Global Regulation of Erythroid Gene Expression by Transcription Factor GATA-1

Running Head: GATA-1 regulated erythroid gene expression

John J. Welch¹, Jason A. Watts², Christopher R. Vakoc², Yu Yao¹, Hao Wang³, Ross C. Hardison³, Gerd A. Blobel¹, Lewis A. Chodosh⁴, and Mitchell J. Weiss¹,⁵

1. The Children’s Hospital of Philadelphia, Division of Hematology, and the University of Pennsylvania, Philadelphia, PA 19104, USA

2. Cell and Molecular Biology and Combined Degree Graduate Programs, The University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA

3. Department of Biochemistry and Molecular Biology, Center for Comparative Genomics and Bioinformatics, Huck Institutes of Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA

4. Department of Cancer Biology, Abramson Family Cancer Research Institute, University of Pennsylvania, 421 Curie Boulevard, Philadelphia, PA 19104, USA

Financial Support: NIH NIDDK R01 DK065806 (R.C.H. and M.J.W.), Abramson Cancer Center of the University of Pennsylvania Pilot Project Grant and Johnson and Johnson Focused Giving Award (M.J.W.). NIH R33 CA94393 (L.A.C.)

5. Corresponding author.
The Children’s Hospital of Philadelphia
Division of Hematology
34th & Civic Center Blvd
Abramson Research Center
Philadelphia, PA 19104
weissmi@email.chop.edu
Fax: (215) 590-4834
Phone:(215) 590-0565

Word counts: Abstract: 199 Text: 4981

Scientific Heading: Hematopoiesis
Abstract

Transcription factor GATA-1 is required for erythropoiesis, yet its full actions are unknown. We performed transcriptome analysis of G1E-ER4 cells, a GATA-1-null erythroblast line that undergoes synchronous erythroid maturation when GATA-1 activity is restored. We interrogated over 9,000 transcripts at six timepoints representing the transition from late BFU-E to basophilic erythroblast stages. Our findings illuminate several new aspects of GATA-1 function. First, the large number of genes responding quickly to restoration of GATA-1 extends the repertoire of its potential targets. Second, many transcripts were rapidly downregulated, highlighting the importance of GATA-1 in gene repression. Third, upregulation of some known GATA-1 targets was delayed, suggesting that auxiliary factors are required. For example, induction of the direct GATA-1 target gene β major globin was late, and surprisingly, required new protein synthesis. In contrast, the gene encoding Fog1, which cooperates with GATA-1 in β globin transcription, was rapidly induced independently of protein synthesis. Guided by bioinformatic analysis, we demonstrated that selected regions of the Fog1 gene exhibit enhancer activity and in vivo occupancy by GATA-1. These findings define a regulatory loop for β globin expression, and more generally, demonstrate how transcriptome analysis can be used to generate testable hypotheses regarding transcriptional networks.
Introduction

Studies of red blood cell development provide an important basis for understanding general mechanisms of gene regulation and tissue formation. Erythrocyte maturation is largely dedicated to producing hemoglobin and supporting its major function of oxygen delivery. Acquisition of this specialized phenotype is controlled by a concerted differentiation program that serves both generalized and cell type-specific requirements. Production of hemoglobin is coordinated to avoid the toxic accumulation of globins or heme synthetic intermediates. Special adaptations are required for iron import and defense against reactive oxygen species. The cell membrane must undergo remodeling to present developmentally appropriate cytokine receptors and ion transporters. In addition, erythropoiesis requires active mechanisms to prevent apoptosis and coordinate cell cycle arrest with terminal maturation.

Erythroid differentiation is highly regulated at the level of mRNA transcription and factors controlling the expression of numerous erythroid genes, particularly globins, have been extensively characterized. Recent studies of primary hematopoietic precursors 1-6 and immortalized cell lines 7,8 are beginning to define global gene expression patterns during hematopoietic maturation, although a comprehensive view of erythropoiesis is lacking. This information is essential for new gene discovery and to better understand the regulatory networks that specify the formation and function of red blood cells. One general approach to understanding gene regulation during development is to identify lineage-determining transcription factors and then define the genes that they control. In this regard, studies of the nuclear protein GATA-1 have provided numerous insights into erythropoiesis 9-11.
GATA-1 is a transcription factor essential for erythrocyte, megakaryocyte, mast cell and eosinophil differentiation. Ablation of the GATA-1 gene in mice causes embryonic death from severe anemia due to maturation arrest and apoptosis of committed erythroid precursors. In humans, missense mutations in GATA-1 cause dyserythropoietic anemia and/or thalassemia. While virtually all erythroid genes examined are positively regulated directly by GATA-1, the full extent of its transcriptional actions is currently unknown. Defining the genes regulated by GATA-1 more comprehensively should better elucidate the molecular pathways that control erythropoiesis.

To examine the role of GATA-1 in erythroid development, we created G1E cells (for GATA-1–null line derived from gene-targeted embryonic stem cells). G1E cells proliferate in culture as immature erythroblasts and undergo terminal erythroid maturation when GATA-1 function is restored. Thus, complementation of a defined loss-of-function mutation generates a synchronized cohort of differentiating erythroid precursors, an ideal system in which to delineate gene expression at specific stages of erythropoiesis and to investigate the transcriptional effects of GATA-1 in a physiologic context. Previously, we used subtractive hybridization to analyze transcripts upregulated by GATA-1 in G1E cells. These studies identified known GATA-1 targets and also new genes important for erythroid maturation, but examined a limited number of transcripts at only one point in differentiation.

Here we report transcriptome analysis using Affymetrix oligonucleotide microarrays to simultaneously interrogate 12,451 murine transcripts corresponding to 9266 unique known genes and ESTs. RNA from G1E cells was examined at baseline and at five timepoints after
Welch et al., GATA-1-regulated erythroid gene expression

synchronous terminal maturation was induced by an estradiol-activated form of GATA-1. This model system allowed us to define both immediate and delayed actions of GATA-1 at specific developmental stages of erythropoiesis.

Numerous erythroid markers and known GATA-1 target genes were induced, which, together with prior studies\textsuperscript{22,24-29}, validate the use of G1E cells as a model system for erythropoiesis. Our work illustrates several new aspects of GATA-1 function and erythroid biology. A surprising number of ESTs and transcripts encoding known proteins were altered at early timepoints, identifying genes likely to be regulated directly by GATA-1 and providing new insights into the control of erythropoiesis. Moreover, GATA-1 rapidly inhibited as many transcripts as it induced, highlighting an underappreciated role in gene repression. In addition, upregulation of a subset of established GATA-1 targets was relatively delayed, suggesting that their activation requires induction of accessory factors. One important example was $\beta$ major globin, whose mRNA accumulated only at late timepoints. $\beta$ globin induction required new protein synthesis whereas the gene encoding the GATA-1 cofactor, Fog-1 (\textit{Fog1, Zfpm1}), was induced rapidly and independent of protein synthesis. Here we show that GATA-1 also binds the \textit{Fog1} locus \textit{in vivo} at a site within an enhancer, suggesting a direct mechanism for gene activation. The latter tests were guided by bioinformatic predictions based on genomic sequence alignments and knowledge of factor-binding sites, illustrating the power of transcriptome analysis coupled with comparative genomics techniques. Together, these studies identify a regulatory hierarchy in which GATA-1 induces its transcriptional cofactor to cooperate in activation of $\beta$ major globin gene expression.

Hence, detailed kinetics of GATA-1-regulated gene expression provided by the current studies
forms a basis to generate experimentally testable hypotheses regarding the transcriptional cascades that coordinate terminal erythrocyte maturation.

**Methods**

**Cell Culture**

G1E-ER4 cells were cultured as described previously.

**Western Blotting**

GATA-1 was detected in Western blots with antibody N-6 from Santa Cruz Biotechnology (Santa Cruz, CA) at 1:10000 dilution.

**Microarray Experiments**

In three independent experiments, G1E-ER4 cells growing in log phase were induced for 0, 3, 7, 14, 21 or 30 hours with $10^{-7}$ M $\beta$-estradiol. RNA from $5 \times 10^7$ G1E-ER4 cells was extracted using Trizol reagent (Invitrogen, Carlsbad, CA), and processed for hybridization to Affymetrix MG-U74Av2 GeneChips.

Analysis was performed with Affymetrix MAS 5.0 software using the statistical algorithm. Chip-wide signal was scaled to a value of 150 using all probesets, without additional normalization, using default parameters. The signal for the three replicates of each probeset at each timepoint was averaged. Average signal log$_2$ ratios (SLRs) were computed for comparative analyses by averaging the linearized SLR values for each probeset triplicate, followed by reconversion to a base 2 logarithm. Annotations were extracted from the Unigene and GO.
Welch et al., GATA-1-regulated erythroid gene expression databases using Genbank accessions numbers assigned by Affymetrix. Results and annotations were merged into a Microsoft Access database.

For comparative analyses between two timepoints, we adopted a stringent numerical filtering strategy based on Cammenga et al. \(^4\): all p-values for each comparison in a triplicate had to satisfy the condition that \((p \text{ value } < 0.005 \text{ or } p \text{ value } > 0.995)\), and at least one of the two average signals being compared (baseline or experimental) had to be above a threshold value of 100, with an \(|\text{SLR}| > 1.5\).

Hierarchical clustering of unfiltered data was performed using dCHIP \(^{34}\). After normalization to the median intensity array, “PM only” model-based expression was calculated with default settings. Triplicates were pooled for each timepoint and genes were filtered using the following criteria: variation across samples after pooling: \(0.6 < \text{standard deviation/mean} < 10\); % Present calls in the arrays used >= 15%; replication variation such that \(0 < \text{median} < 0.5\) and signal level >= 100 in >= 15% of samples. These criteria were selected to obtain a visual representation of about 100 genes that changed the most over the thirty hour timecourse.

**Gene expression Analysis**

Northern blotting was performed as previously described \(^{23}\). To detect \(\beta\) major globin and Fog1 RNAs, total RNA was converted to cDNA using reverse transcriptase according to standard methods. Real time PCR was performed in 25 \(\mu\)l reactions (1X SYBR Green Master Mix (QIAGEN), 10 ng cDNA, 0.6 \(\mu\)M primer) using an ABI prism 7000 with standard amplification conditions. mRNA levels were normalized to those of glyceraldehyde phosphate dehydrogenase
Welch et al., GATA-1-regulated erythroid gene expression

(Gapdh). The primers were as follows: β major globin: forward, 5’TCCCGTCAACTTCAAGCTCCT3’, reverse, 5’GGAATTTGTCCAGAGAGGCATG3’; Fog1: forward, 5’ATCCCCTGAGAGAGAAGAACCG3’, reverse, 5’GGCGTCATCCTCCTCTGTAGATC3’; Gapdh: forward, 5’GATGCCCCCATGTTTGTGAT3’, reverse, 5’GGTCATGAGCCCTTCCACAAT3’.

Predicting cis-regulatory modules in Fog1 by comparative genomics.

All bioinformatic predictions used whole-genome sequence alignments of mouse and human generated by blastZ 35 or mouse, rat and human generated by multiZ 36. For a measure of selection that indicates biological function, we used L-scores, which give a log-likelihood estimate of the probability that an alignment would not be generated at the locally adjusted neutral rate 37. The conservation score data were plotted in the UCSC Browser display 38. The regulatory potential score (RP score) is a log-likelihood measure that estimates a probability that the function under selection is a cis-acting gene regulatory element 39. In this method, alignments are scored for how well they match patterns that distinguish alignments in regulatory regions from those in neutral DNA. The database of Genome Alignments and Annotations (GALA, 40 was searched to find DNA intervals in the Fog1/Zfpm1 locus whose alignments had an RP score of at least 2.3, an emperically derived threshold based on identification of known regulatory elements in the Hbb locus 41. GALA was also queried for all mouse sequences in the Fog1 locus that match weight matrices for GATA-1 binding sites in TRANSFAC 42 and for blocks in the mouse-rat-human three-way alignments containing conserved matches to GATA-1 weight matrices (computed by tffind) 43. The DNA intervals exceeding the RP score threshold that also contain a conserved, predicted GATA-1 binding site were identified by intersection operations in
GALA. Data from GALA were displayed as custom tracks in the UCSC Genome Browser to provide a consolidated view.

**Generation of G1E-ER4 cells overexpressing bcl-xL**

The coding region of mouse bcl-xL was cloned into the retroviral vector pGD and G1E-ER4 cells were transduced as described. A polyclonal population of cells expressing ectopic bcl-xL was generated by culturing the transduced cells for 24 hours in the absence of erythropoietin and kit ligand. Under these conditions virtually 100% of G1E-ER4 cells transduced with vector alone died, while about 50% of the bcl-xL–transduced cells survived. After this selection step, all of the bcl-xL–transduced cells survived 24 hour cytokine deprivation and underwent erythroid differentiation in response to estradiol-induced activation of GATA-1.

**Chromatin Immunoprecipitation.**

Chromatin immunoprecipitation (ChIP) assays were performed as described. PCR products were quantified using SYBR green dye on an ABI 7000 real-time PCR machine. PCR product signals were referenced to a dilution series of the relevant input. Primer pairs used for PCR: β globin HS3: forward, 5’ ATGGGACCTCTTGATAGACACATCTT3’; reverse, 5’ CTAGGGACTGAGAGGCTGCTT3’. *Fog1* upstream region: forward, 5’ GGCAGATGTTCACTGTGGCA3’, reverse, 5’ GGGAGGAGCCAGAGGTCAG3’; *Fog1* preCRM1: 5’ TGCAAGTCCCCATCTGGATAAGA3’, reverse, 5’ GCACGAGGACTAGACACATCTT3’. *Fog1* preCRM4: forward, 5’ GCACGACTAGACACATCTT3’. *Fog1* preCRM4: forward, 5’ GCGATAACGGGCACCTAGAGC3’, reverse, 5’ AAGCGAGCGGAGCCG3’.
Enhancer assays by transient transfection

The ability of predicted cis-regulatory module 1 (preCRM1 or R1) to enhance the level of expression from an erythroid promoter was tested by measuring its effect on expression of a gamma-globin-promoter-luciferase reporter gene plasmid, gammaLuc \(^{46}\) after transient transfection of uninduced K562 cells, using methods described in Elnitski et al. \(^{32}\). The DNA segment from mouse Chromosome 8, positions 121975271-121975514 (February 2003 assembly, mm3 \(^{37}\)) contains preCRM1, located in the first intron of Fog1. It was added upstream of the promoter in the plasmid gammaLuc to make R1gammaLuc. The parental and test plasmids were transfected in triplicate into K562 cells using Tfx50 (Promega), along with a control plasmid with a Renilla luciferase reporter gene. The firefly luciferase activity encoded by gammaLuc plasmids and the Renilla luciferase activity from the control were measured using the Promega Dual-Luciferase Reporter (DLR) Assay kit, and the ratio of these activities (FF/Ren) were calculated for each replicate. The experiment was performed twice. In order to present the results on a similar scale, the FF/Ren ratios were normalized by dividing by the mean FF/Ren value for the parental gammaLuc samples in each experiment.

Results and Discussion

The complete microarray dataset described below was deposited at the GEO (accession GSE628). “Supplementary data” refers to analysis, raw data, and the G1EDb database containing results and annotations found at the URL http://stokes.chop.edu/web/weiss/G1Eindex.html and its subdirectories.
We used G1E-ER4 cells, a G1E subclone that constitutively expresses a conditional, estradiol-activated form of GATA-1 (GATA-1-ER, GATA-1 fused to the estrogen receptor ligand binding domain)\textsuperscript{25,26,47}. Upon addition of estradiol, G1E-ER4 cells undergo synchronous and homogeneous erythroid maturation (\textbf{Figure 1A}). G1E-ER4 cells express GATA-1-ER at a level similar to that of endogenous GATA-1 in murine erythroleukemia (MEL) cells (\textbf{Figure 1B}) and estradiol has no apparent effects on parental G1E cells that do not express GATA-1-ER (references \textsuperscript{25,47} and not shown). These control experiments indicate that erythroid maturation of G1E-ER4 cells is caused by physiologic restoration of GATA-1 and is not due to its overexpression or irrelevant estradiol effects.

We used Affymetrix GeneChips (MG-U74Av2) to profile mRNA expression in G1E-ER4 cells at 0, 3, 7, 14, 21 and 30 hours after activation of GATA-1. These timepoints were selected to examine both early and late effects of GATA-1 and to encompass several key aspects of erythroid maturation, including loss of proliferative capacity, extensive hemoglobin synthesis and characteristic morphological changes\textsuperscript{25,47} (\textbf{Figure 1A}).

Three biologic repeats were performed for each timepoint to minimize random variations in gene expression. There was strong agreement in datasets from identical timepoints of separate experiments (\textbf{Supplementary Data}). In filtering genes which changed versus time zero, we favored specificity over sensitivity, employing stringent criteria requiring large fold induction and highly significant p-values; for a full listing of genes considered to be induced/repressed at each timepoint refer to \textbf{Supplementary Data}. 
Among the 12,451 murine probesets printed on the microarray, 3419 (27.4%) were designated present (in triplicate) in at least one timepoint. In general, activation of GATA-1 triggered numerous rapid changes in gene expression: at 3 hours, 1.3% of the probesets represented on the microarray were increased and 0.8% were decreased versus time zero. By 30 hours, 5.6% were increased, and 10.5% were decreased, consistent with a more restricted pattern of gene expression during cellular maturation.

Data were compiled into a custom database, G1EDb, available for public download (Supplementary Data). Expression data and corresponding annotations can be retrieved gene-by-gene in a visually intuitive manner and filtered according to present/absent calls, numerical criteria, and functional categories. The G1EDb also includes hypertext links to public and Affymetrix databases. Our microarray data have also been deposited at the NCBI GEO 48, accession GSE628.

Validating the experimental system

We addressed two important validation concerns: does the experimental system recapitulate GATA-1-dependent erythropoiesis and does the microarray accurately reflect transcript levels? Regarding the first issue, numerous erythroid-expressed genes, many of which are known GATA-1 targets, were upregulated after GATA-1 activation (Figure 2). In addition, our dataset is in agreement with an earlier study that examined expression of 600 genes during erythroid maturation of primary fetal liver cells ³ (Figure 3). Although erythroid and megakaryocytic lineages arise from a common precursor and are both GATA-1 dependent, platelet-expressed
transcripts such as platelet factor 4, glycoprotein 1b, coagulation factors 8 and 10, and integrin β3 were not detected in G1E-ER4 cells (Supplementary Data). Microarray transcript profiles were in agreement with the kinetics of RNA expression demonstrated by Northern blotting, as described previously 23,25,47 and in Figure 4. In all cases examined, the microarray data accurately reflected trends in gene expression. For both α and β globins, the microarrays tended to underestimate the magnitude of RNA induction, possibly due to saturation of the detection system by these extremely abundant transcripts.

**Developmental Stages of G1E cell maturation defined by gene profiling**

Erythroid precursors develop through distinct stages that are defined by proliferative capacity, cytokine requirements, cell surface markers, morphology and patterns of gene expression 49. G1E cells without GATA-1 most closely resemble the late burst forming unit erythroid (BFU-E) stage 25,27,47. They exhibit an immature morphology, are negative for the proerythroblast marker Ter 119 50, and require kit ligand (KL, stem cell factor) for proliferation and survival. By 24 hours after GATA-1 activation, G1E cells become division-arrested, down-regulate c-Kit, the receptor for KL (see below), and require erythropoietin to prevent apoptosis. By 30 hours, most of the cells stain for hemoglobin and resemble basophilic erythroblasts. The maturation of G1E cells in relation to designated developmental stages of erythropoiesis is further defined by the current gene expression studies (Figure 5). For example, expression kinetics of erythrocyte membrane protein Rh50 (Rhag), glycophorin A (Gypa) and band III (Slc4a1) mRNAs during G1E cell maturation resembled those observed in primary erythroid cultures, with Rh50 appearing first and band III appearing last on basophilic erythroblasts 51. Several developmental
switches in protein isoforms that occur during normal erythroid maturation, including the shifts from carbonic anhydrase 1 (Car1) to carbonic anhydrase 2 (Car2) \(^52\) and transferrin receptor 2 (Trfr2) to transferrin receptor 1 (Trfr1) \(^53\), were recapitulated during GATA-1-induced maturation of G1E cells (Figure 5).

**Global GATA-1-induced changes in gene expression**

Analysis of transcripts according to their expression kinetics provides a basis to define the regulatory hierarchy through which GATA-1 controls erythroid development. A broad overview of transcripts which change during GATA-1-induced maturation of G1E cells is illustrated by hierarchical clustering (Figure 6). This method selects the genes which change the most over the entire 30 hour time course and graphically sorts them according to the similarity of their transcriptional profiles. Since each transcript is scaled relative to its own median value for the experiment, all transcripts can be viewed at once, despite a range of absolute signal intensities. As a complementary approach, we identified the most significant changes in gene expression at each timepoint relative to the baseline level at time zero. This pairwise comparison was particularly useful to identify rapidly modulated genes (Table I, and Supplementary Data). Together, these studies define groups of genes that are regulated by GATA-1 at early and late timepoints. These genes represent various functional classes including transcription factors, signaling molecules, enzymes and structural proteins. Insights into GATA-1 actions and erythroid development deriving from these analyses are discussed below.
GATA-1 as a repressor of transcription. One of the most striking findings is that numerous transcripts were rapidly downregulated, highlighting a potential role for GATA-1 in gene repression. Traditionally, GATA-1 is viewed as an activator of erythroid-specific transcripts; potential roles in gene repression have been implicated in prior studies, but are poorly understood \[25,54-57\]. The current findings indicate that gene repression by GATA-1 is more prevalent than previously recognized and likely to play a critical role in hematopoiesis. We recently demonstrated that repression of Gata2 \[58\] and Myc \[25\] in erythroid cells is associated with GATA-1 binding to the respective genes, indicating potentially direct mechanisms for transcriptional inhibition. GATA-1 may repress gene expression through its interaction partner Fog-1 (Zfpm1, “Friend of GATA-1”), which binds co-repressors and inhibits GATA-1-mediated transcription in specific promoter and cellular contexts \[45,57,59\].

Transcription factors were prominent among the most rapidly repressed transcripts, for example: Brf2, Eto-2 (Chfa2t3h), Gata2, Lyl1, Nab2, Ier2, and Tieg (Table I). It is likely that many pleiotropic effects of GATA-1 are transmitted through modulation of some or all of the above nuclear proteins, although their specific roles in erythropoiesis require further study. GATA-2 is of particular interest; this GATA-1 relative is expressed in hematopoietic stem cells and/or early progenitors \[60\] and is derepressed in GATA-1\(^{-}\) erythroblasts, which may account for their baseline transcription of GATA-dependent target genes \[13,61\]. These findings suggest functional overlap between GATA-1 and GATA-2 \textit{in vivo}. However, distinct roles for these proteins are illustrated by the ability of GATA-1, but not GATA-2 to repress transcription through a defined \textit{cis} element in the GATA-2 gene \[45,58,59\]. Hence, it is possible that competition between GATA-1 and GATA-
2 regulates the transcriptional output of some GATA-regulated genes and reflects a requirement for GATA-2 repression during normal erythropoiesis.

Also of note, several genes associated with oncogenic transformation were downregulated after GATA-1 activation. These include Myc, Myb, Kit, and Nab2. GATA-1 binding to the Myc gene is associated with its transcriptional repression during erythroid maturation. GATA-1 may also repress Myb through cognate elements in the proximal promoter. Hence, GATA-1 may promote hematopoietic maturation by repressing mitogenic genes. This hypothesis is consistent with recent findings that link GATA-1 mutations to a subset of megakaryocytic leukemias.

**Genes that are rapidly induced by GATA-1 (Table I).** Some of the earliest upregulated genes might encode proteins that participate in the initial cascade of GATA-1-induced maturation. For example, transcripts encoding EKLF (Klf1) and Fog-1 (Zfpm1), nuclear proteins that augment GATA-1 activities, were induced at early timepoints (see also below). Another rapidly upregulated gene was Bach1, which encodes a CNC-type transcription factor that interacts with control elements in the β globin gene locus to coordinate transcription with heme availability. Surprisingly, cytokine receptor genes Csf2rb1 and Csf2rb2 were strongly upregulated within the first three hours (Table I). Csf2rb2 encodes a shared β subunit (βc), which pairs with unique α chains to form IL-3, IL-5 or GM-CSF receptors, while Csf2rb1 encodes βIL-3 which pairs only with the IL-3α subunit. As IL-3 is known to stimulate erythropoiesis in vitro at the BFU-E-stage, it is unexpected to find that its receptor subunits are upregulated at later stages of erythropoiesis in G1E-ER4 cells. Nonetheless, similar findings were observed during maturation of fetal liver primary erythroid precursors. This suggests potential functions for
Csf2rb1 and Csf2rb2 in erythropoiesis. Supporting this, Csf2rb2 was observed to potentiate erythropoietin signaling through direct physical interaction with the erythropoietin receptor 68, although no defects in steady-state erythropoiesis or erythropoietin signaling were demonstrated in mice with targeted mutations in Csf2rb1 and Csf2rb2 69. In addition to the above examples, numerous other signaling molecules and transcription factors whose role in erythroid development is currently unexplored were rapidly upregulated (Table I).

Facilitating new insights into erythroid function and physiology. Many of the genes upregulated by GATA-1 at early and late timepoints are known to participate in specialized aspects of terminal erythropoiesis (Figure 2, above). Examination of additional induced and repressed genes not previously implicated in erythroid development or function should provide further insight into these processes (Table I, Figure 6 and Supplementary Data). To facilitate the discovery of genes that participate in erythrocyte biology, we annotated transcripts in the G1EDb according to functional classes as predicted by prior studies or the presence of signature protein motifs. These include transcription factors, signaling molecules, antioxidants and others. Data retrieval according to gene function and expression patterns provides a broad overview of specific aspects of erythropoiesis and illustrates potential new pathways that can be tested experimentally in future studies. Analysis of GATA-1-regulated gene expression according to functional categories is discussed in Supplementary Data. In addition, several poorly characterized ESTs were upregulated by GATA-1 in G1E-ER4 cells. As indicated by our prior studies 23,24, characterization of these ESTs should identify novel proteins that participate in erythroid development or function.
Regulatory hierarchies defined by the kinetics of GATA-1-regulated gene expression.

Numerous established erythroid GATA target genes including those encoding alpha globin \((Hba-a1)\), band 3 \((Slc4a1)\) and \(\beta\) major globin \((Hbb-b1)\) exhibited relatively delayed upregulation (Figures 4, 6 and 7A). Therefore, it is possible that some GATA-1 targets require modulation of additional transcription factors and/or intracellular signaling pathways for full activation. We examined this possibility experimentally, focusing on \(\beta\) major globin.

Delayed kinetics of \(\beta\) major globin induction predicted by the microarrays is shown in Figure 7A, left panel. In contrast, RNA encoding Fog-1, a cofactor that participates with GATA-1 in \(\beta\) major globin transcription \(^{26,45,70}\), was activated more rapidly (Figure 7A, right panel). These data suggest that GATA-1 could facilitate \(\beta\) major globin production by first stimulating transcription of \(\text{Fog1}\). To explore this further, we investigated whether either of these genes are induced by GATA-1 independently of new protein synthesis. Initially, it was not possible to examine gene expression in G1E-ER4 cells treated with the protein synthesis inhibitors cycloheximide or anisomycin, as these drugs induced rapid apoptosis (data not shown). To circumvent this problem, we transduced G1E-ER4 cells with retrovirus encoding the anti-apoptotic protein bcl-\(x_L\). In response to GATA-1 activation, G1E-ER4-bcl-\(x_L\) cells underwent erythroid maturation similar to the G1E-ER4 parental line, but survived for at least 12 hours in the presence of cycloheximide (not shown). We used real-time PCR to determine whether cycloheximide inhibits activation of \(\text{Fog1}\) or \(\beta\) major globin by GATA-1 in G1E-ER4-bcl-\(x_L\) cells (Figure 7B). After treatment with estradiol alone, both mRNAs were strongly induced. Cycloheximide treatment, which blocked >95% of protein synthesis (not shown), completely
inhibited estradiol-induced β major globin RNA upregulation, indicating that additional downstream GATA-1 targets are required. In contrast, *Fog1* mRNA was fully induced after activation of GATA-1 in the presence of cycloheximide (Figure 7B). These findings raise the possibility that GATA-1 activates *Fog1* transcription directly and that this might be necessary for full β major globin gene induction.

To test whether *Fog1* is a direct GATA-1 target gene, we first applied bioinformatic predictions to find candidate GATA-1 binding sites in *Fog1* that are likely to be involved in transcriptional regulation (Figure 8A). Whole-genome sequence alignments between mouse and human were analyzed to find sequences likely to be functional (conservation score) and sequences more likely to be involved in regulation of gene expression (regulatory potential 39). Also, matches to weight matrices for GATA-1 binding sites that are conserved in human, mouse and rat were identified. As shown in Figure 8A, 17 intronic segments have prominent peaks for both regulatory potential and conservation and all of these contain one or more conserved GATA-binding sites. The presence of potential GATA sites in regulatory regions of erythroid-expressed genes is not unexpected. However, the *Fog1* locus is remarkable in the large number of predicted regulatory regions which all contain conserved GATA motifs. This suggests that GATA-1 may be a critical protein in regulating *Fog1* and that the multiple binding sites could mediate a concentration-dependent effect.

Two regions in the *Fog1* locus, designated preCRM1 and preCRM4 (for predicted *cis*-regulatory module), were examined for *in vivo* GATA-1 occupancy by chromatin immunoprecipitation (ChIP) (Figure 8B). An upstream region of the *Fog1* promoter without any conserved GATA
sites was selected as a negative control (labeled as “U” in Figure 8A and as “upstream” in Figure 8B). ChIP was performed in G1E-ER4 cells before and 24 hours after estradiol treatment. Occupancy of the GATA-1-ER fusion protein was measured using two different antibodies, one directed against GATA-1 and one against the ER moiety. Both CRM1 and CRM4 regions exhibited significant GATA-1 occupancy after estradiol treatment, as detected with both antibodies; roughly 50-100 fold greater than isotype control or irrelevant antibodies and comparable to GATA-1 occupancy at hypersensitive site (HS) 3 of the β major globin locus. No GATA-1 occupancy was detected at the negative control site upstream of Fog1 exon 1.

Next, we tested preCRM1 for its ability to augment gene expression in erythroid cells. We inserted this region upstream of a minimal γ-globin gene promoter linked to a luciferase reporter gene and measured the level of expression after transient transfection into erythroid cells. As illustrated in Figure 7D, preCRM1 augmented expression of the linked reporter gene an average of about 7-fold in erythroid cells in two independent experiments. Together these results demonstrate that GATA-1 occupies functionally important regions of the Fog-1 locus and is likely to play a direct role in activating its transcription. More generally, the results verify the utility of our microarray database in combination with bioinformatics to predict genomic sites where GATA-1 acts to facilitate the erythroid program of transcription.

Considering that Fog-1 interacts with GATA-1 both functionally and physically and that both nuclear proteins are required for optimal β globin synthesis, our data are consistent with a feed-forward loop in which GATA-1 induces Fog1 to activate β major globin transcription (Figure 8D). This is in accord with prior studies demonstrating that Fog1 is induced by GATA-1
Welch et al., GATA-1-regulated erythroid gene expression independent of its interaction with the Fog-1 protein. The model predicts that GATA-1 binds the Fog1 locus at early timepoints prior to β major globin induction during induced maturation of G1E-ER4 cells; detailed kinetic studies are underway to test this. The feed-forward loop is a common mode of gene regulation described in prokaryotes and in the development of muscle, liver and pancreas. The current example illustrates how a tissue-specific transcription factor (GATA-1) can positively regulate the production of its own cofactor (Fog-1) which functions in both gene activation and repression. Of note, low levels of β globin mRNA are expressed without GATA-1 in G1E-ER4 cells and in primary GATA-1– erythroblasts. Also, Fog1 mRNA is expressed basally in G1E-ER4 cells prior to activation of GATA-1. Therefore, it seems likely that GATA-1 and Fog-1 act in a dose-dependent fashion and that relatively high threshold levels are required for optimal β major globin gene activation. Specific dosage dependency thresholds also probably exist for genes that are repressed by GATA-1 in a Fog-1-dependent manner. Finally, our data do not exclude the possibility that additional GATA-1-induced signaling pathways or transcription factors are required for β major globin transcription. If so, these additional genes should be represented among the cohort of rapidly induced transcripts in G1E-ER4 cells. One such candidate is EKLF, which is activated by GATA-1 in G1E-ER4 cells (reference 73, Figure 2 and Table I), is known to be a direct GATA-1 target, and cooperates with GATA-1 to activate β major globin transcription.

Defining Erythroid Development Through GATA-1 Gene Complementation

Use of microarrays for mRNA profiling is a powerful tool for examining gene expression during hematopoiesis. The current study validates the utility of G1E cells as a biologically
relevant model system and provides a comprehensive transcriptome analysis of the transition through several well-defined stages of erythropoiesis, illustrating new aspects of its metabolic and developmental control. In addition, our experimental system provides a unique view into erythroid maturation from the perspective of GATA-1, further defining its regulatory hierarchy and mechanisms of action. In this regard, the kinetics of gene expression in G1E cells, which are arrested by loss of GATA-1, may differ somewhat from that in maturing primary erythroid cells where GATA-1 is induced in a more gradual and physiologic manner. Therefore, it will be interesting and important to contrast the current findings to more comprehensive transcriptome studies performed using other model systems in which the GATA-1 gene is intact and under endogenous control. Such comparisons should better define the developmental block conferred by loss of GATA-1. Ultimately, it is likely that these complementary experimental approaches will intersect to provide a comprehensive understanding of the transcriptional program in erythropoiesis. In the future, it will also be important to analyze G1E-ER4 cells and other model systems using proteomic approaches, as many aspects of erythroid gene expression are regulated at the level of protein translation, modification or degradation.
Acknowledgements

We thank Ernst Müller and Jeffrey Miller for sharing their cDNA and subtraction hybridization data. We are also grateful to Mortimer Poncz for reviewing this manuscript.
References

Welch et al., GATA-1-regulated erythroid gene expression


Welch et al., GATA-1-regulated erythroid gene expression


54. Li Q, Peterson K, Stamatoyannopoulos G, Raich N. GATA1 and YY1 are developmental repressors of the human epsilon-globin gene. Blood. 1994;84:4344-4353


71. Tsang AP, Fujiwara Y, Hom DB, Orkin SH. Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. Genes Dev. 1998;12:1176-1188


<table>
<thead>
<tr>
<th>Probe</th>
<th>Gene</th>
<th>Description</th>
<th>Function</th>
<th>On 3h</th>
<th>Off 3h</th>
<th>SLR</th>
</tr>
</thead>
<tbody>
<tr>
<td>94246</td>
<td>Cdhl1</td>
<td>GATA-1-regulated factor 2 receptor, beta 1</td>
<td>Signaling</td>
<td>2 (2.7)</td>
<td>4.24</td>
<td>1.7</td>
</tr>
<tr>
<td>95020</td>
<td>SoxA1</td>
<td>soluble carrier family 9 (octaspan/160-exchange family)</td>
<td>Membrane transporter</td>
<td>66 (54-65)</td>
<td>3.9</td>
<td></td>
</tr>
<tr>
<td>94415</td>
<td>Btf3</td>
<td>BTB and CNC homology</td>
<td>Transcription factor</td>
<td>9 (8-8)</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>94747</td>
<td>Cdhl1</td>
<td>GATA-1-regulated factor 2 receptor, beta 1</td>
<td>Signaling</td>
<td>1 (1.1)</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>103210</td>
<td>Rhdx</td>
<td>HoxB9/HoxC9 regulatory genes</td>
<td>Signaling</td>
<td>65 (2.7)</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>103300</td>
<td>Rhdx</td>
<td>HoxB9/HoxC9 regulatory genes</td>
<td>EST</td>
<td>190 (1.2)</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>94628</td>
<td>Rhdx</td>
<td>HoxB9/HoxC9 regulatory genes</td>
<td>EST</td>
<td>129 (9.2)</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>94109</td>
<td>Lef1</td>
<td>Lef1</td>
<td>Transcription factor</td>
<td>512 (10.5)</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>93000</td>
<td>Atp5f3</td>
<td>ATP synthase subunit A, mitochondrial</td>
<td>Transcription factor</td>
<td>108 (5.2)</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>94315</td>
<td>Rhdx</td>
<td>HoxB9/HoxC9 regulatory genes</td>
<td>EST</td>
<td>117 (3.5)</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>102390</td>
<td>Rhdx</td>
<td>HoxB9/HoxC9 regulatory genes</td>
<td>EST</td>
<td>58 (2.1)</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>100210</td>
<td>Rfxn4</td>
<td>RFXN4</td>
<td>Signaling</td>
<td>142 (9.2)</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>100262</td>
<td>Nkx2</td>
<td>Nkx2</td>
<td>Signaling</td>
<td>58 (2.1)</td>
<td>1.6</td>
<td></td>
</tr>
</tbody>
</table>

**Table I. Early GATA-1-regulated genes.** Microarray results were filtered at an SLR of 1.5, P value < 0.005, and threshold of 100 for the three and seven hour timepoints versus time zero. The penultimate column indicates signal values for the indicated timepoints (average for triplicate experiments).
Figure Legends

Figure 1. GATA-1-mediated erythroid maturation of G1E-ER4 cells.

A) Morphologic maturation and accumulation of hemoglobin after activation of GATA-1. GATA-1 was activated by estradiol at time zero and cells were examined by cytocentrifugation and histological staining at the indicated timepoints. MGG, May-Grunwald Giemsa stain; BZ, benzidine stain for hemoglobin. B) GATA-1-ER expression in G1E-ER4 cells. Western blots of whole cell lysates at various timepoints after activation of GATA-1-ER by estradiol. MEL cell lysates at 0 and 96 hours after induction of differentiation with 2% DMSO were used for comparison. 10 µg protein per lane.

Figure 2. Activation of erythroid genes by GATA-1 in G1E-ER4 cells. Each column represents a timepoint from 0 to 30 hours after GATA-1 activation, and each row is assigned to a transcript. The color scale ranges from blue to red, corresponding to values ranging from -2.5 to +2.5 standard deviations about each transcript’s mean, with white representing the mean value. Hence, a shift from blue to red indicates induction. Signal values (average of three triplicates) at 0 and 30 hours are shown in parentheses; these correspond roughly to transcript abundance. Transcripts shown here underwent at least a two-fold change from time zero, and had an average signal value of at least 100 in at least one timepoint.

Figure 3. Concordance of gene expression during terminal maturation of G1E-ER4 cells and primary fetal liver-derived erythroid cells (FLDECs). Signal log ratio (SLR) of transcript levels in differentiated versus undifferentiated G1E-ER4 cells (Y axis, 0 versus 30 hours) and FLDECs induced to undergo erythroid maturation (X axis, 0
Welch et al., GATA-1-regulated erythroid gene expression versus 48 hrs). Each symbol represents one gene that met nonbiased filtering criteria (SLR > 2, P < 0.005, threshold 100 in both systems). Upper right and lower left quadrants show transcripts with concordant expression trends in both experimental systems: Lower left quadrant (down-regulated in both): Brca2, Cdc20, Hsp60, Myb12, Gstp2, Msh2, Akt (Pkb), Myb, Myc, Kit, Pcna, Mre11a, Nme2. Upper right quadrant (up-regulated in both): Ccng2, Pim1, Ube2b (Hr6b), Zfp36, Gadd45a, Prkcd, Bcl2l (Bcl-XL), Ddit3 (Chop10), Mafk, U2af1-rs1, Cdkn1b, Mad, Csf2rb2. Transcripts not expressed concordantly in both systems: Lower right quadrant: Slc2a1, Pold1, Il4ra. Upper left quadrant: Myln, Ccnd3. An expanded description of these genes is available in Supplementary Data.

Figure 4. The microarray studies reflect gene expression kinetics accurately. Genes previously shown to exhibit altered expression during G1E-ER4 maturation include Gata-2, Myc, Abcb10 (Abc-me) and Bcl-xL. Northern blots are compared to transcript levels predicted by microarrays. The Y axis of the graphs show absolute signal values from the microarrays. Hours after estradiol treatment are indicated.

Figure 5. Developmental stages recapitulated during G1E-ER4 cell maturation. The histograms represent the average signal level of selected erythroid transcripts at indicated timepoints, as predicted by the current microarray study. The shaded envelopes drawn around the histograms are based on previously reported expression kinetics determined in other models of erythropoiesis. The expression pattern observed during G1E-ER4 cell maturation is consistent with the transition from BFU-E to CFU-E stages.
Figure 6. Overview of gene expression by hierarchical clustering. The color scale is identical to that in Figure 2. Filtering parameters were selected to illustrate approximately 100 genes whose expression changed most markedly over the entire timecourse (see Methods). Full listings of all transcripts is provided in Supplementary Data.

Figure 7. Induction of β major globin and Fog1 by GATA-1. A. Messenger RNA levels in G1E-ER4 cells predicted by microarray experiments. B. Messenger RNA levels quantitated by real-time polymerase chain reaction (PCR) in G1E-ER4 cells overexpressing bcl-xL to overcome cycloheximide-induced death. Each datapoint represents six separate PCR reactions in two independent experiments. Cycloheximide (0.5 mM) was added as indicated. Note that the magnitude of β globin RNA induction appears to be underestimated by the microarrays, as was also observed for α globin (see Figure 4, above). This is likely due to saturation of the microarray system by these abundant transcripts.

Figure 8. GATA-1 regulates Fog1 directly. A. Cis-regulatory modules in Fog1 as predicted by comparative genomics. The mouse Fog1 gene (Zfpm1) is plotted in the middle, using base positions on Chromosome 8 from the February 2003 (mm3) assembly. Coding exons are taller boxes, untranslated regions are shorter boxes, and introns are shown as lines with arrowheads pointing in the direction of transcription. Positions of CpG islands are on the line underneath the gene map. The Conservation Score estimates a log-likelihood that an alignment is in functional sequences. The RP score is a log-likelihood measurement that estimates the probability that a sequence is involved in regulating expression. Matches to weight matrices for GATA-1 binding sites were identified in the mouse sequence (track labeled “all GATA-1_BS”), and these
were filtered to find those that are conserved in mouse, rat and human alignments (track labeled “conserved GATA-1_BS”). The DNA segments whose alignments exceed a calibrated threshold for RP score and also contain at least one conserved, predicted GATA-1 binding site are shown in the track labeled “hi RP & cGATA-1_BS”; these are the predicted cis-regulatory modules for the Fog1 locus. Two were tested for occupancy by GATA-1 in vivo (panel B); these are shown as ChIP amplicons R1 (preCRM1 in intron 1) and R2 (preCRM4 in intron 2) along with a negative control (“U”) located upstream of the Fog1 start site for transcription. The information in the first three tracks was obtained by queries to GALA and the data were visualized in the UCSC Genome Browser. B. Chromatin immunoprecipitation (ChIP) experiments showing GATA-1 occupancy at the Fog1 locus in G1E-ER4 cells at 24 hours after GATA-1 activation. The regions of the Fog1 gene examined are indicated as “ChIP amplicons” in line 4 of panel A. Antibodies used are indicated below the figure: G-1, anti-GATA-1 and ER, anti-estrogen receptor (used to detect the GATA-1/ER fusion protein). Control antibodies: rIgG, rat IgG and mIgG, mouse IgG. Each datapoint represents the average value obtained from two independent ChIP experiments. C. Predicted cis-regulatory module 1 enhances expression from an erythroid promoter. The plasmid gammaLuc contains a firefly (FF) luciferase gene driven by the γ globin gene promoter, and preCRM1 (R1) was added upstream of the promoter to make the plasmid R1 gammaLuc. The normalized levels of expression are plotted for two experiments, with each sample tested in triplicate in both. The activity from R1 gammaLuc is significantly higher than that from gammaLuc, with a p-value no greater than 0.005 for either by Student’s t-test. D. GATA1 and Fog1 act in a feed-forward loop to activate the β major globin gene.
Figure 1. Welch et al.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eph7.2</td>
<td>erythrocyte protein band 7.2</td>
</tr>
<tr>
<td>Zipm1</td>
<td>zinc finger protein, multitype 1</td>
</tr>
<tr>
<td>Klf1</td>
<td>Kruppel-like factor 1 (erythroid)</td>
</tr>
<tr>
<td>Ppox</td>
<td>protoporphyrinogen oxidase</td>
</tr>
<tr>
<td>Hbb-b1</td>
<td>hemoglobin, beta adult major chain</td>
</tr>
<tr>
<td>Abcb10</td>
<td>ATP-binding cassette, sub-family B (MDR/TAP), member 10</td>
</tr>
<tr>
<td>Alad</td>
<td>aminolevulinate, delta-, dehydratase</td>
</tr>
<tr>
<td>Mafk</td>
<td>v-maf musculoaponeurotic fibrosarcoma oncogene, protein K</td>
</tr>
<tr>
<td>Trfr</td>
<td>transferrin receptor</td>
</tr>
<tr>
<td>Gypa</td>
<td>glycoprotein A</td>
</tr>
<tr>
<td>Ephb4</td>
<td>erythrocyte protein band 4.9</td>
</tr>
<tr>
<td>Ank1</td>
<td>ankryin 1, erythroid</td>
</tr>
<tr>
<td>Htrb</td>
<td>heme-binding protein</td>
</tr>
<tr>
<td>Uros</td>
<td>uroporphyrinogen III synthase</td>
</tr>
<tr>
<td>Hbb-b2</td>
<td>hemoglobin, beta adult minor chain</td>
</tr>
<tr>
<td>Urod</td>
<td>uroporphyrinogen decarboxylase</td>
</tr>
<tr>
<td>Hmbs</td>
<td>hydroxymethylbilane synthase</td>
</tr>
<tr>
<td>Bach1</td>
<td>BTB and CNC homology 1</td>
</tr>
<tr>
<td>Hba-a1</td>
<td>hemoglobin alpha, adult chain 1</td>
</tr>
<tr>
<td>Nfe2</td>
<td>nuclear factor, erythroid-derived 2, 45 kDa</td>
</tr>
<tr>
<td>Slc4a1</td>
<td>solute carrier family 4 (anion exchanger), member 1 (Band 3)</td>
</tr>
<tr>
<td>Hba-x</td>
<td>hemoglobin X, sloth-like embryonic chain in Hba complex</td>
</tr>
<tr>
<td>Alas2</td>
<td>aminolevulinate acid synthase 2, erythroid</td>
</tr>
<tr>
<td>Gtx2</td>
<td>glutathione S-transferase, theta 2</td>
</tr>
<tr>
<td>Mafg</td>
<td>v-maf musculoaponeurotic fibrosarcoma oncogene, protein G</td>
</tr>
<tr>
<td>Fli1</td>
<td>fliki light chain 1</td>
</tr>
<tr>
<td>Hbb-y</td>
<td>hemoglobin Y, beta-like embryonic chain</td>
</tr>
</tbody>
</table>

**Figure 2. Welch et al.**
Figure 3. Welch et al.
Figure 4. Welch et al.
Figure 5. Welch et al.
Figure 7. Welch et al.
Figure 8. Welch et al.
Global regulation of erythroid gene expression by transcription factor GATA-1

John J Welch, Jason A Watts, Christopher R Vakoc, Yu Yao, Hao Wang, Ross C Hardison, Gerd A Blobel, Lewis A Chodosh and Mitchell J Weiss