Supplementary Methods, Figures, and Tables:

Gene Expression-Based High Throughput Screening (GE-HTS) and Application to Leukemia Differentiation

Kimberly Stegmaier[1], Kenneth N. Ross[2], Sierra A. Colavito[1], Shawn O'Malley[2], Brent R. Stockwell[2], and Todd R. Golub[1,2,3,*]

[1] Department of Pediatric Oncology, Dana-Farber Cancer Institute and Children's Hospital Boston, Harvard Medical School, Boston, MA 02115

[2] Whitehead Institute for Biomedical Research, Cambridge, MA 02139

[3] Howard Hughes Medical Institute

[*] To whom correspondence should be addressed. Email: golub@broad.mit.edu

Table of Contents

Figures

Tables

Part I: Screen Design

A. Characterization of Expression Profiles in Primary Cells

Isolation of Cells and RNA

Normal peripheral blood monocytes and neutrophils were isolated using Ficoll-Paque (Amersham Pharmacia Biotech) separation from three different donor leukopack samples provided by the Dana Farber Cancer Institute Blood Bank. The isolation of monocytes was confirmed with FACS analysis for CD14. We confirmed the isolation of neutrophils and monocytes with morphological analysis after May-Grunwald-Giemsa staining (Sigma). RNA was obtained using Trizol (GIBCO/BRL) as per the manufacturer's guidelines. Quality was assessed by the presence of undegraded 18S and 28S ribosomal RNA bands by denaturing gel electrophoresis.

Adult AML samples were obtained from the Cancer and Leukemia Group (CALGB) leukemia bank and processed as previously reported[1]. Samples were selected without regard to immunophenotype, cytogenetics, or other molecular features. After informed consent was obtained, mononuclear cells were isolated by Ficoll sedimentation and RNA was obtained from three primary patient AML cell samples as described above.

Target Preparation and Hybridization to Microarrays

RNA from the three patient AML samples, three normal peripheral blood monocytes and 3 neutrophil samples (10 μg per sample) was used to create target for hybridization to DNA microarrays. First strand cDNA synthesis was generated using a T7-linked oligo-dT primer, followed by second strand synthesis. An in vitro transcription reaction was performed to generate cRNA containing biotinylated UTP and CTP, which was subsequently chemically

fragmented at 95° C for 35 minutes. Ten micrograms of the fragmented, biotinylated cRNA was hybridized in MES buffer (2-[N-Morpholino]ethansulfonic acid) containing 0.5 mg/ml acetylated bovine serum albumin (Sigma, St. Louis) to Affymetrix (Santa Clara, CA) HuFL arrays at 45°C for 16 hours. HuFL arrays contain 5920 known genes and 897 expressed sequence tags. Arrays were washed and stained with streptavidin-phycoerythrin (SAPE, Molecular Probes). Signal amplification was performed using a biotinylated anti-streptavidin antibody (Vector Laboratories, Burlingame, CA) at 3 μg/ml. This was followed by a second staining with SAPE. Normal goat IgG (2 mg/ml) was used as a blocking agent. Scans were performed on Affymetrix scanners and average differences (expression values) were calculated using GeneChip MAS4 Software (Affymetrix). Minor differences in microarray intensity were corrected using a scaling method as detailed in the next section.

Microarray Data Analysis

Criteria for scan rejection included fewer than 1000 genes receiving "Present" calls or visible microarray artifacts. No scans met criteria for exclusion. The raw expression data as obtained from Affymetrix's GeneChip were scaled to account for differences in chip intensities. We calculated the mean expression level (E) for all genes on each array. All scans within an experiment were scaled to the array with the median E value (all expression values are multiplied by E_{median}/E). These scaled data are contained in **Res File 1, Myeloid_primarycells.res**.

Next, we preprocessed the data by applying thresholds and filtering. A ceiling of 16,000 units was chosen because we observed fluorescence saturation of the scanner at this level. A floor was set at 100 to minimize noise and remove negative values. After this preprocessing, gene expression values were subjected to a variation filter that excluded genes showing minimal variation across the samples being analyzed. We limited our analysis to genes that

demonstrated at least a 5-fold change in relative expression level across the dataset and an absolute change of at least 400 units.

For marker gene selection, we used the signal-to-noise (SNR) statistic to rank the genes that correlated with the AML vs. neutrophil and the AML vs. monocyte distinction [1,2]. SNR= $(\mu_o - \mu_1)/(\sigma_o + \sigma_1)$ where μ and σ represent the mean and the standard deviation of the expression, respectively, for each class. We next needed to identify statistically significant marker genes. Permutation of the sample labels was performed to compare these correlations to what would be expected by chance as described below.

The permutation test procedure for a given comparison of interest (e.g. markers high in class 0 and low in class 1) is as follows:

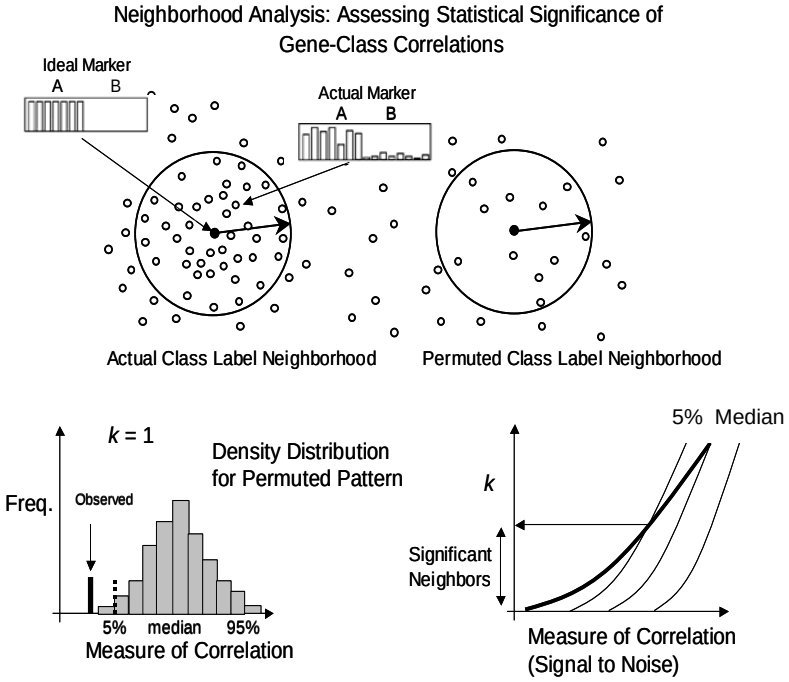
- Generate signal-to-noise scores for all genes that pass a variation filter using the actual class labels (phenotype) and sort them accordingly. The best match (k=1) is the gene "closer" or more correlated to the phenotype using the signal-to-noise as a correlation function. In fact, one can imagine the reciprocal of the signal-to-noise as a "distance" between the "phenotype" and each gene as shown in the **Supplementary Fig. 1** top diagrams. One can also use a *t*-statistic $(\mu_0 - \mu_1)/\sqrt{(\sigma^2_0 + \sigma^2_1)}$ and obtain very similar results.

- Generate 2500 random permutations of the class labels (phenotype). For each case of randomized class labels generate signal-to-noise scores and sort genes accordingly.

- Build a histogram of signal-to-noise scores for each value of k. For example, build one for the 500 top markers (k=1), another one for the 500 second best (k=2), etc. These histograms represent a reference statistic for the best match, second best, etc. where many genes contribute to a given value of k. Notice that the correlation structure of the data is

preserved by this procedure. For each value of k, determine different percentiles (1%, 5%, 10%, 50% etc.) of the corresponding histogram. (See the bottom diagrams in **Supplementary Fig. 1**.)

- Compare the actual signal-to-noise scores with the different significance levels obtained for the histograms of permuted class labels for each value of k. This test helps to assess the statistical significance of gene markers in terms of the distribution of class-gene scores using permuted labels.



Supplementary Figure 1: Neighborhood Analysis

In **Sheets A and B (AML vs. Neut. (HuFL) and AML vs. Mono. (HuFL), respectively) of the Supplementary Data Excel Worksheet**, the values for permutation tests of marker genes are reported in tables with the following format. The Distinction column represents the class for which the markers are high (low in the other classes). The Distance column is the signal-to-noise to the actual phenotype. The Permutation 20% columns represent the percentile

(significance level) in the histograms of signal to-noise scores for permuted labels for a given value of k. The Feature column is the gene accession number and the Description column is the gene name.

Two thousand five hundred permutations were performed to identify the differentiation signature genes that met statistical significance at the 80th percentile using GeneCluster2 (http://www.broad.mit.edu/cancer/software/software.html). We chose marker genes from amongst this list; achievement of higher levels of statistical significance was restricted by data set size. In addition, *Glyceraldehyde 3-phosphate dehydrogenase* (*GAPD*) was selected as the control gene based on the Affymetrix expression data and historical use of *GAPD* as a control. It showed minimal variation across the microarray data set. In the actual screen, this control gene was used to filter out chemicals that killed the cells, to filter out wells where, for technical reasons the experiment did not work, and to normalize for well to well and plate to plate variability in total amount of material as discussed below in Part II Small Molecule Library Screen: Data Analysis.

B. Confirmation of Signatures in an HL-60 Cell Line

In Vitro Differentiation and Preparation for Microarray Analysis

HL-60 cells (American Type Culture Collection) were maintained in culture in RPMI 1640 (Cellgro) with 10% fetal bovine serum (Sigma) and 1% penicillin-streptomycin (Cellgro). In a duplicate experiment, HL-60 cells were differentiated to neutrophils with 1 μM all *trans* retinoic acid (ATRA) (Sigma) for 0, 24, 72, and 120 hours. Differentiation was confirmed with examination of morphology after May-Grunwald-Giemsa staining**.** Maximum differentiation was seen at 5 days. HL-60 cells were stimulated to differentiate to a monocyte-macrophagelike cell, in duplicate, with 10 nM phorbol 12-myristate 13-acetate (PMA) (Sigma) for 0, 4, 12, and 24

hours. Differentiation was confirmed by examination under light microscopy. PMA differentiated cells became flattened, adherent to the cell culture dish, and developed pseudopods. Nearly all cells appeared to be differentiated by 24 hours. RNA was extracted from the cells at each point in the time course. This RNA was then prepared for hybridization to Affymetrix HuFL microarrays. Data were scaled and filtered as described above; the scaled data are contained in **Res File 2, HL60_undiff_PMA_ATRA.res**.

Microarray Data Analysis

We confirmed that the expression signatures characterized for primary AML cells versus normal monocytes or neutrophils could also distinguish an undifferentiated HL-60 cell from a PMA differentiated or ATRA differentiated HL-60 cell, respectively. We used the SNR statistic to rank the genes that distinguished the untreated HL-60 cells from the ATRA or PMA treated HL-60 cells. Genes meeting statistical significance in both the primary cells and the HL-60 model of differentiation were chosen as candidate markers. To further refine this list, we performed the following analysis. A supervised vector was created representing an idealized signature gene, whereby expression was low in the primary AML cells, high in purified neutrophils, and showed increasing levels of expression in 5 day time-course of ATRA-treated HL-60 cells. Next, the genes were ranked according to their similarity to this supervising vector, using normalized gene expression values and Euclidean distance as the metric. The genes topping this list represented potential candidate genes for use in the screen. Next, these candidates were examined by eye for their being expressed at low absolute levels in the undifferentiated state because some of the markers showed relative up-regulation, but were nonetheless highly expressed in the undifferentiated cells. Candidate genes with robust induction (i.e. high fold-induction) were similarly given highest priority. Such high priority genes were then brought forward to multiplexed RT-PCR testing. The 5 genes constituting the differentiation signature

represented the first 5 genes that exhibited robust differential expression as detected by multiplexed RT-PCR and SBE/mass spectrometric detection. The gene selection process was therefore somewhat *ad hoc*, and future screens would likely benefit from a more systematic and automated approach to marker selection and testing. The genes selected include *Interleukin 1 receptor antagonist* (*IL1RN*) and *Secreted phosphoprotein 1* (*SPP1*) for the monocyte signature genes and *Autosomal chronic granulomatous disease protein* (*NCF1*) and *Orosomucoid 1*, or α*1acid glycoprotein*, (*ORM1*) for the neutrophil signature genes. **(See Sheets C and D (Undiff. HL60 vs. ATRA and Undiff. HL60 vs. PMA, respectively) of the Supplementary Data Excel Worksheet)**

C. High Throughput RT-PCR Signature Gene Amplification

Cell culture and RT-PCR were converted to a 384-well format. HL-60 cells were grown in 384-well culture plates in 40 μl of medium (RPMI 1640 with 10% fetal bovine serum and 1% penicillin-streptomycin) at 0.45 x 10^6 cells/ml concentration. Cells were lysed with 45 μl/well of a mixture containing a hypotonic, detergent-containing solution, DTT, and RNAse inhibitor. We assessed for genomic DNA contamination by PCR for genomic DNA sequence in the absence of a reverse transcription reaction and found no evidence of contamination (data not shown). 15 μl of lysate and 6 μl of a 2.5X binding buffer were then transferred to a 384-well plate coated with oligo-dT. The polyA tails of mRNA bind to the oligo-dT on the plate during a 15 minute incubation. The wells were washed twice with a low salt buffer and reverse transcription was performed using the oligo-dT as a primer in a 20 μl M-MuLV reaction at 37^o C for 1.5 hours. Thus, the single stranded cDNA was covalently linked to the plate via the oligo-dT primer. These cDNAs were subsequently used in a multiplexed PCR reaction. Lysis buffers, 384-well custom coated oligo-dT plates, wash buffers, and M-MuLV were purchased from Pierce and used in a modified version of their Express Direct mRNA Capture and RT-PCR system. We

currently obtain oligo-dT coated 384-well plates and lysis buffers from RNAture (Irvine, CA).

PCR primers were designed with Primer 3 software (http://www.broad.mit.edu/cgi-bin/primer/primer3_www.cgi). To eliminate the possibility of amplifying contaminating genomic DNA, PCR primers were designed to span a large intron. Primers contained 19-22 sequence specific nucleotides and a tag of nonspecific sequence (9-23 nucleotides). The addition of a tag prevents PCR primers from interfering with the assessment of SBE/MALDI-TOF data (see below). Amplicons were 120-385 nucleotides in size. PCR was performed with the following reagents and conditions: 1X PCR buffer (Perkin Elmer), 5 mM $MgCl_2$, 2.5 mM dNTPs, 0.05 uM each primer, and 0.15 units/rx Taq polymerase (AmpliTaq Gold, Perkin Elmer). In an MJ 384-well thermocycler, samples were incubated at 92^o C for 9 minutes and then 30 cycles of 92^o C for 30 sec, 65^o C for 30 sec, and 72^o C for 1 min were performed. A final extension of 72^o C for 5 minutes completed the PCR.

Supplementary Table 1: Marker Gene PCR Primer Sequence

PCR Primers:

Gene Name (GenBank number)	Primer Sequence
Glyceraldehyde 3-dehydrogenase (M33197)	
GAPD FT7	TAATACGACTCACTATAGGGAGAAGCCACATCGCTCAGACAC
GAPD RT3	AATTAACCCTCACTAAAGGGAGACTCCATGGTGGTGAAGACG
Interleukin 1 receptor antagonist (X53296)	
IL1RN FT7	TAATACGACTCACTATAGGGAGACTGGGATGTTAACCAGAAGACC
IL1RN RT3	AATTAACCCTCACTAAAGGGAGAAGCTGGAGTCTGGTCTCATCA
Secreted phosphoprotein 1 (U20758)	
SPP1 FT7	AGCGGATAACGCCTTCTCAGCCAAACGCCG
SPP1 RT3	AGCGGATAACGCCTTGGAAGGGTCTGTGGGGC
47 kD Chronic granulomatous disease protein (M55067)	
NCF1 FT7	AGCGGATAACAGTCCTGACGAGACGGAAGA
NCF1 RT3	AGCGGATAACCGTCCAGGAGCTTGTGAATTA
Orosomucoid (X02544)	
ORM1 FT7	TAGGTTGACAAGCTCTCGACTGCTTGTGC
ORM1 RT3	TAGGTTGACCTCTCCTTCTCGTGCTGCTT

D. Signature Gene Detection

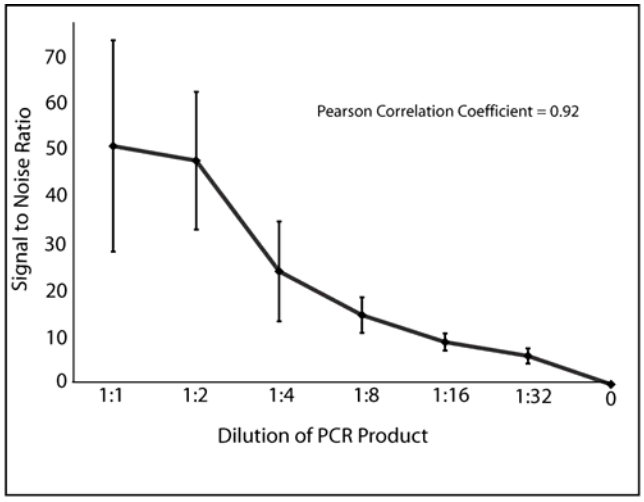
PCR amplicon detection was accomplished using single base extension (SBE) matrix-assisted laser desorption/ionization time-of–flight (MALDI-TOF) mass spectrometry. 5 μl of PCR product was treated with 0.3 units of shrimp alkaline phosphatase (Sequenom) to inactivate any remaining dNTPs (34^{o} C X 20 min; 85^{o} C X 5 min). SBE probes were 16-21 nucleotides in size with an annealing temperature of 50-55^{o} C and sequence complementary to the PCR amplicon

of interest. We then performed a 5-plex SBE reaction in a 9 μl reaction volume with 1x Thermosequenase buffer, 2.7 μM of each primer, 0.2 mM of each ddNTP (Sequenom), and 0.58 units/rx of Thermosequenase (Sequenom) in an MJ 384-well Thermocycler (92^{o} C X 2 min, 40 cycles of 92^{o} C X 20 sec, 50^{o} C X 30 sec). The SBE product was then treated with a cation resin (Sequenom) to remove residual salt from the reaction. The purified extension product was then loaded onto a matrix pad (3-hydroxypicolinic acid) of a SpectroCHIP (Sequenom) with a Spectropoint robot (RoboDesign). SpectroCHIPs were analyzed using a Bruker Biflex III MALDI-TOF mass spectrometer (SpectroREADER) and spectra processed using SpectroTYPER software (Sequenom). For each extension fragment, there is a peak intensity at the expected mass corresponding to the amount of that fragment and a signal-to-noise ratio (SNR) is calculated correcting this intensity for background noise.

Supplementary Table 2: SBE Probe Sequence

SBE Probe	Probe Sequence	Terminator
GAPD_T	ATGGGGAAGGTGAAGG	T
IL1RN_T	CATTGAGCCTCATGCTC	T
SPP1_G	TACAACAAATACCCAGATGCT	G
NCF1_G	AAGGCCTACACTGCTGTG	G
ORM11_C	CCCAGGTCAGATGTCATGTA	C

Supplementary Fig. 2 below illustrates the correlation between the PCR amplicon abundance and that estimated by mass spectrometry. PCR product from *GAPD* was serially diluted 2-fold into an SBE reaction and then evaluated by MALDI-TOF mass spectrometry. Each dilution point represents 19 replicates.



Supplementary Figure 2: Linear Relationship Between PCR Amplicon and Mass Spectrometric Measurement of SBE Product

Part II: Small Molecule Library Screen

A. Library Description

We used a library enriched with compounds with known mechanisms of action and containing many FDA approved drugs[3]. The library contained 1739 compounds and was assembled from three sources i) compounds from Sigma Corporation with annotation describing biological activity, ii) a set of FDA-approved small molecule drugs, and iii) a library of 640 pharmacologically active compounds (LOPAC) sold by RBI, a subdivision of Sigma Corporation. Compounds in group (i) were selected by attempting to identify compounds in the Sigma catalogue documented to have biological activity. In cases where dozens of close analogs or salt forms of a compound were available, one to three representative members of the compound class were selected. The FDA-approved drugs in group (ii) were selected with the Electronic Orange Book list that is maintained by the FDA and lists all approved drug products [4]. As of November 1999, there were 19,299 products approved by CDER (Center for Drug

Evaluation and Research within the FDA), including over the counter medications, prescription drugs and discontinued products. However, many of these products are alternative doses of the same active ingredient or different salt forms of the same active compound. By eliminating such redundancies, the list was reduced to 1,320 distinct small molecule FDA-approved drugs. 708 of these drugs (54%) were purchased from Sigma, Aldrich, RBI and Fluka. Compounds in group (iii) were purchased from RBI as a set of 640 compounds. The compounds were dissolved in dimethylsulfoxide (DMSO) at a concentration of 4 mg/ml and formatted in 384-well plates for screening. The library also contained 181 wells with 0.1% DMSO only for a total of 1920 wells. A list of the compounds in the library is located in **Supplementary Data Excel File, Sheet E, Compound Library**.

B. Screening Methods

HL-60 cells were grown at 0.45 x 10^6 cells/ml in 40 μl, in 384-well culture plates. On each plate, there were 16 wells with the following controls: medium only, undifferentiated, 10 nM PMA differentiated, and1 μM ATRA differentiated HL-60 cells. The remaining wells each contained a compound from the library. 40 nl of compound was transferred from a stock collection of 4 mg/ml for a final concentration of 4 μg/ml. For an average compound with a MW=400, the final concentration would be 10 μM. At three days, RNA was extracted and 20 μl RT-PCR performed in high throughput as described above with 5 primer pairs: GAPD, IL1RN, SPP1, NCF1, and ORM1. Five percent of the wells from each plate were evaluated by gel electrophoresis to confirm that negative control wells were appropriately negative and that the PCR had worked as expected in the positive controls. Then, PCR product was detected after a 5-plex single base extension reaction and MALDI-TOF mass spectrometry as described above.

C. Data Analysis

We developed an analysis pipeline containing several algorithms to identify and prioritize likely differentiating chemicals. These algorithms combined the data across triplicate replicates for each chemical well in the library and utilized the controls in each plate. The SNR generated by SBE MALDI-TOF mass spectrometry, as described above, was used as a proxy for the gene expression level of each of the five genes. We used the expression ratio, g_i/g_{GAPD} where g_i = the SNR level of the signature gene and g_{GAPD} = the SNR level of *GAPD*, to achieve maximal consistency across and within plates. The analysis pipeline consisted of several steps as detailed below: 1. Filtering of wells containing significant cell death, 2. Normalization of plate to plate expression levels, 3. Threshold-based analysis, 4. Probability-based analysis, and 5. Final Neutrophil and Monocyte Score calculation.

Filtering

This step was designed to eliminate dead wells from further analysis. We used *GAPD* values as a proxy for cell viability. Because ratios with *GAPD* as the denominator were used for subsequent analysis, we needed to eliminate wells where *GAPD* was nominally zero with some measurement noise. Forming a ratio with such a low denominator would potentially falsely flag a well where the cells were merely dead or dying.

One standard deviation above the mean of the *GAPD* SNR for the water control wells was calculated. Wells falling below this value were eliminated from further analysis. After filtering was complete, all subsequent analysis used a ratio of the readout genes (*NCF1* and *ORM1* for neutrophil signature and *IL1RN* and *SPP1* for the monocyte signature) to the reference gene (*GAPD*), the expression ratio.

Normalization

This step was applied to correct for plate to plate variability. We found the median expression ratio for each marker gene for the positive controls on each plate (X_{ij} where i=plate number and j=marker gene). Next, we scaled all of the plates by multiplying each expression ratio j on plate i by $1.5/X_{ij}$.

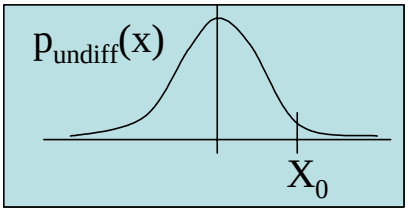
Threshold-Based Analysis

We explored a combination of two methods to identify candidate differentiating agents. The first method used a threshold-based analysis. For each marker gene expression ratio, a threshold distinguishing undifferentiated versus differentiated was established using the untreated and monocyte or neutrophil differentiated positive controls. These thresholds were optimized using a recursive algorithm to correctly identify the controls as undifferentiated, monocyte differentiated, or neutrophil differentiated. Each compound was scored for each marker gene to determine the number of measurements for the triplicate replicate above the established threshold.

Probability-Based Analysis

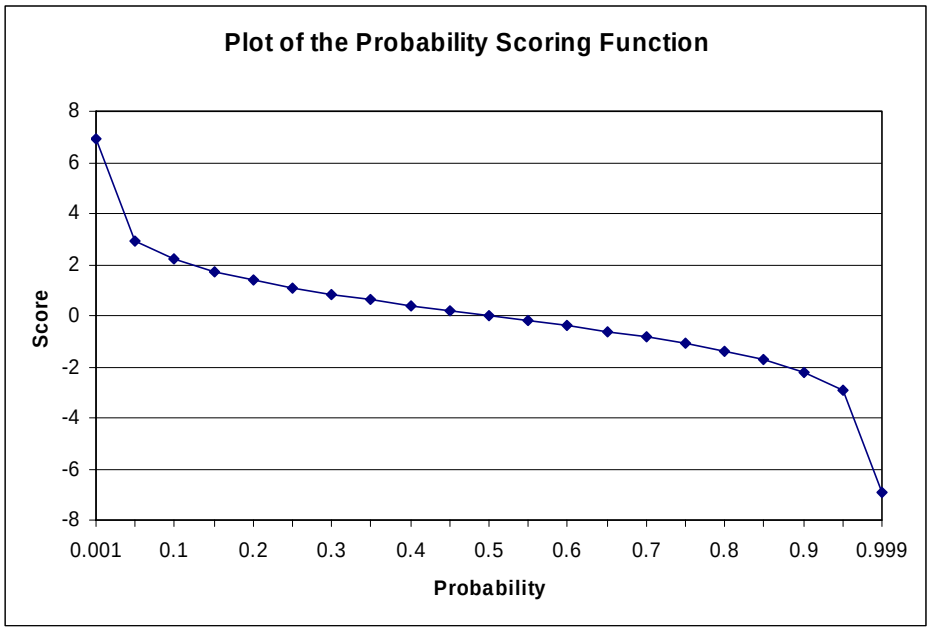
This approach converted the normalized expression ratios into a measure of the likelihood of the well containing differentiated cells. The basis of the probability model was a Gaussian density model for each gene where the Gaussian density parameters (mean and standard deviation) were learned using the measured log-gene expression ratios from the untreated HL-60 control wells. We used log-expression ratios because they better fit the symmetric Gaussian distribution. Our analysis focused on the cumulative probability that a well was undifferentiated. This was calculated by integrating the Gaussian density from the observed log-expression ratio (X_0) for each chemical well to plus infinity (see **Supplementary Fig. 3**). For each gene, this procedure gave a value that a given well was undifferentiated. Rather than work directly with

probabilities, we calculated the following probability score metric: $S = \ln\left(\frac{1 - P_{undiff}(x_0)}{P_{undiff}(x_0)}\right)$



Supplementary Figure 3: Cumulative Probability Distribution

This probability score gives "votes" from each piece of information and varies between negative infinity and plus infinity. In practice, we applied a maximum (0.999) and minimum (0.001) limit to avoid infinity. A probability that a well is undifferentiated gives a highly negative score, a probability near 0 gives a high positive score, and a probability of 0.5 gives a score of 0. The **Supplementary Fig. 4** below shows a plot of the score as a function of probability. For each marker, the average probability score across three replicates was calculated.



Supplementary Figure 4: Plot of the Probability Scoring Function

Final Monocyte and Neutrophil Score Calculation

This metric sought to capture both the threshold-based and the probability-based statistics in one score. A final Monocyte Score was calculated for each well by first taking the sum of the *IL1RN* and *SPP1* probability scores. This value was then multiplied by the total fraction of measurements above threshold for *IL1RN* and *SPP1*.

$$\text{Monocyte Score} = \frac{C}{6-F} * \left(S_{IL1RN} + S_{SPP1}\right)$$

where C is the total number of *IL1RN* and *SPP1* expression ratios above threshold, F is the number of replicates filtered for this compound, and Sj is the probability score metric for the marker gene j. In a similar manner, the Neutrophil Score was calculated for the marker genes *NCF1* and *ORM1*.

D. Compound Hit Selection

The distribution of the Neutrophil Scores for the test compounds was compared to that of the 96 negative control HL-60 wells. Pearson's chi-square analysis was used to evaluate whether the tails of the Neutrophil Score distribution were significantly different for the compound treated and negative control wells. A 2 x 2 contingency table was created by categorizing both untreated and treated samples according to their Neutrophil Scores by designating wells with scores less than 0.4 as undifferentiated and scores greater than or equal to 0.4 as differentiated. The contingency table was evaluated in S-Plus (http://www.mathsoft.com/splus) using the Pearson's chi-square test with Yates' continuity correction. We used the chi-square test to test the hypothesis that the proportion of differentiated wells is the same for the treated and negative control wells. A similar analysis was performed for the Monocyte Score using a cutoff of 0.1 for the differentiated designation **(Supplementary Data Excel File, Sheet F, Score Distributions)**.

We prioritized 15 top scoring compounds by using a combination of selection strategies. Eight compounds were selected from those with top scores based on the Monocyte Score and Neutrophil Score. We also evaluated compounds with high performance at a single marker gene. Six compounds were selected from those with two or more replicates passing threshold and high probability-based scores for the marker gene *NCF1*. For several compounds, all three replicates passed threshold.

One compound, 4,5-dianilinophthalimide, was selected based on high performance utilizing an alternative signature detection method. In the process of developing the mass spectrometric detection method, we explored other detection methods. Although we believe the mass spectrometric detection is a superior system, one of the alternative methods yielded the compound 4,5-dianilinophthalimide. This method is the reverse of traditional cDNA arrays in

which probes are spotted onto a slide and the samples hybridized to the probes. Here, unpurified, multiplexed PCR amplicons were spotted via a Genetix spotter onto an aminosilane-coated microscope slide in duplicate. The spotted DNA was UV cross-linked and the slides boiled in sterile water for two minutes to denature the PCR duplex. The genes specific for each differentiated phenotype, and the control gene *GAPD*, were detected with a 2-step fluorescence signal amplification staining procedure using DNA dendrimer probes. The 3DNA dendrimer is a complex of DNA duplexes with an end-labeled fluorescent moiety (ALEXA, CY3, or CY5). The 3DNA dendrimer contained a sequence that is captured by a bipartite probe. The bipartite probe had sequence complementary to the PCR amplicon of interest and a dendrimer capture sequence. In the first hybridization step, 4 bipartite probes per gene were hybridized to the microarray and the slide incubated at 45º C for 45 minutes with a coverslip in a humidifying chamber. The slides were washed and hybridized with the 3DNA dendrimer during a 45 minute incubation at 45º C. The slides were then washed, dried by centrifugation, and scanned using a GSI 5000 scanner. Arrayvision software extracted the scanned image. Data was processed in a similar manner to that generated by mass spectrometry. While this method could be performed at a lower cost, there was higher background noise and lower reproducibility than mass spectrometric PCR amplicon detection.

The complete list of prioritized compounds further characterized is shown in **Supplementary Table 3**.

Supplementary Table 3: Compound Hits

Aminopterin
R-(-)-Apomorphine HCl
8-(3-Chlorostyryl)caffeine
Cyclazosin HCl
4,5-Dianilinophthalimide
Dimaprit dihydrochloride
Erythro-9-(2-hydroxy-3-nonyl)adenine HCl
5-Fluorouracil
5-Fluorouridine
16-Ketoestradiol
α-Methyl-L-p-tyrosine
Pergolide methanesulfonate
1,10- Phenanthroline
(-)Scopolamine methyl bromide
Sulmazole

Part III: Confirmation of Hits

A. Signature Gene Evaluation

HL-60 cells, in triplicate, were treated for 5 days with compound hits at the following concentrations based on preliminary experimental evaluation of differentiation and growth inhibition (data not shown): 5 μM R-(-)-apomorphine HCl, 75 μM 8-(3-chlorostyryl) caffeine, 7.5 μM cyclazosin HCl, 30 μM 4,5-dianilinophthalimide, 70 μM erythro-9-(2-hydroxy-3-nonyl) adenine HCl, 1 μM 5-fluorouracil, 0.1 μM 5-fluorouridine, 70 μM 16-ketoestradiol, 100μM α-methyl-L-p-tyrosine, 50 μM pergolide methanesulfonate, 0.8 μM 1,10-phenanthroline, 75 μM (-) scopolamine methyl bromide, and 70 μM sulmazole. Compounds were all purchased from Sigma with the exception of cyclazosin that was kindly provided by Dr. Dario Giardina. We also included untreated, 0.1% DMSO, 1 μM ATRA, 10 nM PMA, and 0.1 μM 1,25-dihydroxyvitamin

D3 treated controls. RNA was extracted and prepared for hybridization to Affymetrix HG-U133A microarrays containing 22,283 genes and ESTs as described above. Nine primary patient AML samples obtained from the Cancer and Leukemia Group (CALGB) leukemia bank and three normal monocyte and neutrophil samples (described above) were also included in this data set. Expression values were calculated using GeneChip MAS5 software (Affymetrix). All expression files in a given experiment were scaled to a reference file (the file found to have the median value of expression) based upon the mean expression value for all genes present on the microarrays. A floor of 10 and a ceiling of 16,000 were used. Only genes with a 5-fold variation in expression across the data set and a minimum absolute difference of 50 were considered. These scaled data are available in **Res File 3, Myeloid_Screen_Compound_Eval.res**.

HuFL Affymetrix accession numbers for the marker genes were mapped to Affymetrix U133A accession numbers using Affymetrix's Array comparison worksheets (available at http://www.affymetrix.com/support/technical/comparison_spreadsheets.affx). Because there was no direct map from HuFL to U133A, we mapped HuFL accession numbers to U95A accession numbers using the 'HuGeneFL to Human Genome U95A' map and then mapped the U95A accession numbers to U133A accession numbers using the 'Human Genome U95 to Human Genome U133, Best Match' map (see **Supplementary Table 4**). We compared the mean expression value for each marker gene in the negative controls (untreated and 0.1% DMSO vehicle treated HL-60 cells) to that in the chemical treated HL-60 samples. For each marker gene, we evaluated the fold induction and estimated statistical significance with a one-tailed T-test assuming two samples with unequal variance. Chemicals were considered confirmed hits if the mean fold change was 2-fold or greater and if the *P*-value was less than 0.05 on T-test analysis for the marker gene **(Supplementary Data Excel File, Sheet G, Marker Gene Confirm)**.

Supplementary Table 4: Affymetrix Accession Number Mapping

Marker Gene	HuFL Probe Set	U95A Probe Set	U133A Probe Set
IL1RN	X53296_at	37603_at	212657_s_at
SPP1	U20758_at	2092_s_at	209875_s_at
NCF1	M55067_at	40159_r_at	214084_x_at
ORM1	X02544_at	35315_at	205040_at

B. Analysis of Whole Genome Effects of Chemicals

Mantel Test Analysis

We used a Mantel test to assess whether the chemicals induced changes on a whole genome level consistent with differentiation. A Mantel test is a non-parametric, randomization-based procedure that estimates the correlation between two distance matrices [5]. The Mantel test was initially used to study the correlation in the temporal and spatial distributions of cancer incidences. Since then, there has been a large body of work that has employed this statistic in the analysis of autocorrelated interactions, especially in the fields of ecology, vegetation science, and epidemiology. Because the Mantel test is non-parametric, it can be used to test relationships between data sets that may not be totally independent, unlike more commonly used measures such as the least squares regression and the chi-square test. This characteristic of the Mantel test makes it particularly useful in the analysis of biological data sets where many factors can influence observed phenotypes.

For this work, we used the Mantel test to compare sets of primary AML versus the normal mature myeloid samples to sets of undifferentiated versus compound treated HL-60 cells according to their level of expression to see if genes across the whole genome were being up-

regulated and down-regulated similarly. Specifically, we compared the expression patterns (measured on Affymetrix's U133A microarray) observed in data sets composed of 9 primary AML versus three mature myeloid samples to those expression patterns observed in data sets composed of 12 untreated samples versus the three compound treated samples for each of the selected compounds. For a given gene expression data set X and its corresponding class labels, the distance of each feature from the class labels was calculated using the signal-to-noise statistic. The signal-to-noise statistic is calculated as follows:

$$X_i = \frac{\mu_{i1} - \mu_{i2}}{\sigma_{i1} + \sigma_{i2}}$$

where μ_{i1} represents the mean expression of samples from class 1 for feature *i* and σ_{i1} represents the standard deviation of class 1 for feature *i*. Similarly, the signal-to-noise statistic is calculated for the second set of samples *Y*. The elements of vector *X* and vector *Y* correspond to the same set of objects (U133A probe sets). The Pearson correlation was computed between the corresponding elements of the two vectors to produce the Mantel correlation R_m. The Pearson correlation was calculated as follows:

$$R_m = \frac{\sum_{i=1}^{N} X_i Y_i - \frac{\sum_{i=1}^{N} X_i \sum_{i=1}^{N} Y_i}{N}}{\sqrt{\left(\sum_{i=1}^{N} X_i^2 - \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N}\right)\left(\sum_{i=1}^{N} Y_i^2 - \frac{\left(\sum_{i=1}^{N} Y_i\right)^2}{N}\right)}}.$$

where X_i is the signal-to-noise statistic for feature *i* of sample set *X* and Y_i is the signal-to-noise statistic for feature *i* of sample set *Y*. The Mantel correlation R_m was used as the reference

value in the Mantel test. To calculate the significance level, the elements of one of the vectors were randomly permuted to produce a permuted vector *X**. As before, the Mantel statistic R_m* was computed between *X** and *Y*. The permutation-computation steps were repeated 2500 times and the resulting distribution was used to estimate the *P*-value by examining the proportion of R_m* values that are greater than R_m. This procedure was repeated for each of the selected compounds **(Supplementary Data Excel File, Sheet H, HL-60 Mantel Test)**.

Identification of Gene Induction Associated with Differentiation

Using the above data set, the signal-to-noise ratio was used to rank the genes distinguishing the 9 primary AML samples from the mature monocytes and the mature neutrophils. The top 100 genes with a *P*-value of 0.01 on permutation testing (2500 permutations) were identified **(Supplementary Data Excel File, Sheets I and J, AML vs. Neut. (U133A) and AML vs. Mono. (U133A))**. We then projected this list of genes in the space of the undifferentiated versus chemical treated HL-60 cells positively correlated with the differentiated state by the Mantel test. We used the signal-to-noise ratio to rank order these genes. We projected the top 25 genes in each direction onto a heat map. We then projected these genes in the space of undifferentiated HL-60 cells versus ATRA, 1,25-dihyroxyvitamin D3, and PMA treated HL-60 cells.

Gene Overlap Analysis in Selected Compounds

Because the Mantel test does not reflect the potency of gene expression changes induced by the candidate compounds, we also performed an alternate method of comparing whole genome effects that compared the regulation of individual genes for the 8 selected compounds to that of the reference compounds and primary cells. **Supplementary Data Excel File**, **Sheet K, Gene Overlap Analysis,** summarizes the results from this whole genome comparison. This analysis

used the data file **Myeloid_Screen_Compound_Eval.res** that contained all samples. We first selected the subset of three normal human neutrophil and 9 primary patient AML samples, applied thresholding (10 minimum and 16,000 maximum) and filtering (5-fold minimum difference and 50 minimum absolute difference), and found the genes that were significantly regulated at the 1% level using GeneCluster2 with 1000 permutations and the mean-based signal-to-noise statistic (see Part 1 - Section A for a description of the calculation of the signal-to-noise statistic). 10,048 of 22,283 genes were significant at the 1% level. We then followed a similar procedure to find the subset of those 10,048 genes that were significant at the 1% level in the 12 untreated versus three ATRA treated HL-60 samples. 1143 genes were significant at the 1% level in both the AML versus neutrophil distinction and the untreated versus ATRA treated distinction. We then performed a two-tailed T-test in Microsoft Excel for each of the 8 sets of compound treated samples versus the untreated HL-60 samples for each of the 1143 genes. **Supplementary Data Excel File, Sheet K, Gene Overlap** shows which genes were significantly regulated for each of the selected compounds and summarizes the fraction of the 1143 genes that were significant at the 5% level for each of the selected compounds.

C. NBT Reduction Assay

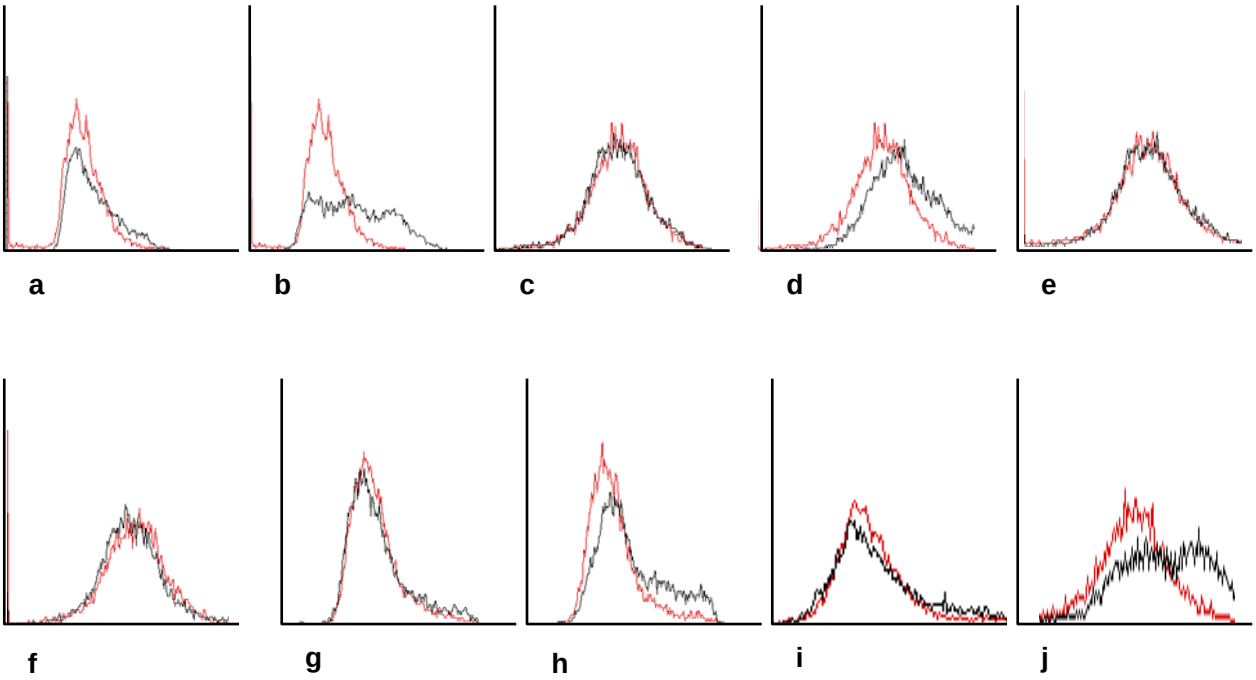
HL-60 cells were exposed to chemicals confirmed to overexpress the differentiation marker genes at the concentrations described above. Experiments were performed in triplicate. Nine untreated and 9 DMSO treated HL-60 cell negative control samples were evaluated. At 6 days, an NBT reduction assay was performed. Cells were incubated at 37º C for 1 hour in a mixture containing total medium, 0.1% NBT (Sigma), and 1μg/ml TPA (Sigma). The percentage of blue cells was counted by light microscopy for at least 200 cells per sample. Untreated cells were compared to chemical treated cells with a one-tailed T-test analysis assuming two samples with unequal variance. The untreated cells were not statistically different from the DMSO treated

samples.

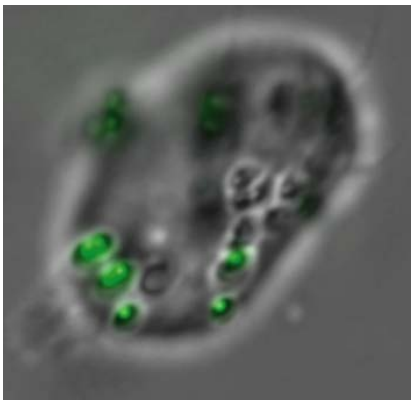
D. Phagocytosis Assay

HL-60 cells were exposed for 5 days to chemicals confirmed to overexpress the differentiation marker genes at the concentrations described above. Experiments were performed in replicates of 5. Untreated and vehicle treated HL-60 cell negative controls were also evaluated. Cells were incubated for one hour with 0.026% fluorescent latex beads (Fluoresbrite Carboxylate 0.75 micron microspheres, Polysciences) and then washed three times with PBS. Fluorescent uptake was then analyzed by FACS with a Becton Dickinson FACScan and CELLQuest analytical software. Laser excitation of 488nm was used and fluorescent emission of 530/30 nm band pass detected. First, cells without beads established the gate for live cells using forward and side scatter patterns. Next, cells incubated with beads were used to establish the fluorescent intensity background. Compound treated cells were then compared against this background. An M1 gate was set at 5% for the untreated and vehicle treated control cells. We then evaluated the percent of cells above this gate for the compound treated cells. An average of the 5 replicates was taken. In a one-tailed T-test the mean of the untreated or vehicle treated controls was compared to the chemically treated cells assuming two groups with unequal variance. Results for candidate compounds and ATRA and 1,25-dihydroxyvitamin D3 controls are shown in **Supplementary Fig. 5a** below (a = (R)-(-)-apomorphine HCl, b = 4,5-dianilinophthalimide, c = erythro-9-(2-hydroxy-3-nonyl) adenine HCl, d = 5-fluorouridine, e = 16-ketoestradiol, f = pergolide methanesulfonate, g = cyclazosin HCl, h = 1,10-phenanthroline, i = ATRA, j = 1,25-dihydroxyvitamin D3) .The red spectra represents the untreated controls and the black the chemically treated cells.

Supplementary Figure 5a: Phagocytosis Analyzed by FACS



Deconvolution microscopy confirmed that the fluorescent beads were indeed intracellular and not simply decorating the cell. Shown in **Supplementary Fig. 5b** is an HL-60 cell treated with 25 uM 4,5-dianilinophthalimide for three days and incubated with fluorescent beads. Images were obtained with a Zeiss Axiovert microscope and fluorescent image deconvolution with SlideBook 3I software. 3-D reconstruction was performed with Volocity software using two channels (FITC and Brightfield).



Supplementary Figure 5b: Phagocytosis Analyzed by Fluorescent Microscopy

E. Propidium Iodine (PI) Assay Determination of IC50

A PI cell growth and cytotoxicity assay was performed according to the protocol described in Cell Biology: A Laboratory Handbook[6] to determine the inhibitory concentration 50% (IC50) of the chemical hits. When possible, chemical stocks were suspended in water. Otherwise, they were suspended in the minimum concentration of ethanol or DMSO to obtain solubility. DMSO and ethanol had no effects on cell growth in the concentration of diluent used in the subsequent experiments (data not shown).

HL-60 cells were plated in 100 μl of RPMI 1640 (Cellgro) with 10% fetal bovine serum (Sigma) and 1% penicillin-streptomycin (Cellgro) at 100,000 cells/ml. Chemical hits were evaluated in triplicate in a 2-fold dilution series starting at a maximum concentration of 100 μm. Chemical + medium only control wells (no cells) were included as controls. Chemical treated HL-60 cells were compared to 7 untreated HL-60 cell control wells. HL-60 cells were incubated for 0 and 5 days in a 37º C incubator. They were then frozen for a minimum of 12 hours at -20º C wrapped in Parafilm. At the time of assessment, plates were thawed at 50º C for 15 minutes. 50 μl of 200 μg/ml propidium iodine solution (Sigma) was added for a final concentration of 40 μg/ml. Plates were incubated in the dark for one hour at room temperature. PI fluorescence was then read in

a BioLum 960 plate reader using a 530-nm excitation filter and a 620-nm emission filter. To determine the IC50, the following calculations were performed:

1. For time 0, the average no cell control values of fluorescence were subtracted from the cell values to determine the amount (Z) of cellular polynucleic acid (PNA) (DNA + RNA) present at the beginning of the chemical incubation period.
2. For time 5 days, the average no cell control values were subtracted from the average untreated cell values to determine the amount (C) of cellular PNA present in the untreated cells at the end of chemical incubation period.
3. For time 5 days, the average chemical blank value (no cells) was subtracted from the average end of assay test values (cells incubated with chemicals) to determine the amount (T) of cellular PNA present in the test cultures at the end of the chemical incubation period.
4. The IC50 is the growth inhibitory concentration of a chemical that reduces the (T-Z) to 50% of (C-Z)

Supplementary Table 5: Compound IC50

Compound Name	IC50
R (-) - apomorphine HCl	3.8 μM
8-(3-chlorostyryl) caffeine	> 100 μM
cyclazosin HCl	7.3 μM
4,5-dianilinophthalimide	5.5 μM
erythro-9-(2-hydroxy-3-nonyl) adenine HCl	> 100 μM
5-fluorouracil	6.6 μM
5-fluorouridine	29.8 nM
16-ketoestradiol	> 100 μM
α-methyl-L-p-tyrosine	> 100 μM
pergolide methanesulfonate	62.9 μM
1,10-phenanthroline	2.4 μM
(-) -scopolamine methyl bromide	> 100 μM
sulmazole	> 100 μM

F. U937 Cell Analysis

U937 cells (American Type Culture Collection) were treated in duplicate with compounds as follows: 5 μM R-(-)-apomorphine HCl, 7.5 μM cyclazosin HCl, 70 μM erythro-9-(2-hydroxy-3-nonyl) adenine HCl, 0.1 μM 5-fluorouridine, 70 μM 16-ketoestradiol, 50 μM pergolide methanesulfonate, and 0.8 μM 1,10-phenanthroline. 1μM ATRA and 0.1 μM 1,25-dihydroxyvitamin D3 were used as positive controls. At 5 days, May Grunwald Giemsa staining was performed after cytospin preparation. At the above tested concentrations, cyclazosin HCl and 1,10-phenanthroline killed the cells. Erythro-9-(2-hydroxy-3-nonyl) adenine HCl and 5-fluorouridine induced morphological changes consistent with macrophage differentiation and ATRA induced changes consistent with neutrophil differentiation. The remaining compounds did

not induce striking morphology changes. We next performed a phagocytosis assay in duplicate at 6 days as described above. Erythro-9-(2-hydroxy-3-nonyl) adenine HCl and 5-fluorouridine induced significant phagocytosis ($P < 0.001$) as did 1,25-dihydroxyvitamin D3 ($P = 0.02$). Data are shown in **Supplementary Data Excel File, Sheet L, U937 Phagocytosis**. NBT reduction was induced only for the ATRA treated cells.

G. Primary Patient AML Cell Analysis

Patient-derived AML samples from the Children's Hospital of Boston were obtained with Internal Review Board (IRB) approval and parent/patient informed consent at time of diagnosis. One sample was from a leukopharesis of a patient with acute promyelocytic leukemia (APL) with a t(15;17) and a white blood cell count of 185,000 with 97% myeloblasts at diagnosis. The other sample was from the peripheral blood of a patient with M1-AML with a white blood cell count of 37,200 with 74% percent myeloblasts. Samples were processed with Ficoll-Paque separation. For the primary APL sample, cells were treated in duplicate with 1 μM ATRA, 70 μM 16-ketoestradiol, 25 μM 4,5-dianilinophthalimide, 70 μM erythro-9-(2-hydroxy-3-nonyl) adenine HCl, and 1 μM 1,10-phenanthroline and evaluated with May-Grunwald-Giemsa staining daily. At 5 days, RNA was extracted from these samples along with two untreated controls and prepared for hybridization to HG-U133A Affymetrix microarrays as described above. Preprocessing of the data included establishing a floor of 75 and a ceiling of 16,000; a variation filter with a minimum of 5-fold change was applied. The Mantel test was performed to evaluate the whole genome effects of the compounds on primary APL cells (Scaled data are in **Res File 4, Myeloid_APL_compound_eval.res;** Mantel test results are in **Supplementary Data Excel File, Sheet M, APL Mantel Test**).

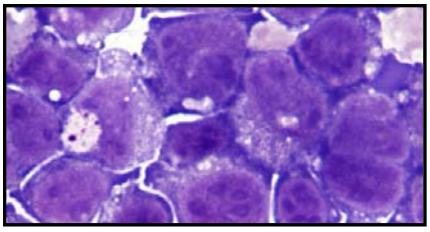
For the M1-AML samples, cells were set up in triplicate and incubated with compounds for 5

days (10 μM ATRA, 5 μM R-(-)-apomorphine HCl, 70 μM erythro-9-(2-hydroxy-3-nonyl) adenine HCl, 70 μM 16-ketoestradiol, 7.5 μM cyclazosin HCl, 1 μM 1,10-phenanthroline, and 50 μM pergolide methansulfonate). Untreated cells were also evaluated. Cells were stained with May-Grunwald-Giemsa and examined with light microscopy. Cells were also evaluated with NBT reduction assay as described above. Cells were then evaluated at five days with lower concentrations of confirmed compounds: 1μM 16-ketoestradiol, and 1μM pergolide methanesulfonate. Negative controls were treated with 0.1% DMSO. May-Grunwald-Giemsa staining and the NBT reduction assay were performed as above **(Supplementary Data Excel File, Sheet N, NBT Reduction Assay)**.

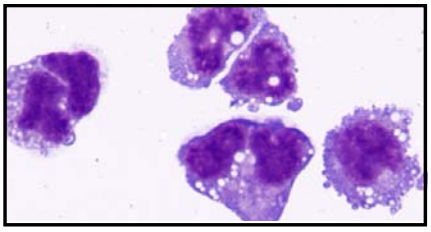
H. Estrogen Derivative Testing

HL-60 cells were treated with estrogen derivates at 10 uM for 6 days: 17 α-estradiol, β-estradiol, and 17-α-ethynylestradiol. NBT reduction was performed at 3 and 6 days as described above in triplicate and May Grunwald Giemsa staining at 5 days after cytospin preparation. Morphological evidence of differentiation consistent with neutrophil maturation was seen as shown below in **Supplementary Fig. 6**. These compounds also induced NBT reduction as shown in **Supplementary Data Excel File, Sheet O, Estrogen Testing.**

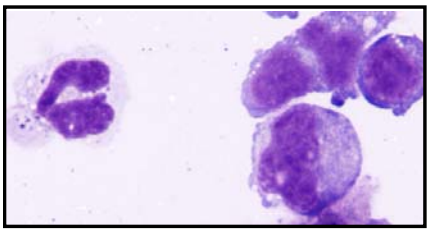
Supplementary Figure 6: HL-60 Treatment with Estrogen Derivatives



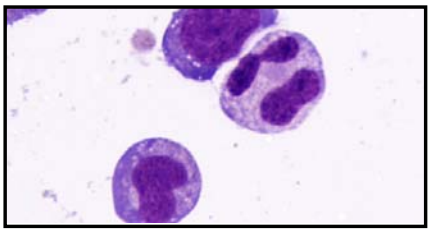
Undifferentiated HL-60



10 μM17-α-ethynylestradiol treated



10 μM 17-α-estradiol



10 μM 17-β-estradiol

Raw microarray data are available at either

http://www.broad.mit.edu/cancer/pub/GE-HTS_leuk or

http://www.ncbi.nlm.nih.gov/geo/

References

[1] Golub, T.R. *et al*. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science* **286**, 531-7 (1999).

[2] Pomeroy, S.L. *et al*. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* **415**, 436-442 (2002).

[3] Root, D. E., Flaherty, S. P., Kelley, B. P. & Stockwell, B. R. Biological Mechanism Profiling Using an Annotated Compound Library. *Chem. Biol.* **10**, 882-891 (2003).

[4] Knoben, J.E., Scott, G.R., & Tonelli, R.J. An overview of the FDA publication Approved Drug Products with Therapeutic Equivalence Evaluations. *Am. J. Hosp. Pharm.* **47**, 2696-2700 (1990).

[5] Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res*. **27**, 209-220 (1967).

[6] Skehan, P. *Cytotoxicity and Cell Growth Assays*. In Cell Biology A Laboratory Handbook, Vol 1, Cellis, Julio E., Ed., Academic Press, San Diego, CA, 311-18 (1998).