulation of the aromatase activity of human placental cytotrophoblasts by insulin, insulin-like growth factor-I, and -II. *J. Steroid Biochem. Mol. Biol.* **44**: 449–457.
- [12] Rheume, E., Lachance, Y., Zhao, H. F., Breton, N., Dumont, M., de Launoit, Y., Trudel, C., Luu-The, V., Simard,

- J., Labrie, F. (1991) Structure and expression of a new complementary DNA encoding the almost exclusive 3 beta-hydroxysteroid dehydrogenase/delta 5-delta 4-isomerase in human adrenals and gonads. *Mol. Endocrinol.* **5**: 1147–1157.
- [13] Dupont, E., Labrie, F., Luu-The, V., Pelletier, G. (1992) Immunocytochemical localization of 3 beta-hydroxysteroid dehydrogenase/delta 5-delta 4-isomerase in human ovary. *J. Clin. Endocrinol. Metab.* **74**: 994–998.
- [14] Juengel, J. L., Guy, M. K., Tandeski, T. R., McGuire, W. J., Niswender, G. D. (1994) Steady-state concentrations of messenger ribonucleic acid encoding cytochrome P450 side-chain cleavage and 3 beta-hydroxysteroid dehydrogenase/delta 5,delta 4 isomerase in ovine corpora lutea during the estrous cycle. *Biol. Reprod.* **51**: 380–384.
- [15] Okubo, K., Jinbo, M., Toma, Y., Shimizu, Y., Yanaihara, T. (1996) Aromatase and estrogen 2-hydroxylase activities of human placental microsomes in pregnancy-induced hypertension. *Endocr. J.* **43**: 363–368.
- [16] Okamoto, T., Seo, H., Mano, H., Furuhashi, M., Goto, S., Tomoda, Y., Matsui, N. (1990) Expression of human placenta alkaline phosphatase in placenta during pregnancy. *Placenta* **11**: 319–327.
- [17] Wang, R., Song, L., Ding, L. (1996) Detection of placental alkaline phosphatase in ovarian cancer. *Chung Hua Fu Chan Ko Tsa Chih* **31**: 107–109.
- [18] Yan, C., Wang, H., Boyd, D.D. (2001) KiSS-1 Represses 92-kDa Type IV Collagenase Expression by Down-regulating NF-kappa B Binding to the Promoter as a Consequence of Ikappa Balpha-induced Block of p65/p50 Nuclear Translocation. *J. Biol. Chem.* **276**: 1164–1172.
- [19] West, A., Vojta, P. J., Welch, D. R., Weissman, B. E. (1998) Chromosome localization and genomic structure of the KiSS-1 metastasis suppressor gene (KISS1). *Genomics* **54**: 145–148.
- [20] Lee, J. H., Welch, D.R. (1997) Suppression of metastasis in human breast carcinoma MDA-MB-435 cells after transfection with the metastasis suppressor gene, KiSS-1. *Cancer Res.* **57**: 2384–2387.
- [21] Cocchia, M., Huber, R., Pantano, S., Chen, E. Y., Ma, P., Forabosco, A., Ko, M. S., Schlessinger, D. (2000) PLAC1, an Xq26 gene with placenta-specific expression. *Genomics* **68**: 305–312.
- [22] Jurisicova, A., Antenos, M., Kapasi, K., Meriano, J., Casper, R.F. (1999) Variability in the expression of trophectodermal markers beta-human chorionic gonadotrophin, human leukocyte antigen-G and pregnancy specific beta-1 glycoprotein by the human blastocyst. *Hum. Reprod.* **14**: 1852–1858.
- [23] Pitson, S. M., D'Andrea R, J., Vandeleur, L., Moretti, P. A., Xia, P., Gamble, J. R., Vadas, M. A., Wattenberg, B. W. (2000) Human sphingosine kinase: purification, molecular cloning and characterization of the native and recombinant enzymes. *Biochem. J.* **350 Pt 2**: 429–441.
- [24] Napoli, K. L. (2000) The FTY720 story. *The. Drug. Monit.* **22**: 47–51.
- [25] Payne, S. G., Brindley, D. N., Guilbert, L. J. (1999) Epidermal growth factor inhibits ceramide-induced apoptosis and lowers ceramide levels in primary placental trophoblasts. *J. Cell. Physiol.* **180**: 263–270.
- [26] Vaudin, P., Delanoue, R., Davidson, I., Silber, J., Zider, A. (1999) TONDU (TDU), a novel human protein related to the product of *vestigial* (*vg*) gene of *Drosophila melanogaster* interacts with vertebrate TEF factors and substitutes for Vg function in wing formation. *Development* **126**: 4807–4816.
- [27] Lin, L., Xu, B., Rote, N. S. (2000) The cellular mechanism by which the human endogenous retrovirus ERV-3 env gene affects proliferation and differentiation in a human placental trophoblast model, BeWo. *Placenta* **21**: 73–78.
- [28] Kjellman, C., Sjogren, H. O., Salford, L. G., Widegren, B. (1999) HERV-F (XA34) is a full-length human endogenous retrovirus expressed in placental and fetal tissues. *Gene* **239**: 99–107.

Printed on acid-free paper

© WILEY-VCH Verlag GmbH, D-69469 Weinheim, 2001.
Printed in the Federal Republic of Germany.

Editor-in-Chief: W. Doerfler, Inst. of Genetics, D-50931 Cologne.

Publisher: WILEY-VCH Verlag GmbH, Pappelallee 3, D-69469 Weinheim, Federal Republic of Germany.

Correspondence concerning advertisements should be addressed to WILEY-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Telefax (06201) 606-550, e-mail: sales.journals@VCHGROUP.DE.

Institutional subscription rates 2001 (2002 print only or online only/print & online): Europe € 238,-; (308,-/323,-); Switzerland sFr 428,- (518,-/544,-); Outside Europe US \$ 298,00 (348,-/365,-) incl. postage and

handling. Reduced rates for personal subscribers are available on request. Prices are subject to change.

Rate of publication: 6 times a year.

Cancellation of subscriptions: The publisher must be notified not later than three months before the end of the calendar year.

Printed by betz-druck, Darmstadt.

All rights reserved (including those of translation into other languages). No part of this journal may be reproduced in any form – by photoprint, microfilm, or any other means – nor transmitted or translated into a machine language without the permission in writing of the publisher. – Only single copies of contributions, or parts thereof, may be reproduced for personal use. – This journal was carefully

produced in all its parts. Nevertheless, authors, editors and publisher do not warrant the information contained therein to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Valid for users in the USA: The appearance of the code at the bottom of the first page of an article in this journal (serial) indicates the copyright owner's consent that copies of the article may be made for personal or internal use, or for the personal or internal use of specific clients. The consent is given on the condition, however, that the copier pay the stated percopy fee through the Copyright Clearance Center, Inc., for copying beyond that permitted by Section 107 or 108 of the U.S. Copyright Law. This consent does not extend to other kinds of copying, such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. For copying from back volumes of this journal see Permissions to Photo-Copy: Publisher's Fee List of the CCC.

