

# Glia Maturation Factor Gamma (GMFG): A Cytokine-Responsive Protein During Hematopoietic Lineage Development and Its Functional Genomics Analysis

Ying Shi<sup>1\*</sup>, Ling Chen<sup>1</sup>, Lance A. Liotta<sup>2#</sup>, Hong-Hui Wan<sup>3</sup>, and Griffin P. Rodgers<sup>1\*</sup>

<sup>1</sup> *Molecular and Clinical Hematology Branch, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institutes of Health (NIH), Bethesda, MD 20892-1822, USA;* <sup>2</sup> *Laboratory of Pathology, National Cancer Institute, NIH, Bethesda, MD 20892-7412, USA;* <sup>3</sup> *Maryland Institute of Genomics and Department of Computer Science, University of Maryland, Silver Spring, MD 20906, USA.*

<sup>#</sup> *Current address: The Center for Applied Proteomics and Molecular Medicine, George Mason University, Manassas, VA 20110, USA.*

Human hematopoiesis was evaluated using the techniques of controlled stem cell differentiation, two-dimensional gel electrophoresis-based proteomics, and functional genomics. We provide the first report that glia maturation factor gamma (GMFG) is a cytokine-responsive protein in erythropoietin-induced and granulocyte-colony stimulating factor-induced hematopoietic lineage development. Results from global functional genomics analysis indicate that GMFG possesses several other features: hematopoietic tissue-specific gene expression, a promoter concentrated with high-score hematopoiesis-specific transcription factors, and possible molecular coevolution with a rudimentary blood/immune system. On the basis of our findings, we hypothesize that GMFG is a hematopoietic-specific protein that may mediate the pluripotentiality and lineage commitment of human hematopoietic stem cells.

**Key words:** GMFG, hematopoiesis, proteomics, functional genomics

## Introduction

Although it is widely accepted that the hematopoietic stem cell (HSC) is capable of both self-renewal and differentiation into all of the peripheral blood cell types, the mechanisms that control these self-renewal and differentiation processes remain a mystery (1, 2). To address this issue, we developed a special culture system in which HSCs cultured for 14 days with erythropoietin (EPO) were recultured for another 14 days with granulocyte-colony stimulating factor (G-CSF) and *vice versa* (that is, G-CSF-stimulated HSCs were recultured with EPO). This unique culture system provides both practical and theoretical platforms for further study in this area. Using this culture system, we observed phenotypic lineage interconversion between erythroid and myeloid cells derived from human hematopoietic AC133<sup>+</sup> stem/progenitor cells (3).

**\*Corresponding authors.**

**E-mail:** yingshi@mail.nih.gov;  
gprod@helix.nih.gov

To focus on the molecular aspects of hematopoiesis, we used proteomics and bioinformatics to further investigate lineage development and interconversion. Upon analysis of protein profiles from two-dimensional gel electrophoresis (2-DE), we discovered a protein spot that was responsive to specific cytokine signals that determine the formation of specific blood cell lineages. Using liquid chromatography-tandem mass spectrometry (LC-MS/MS), the protein was identified as glia maturation factor gamma (GMFG). Later, the cDNA sequence of the GMFG gene was confirmed using the reverse transcriptase-polymerase chain reaction (RT-PCR).

The GMFG gene, also called the glia maturation factor beta homolog (GMFB-h) gene, was first reported by two groups in 1998 (4, 5). Mao *et al* (4) discovered the GMFG gene when sequencing cDNAs from CD34<sup>+</sup> hematopoietic stem/progenitor cells. Asai *et al* (5) found the gene inadvertently when conducting a Northern blot experiment on GMFB in as-

trocytes. The full-length cDNA of GMFG is approximately 0.9 kb and encodes a protein of 142 amino acids. Results of an enzyme-linked immunoassay indicate that the GMFG gene is highly expressed in spleen, thymus, colon, and lung (6). GMFG has also been found in human serum at various ages; it was highly expressed in 21- to 30-year-old individuals and began to decrease rapidly in individuals older than 30 years (6). In a previous study, the promoter analysis identified certain binding sites for transcription factors (TFs) that were reported to be closely related to hematopoiesis (7).

Although the function of GMFG remains unclear, another member of the same protein family, GMFB, has been well studied. GMFB is also a protein of 142 amino acids and was initially identified as a growth/differentiation factor from vertebrate brains (8–12). This protein has several serine/threonine (Ser/Thr) phosphorylation sites and can be phosphorylated with protein kinase A (PKA), protein kinase C (PKC), casein kinase II (CKII), and p90 ribosomal S6 kinase (RSK) (13). The PKA-phosphorylated GMFB is a potent inhibitor of extracellular signal-regulated kinases 1 and 2 (ERK1/ERK2) and is a strong enhancer of p38 (14, 15). Both ERK1/ERK2 and p38 belong to the same mitogen-activated protein kinase (MAPK) family. Thus, GMFB was considered to be a putative intracellular kinase regulator (16) and a modulator of intracellular signal transduction via its phosphorylation (6). In addition, a previous report suggested that GMFB may have played an important role in maintaining stem cell systems as they developed during early metazoan evolution (17).

Here we provide a comprehensive bioinformatic analysis of GMFG, including promoter analysis, tissue/cell distribution analysis, and molecular evolutionary/phylum analysis using available DNA microarray data banks. Our results support the hypothesis that GMFG may play an important role as a maintenance factor for HSC and as a regulation factor for hematopoietic lineage commitment.

## Results

### A cytokine-responsive protein in EPO- and G-CSF-induced HSC differentiation

As shown in a previous report of our unique culture system (3), we cultured AC133<sup>+</sup> human HSCs with EPO for 14 d, which resulted in a definite

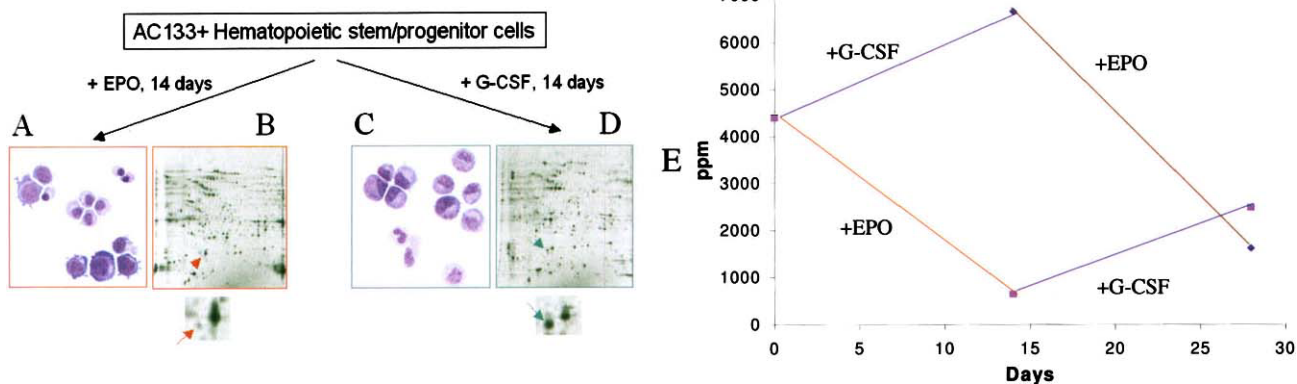
erythroid-lineage cell population that we termed E14 cells (Figure 1A). Simultaneously, we cultured additional AC133<sup>+</sup> cells with G-CSF for 14 d, which resulted in a definite myeloid-lineage cell population that we termed G14 cells (Figure 1C). When the E14 and G14 samples were analyzed by 2-DE, we found an interesting protein spot that exhibited an approximately 10-fold difference in density between the erythroid and myeloid cell populations. This protein was present in untreated AC133<sup>+</sup> cells, termed D0 cells. Its expression decreased with EPO stimulation and increased slightly with G-CSF stimulation (Figure 1B and D, respectively). Furthermore, when we recultured the E14 cells with G-CSF for another 14 d, the erythroid cells switched to a myeloid cell population termed E14→G14, in which the protein spot increased (Figure 1E). Conversely, when we recultured the G14 cells with EPO for another 14 d, the myeloid cells switched to an erythroid cell population termed G14→E14, in which the protein spot decreased (Figure 1E). Therefore, the cytokine-responsive character of this protein can be seen not only in HSCs but also in E14 erythroid cells and G14 myeloid cells. The latter observation is important because it may set up a molecular basis for the lineage switch that we observed in earlier research (3).

### Identification of GMFG as the cytokine-responsive protein

Upon repeating 2-DE on G14 samples, we manually excised and collected the gel spots and analyzed the proteins by LC-MS/MS analysis (Tables 1 and 2). We used GMFB as a reference and thoroughly compared the peptide sequences identified by MS. The peptide difference between GMFG and GMFB permits us to finally identify our protein as GMFG (Figure 2A). At the level of transcription, an RT-PCR validation of the results was performed with both the erythroid and myeloid lineages (Figure 2B).

### Differential expression of GMFG only observed at the protein level

After Zhang *et al* (18) reported the GMFG gene as one of the new genes expressed in CD34<sup>+</sup> HSCs, they designed a gene chip experiment to investigate the transcription of 300 new genes in various hematopoietic cell lines, including NB4, HL60, U937, K562, and Jurkat. Their results revealed no differential expression of the GMFG gene among these cell lines (18).



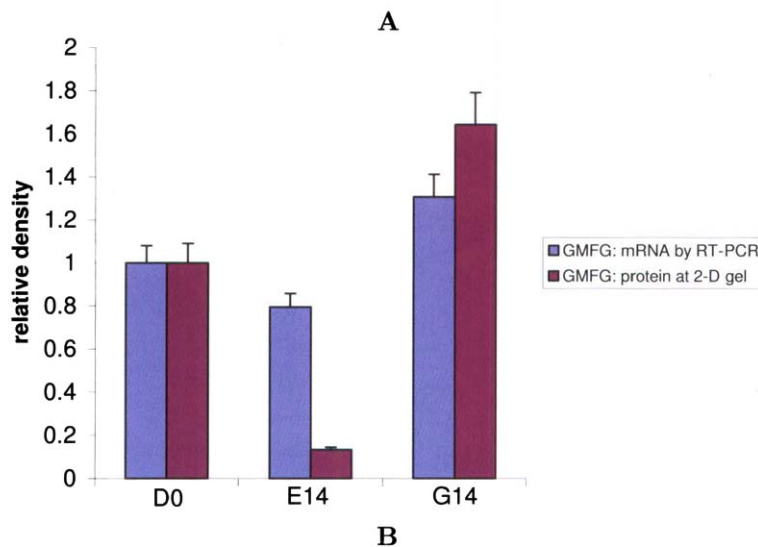
**Fig. 1** A protein spot showing differential expression between erythroid and myeloid lineages. **A** and **C**. Morphology of erythroid and myeloid cells, respectively. **B** and **D**. 2-DE of erythroid and myeloid cells, respectively. The protein spot of interest is indicated by the arrow in the enlargements of the 2-D gels. **E**. Normalized intensity value of the protein spot as it changed during lineage development and switching. The five cell populations, indicated by D0, E14, G14, E14→G14, and G14→E14, are defined in Materials and Methods.

```

MSDSLVVCEV DPELTEKLRK FRFRKETDNA AIIMKVDKDR QMVVLEEEFQ NISPEELKME
MSESLVVCDV AEDLVEKLRK FRFRKETNNA AIIMKIDKDK RLVVLDEELE GISPDELKDE

LPERQPRFVV YSYKYVHDDG RVSYPLCFIF SSPVGCKPEQ QMMYAGSKNR LVQTAELTKV
LPERQPRFIV YSYKYQHDDG RVSYPLCFIF SSPVGCKPEQ QMMYAGSKNK LVQTAELTKV

FEIRTTDDLT EAWLQEKLSF FR-----GMFG
FEIRNTEDLT EEWLREKLGF FH-----GMFB
    
```



**Fig. 2** The protein spot of interest was identified as GMFG. **A**. Protein sequence alignment for GMFG and GMFB. Amino acid differences are underlined within the fragments that were identified by MS in this study. GMFG was confirmed by this comparison. **B**. Comparisons of the GMFG mRNA level and its corresponding protein level during lineage differentiation. Three cell population samples were collected: D0 for AC133<sup>+</sup> hematopoietic stem/progenitor cells, E14 for erythroid cells, and G14 for myeloid cells (see Materials and Methods). The mRNA was evaluated by RT-PCR and the protein was evaluated by 2-DE. For comparison, both the mRNA value and the protein value of D0 were normalized to 1.0. All other values were displayed proportionally.

**Table 1 The General Information from LC-MS/MS Data**

Protein identified	Accession No.	Total score	Mass (Da)	pI	Database	Percentage covered
glia maturation factor gamma (GMFG)	gi 4758440	242	16,801.31	5.1	NCBI (Human)	26.06%

**Table 2 Precursor Mass and Ion Score for Each of the Three Peptides Identified by LC-MS/MS**

Peptide	Residue number	Precursor mass	Ion score
FVVYSYKYVHDDGR	68 – 81	1747.77	79.96
LVQTAELTKVFEIR	111 – 122	1646.88	87.69
MELPERQPR	59 – 67	1155.55	36.59

In the studies herein, our RT-PCR results also demonstrated that the GMFG gene was not significantly differentially expressed between E14 erythroid and G14 myeloid cells (Figure 2B). However, at the level of protein expression, GMFG exhibited a greater than 10-fold difference (Figure 1). The results of many researchers who have conducted simultaneous gene expression and proteomic studies confirm the fact that only a small percentage of genes show a statistically significant correlation between the expression levels of their corresponding mRNAs and proteins (19). It is widely accepted that the primary cause of the discrepancy between gene and protein expression is the post-translational modification of proteins. In our research, we noticed that GMFG is a protein with six consensus phosphorylation sites (16). An experiment is planned to evaluate the potential contributions of post-translational modifications of GMFG to our 2-DE results and to cell lineage formation. The results presented herein provide another example supporting the inevitable trend of researchers to approach a biomedical problem with the combined, comprehensive techniques of genomics, proteomics, and bioinformatics.

Recently, gene profile-based studies of the molecular signature of stem cells (2, 20) have revealed valuable information; however, results from our experiments suggest that the “total picture” might not be obtained until the protein profile is completed. It may not be surprising to find considerable inconsistencies between the gene signature and the protein signature within the same stem cell.

No significant GMFB transcription was observed for both lineages at 28 days’ scale (Figure 3). The 536-bp cDNA product of the RT-PCR was sequenced and confirmed to be GMFG (data not shown). The GMFG gene expression profiling is displayed in Figure 4. The results of GMFG promoter analysis are

demonstrated in Figure 5, and the GMFG’s molecular evolutionary analysis is shown in Figure 6.

## Discussion

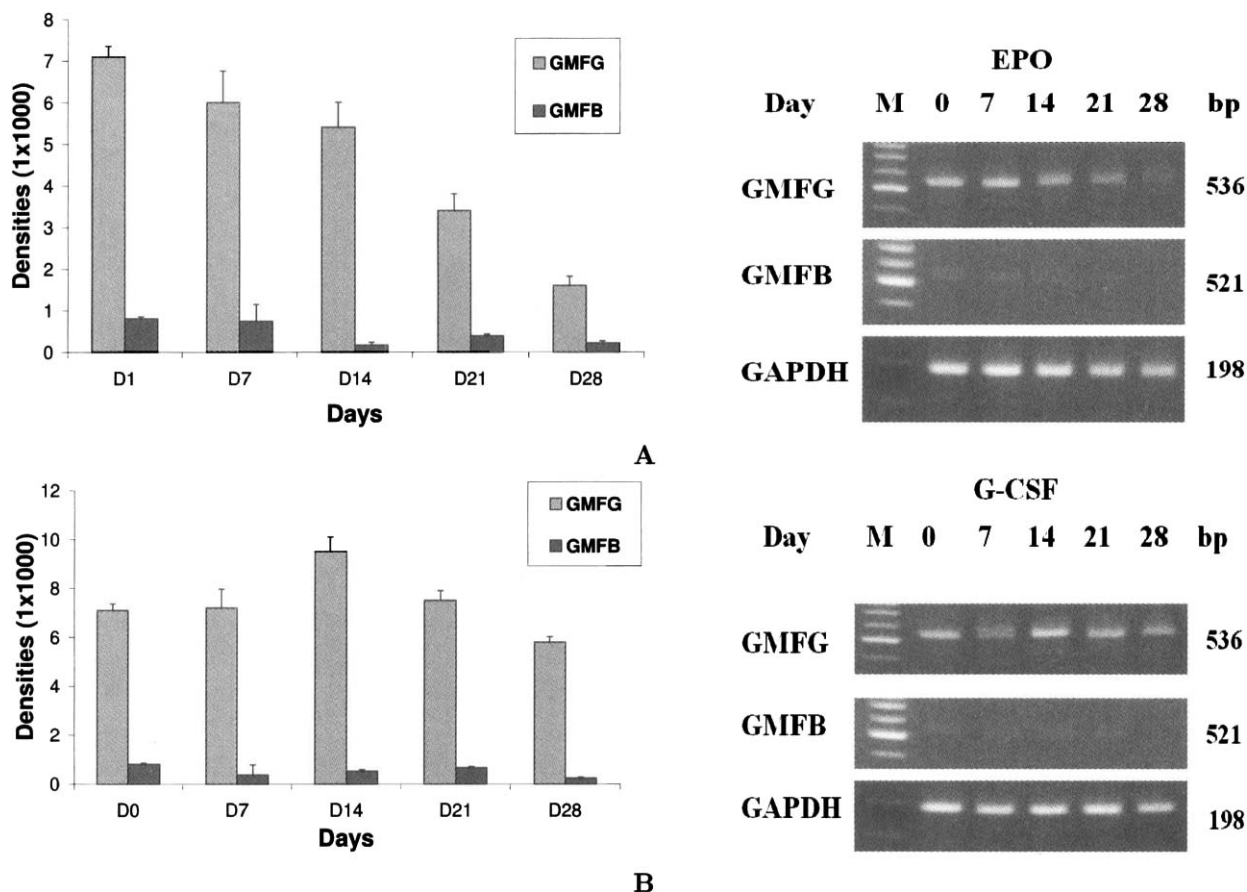
### Possible mechanism of GMFG responsiveness to cytokines

The mechanism of GMFG responsiveness to HSC cytokines and its involvement in regulating HSC differentiation remains unknown. From previous research on GMFB, we know that the PKA-phosphorylated form of GMFB potently inhibits ERK1/ERK2 and strongly enhances p38 (14, 15). These findings led us to consider the possibility that phosphorylation/dephosphorylation may be involved in the mechanism by which GMFG regulates HSC differentiation. Because GMFG and GMFB have the same Ser/Thr phosphorylation molecular characteristics, it is assumed that the major cellular biochemical activities of these two proteins will be similar (21). It has been well documented that during hematopoiesis, Ser/Thr phosphorylation plays an important role in signal transduction in HSC differentiation and proliferation. For example, both the inhibition of MAP1/2 (22) and the enhancement of p38 (23) can increase erythropoiesis. PKA-phosphorylated GMFB is known to be an inhibitor of MAP1/2 and an enhancer of p38 (14, 15), and there is evidence that PKA is involved in erythropoiesis (24). We have begun experiments to investigate the roles of GMFG in these processes.

### GMFG and hematopoietic development

#### *Tissue-specific distribution*

By using a global bioinformatics profiling strategy, we collected all accessible microarray data about GMFG



**Fig. 3** Validation of GMFG by RT-PCR. **A.** RT-PCR results for GMFG and GMFB during erythroid differentiation induced by EPO. Samples were collected at 0, 7, 14, 21, and 28 d, respectively. **B.** RT-PCR results for GMFG and GMFB during myeloid differentiation induced by G-CSF. Samples were collected at 0, 7, 14, 21, and 28 d, respectively.

gene expression. The expression level distribution of the GMFG gene in human tissues and cell lines is shown in Figure 4. It can be seen that the GMFG gene is highly expressed in blood (including in myeloid leukemia and lymphoid leukemia cell lines), thymus, spleen, fetal liver, and lung. It is noteworthy that most of the tissues with high levels of GMFG expression are related to the blood and immune systems. The distribution pattern is consistent with the results of our recent Northern blot experiment (data not shown) and is also very similar to the findings of a recent study that used immunoassays to determine the distribution of the GMFG protein (6). Our bioinformatics data also indicates that the expression of the GMFG gene in nervous/brain tissue is not higher than that in any other tissue, which confirmed Walker's finding (25), and thus the expression distribution pattern of the GMFG gene differs from that of the GMFB gene. It is interesting to note that the Burkitt's lymphoma cell line Daudi differs greatly from the Burkitt's lymphoma cell line Raji with respect to GMFG gene expression (Figure 4), with the

latter cell line having a much lower level of expression of the GMFG gene.

**High-score hematopoietic TF binding sites of the GMFG promoter**

The continual replacement of blood cells is essential for vertebrates. This process is largely dependent on hematopoietic, lineage-specific TFs (26). We evaluated the 750-bp GMFG promoter region for TF binding sites using our new scoring system. By setting a threshold score of 0.900, we found 20 putative TF binding sites with scores ranging from 0.900 to 1.000 (Figure 5). Among these high-score sites, there are three GATA-1 binding sites, with scores of 0.902, 0.906, and 0.963, respectively; one GATA-2 binding site, with a score of 0.921; two GATA-3 binding sites, with scores of 0.909 and 0.922, respectively; two AML-1a (27) binding sites, each with a perfect score of 1.000; three MZF1 (28) binding sites, with scores of 0.904, 0.930, and 0.957, respectively; one Lyf-1 binding site, with a score of 1.000; and one c-Ets-1 binding

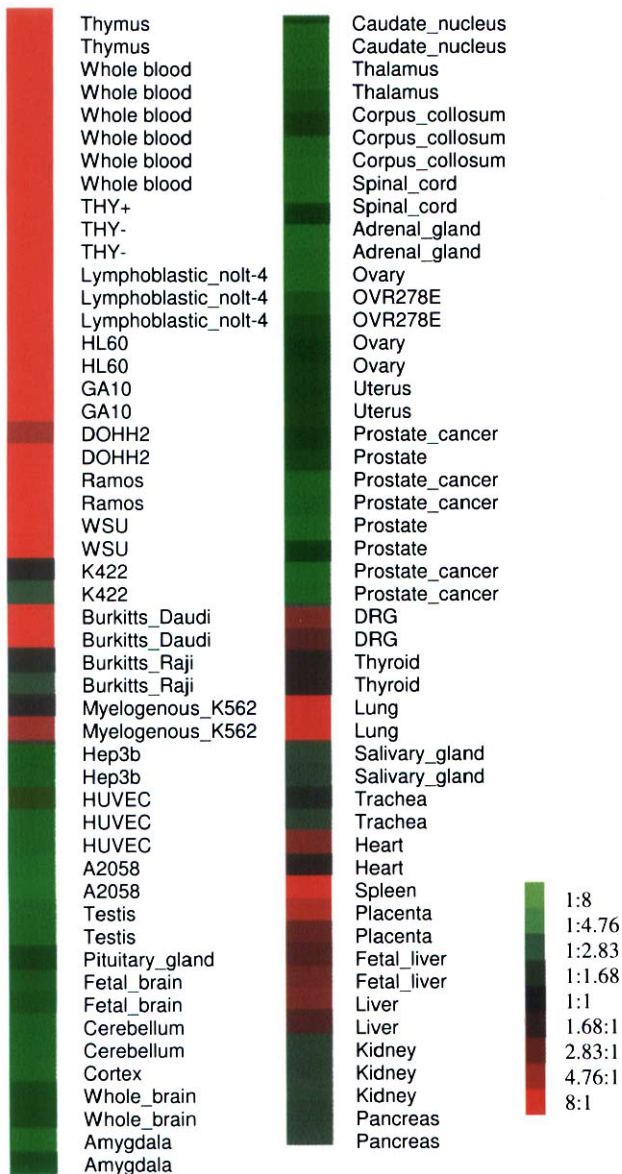


Fig. 4 Global GMFG gene expression profile.

site, with a score of 0.922. All of these TFs are related to erythroid, myeloid, T cell, and B cell differentiation and proliferation. In particular, AML-1a is considered to be a critical (master) regulator of hematopoietic cell development (27). Many hematopoiesis-specific genes are activated by AML-1a, including interleukin-3, granulocyte-macrophage-colony stimulating factor (GM-CSF), CSF-1R, and the T cell receptor (27). AML-1a may also be involved in HSC maintenance and renewal (26). Its tissue-specific expression occurs mostly in thymus and spleen (29), which is consistent with the tissue specificity expression of the GMFG gene, indicating that the latter might be a downstream regulator for AML-1a.

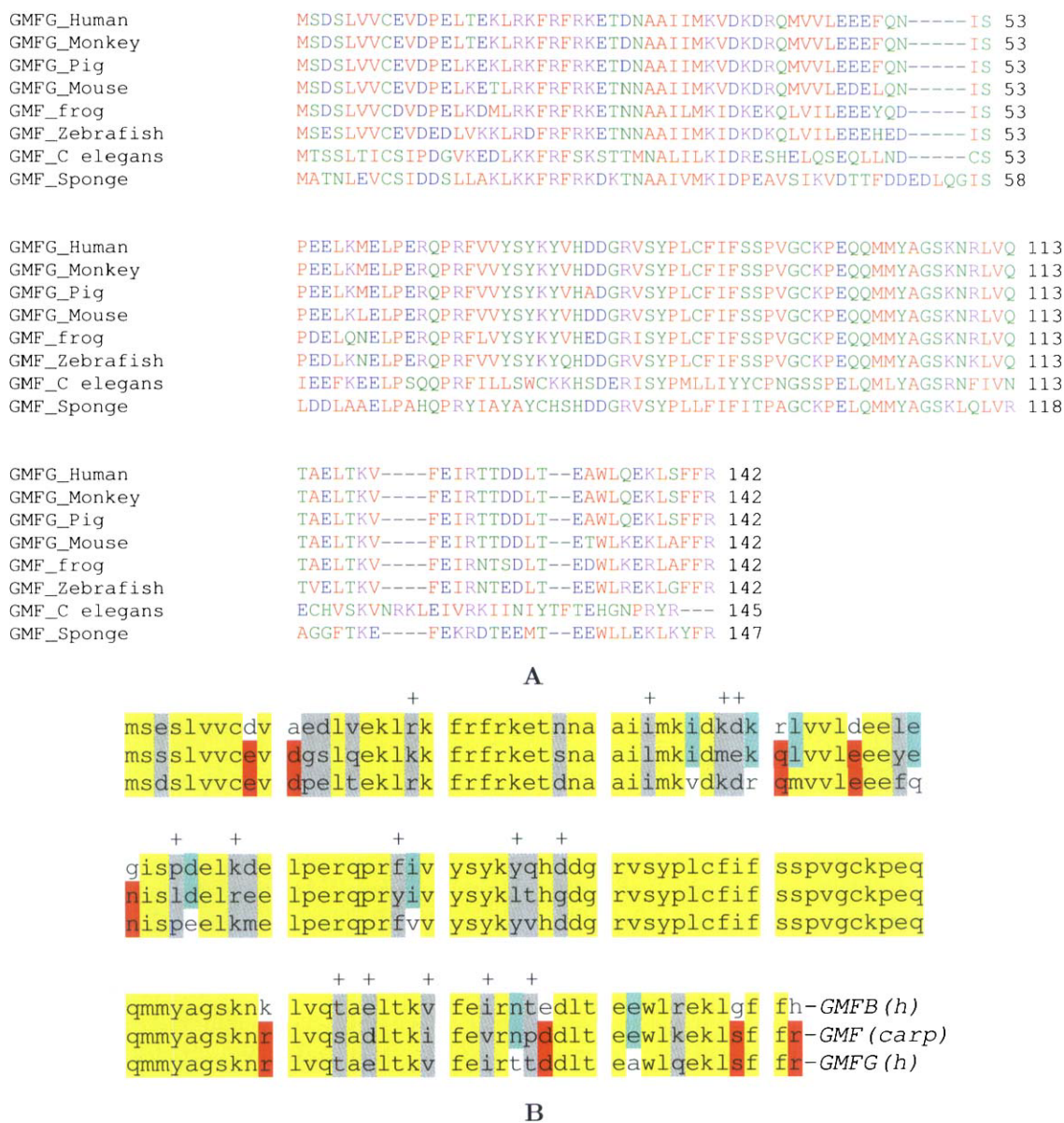
### GMFG protein coevolved with increasing complexity of the blood/immune system

In addition to the GMFGs from human, mouse, and rat that were reported by others (16), we have detected the protein sequences of GMFG in chimpanzee, monkey, pig, and cow using our GeneKey software. We also detected GMFGs in frog, zebrafish, the spotted green pufferfish *Tetraodon nigroviridis*, the filarial worm *Brugia malayi*, mosquito, and the nematode *Caenorhabditis briggsae*. GMFG in the sponge was reported by Muller *et al* (17). Selected multiple sequence alignments are shown in Figure 6A. From an evolutionary perspective, we can see that GMFG genes that exist in Metazoa probably originated via duplication of the actin-depolymerizing factor (ADF) gene (data not shown). Late in the evolution of the Vertebrata, we find the two family members GMFG and GMFB. Using our Sandwich Sequence Alignment software, we also found a GMFG in the carp (*Cyprinus carpio*) (30) that is very similar to human GMFB and GMFG, with slightly more similarity to GMFG (Figure 6B). This observation might imply that animals of the early stages of vertebrate evolution may have a single GMFG protein that shares the functions of GMFB and GMFG for development of the nervous system and the blood/immune system. As the complexity of the nervous system and the blood/immune system increased in the later stages of vertebrate evolution, especially in mammals, it is conceivable that the single GMFG gene was duplicated and subsequently modified to give rise to the GMFB gene and the GMFG gene. The observation may provide one possible explanation for why the overlapping genes, when comparing the stem cell molecular signatures, are “negligible” among HSCs, neural stem cells, and embryonic stem cells (31). It seems that different members of the same gene family are preferentially expressed in different systems/tissues to perform virtually the same task. Although these members are structurally and functionally similar, they have distinct characteristics in cDNA microarray analysis.

## Conclusion

There are two primary features of HSCs: (1) their ability to replicate without differentiation, and (2) their ability to differentiate, under specified conditions, to become different types of blood cells. Because GMFG (both the mRNA and protein forms) is present in HSCs and exhibits a differential pattern at





**Fig. 6** Molecular evolutionary analysis of GMFG. **A.** Amino acid sequence alignment of GMFG for selected species. For certain species, the protein is indicated as “GMF” instead of “GMFG” because only one GMF gene has been identified in these species. **B.** The result of Sandwich Sequence Alignment used for GMF of carp (*Cyprinus carpio*). The green color indicates the identical residues between carp GMF and human GMFB. The red color indicates identical residues between carp GMF and human GMFG. The yellow color indicates residues shared by all three proteins. “+” indicates residues that are identical between human GMFB and GMFG, but differ from carp GMF.

0.5% (vol/vol) IPG buffer, pH 4–7, in a Reswelling Tray (Amersham Bioscience, Uppsala, Sweden) on an 11-cm, pH 4–7 IPG strip (Immobiline DryStrip; Amersham Bioscience). Isoelectric focusing of samples was performed on a Multiphor II Electrophoresis Unit (Amersham Bioscience) for 45,000 Vh using the following protocol: 30 min at 150 V, 1 h at 300 V, 1 h at 1,500 V, and 12 h 20 min at 3,500 V. Subsequently, IPG strips were equilibrated for 15 min in equilibration buffer [6 M urea, 30% (wt/vol)

glycerol, 2% sodium dodecyl sulfate (SDS) in 0.05 M Tris-HCl buffer, pH 8.8] containing 1% (wt/vol) DTT and 0.001% (wt/vol) bromophenol blue. Next, IPG strips were equilibrated for 15 min in the equilibration buffer containing 250 mM iodoacetamide.

IPG strips were further processed for second-dimension polyacrylamide gel electrophoresis on ExcelGel SDS XL 12–14 according to procedures recommended by the manufacturer (Amersham Bioscience). ExcelGels were silver-stained with Hoefer Processor

Plus automatic stainer according to the protocol provided by the manufacturer (Amersham Bioscience).

## 2-DE image analysis

Scanning of gels was performed on a BioRad GS-800 Calibrated Imaging Densitometer (BioRad, Venendaal, Netherlands). Scanned TIFF images were analyzed using PDQuest 2D Gel Analysis Software version 7.0 (BioRad). Spots were automatically detected and images were checked by eye for undetected or incorrectly detected spots. Both spot volume and normalized spot volume datasets were used for further analysis. Average gels were obtained using the Create Average Gel Option of the software.

## Protein identification by MS

Silver-stained gel spots were destained, and individual protein gel spots were subjected to reduction and alkylation, followed by *in situ* digestion with sequencing-grade modified trypsin (Promega, Madison, USA). Peptides from in-gel digests were analyzed by capillary LC-MS/MS and matrix-assisted laser desorption/ionization (MALDI). A nano high-performance liquid chromatography system (LC Packings Inc., San Francisco, USA) with a Füsica column (0.075×150 mm; packed with PepMap™C18, 5 µm, 100Å; LC Packings, Inc.) was interfaced to a QSTAR mass spectrometer (Applied Biosystems, Foster City, USA). Peptides were eluted with a 10-min gradient of 5%–40% (vol/vol) followed by another 10-min gradient of 40%–95% (vol/vol) acetonitrile containing 0.1% formic acid/0.01% trifluoroacetic acid at a rate of 0.2 µL/min. The QSTAR mass spectrometer was set to iteratively acquire a positive time-of-flight (TOF) MS scan at 1-s accumulation time between 400 and 1,700 *m/z* followed by MS/MS scans at 5-s accumulation time between 50 and 2,500 *m/z* of the three most abundant ions from the preceding MS scan. Unprocessed data files containing MS/MS spectra from the QSTAR instrument were submitted to the Mascot search engine (Matrix Science Ltd., London, UK) for database searching and protein identification using the Mascot Daemon application (32). The SwissProt-Trmbl database and the National Center for Biotechnology Information (NCBI) nonredundant database were searched using *Homo sapiens* as a taxonomic restrictor.

## RT-PCR

RT-PCR was carried out on extracted total RNA as described previously (3). Total RNA from cells was extracted using TRI Reagent® (Molecular Research Center, Inc., Cincinnati, USA). 2 µg of total RNA was reverse transcribed with oligo (dT) and murine leukemia virus (MuLV) reverse transcriptase according to the protocol supplied with the GeneAmp RNA PCR Core Kit (PE Applied Biosystems) and amplified using Tag polymerase. PCR [28 cycles; melting temperature ( $T_m$ )= 58°C] was performed with the GMFG and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) primers. GMFG: forward primer = 5'-AAAGAAGAGGCCTGTGGACAG-3', reverse primer = 5'-TGGTTGTTTCAGGTCCTAGGG-3'; GAPDH: forward primer = 5'-GTATCGTGGAAGA ACTCATGAC-3', reverse primer = 5'-TGCCAGT GAGCTTCCCGTCAGC-3'. The PCR product size for each gene was determined and matched the expected size.

## Bioinformatic analysis for GMFG tissue distribution

High-throughput gene expression profiling has become an important tool for investigating transcriptional activity of the human GMFG gene in a variety of biological samples. We gathered all published microarray gene expression datasets in which the GMFG gene was expressed. We used Su's dataset (<http://expression.gnf.org/>) as a key base, in which gene expression is profiled from 91 human and mouse samples across a diverse array of tissues, organs, and cell lines; 101 unique specimens representing 47 tissue/cell lines are represented. We carried out an integrated bioinformatic analysis on GMFG using Perou's dataset on responses of human mammary epithelial cells to EGF, TGF-beta 1, interferon, and growth on Matrigel; Mao's dataset (4) on human CD34<sup>+</sup> hematopoietic stem/progenitor cells; and Zhang's dataset (18) on human CD34<sup>+</sup> hematopoietic stem/progenitor cells.

## Bioinformatic analysis for the promoter of the GMFG gene

We used two computational prediction tools to search for TF binding sites: the TFSEARCH program (33), which employs the Position Weight Matrix method, and our GeneKey software (34), which uses a novel

Shuffled Matrix method. By averaging the two scores obtained from TFSEARCH and GeneKey, we defined a new score, ranging from 0.0 to 100.0, with a higher score indicating a higher probability of a putative TF binding site. The 5'-flanking 750-bp promoter region of the GMFG gene was analyzed.

## Bioinformatic analysis for GMFG molecular evolution

To better understand the molecular evolution of GMFG, we collected the GMFG gene sequences from multiple species using our GeneKey software (34) on data from the NCBI UniGene and expressed sequence tag (EST) databases, NCBI Human Genome Resources (<http://www.ncbi.nlm.nih.gov/genome/guide/human/>), Ensembl Genome Browser (<http://www.ensembl.org/>), and UCSC Human Genome Browser (<http://genome.ucsc.edu/cgi-bin/hgGateway>). GeneKey software was designed to assemble ESTs to cDNA/mRNA sequences and translate the mRNA to corresponding protein sequences. To establish a phylogenetic tree for GMFG, our Maligner software (35) was used. Another software program, Sandwich Sequence Alignment, was created and used to evaluate the similarities of the third sequence to two, instead of one, family members.

## Acknowledgements

We thank Drs. Rafael Daniel Camerini-Otero and Peggy Hsieh for providing 2D image equipment and 2D analysis software, Dr. Chu-Xia Deng for sharing 2D analysis software, and Dr. Heidi Hoffman for assisting with protein identification.

## Authors' contributions

YS carried out proteomics experiments and drafted the manuscript. LC carried out the cell cultures and PCR experiments. LAL participated in the 2D data analysis and helped to draft the manuscript. HHW performed the bioinformatics study. GPR coordinated the study, participated in its design, and helped to draft the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors have declared that no competing interests exist.

## References

1. Krause, D.S. 2002. Regulation of hematopoietic stem cell fate. *Oncogene* 21: 3262-3269.
2. Ivanova, N.B., et al. 2002. A stem cell molecular signature. *Science* 298: 601-604.
3. Chen, L., et al. 2002. Influence of lineage-specific cytokines on commitment and asymmetric cell division of haematopoietic progenitor cells. *Br. J. Haematol.* 118: 847-857.
4. Mao, M., et al. 1998. Identification of genes expressed in human CD34<sup>+</sup> hematopoietic stem/progenitor cells by expressed sequence tags and efficient full-length cDNA cloning. *Proc. Natl. Acad. Sci. USA* 95: 8175-8180.
5. 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.
6. Inagaki, M., et al. 2004. Sensitive immununoassays for human and rat GMFB and GMFG, tissue distribution and age-related changes. *Biochim. Biophys. Acta* 1670: 208-216.
7. Kawai, Y., et al. 2003. Structure and promoter activity of the human glia maturation factor-gamma gene: a TATA-less, GC-rich and bi-directional promoter. *Biochim. Biophys. Acta* 1625: 246-252.
8. Lim, R., et al. 1987. Endogenous immunoreactive glia maturation factor-like molecule in astrocytes and glioma cells. *Brain Res.* 430: 49-57.
9. Kato, T., et al. 1987. Functional dissociation of dual activities of glia maturation factor: inhibition of glial proliferation and preservation of differentiation by glial growth inhibitory factor. *Brain Res.* 430: 153-156.
10. Lim, R., et al. 1989. Purification and characterization of glia maturation factor beta: a growth regulator for neurons and glia. *Proc. Natl. Acad. Sci. USA* 86: 3901-3905.
11. Lim, R., et al. 1990. Complete amino acid sequence of bovine glia maturation factor beta. *Proc. Natl. Acad. Sci. USA* 87: 5233-5237.
12. Kaplan, R., et al. 1991. Molecular cloning and expression of biologically active human glia maturation factor-beta. *J. Neurochem.* 57: 483-490.
13. Lim, R. and Zaheer, A. 1995. Phorbol ester stimulates rapid intracellular phosphorylation of glia maturation factor. *Biochem. Biophys. Res. Commun.* 211: 928-934.
14. Lim, R. and Zaheer, A. 1996. *In vitro* enhancement of p38 mitogen-activated protein kinase activity by phosphorylated glia maturation factor. *J. Biol. Chem.* 271: 22953-22956.

15. Zaheer, A. and Lim, R. 1996. *In vitro* inhibition of MAP kinase (ERK1/ERK2) activity by phosphorylated glia maturation factor (GMF). *Biochemistry* 35: 6283-6288.
16. Bourgeois, F., *et al.* 2001. Identification and isolation of a full-length clone of mouse GMFB (Gmfb), a putative intracellular kinase regulator, differentially expressed in telencephalon. *Cytogenet. Cell Genet.* 92: 304-309.
17. Muller, W.E., *et al.* 2004. The chemokine networks in sponges: potential roles in morphogenesis, immunity and stem cell formation. *Prog. Mol. Subcell. Biol.* 34: 103-143.
18. Zhang, Q.H., *et al.* 2000. Cloning and functional analysis of cDNAs with open reading frames for 300 previously undefined genes expressed in CD34<sup>+</sup> hematopoietic stem/progenitor cells. *Genome Res.* 10: 1546-1560.
19. Hanash, S. 2003. Disease proteomics. *Nature* 422: 226-232.
20. Ramalho-Santos, M., *et al.* 2002. "Stemness": transcriptional profiling of embryonic and adult stem cells. *Science* 298: 597-600.
21. Schlee, M., *et al.* 2004. Identification of Epstein-Barr virus (EBV) nuclear antigen 2 (EBNA2) target proteins by proteome analysis: activation of EBNA2 in conditionally immortalized B cells reflects early events after infection of primary B cells by EBV. *J. Virol.* 78: 3941-3952.
22. Uchida, M., *et al.* 2001. A functional role of mitogen-activated protein kinase, Erk1 and Erk2, in the differentiation of a human leukemia cell line, UT-7/GM: a possible key factor for cell fate determination toward erythroid and megakaryocytic lineages. *Int. J. Hematol.* 73: 78-83.
23. Nagata, Y., *et al.* 1998. Activation of p38 MAP kinase and JNK but not ERK is required for erythropoietin-induced erythroid differentiation. *Blood* 92: 1859-1869.
24. Boer, A.K., *et al.* 2003. cAMP/PKA-mediated regulation of erythropoiesis. *Leuk. Lymphoma* 44: 1893-1901.
25. Walker, M.G. 2003. Gene expression versus sequence for predicting function: Glia Maturation Factor gamma is not a glia maturation factor. *Genomics Proteomics Bioinformatics* 1: 52-57.
26. Cantor, A.B. and Orkin, S.H. 2002. Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene* 21: 3368-3376.
27. Lutterbach, B. and Hiebert, S.W. 2000. Role of the transcription factor AML-1 in acute leukemia and hematopoietic differentiation. *Gene* 245: 223-235.
28. Gaboli, M., *et al.* 2001. Mzf1 controls cell proliferation and tumorigenesis. *Genes Dev.* 15: 1625-1630.
29. Miyoshi, H., *et al.* 1995. Alternative splicing and genomic structure of the AML1 gene involved in acute myeloid leukemia. *Nucleic Acid Res.* 23: 2762-2769.
30. Fujiki, K., *et al.* 2000. Molecular cloning of carp (*Cyprinus carpio*) leucocyte cell-derived chemotaxin 2, glia maturation factor beta, CD45 and lysozyme C by use of suppression subtractive hybridization. *Fish Shellfish Immunol.* 10: 643-650.
31. Evsikov, A.V. and Solter, D. 2003. Comment on "stemness": transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature". *Science* 302: 393.
32. Perkins, D.N., *et al.* 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20: 3551-3567.
33. Heinemeyer, T., *et al.* 1998. Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. *Nucleic Acids Res.* 26: 362-367.
34. Zhou, Y., *et al.* 2004. Prediction of eukaryotic gene structures based on multilevel optimization. *Chin. Sci. Bull.* 49: 321-327.
35. Hunt, F.Y., *et al.* 2003. An optimization approach to multiple sequence alignment. *Appl. Math. Lett.* 16: 785-790.