Gene Expression Versus Sequence for Predicting Function: Glia Maturation Factor Gamma Is Not A Glia Maturation Factor

Michael G. Walker
Incyte Genomics, Inc., CA 94087, USA

It is standard practice, whenever a researcher finds a new gene, to search databases for genes that have a similar sequence. It is not standard practice, whenever a researcher finds a new gene, to search for genes that have similar expression (co-expression). Failure to perform co-expression searches has lead to incorrect conclusions about the likely function of new genes, and has lead to wasted laboratory attempts to confirm functions incorrectly predicted. We present here the example of Glia Maturation Factor gamma (GMF-gamma). Despite its name, it has not been shown to participate in glia maturation. It is a gene of unknown function that is similar in sequence to GMF-beta. The sequence homology and chromosomal location led to an unsuccessful search for GMF-gamma mutations in glioma. We examined GMF-gamma expression in 1432 human cDNA libraries. Highest expression occurs in phagocytic, antigen-presenting and other hematopoietic cells. We found GMF-gamma mRNA in almost every tissue examined, with expression in nervous tissue no higher than in any other tissue. Our evidence indicates that GMF-gamma participates in phagocytosis in antigen presenting cells. Searches for genes with similar sequences should be supplemented with searches for genes with similar expression to avoid incorrect predictions.

Key words: gene function prediction, expression analysis, sequence analysis, GMF-gamma

Introduction

When a researcher finds a new gene they search Genbank or other databases for genes that have a similar sequence. It is not standard practice, when a researcher finds a new gene, to search for genes that have similar expression (co-expression). Failure to perform co-expression searches has led to incorrect conclusions about the likely functions of new genes, and has led to wasted laboratory attempts to confirm functions incorrectly predicted by sequence similarity.

We describe here the case of Glia Maturation Factor gamma. Despite its name, Glia Maturation Factor gamma (GMF-gamma) has not been shown to participate in glia maturation. It is a recently-identified gene of unknown function. It was first described by Asai (*1*), who named it based on its homology to GMF-beta (82% amino acid identity, 70% DNA identity), a neural growth and maturation factor. Asai and colleagues examined GMF-gamma expression in eight human tissues and found it predominantly in lung, heart and placenta, with trace expression in brain, liver, skeletal muscle, kidney, and pancreas. (Their libraries did not include hemic or immune system tissues.) Peters and colleagues (*2*) determined that GMF-gamma is located on human chromosome 19 at band q13.2, a region that is frequently deleted in malignant glioma. The possible role of GMF-gamma in glial differentiation (based on its homology to GMF-beta) and its chromosomal location in a region linked to malignant glioma led Peters to study its potential role in this disease. They examined 41 gliomas, but found no mutations in GMF-gamma, indicating that it is not the 19q13.2 glioma gene. Thus, the function of GMF-gamma is currently unknown.

E-mail: mwalker@stanfordalumni.org

Results and Discussion

We detected GMF-gamma mRNA in 292 of 1432 cDNA libraries (Table 1), in every tissue category except stomatognathic. It is most abundant in hemic and immune tissues (87 of 179 libraries). These 179 hemic and immune libraries were derived from peripheral blood, umbilical cord blood, lymph nodes,

thymus, spleen, bone marrow, tonsil, and Jurkat cells. We detected GMF-gamma in monocytes, macrophages, dendrites, B-cells, T-cells, blast cells, eosinophils, and mast cells in the blood samples. We also detected it in lymph node, spleen, thymus, and tonsil tissue. It was not detected in bone marrow samples. The high expression in hemic and immune tissues, absence in bone marrow, and ubiquitous but low-level expression in other tissue samples are consistent with expression in mature blood cells.

Table 1 Distribution of GMF-gamma EST's by Tissue Type

Tissue Category	Number of GMF gamma ESTs	Number of ESTs in libraries from this tissue	Percent GMF gamma ESTs	Number of Libraries	Number of libraries in which GMF gamma is detected
Cardiovascular System	17	272986	0.006	74	11
Connective Tissue	8	151678	0.005	54	6
Digestive System	35	521762	0.007	155	28
Embryonic Structures	8	108468	0.007	24	5
Endocrine System	14	233683	0.006	63	7
Exocrine Glands	29	258383	0.011	67	18
Genitalia, Female	21	456353	0.005	117	14
Genitalia, Male	21	463016	0.005	120	14
Germ Cells	7	48181	0.015	5	2
Hemic and Immune	721	725942	0.099	179	87
Liver	5	115620	0.004	37	5
Musculoskeletal	22	162801	0.014	50	13
Nervous System	55	995533	0.006	231	31
Pancreas	6	111771	0.005	25	4
Respiratory System	41	412898	0.010	96	28
Sense Organs	1	25345	0.004	10	1
Skin	4	72732	0.006	18	2
Stomatognathic	0	14712	0.000	17	0
Unclassified/Mixed	14	159180	0.009	22	8
Urinary Tract	30	295517	0.010	68	8
Totals	1059	5606561	0.019	1432	292

Table 2 shows genes with expression patterns similar to that of GMF-gamma by co-expression analysis (see Methods). For all the co-expressed genes in Table 2, the probability that the co-expression with GMF-gamma is due to chance is less than 1.0×10^{-15} by the Fisher Exact test. What do the genes with which GMF-gamma is co-expressed indicate about its likely function? Cathepsin S (CTSS) is a cysteine protease expressed in antigen presenting cells whose function is to cleave the MHC class II invariant chain that is required for antigen presentation (*3 – 6*). It is found at the highest levels in spleen, heart and lung, in the lung being detected only in macrophages (*6*). CD53 is a member of the transmembrane-4 superfamily (TM4SF) of proteins that act as linker molecules, recruiting protein kinase C (PKC) into a signaling complex with beta integrins (*7*). It is expressed predominantly in lymphoid-myeloid lineage

cells (*8*), participates in LFA-1 dependent pathway of lymphocyte activation and cell adhesion (*9*), and associates with MHC class II molecules at the plasma membrane of lymph cells (*10*). HLA-DR alpha is an MHC class II protein expressed specifically by antigen presenting cells; it regulates antigen presentation. The Fc-epsilon-receptor gamma is a subunit of several Fc receptors, and is expressed in antigen presenting cells and other leukocytes. Fc receptors phagocytose antigen-antibody complexes, induce antigen presentation by MHC II proteins, and participate in signal transduction (*11* – *17*). PKC isoforms are important in phacocytosis by Fc receptors (*18*). Lysosomal protein (LAPTm5) is expressed preferentially in hematopoietic cell lines, is localized to lysosomes, and binds to ubiquitin (*19*). Adra and colleagues found high expression of LAPTm5 in peripheral blood leukocytes, lung, thymus, and spleen in human adult tissues. Leukocyte factor associated protein 1 (LFA-1) integrin beta subunit (alternate name CD18/CD11) is a leukocyte cell surface adhesion molecule that mediates signal transduction and co-operates with Fc receptors (*20* – *23*). L-plastin (LPL) is a calcium-regulated actin-binding protein expressed in leukocytes, fibroblasts and in diverse set of solid tumors (*24* – *26*). LPL is phosphorylated in phagocytes in response to inflammatory stimuli, which also increase actin polymerization (*24*, *27*). Pleckstrin is a protein kinase C substrate expressed in macrophages and localizes to the phagosomal membranes upon ingestion of opsonized antigen (*18*, *28*). It is found in lymphocytes, monocytes, granulocytes, and platelets but is not detected in non-hematopoietic cells (*29*). It is induced upon differentiation of hematopoietic cells and is thought to participate in signaling (*30*). p40hox is expressed specifically in phagocytic cells, is a PKC substrate, and participates in the oxidation of phagocytosed particles (*31* – *34*). Upstream stimulation factor 1 (USF1) is a transcription factor that regulates differentiation of hematopoietic cells and activates MHC genes, among others (*35* – *37*). Co-expression of GMF-gamma with these genes indicates that it is expressed particularly in phagocytic / antigen processing cells. GMF-gamma is co-expressed at lower levels with many other MHC, proteasome, and blood-cell specific genes.

Is expression of GMF-gamma in these cell types consistent with the types of tissues in which it is most abundant? Asai found GMF-gamma at the highest levels in human lung, heart and placenta. In the LifeSeq collection of 1432 cDNA libraries, we find it predominantly in hemic and immune tissues, and most specifically in hematopoietic cells, but detect it in every tissue but bone marrow. Macrophages are particularly abundant in lung and placenta. Coronary vascular epithelial cells perform phagocytosis and antigen presentation, and genes with which GMF-gamma is co-expressed, such as cathepsin S, are expressed at high levels in heart tissue. Thus, the tissues in which other researchers and we detect GMF-gamma are consistent with expression in phagocytic and antigen processing cells. We would expect that GMF-gamma is absent from bone marrow because immature blood cells do not perform phagocytosis or antigen processing, and therefore do not express these genes. We would expect that GMF-gamma is present at low levels in virtually every other tissue because mature blood cells are present in virtually every tissue.

Table 2 Genes Co-expressed with GMF-gamma (*p*-values from Fisher Exact Test)

Gene	-log of *p*-value
LAPTm5	30
HLA-DR alpha	29
Fc-epsilon R1 gamma	26
CD53	25
LFA-1/integrin beta	25
LCP1 / L-plastin / p65	24
pleckstrin p47	23
p40phox	23
USF1	23
CTSS cathepsin S	23

Tsuiki and colleagues cloned a rat gene with 91% amino acid identity to human GMF-gamma, which they named rGMF-gamma (*38*). They examined rGMF-gamma expression in eight rat tissue samples. Northern blots indicated the strongest expression in thymus, testis, and lung. Western blots indicated the strongest expression in spleen and thymus. Expression was low or absent in brain, skin, small intestine, and stomach. In rat testis, they found rGMF-gamma mRNA in spermatids, while in brain it was localized around pyramidal cells within CA3 of the hippocampus. Four phosphorylation sites that are targets of PKC, PKA, and other kinases *in vitro* are conserved in rGMF-beta and rGMF-gamma, and Tsuiki con-

curred with the view that these genes may be involved in cell differentiation and growth via signal transduction. Nishiwaki and colleagues investigated the roles of rGMF-beta and rGMF-gamma in development and growth of the rat retina (*39*). They reported that rGMF-gamma is synthesized and localized mainly in Muller glial cells in the rat retina during fetal development. Earlier research has shown that retinal Muller glial cells are phagocytic, can express MHC class II determinants and function as antigen presenting cells (*40*, *41*).

Expression analysis provides hypotheses about the likely cell-specificity and function of GMF-gamma, 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 co-expression analysis makes several assumptions that are violated to greater or lesser degrees by aspects of the library selection and preparation. For example, libraries are not completely independent, because more than one library may be obtained from a single patient. Normalizing or subtracting makes the detection of associations between genes expressed at different levels more difficult. The cDNA libraries used in this analysis were prepared at different times and with different methods, and may not be consistent. The effects of different cDNA library samples, different normalization, different preparation methods, or preparation at different times are most likely to obscure true relationships. Such differences will make the calculated probability of association less accurate. However, it is unlikely that a pattern that is consistent across 1432 libraries, has good p-values, and is consistent with known biological relationships would be introduced by the cumulative random effects of such differences. Thus, co-expression analysis will yield false negatives, but it is unlikely to yield false positive results with a database of this size.

We have observed that human GMF-gamma is expressed predominantly in phagocytic cells, is absent from immature blood cells, and is co-expressed with phagocytosis and antigen processing genes. The evidence indicates that GMF-gamma is involved in phagocytosis and antigen presentation. Despite its name, Glia Maturation Factor gamma is unlikely to be a glia maturation factor. These results indicate that searches for genes with similar sequences should be supplemented with searches for genes with similar expression, to avoid incorrect predictions of putative gene function.

Methods

We examined the expression of GMF-gamma in 1432 human cDNA libraries from diverse anatomic and pathologic states. Some libraries were subtracted or normalized to enrich rare mRNA. Approximately 5000 cDNA's from each library were sequenced by gel electrophoresis, assembled, and aligned against known genes. All genes that were detected in at least five of the 1432 libraries were included in the analysis described here, which yielded 37,071 genes, gene fragments, or splice variants.

To identify genes with a similar expression pattern to GMF-gamma, we performed co-expression analysis using the Guilt-by-Association (GBA) algorithm (*42*). Briefly, in a GBA analysis, we consider a gene to be present (expressed) in a library if cDNA corresponding to that gene is detected in the library. We consider a gene to be absent (not expressed) in a library when no cDNA for that gene is detected. For a given pair of genes, the co-expression data can be summarized in a two-by-two contingency table, where the contingency table entries indicate one of four categories: both genes detected, neither detected, and one or the other gene detected. From the contingency table, we determine the probability that the co-expression occurs by chance using a chi-square test or a Fisher Exact test (43).

Acknowledgements

We thank our colleagues at Incyte Genomics, Inc. for their contributions to this work.

References

1. Asai, K., *et al.* 1998. Isolation of novel human cDNA (hGMF-gamma) homologous to Glia Maturation Factor-beta gene. *Biochim. Biophys. Acta.* 1396: 242-244.
2. Peters, N., *et al.* 1999. The human glia maturation factor-gamma gene: genomic structure and mutation analysis in gliomas with chromosome 19q loss. *Neurogenetics* 2: 163-166.
3. Driessen, C., *et al.* 1999. Cathepsin S controls the trafficking and maturation of MHC class II molecules in dendritic cells. *J. Cell. Biol.* 147: 775-790.
4. Kos, J., *et al.* 2001. Cathepsin S in tumours, regional lymph nodes and sera of patients with lung cancer: relation to prognosis. *Br. J. Cancer* 85: 1193-1200.

5. Riese, R.J., *et al.* 1998. Cathepsin S activity regulates antigen presentation and immunity. *J. Clin. Invest.* 101: 2351-2363.
6. Shi, G.P., *et al.* 1994. Human cathepsin S: chromosomal localization, gene structure, and tissue distribution. *J. Biol. Chem.* 269: 11530-11536.
7. Zhang, X.A., *et al.* 2001. Transmembrane-4 superfamily proteins associate with activated protein kinase C (PKC) and link PKC to specific beta (1) integrins. *J. Biol. Chem.* 276: 25005- 25013.
8. Hernandez-Torres, J., *et al.* 2001. Differential cooperation between regulatory sequences required for human CD53 gene expression. *J. Biol. Chem.* 276: 35405-35413.
9. Cao, L., *et al.* 1997. Anti-CD53 monoclonal antibody induced LFA-1/ICAM-1-dependent and -independent lymphocyte homotypic cell aggregation. *Immunobiology* 197: 70-81.
10. Engering, A. and Pieters, J. 2001. Association of distinct tetraspanins with MHC class II molecules at different subcellular locations in human immature dendritic cells. *Int. Immunol.* 13: 127-134.
11. Amigorena, S. and Bonnerot, C. 1999. Fc receptor signaling and trafficking: a connection for antigen processing. *Immunol. Rev.* 172: 279-284.
12. Amigorena, S. and Bonnerot, C. 1999. Fc receptors for IgG and antigen presentation on MHC class I and class II molecules. *Semin. Immunol.* 11: 385-390.
13. Heiken, H., *et al.* 1996. T lymphocyte development in the absence of Fc epsilon receptor I gamma subunit: analysis of thymic-dependent and independent alpha beta and gamma delta pathways. *Eur. J. Immunol.* 26: 1935-1943.
14. Huber, V.C., *et al.* 2001. Fc receptor-mediated phagocytosis makes a significant contribution to clearance of influenza virus infections. *J. Immunol.* 166: 7381-7388.
15. Kraft, S. and Bieber, T. 2001. Fc epsilon ri-mediated activation of transcription factors in antigen-presenting cells. *Int. Arch. Allergy Immunol.* 125: 9-15.
16. Kraft, S., *et al.* 2001. Unexpected functions of FcepsilonRI on antigen-presenting cells. *Int. Arch. Allergy Immunol.* 124: 35-37.
17. Minskoff, S.A., *et al.* 1998. Fc gamma RII-B1 regulates the presentation of B cell receptor-bound antigens. *J. Immunol.* 161: 2079-2083.
18. Brumell, J.H., *et al.* 1999. Expression of the protein kinase C substrate pleckstrin in macrophages: association with phagosomal membranes. *J. Immunol.* 163: 3388-3395.
19. Adra, C.N., *et al.* 1996. LAPTM5: a novel lysosomal-associated multispanning membrane protein preferentially expressed in hematopoietic cells. *Genomics* 35: 328-337.
20. Bianchi, E., *et al.* 2000. Integrin LFA-1 interacts with the transcriptional co-activator JAB1 to modulate AP-1 activity. *Nature* 404: 617-621
21. Serrander, L., *et al.* 1999. Particles binding beta (2)-integrins mediate intracellular production of oxidative metabolites in human neutrophils independently of phagocytosis. *Biochim. Biophys. Acta.* 1452: 133-144.
22. van Spriel, A.B., *et al.* 2001. Mac-1 (CD11b/CD18) is essential for Fc receptor-mediated neutrophil cytotoxicity and immunologic synapse formation. *Blood* 97: 2478-2486.
23. VanCompernolle, S.E., *et al.* 2001. Anti-CD81 activates LFA-1 on T cells and promotes T cell-B cell collaboration. *Eur. J. Immunol.* 31: 823-831.
24. Jones, S.L., *et al.* 1998. A role for the actin-bundling protein L-plastin in the regulation of leukocyte integrin function. *Proc. Natl. Acad. Sci. USA* 95: 9331-9336.
25. Lin, C.S., *et al.* 1993. Characterization of the human L-plastin gene promoter in normal and neoplastic cells. *J. Biol. Chem.* 268: 2793-2801.
26. Lin, C.S., *et al.* 2000. Upregulation of L-plastin gene by testosterone in breast and prostate cancer cells: identification of three cooperative androgen receptor-binding sequences. *DNA Cell Biol.* 19: 1-7.
27. Jones, S.L. and Brown, E.J. 1996. FcgammaRII-mediated adhesion and phagocytosis induce L-plastin phosphorylation in human neutrophils. *J. Biol. Chem.* 271: 14623-14630.
28. Tyers, M., *et al.* 1988. Molecular cloning and expression of the major protein kinase C substrate of platelets. *Nature* 333: 470-473.
29. Gailani, D., *et al.* 1990. P47 phosphoprotein of blood platelets (pleckstrin) is a major target for phorbol ester-induced protein phosphorylation in intact platelets, granulocytes, lymphocytes, monocytes and cultured leukaemic cells: absence of P47 in non-haematopoietic cells. *Br. J. Haematol.* 74: 192-202.
30. Kienzle, N., *et al.* 1997. Evidence that the expression and phosphorylation status of pleckstrin is modulated by Epstein-Barr virus in human B lymphocytes. *Blood* 89: 3488-3490.
31. Cross, A.R. 2000. p40 (phox) Participates in the activation of NADPH oxidase by increasing the affinity of p47 (phox) for flavocytochrome b (558). *Biochem. J.* 349: 113-117.
32. Grandvaux, N., *et al.* 1999. The Ku70 autoantigen interacts with p40phox in B lymphocytes. *J. Cell. Sci.* 112 (Pt 4): 503-513.
33. Hua, J., *et al.* 2001. Evaluation of the expression of NADPH oxidase components during maturation of HL-60 clone 15 cells to eosinophilic lineage. *Inflamm. Res.* 50: 156-167.
34. Zhan, S., *et al.* 1996. Genomic structure, chromo-

somal localization, start of transcription, and tissue expression of the human p40-phox, a new component of the nicotinamide adenine dinucleotide phosphate-oxidase complex. *Blood* 88: 2714-2721.

35. Giannola, D.M., *et al.* 2000. Hematopoietic expression of HOXB4 is regulated in normal and leukemic stem cells through transcriptional activation of the HOXB4 promoter by upstream stimulating factor (USF)-1 and USF-2. *J. Exp. Med.* 192: 1479-1490.
36. Kiermaier, A., *et al.* 1999. DNA binding of USF is required for specific E-box dependent gene activation *in vivo*. *Oncogene* 18: 7200-7211.
37. Zhong, G., *et al.* 2001. Identification of a chlamydial protea se-like activity factor responsible for the degradation of host transcription factors. *J. Exp. Med.*193: 935-942.
38. Tsuiki, H., *et al.* 2000. Cloning of a rat glia maturation factor-gamma (rGMFG) cDNA and expression of its mRNA and protein in rat organs. *J. Biochem. (Tokyo)* 127: 517-523.
39. Nishiwaki, A., *et al.* 2001. Expression of glia maturation factor during retinal development in the rat. *Mol. Brain. Res.* 95: 103-109.
40. Francke, M., *et al.* 2001. Retinal pigment epithelium melanin granules are phagocytozed by Muller glial cells in experimental retinal detachment. *J. Neurocytol.* 30: 131-136.
41. Roberge, F.G., *et al.* 1988. Glial retinal Muller cells produce IL-1 activity and have a dual effect on autoimmune T helper lymphocytes. Antigen presentation manifested after removal of suppressive activity. *J. Immunol.* 140: 2193-2196.
42. Walker, M.G., *et al.* 1999. Prediction of gene function by genome-scale expression analysis: prostate cancer-associated genes. *Genome Res.* 9: 1198-1203.
43. Agresti, A. 1990. *Categorical data analysis.* Wiley, New York, USA.

Received: 28 December, 2002
Accepted: 7 January, 2003