Gene expression-based screening of inhibitors of signal transduction

By

Alena A. Antipova

B.S. Mathematics and Physics
Moscow Institute of Physics and Technology, 1994

M.S. Molecular Biology
Moscow Institute of Physics and Technology, 1996

SUBMITTED TO THE DEPARTMENT OF CHEMISTRY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN PHYSICAL CHEMISTRY

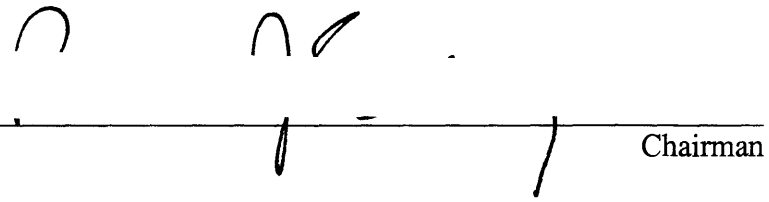
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
[February 2006]
SEPTEMBER 2005

Signature of Author _____
September 2005
Department of Chemistry

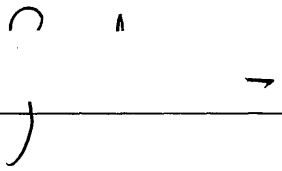
Certified by _____
Robert J. Silbey
Dean of Science
Thesis Supervisor

Accepted by _____
Robert W. Field
Chairman, Departmental Committee on Graduate Students

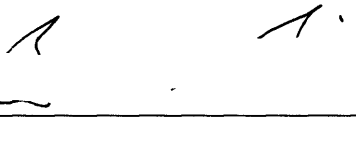
This doctoral thesis has been examined by a Committee of the Department of Chemistry as follows:



Professor Robert J. Silbey _____
Chairman



Professor Todd R. Golub _____
Research Supervisor



Professor Bruce Tidor _____

Gene expression-based screening of inhibitors of signal transduction

By

Alena A. Antipova

B.S. Mathematics and Physics
Moscow Institute of Physics and Technology, 1994

M.S. Molecular Biology
Moscow Institute of Physics and Technology, 1996

Submitted to the Department of Chemistry
in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in
Physical Chemistry, September 2005

ABSTRACT

Advances in our understanding of cellular function and signal transduction have resulted in ever increasing number of therapeutical targets and, consequently, underscored a demand for simple and universally applicable markers to monitor activity of the biological system. Gene expression-based screening is based on the assumption that introduction of a compound into biological system would affect directly or indirectly the gene expression, initiating an elaborate expression response to the disturbance of the system. In my thesis work I investigated new strategies to create a limited expression signature accurately representing the activation state of a biological system and to decrease redundancy in the complex expression pattern. To test the validity of our approach, I generated an expression signature for PDGF activation of ERK pathway in human fibroblasts and used this signature to screen a chemical compound library for an inhibitor of PDGFR/ERK pathway. As a result of the screen, I identified aurintricarboxylic acid, ATA, as a new inhibitor of PDGF receptor. The success of this study suggests that endogenous mRNA expression signature is an effective tool to monitor activation state of a cell signaling pathway and can be broadly used to identify compounds modulating condition of a biological system.

Thesis Supervisors: Eric S. Lander; Todd R. Golub
Titles: Director of the Broad Institute; Director of Cancer Program, Broad Institute at MIT

ACKNOWLEGEMENTS

This work would not be possible without enormous help from many people, which I was fortunate to receive along the way.

First and utmost I would like to thank my Ph.D. adviser, Dr. Todd R. Golub, for his tremendous support and invaluable guidance in pursuing my scientific interests, and for research and career advice he offered me.

I would like to thank Dr. Robert J. Silbey and Dr. Bruce Tidor, members of my Thesis Committee, for their interest in my work, for giving me a valuable feedback, and for their understanding regarding my graduation timeline.

I am very grateful to Pablo Tamayo and Bonnie Berger for showing me the cool side of bioinformatics and pointing me to Cancer Genomics Program.

I received tremendous amount of support from researches in Dr. Todd Golub's laboratory at Dana Farber Cancer Institute and the Broad Institute Cancer Program, for which I am very grateful. I would like to especially thank Koichi Sasaki, Steven Lessnick, Christine Ladd-Acosta, Kimberly Stegmaier, and David Peck for teaching me and assisting me with different aspects of molecular biology parts of the project. I greatly appreciate all the programming help and advice I received from Michael Angelo and Michael Reich. I would like to thank Jill Mesirov, Stefano Monti and Nick Patterson for their advice on

statistical aspects of the project. I would like to especially thank Jadwiga Grabarek for her scientific and career advice, as well as her personal support during my time at DFCI.

I would like to thank Dr. Brent Stockwell from Whitehead Institute, my collaborator on small-molecule screening project, for his help and advice.

I am very grateful to Hans Widlund and Vivien Igras from Dana Farber Cancer Institute for their help and advice on cell culture methods.

I would like to thank Prof. Eric Lander for his advice and helpful discussions over the years.

Finally, I would like to thank my family for their love, support and understanding.

To my teachers, family and friends.

Parts of this thesis are reproduced from Antipova A.A., Tamayo P., and Golub T.R. "A strategy for oligonucleotide microarray probe reduction." Genome Biol. 2002; 3(12):RESEARCH0073. Epub 2002 Nov 25 with kind permission from Pablo Tamayo and Todd R. Golub.

INTRODUCTION

Endogenous gene expression based approach for HTS.

High-throughput screening (HTS) involves systematic testing thousands of diverse chemical structures against disease targets to find candidates for lead optimization. HTS is characterized by its simplicity, rapidness, low cost, and high efficiency. Currently, HTS primarily involves exposure of purified proteins to arrayed collections of small molecules ([1], [2]). More complex cell-based or whole-organism assays are typically developed for specific systems and pathways, require expression vectors or reporters, or rely on monitoring a particular phenotype ([3], [4]). While cell- and organism-based assays have increasingly been used as a primary screening platform for annotating small molecules, there is a need to develop simple, phenotype and reagent-independent methods that can be easily adopted for different biological systems.

Small organic molecules have been widely used to study gene product function in a cellular or organismal context ([5], [6]). By binding to their protein targets, these exogenous ligands can act like either gain-of-function or loss-of-function mutations to activate or inhibit the function of their target proteins. Thus for a particular phenotype of interest, it is possible to identify new reagents that act as conditional mutations ([7], [8]). A comparison of the characteristic gene expression patterns induced by deletions of various genes in yeast to the expression patterns resulting from treatment with known inhibitors seems to indicate that there exists a common pattern of gene expression, or

gene expression signature [9]. Gene expression signature has been successfully used as a surrogate of cellular states in screening a compound library for compounds inducing myeloid differentiation [10]. Taken together, these results suggest that it is possible to select a set of marker genes predicting the ability of a compound to modulate a particular biological pathway. Once identified, this gene set can be used to screen novel compounds for activity related to this specific pathway.

Gene expression based screening approach utilizes endogenous expression of genes to infer activation state of the system. Introduction of a compound affecting the system should detectably change expression of genes in the signature. Thus, if a library is screened for inhibitors of a pathway, and a signature is defined by comparison of activated state (with a ligand) and inactive state (without a ligand), introduction of an inhibitor of a pathway in the presence of a ligand should return the expression of a signature genes to inactive (without a ligand) state. Monitoring the expression of signature genes thus would provide accurate, though indirect, method to find inhibitors of biochemical pathways.

Advantages of the endogenous gene expression-based approach.

There are a number of advantages to screening approach based on expression of endogenous genes. First, the method is generic. It can be easily adapted to almost any pathway, as long as pathway perturbation results in some changes in expression of downstream genes, and the half-life of the transcripts is long enough to allow mRNA

isolation and detection in a high throughput setting. As no additional reagents, such as reporters and expression vectors, are required, the method is suitable for potentially unstable biological systems, where introduction of overexpressed or modified components could shift the equilibrium in the system complicating subsequent analysis of results. In addition, using endogenous RNA to monitor activation of a system is especially beneficial for novel pathways, where there is paucity of tools to probe the pathway status.

Another considerable advantage of an expression-based screening method is that no obvious morphological phenotypic change is required to be observed in the pathway of interest in response to disturbance or pathway activation. This not only makes monitoring activation of the pathway less laborious but also broadens the specter of the observed changes to include transient signaling events.

Signature selection criteria.

As the success of the library screen depends on the expression signature one chooses to monitor activation state of the pathway, a special attention has to be paid to signature selection. The signature has to reflect activation state of the pathway. Preferably, the signature should contain both genes that increase their expression and genes which expression is decreased in response to stimulus. This helps eliminate false positive wells, where total expression is much different from the average well expression on the plate. Uneven total gene expression on the plate is not uncommon for high throughput

screening, when it is hard to obtain comparable total gene expression in each and every one of some thousands wells.

The robustness of the signature is especially important in a high throughput context. All the genes in the signature should have stable expression during the time it takes to isolate RNA in a high throughput setting in numerous plates and replicas. The expression levels of the signature genes have to be high enough to withstand possible variations of cell number, and, consequently, total gene expression, resulting from irregular loading or toxicity of some screened compounds.

The irregularity of cell count, and thus total amount of mRNA, becomes more and more of an issue as one uses plates with smaller and smaller wells in order to boost number of compounds screened. One way to address this issue is introduction of control genes into the signature. These genes maintain a stable level of expression regardless of the activation status of the system. Another purpose of the control genes is to weed out false positives coming from compounds that nonspecifically affect RNA expression, such as DNA intercalators or polymerase inhibitors.

Signature adaptation for the high throughput setting.

After defining the signature that would accurately and robustly reflect activation state of a pathway, the next step is adapting this signature to a high throughput setting. In my

thesis, I discuss two possible approaches to detection of a gene expression signature in a high throughput screen.

Microarray-based approach to signature detection.

First approach is to adapt oligonucleotide microarrays for a high throughput setting. Microarrays are suitable to simultaneously query hundreds and thousands of genes, thus microarrays enable one to survey global changes in gene expression and regulation upon introduction of compounds. Microarrays allow monitoring extensive complex signatures with multiple genes. This is especially beneficial for unstable systems and pathways with small and transient expression changes, as signature redundancy contributes to robustness and accuracy of the method. Additionally, one can envision creating combination signatures that would reflect activation status of multiple pathways and would include genes selectively changing their expression in response to activation of one pathway (or a group of pathways), but not others. These complex multi-pathway signatures would enable one to screen for compounds that simultaneously inhibit a group of pathways, or to just speed up screening of one library for inhibitors of multiple pathways. The high number of genes in the signature could also allow performing a statistically significant clustering of the hits from a screen, facilitating further analysis of the screen results.

While the benefits of using microarrays for screening compound libraries are overwhelming, this technology is still laborious and expensive, which makes it not practical for monitoring expression of genes in thousands of reactions. The challenge is to design simpler and cheaper oligonucleotide arrays, which could be subsequently used in a

high throughput screen. In the first chapter of my thesis I describe one possible solution, reducing complexity of high-density oligonucleotide arrays so that microarrays become affordable and applicable for a high-throughput setting [11]. The chapter focuses on designing and testing oligonucleotide arrays that utilize only 5% of probes (and thus 5% of resources) of the original oligonucleotide arrays, but still provide comparable accuracy in monitoring global gene expression.

Decreasing array size brings another advantage, as smaller volumes of reagents can be used with smaller arrays, further decreasing the costs of the screen. Still, while significant, the twenty-fold reduction of the probe space appeared not sufficient to make microarrays practical for screening applications. The current array hybridization and detection procedures, being laborious and slow, need to be streamlined for microarrays to be suitable for a high throughput setting and a new automated platform has to be developed to allow fast and easy detection of RNA expression with microarrays.

Monitoring expression of a small signature.

As microarrays appeared to be unpractical for a high throughput setting, another approach had to be used, one that relied on monitoring a limited signature of a single pathway. This expression signature had to fulfill an additional set of requirements. The number of genes in the signature had to be high enough to make the signature robust and reliably detectable in thousands of reactions. At the same time, the number of genes in the signature had to be reasonably low so that detection of the signature was feasible and affordable in a high throughput setting.

Measuring gene expression of a small set of candidate genes provides a number of advantages in a high throughput setting. Limiting number of genes in the signature facilitates adaptation of the signature to various detection methods and allows screening large chemical libraries. The challenge of the approach though, is to attain an accurate and robust signature with a small number of genes. The development and application of this strategy for identification of a novel inhibitor of a PDGFR/ERK pathway are described in the second chapter of my thesis.

CHAPTER 1. A STRATEGY FOR OLIGONUCLEOTIDE MICROARRAY PROBE REDUCTION.

Abstract

Background:

One of the factors limiting the number of genes analyzable on high density oligonucleotide arrays is that each transcript is probed by multiple oligonucleotide probes of distinct sequence in order to magnify the sensitivity and specificity of detection. Over the years, the number of probes per gene has decreased, but still no single array for the entire human genome has been reported. To reduce the number of probes required for each gene, a robust systematic approach for choosing the most representative probes is needed. Here, we introduce a generalizable empiric method for reducing the number of probes per gene while maximizing the fidelity to the original array design.

Results:

The methodology has been tested on a dataset comprised of 317 Affymetrix HuGeneFL GeneChips. The performance of the original and reduced probe sets was compared in four cancer classification problems. The results of these comparisons demonstrate that the reduction of the probe set by 95% does not dramatically affect performance, and thus illustrate the feasibility of substantially reducing probe numbers without significantly compromising sensitivity and specificity of detection.

Conclusions:

The strategy described here is potentially useful for designing small, limited-probe genome-wide arrays for screening applications.

Background

DNA microarrays have become commonplace for the genome-wide measurement of mRNA expression levels. The first described microarray for this purpose, the cDNA microarray, involves the mechanical deposition of cDNA clones on glass slides [12]. While this strategy has proven to be highly effective, it has two limitations: 1) cross-hybridization can occur between mRNAs and non-unique or repetitive portions of the cDNA clone, and 2) the maintenance and quality control of large, arrayed cDNA libraries can be challenging. For these reasons, oligonucleotide microarrays have at least theoretical advantages. Short probes (25 nucleotides or longer) can be selected based on their sequence specificity, and either synthesized in situ (by photolithography or inkjet technology) on a solid surface or conventionally synthesized and then robotically deposited.

The first oligonucleotide microarrays contained hundreds of distinct probes per gene in order to maximize sensitivity and specificity of detection [13]. Over the past several years, the number of probes per gene has decreased as increasing amounts of sequence information has become available, probe selection algorithms have improved, feature sizes have decreased and researcher desires to maximize the number of genes assayable on a single microarray have risen. Nevertheless, no single array representing the entire human genome has been described. Furthermore, to date, no systematic high–throughput method has been published that can be used for reducing the number of probes per gene while maximizing the sensitivity and specificity of these reduced probe sets.

Several strategies for probe reduction could be considered. Probes could be selected at random, but given that different probes can have dramatically different hybridization properties, this random method would likely result in failure at least for some genes. Alternatively, one could assess the fidelity of candidate probes by comparison to a gold standard of gene expression measurement such as real-time quantitative PCR or Northern blotting. Such approaches, however, are not feasible at a genome-wide scale. We report here a generalizable, empiric strategy for probe reduction that eliminates 95% of probes, yet maximizes fidelity to the original microarray design.

Results and Discussion

The experiments described here are based on HuGeneFL GeneChips commercially available from Affymetrix, Inc. These arrays contain approximately 282,000 25-mer oligonucleotide probes corresponding to 6817 human genes and ESTs (a total of 7129 probe sets). On average, each gene is represented by 40 probes: 20 'perfect match' probes that are complementary to the mRNA sequence of interest, and 20 'mismatch' probes that differ only by a single nucleotide at the central (13^{th}) base. We refer to the perfect match/mismatch pair as a 'probe pair.' Each gene is thus represented by 20 probe pairs. Normally, these 20 probe pairs are consolidated into a single expression level (known as 'Average Difference') for each gene using GeneChip software (Affymetrix, Inc.) which calculates a trimmed mean of the perfect match minus mismatch differences in order to incorporate some measure of non-specific cross-hybridization to mismatch probes [13]. Alternative methods for estimating message abundance have also been reported ([14], [15]).

In order to reduce the number of probes per gene, we sought to identify the single probe pair for each gene that best approximated the Average Difference, a value that is based on all 20 probe pairs. To accomplish this, we first defined a training set of expression data derived from 141 human tumor samples of diverse cellular origins [16]. For each gene on the array, we generated a vector corresponding to the normalized Average Difference value across the 141 samples. Next, we calculated the perfect match minus mismatch value for each of the 20 individual probe pairs for each gene on the array (referred to

hereafter as delta (Δ). In the final step, the 20 normalized Δs for each gene were ranked according to their degree of correlation with the Average Difference vector across the 141 training samples using Euclidean distance as the metric. The highest-ranking Δ (Δ_h) was chosen for further evaluation in an independent test set. A schematic for this procedure is shown in Figure 1.

The independent test set consisted of expression data derived from 176 tumor samples that were entirely non-overlapping with the training set. We determined the ability of the training set-derived Δ_h values to approximate the Average Difference values in the independent test set, compared to randomly selected Δs. As shown in Figure 2, 79.3% (+/- 3.0%) of Δ_h values were within 2-fold of their respective Average Difference value, as compared to 57.8% (+/- 5.1%) for randomly selected Δs. The relative error of the estimates was 0.8 (+/- 0.1) for Δ_h values and 2.7 (+/- 0.7) for randomly selected Δs. Overall, the distribution of Δ_h accuracies was distinct from randomly selected Δs ($p < 10^{-4}$, chi-squared test). This result indicates that the empirical selection of Δ_h is a better strategy for reducing probe numbers compared to random probe selection.

We next determined whether training-set-derived Δ_h values would be sufficient for pattern recognition and classification of the independent test set of samples. The 176 test samples fall into four binary classification problems: 1) acute myeloid leukemia, AML, vs. acute lymphoblastic leukemia, ALL, (Leukemia Set A; n=35), 2) T-cell ALL vs. B-cell ALL (Leukemia Set B; n=23), 3) diffuse large B-cell lymphoma survival prediction

(n=58), and 4) medulloblastoma brain tumor survival prediction (n=60), as has been described previously ([17], [18], [19]). We used a *k*-Nearest Neighbors (*k*-NN) prediction algorithm [20] and applied it to these four classification problems using either the Average Difference values or the Δ_h values as the starting point. As shown in Table 1, classification accuracy based on Δ_h was nearly identical to that obtained using Average Difference values, despite the fact that 95% fewer probes were utilized. It should be noted that while Δ_h values more accurately approximated the Average Difference compared to random Δs (Figure 2), the random Δs also performed relatively well in these classification problems. It is possible, however, that classification performance would deteriorate when applied to more subtle classification problems, or when applied to samples of different tissue types. These results, taken together, demonstrate the feasibility of substantially reducing probe numbers without dramatically affecting performance.

Conclusions

In conclusion, the empirical approach to probe reduction presented here allows a systematic optimization of individual probe sets. Our studies specifically reinforce the notion that careful selection of probe pairs based on their hybridization behavior is a promising strategy for future chip design. Nevertheless, it remains likely that the use of multiple probes per gene will generate the most accurate and robust detectors. Particularly for diagnostic applications, probe redundancy may significantly improve performance. For screening applications, however, the availability of small, limited-probe, genome-wide arrays could be potentially useful.

Materials and methods

Datasets

The raw data analyzed here has been previously reported([17],[18],[19]) and is available at [16].

Approximation of Average Difference

To estimate the percentage of genes with Δ_h values within 2-fold of the Average Difference, for each gene we compared the value of Δ_h with the Average Difference for this probe set. The percentage of genes within 2-fold of the Average Difference was then averaged over the 176 chips of the training set. To evaluate random probe selection, for each gene a Δ was chosen randomly and the percentage of genes within 2-fold of the Average Difference was similarly calculated. This process was repeated 20 times and then averaged. Values of both Average Difference and selected Δs were normalized and thresholded at 100 units. Relative error for the estimates for Δ_h and randomly selected Δ values was calculated as $|\Delta$ - Average Difference|/ Average Difference.

Rescaling

To account for minor variation in overall chip intensities, Average Difference values were scaled as previously described [19]. For Δ_h values, scaling was adjusted by a slope

and intercept obtained from a least squares linear fit of the Δ_h values for each chip compared to a randomly selected reference chip.

Classification

Average Difference and Δ values were clipped to minimum 20 and maximum 16,000 units. A variation filter was applied that excluded genes that did not vary at least 3-fold and 100 units across the entire dataset. To compare the classification accuracy for Δs and Average Difference, we applied a *k*-Nearest Neighbors (*k*-NN) [20] binary classifier, implemented in the software package GeneCluster 2.0 and available at [21], to each of the four classification problems as previously described [19], Average Difference or Δ feature selection was performed with the signal-to-noise metric [8], ($\mu_{class\ 0} - \mu_{class\ 1}$)/($\sigma_{class\ 0} + \sigma_{class\ 1}$), where μ and σ represent the mean and standard deviation within each class, respectively, and the top ranking features were fed into the k-NN algorithm. Performance was evaluated by leave-one-out cross-validation, whereby for each sample a prediction was made with a model trained on the remaining samples in the problem set, and the number of classification errors was tallied. Classifiers with variable numbers of features (1-100) and nearest neighbors (k = 3 or k = 5) were tested. The best performing classification results are reported.

Acknowledgements

We thank Michael Angelo and Michael Reich for programming help, Sridhar Ramaswamy for providing datasets, and Eric Lander for helpful discussions.

Figure Legends

Figure 1. Schema for selection and evaluation of the single probe pair (Δ_h) that best approximates the Average Difference value derived from all 20 probe pairs. PM indicates Perfect Match. MM indicates Mismatch. Δ = PM – MM.

Figure 2. Comparison of Average Difference values with Δ_h and randomly selected Δs. For each of the datasets shown, the proportion of genes whose Δ_h value is within 2-fold of the Average Difference is shown by the black bars. The same comparison is shown for random Δs (gray bars). Error bars indicate standard deviation. Standard deviation shown reflects variations in the percentage of genes within 2-fold of the Average Difference between the 176 chips of the training set. Note that the Δ_hs better approximate the Average Difference compared to randomly selected Δs.

Abbreviations

PM, perfect match; MM, mismatch; AML, acute myeloid leukemia; ALL, acute lymphoblastic leukemia; k-NN, k-Nearest Neighbors.

Figure 1

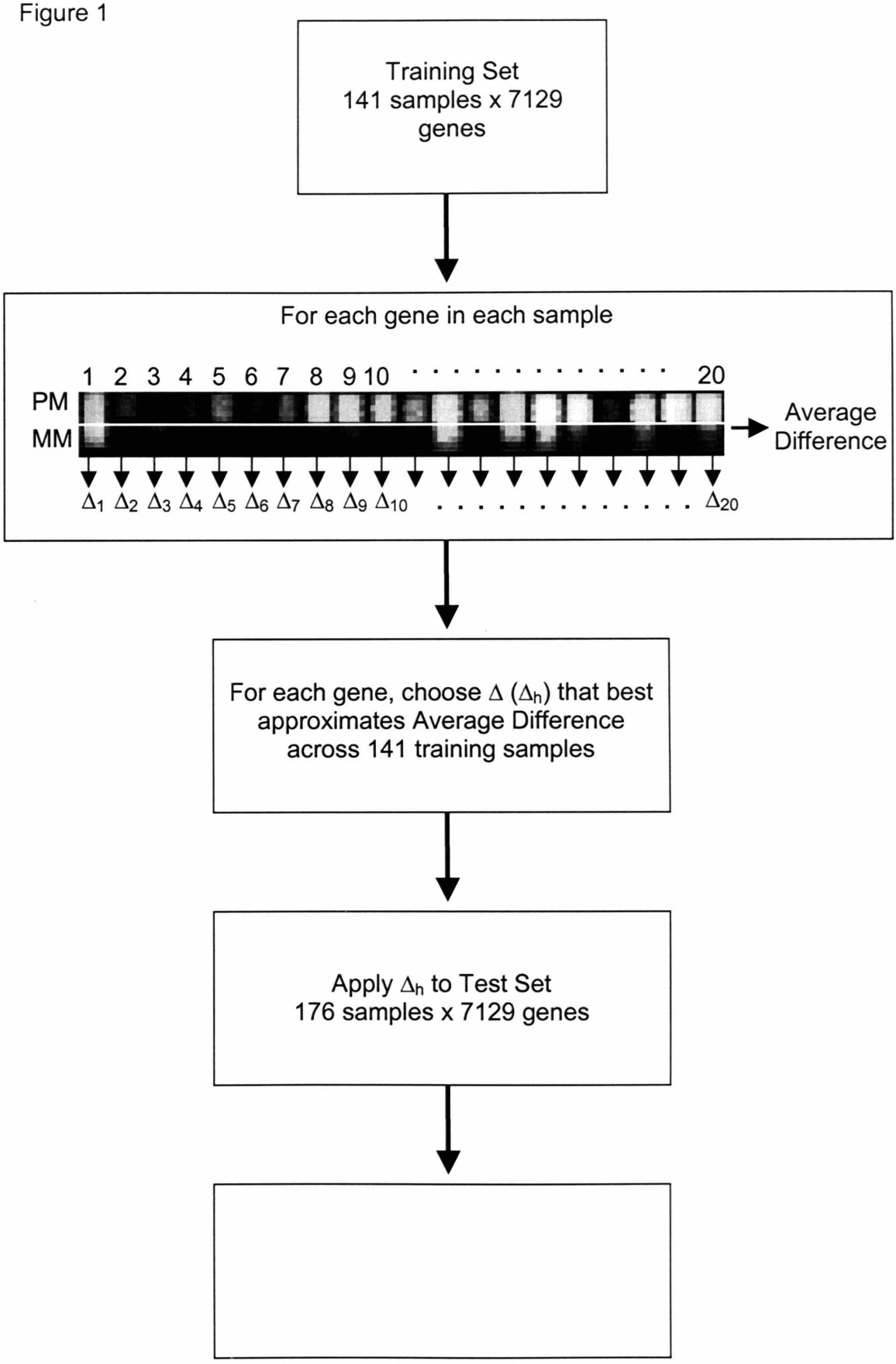
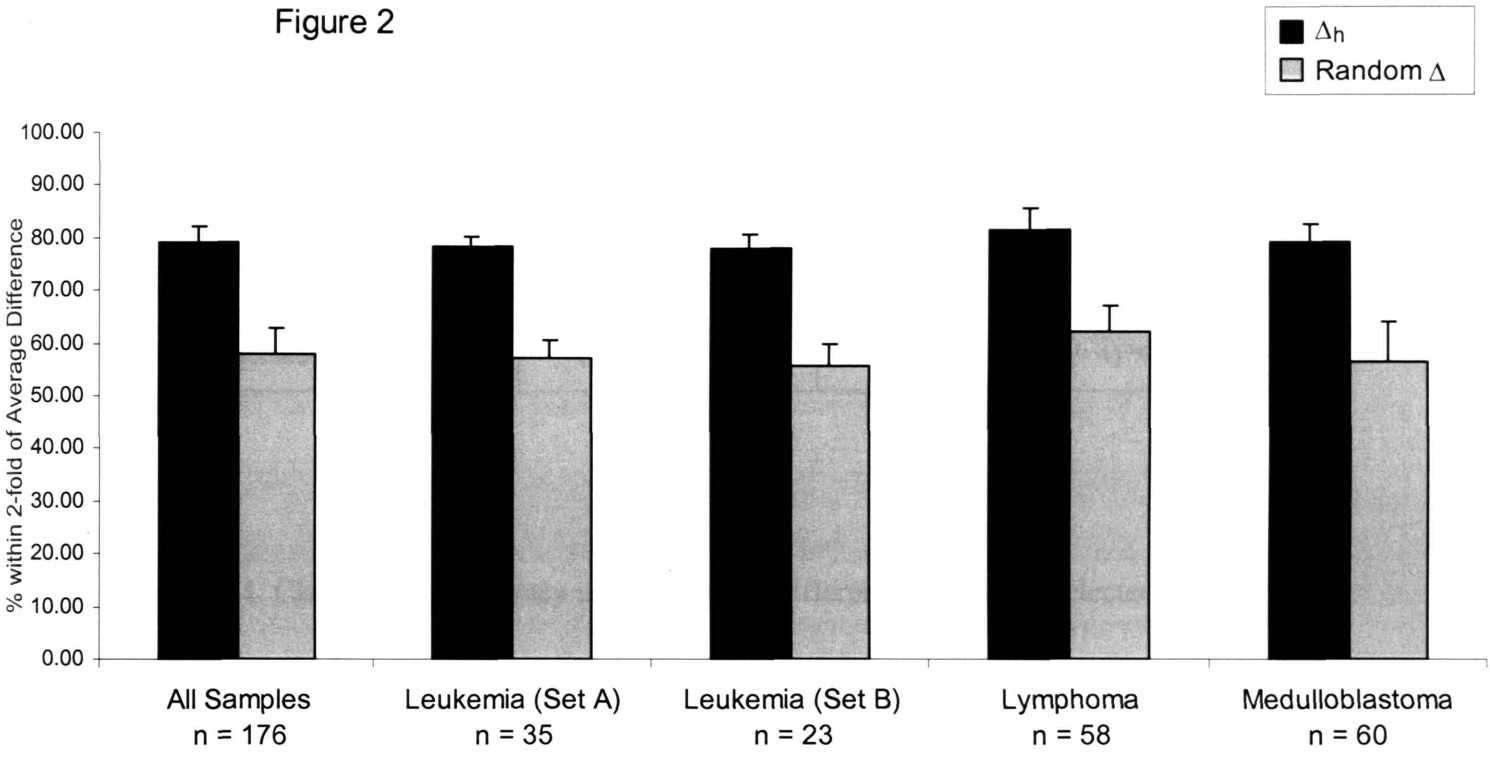

Training Set
141 samples x 7129 genes
For each gene in each sample
1 2 3 4 5 6 7 8 9 10 20
PM
MM
Average Difference
Δ1 Δ2 Δ3 Δ4 Δ5 Δ6 Δ7 Δ8 Δ9 Δ10 . . . . . Δ20
For each gene, choose Δ (Δh) that best approximates Average Difference across 141 training samples
Apply Δh to Test Set
176 samples x 7129 genes

Figure 2


Δh
Random Δ
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
% within 2-fold of Average Difference
All Samples
n = 176
Leukemia (Set A)
n = 35
Leukemia (Set B)
n = 23
Lymphoma
n = 58
Medulloblastoma
n = 60

Dataset	n	Classification problem	Error Rate		
			Δ_h	Random Δs	Average Difference
Leukemia (Set A)	35	ALL vs. AML	3%	2 (+/- 1) %	3%
Leukemia (Set B)	23	T-ALL vs. B-ALL	0%	0 (+/- 1) %	0%
Lymphoma	58	Cured vs. Fatal	26%	29 (+/- 5) %	24%
Medulloblastoma	60	Cured vs. Fatal	18%	26 (+/- 4) %	24%

Table 1. Classification accuracy using Average Difference, randomly selected Δs, and Δ_h values.

CHAPTER 2. GENE EXPRESSION-BASED SCREENING OF INHIBITORS OF SIGNAL TRANSDUCTION.

Abstract

Recent advances in our understanding of cellular function and signal transduction have resulted in ever increasing number of drug targets and, consequently, underscored a demand for simple and universally applicable markers to monitor activity of the biological system. Here we demonstrate that endogenous mRNA expression signature of a cell signaling pathway can be successfully used to monitor short-term activation, and in particular, transient phosphorylation states of a biochemical pathway. As a proof of concept experiment, we utilized an expression signature of a PDGF-activated ERK pathway to infer the phosphorylation status of the cascade in a cell-based high throughput screen for PDGFR/ERK pathway inhibitors. A novel micromolar inhibitor of PDGFR, Aurintricarboxylic Acid (ATA), was found in screening of 1,739 compounds. These results demonstrate that expression signature alone can be sufficient to find modulators of molecular pathways, and thus argue that endogenous mRNA expression can be used as an accurate, though indirect, read-out for an activation state of a biochemical pathway.

Introduction

High throughput screening of combinatorial libraries and small-molecule collections is a well-established and highly productive tool for identification of chemical compounds targeted at a particular protein or biological activity in a living cell. Traditionally, the high throughput screening for modulators of molecular pathways involves biochemical assays, or highly specialized cell-based assays [22]. However, there are situations where alternative approach is beneficial, especially if the targets are poorly characterized and direct measurement of activity of biological pathway is not feasible. Here we demonstrate that it is possible to use only endogenous mRNA expression as a read-out to infer activity of a biochemical pathway, rather than measure this activity directly.

Endogenous mRNA expression has been previously successfully used as a surrogate of cellular states in a high-throughput screening for compounds inducing differentiation of acute myeloid leukemia cells [10]. The expression signature was developed to correlate with established phenotype 72 hours after introduction of compounds to cells. For a fast signaling pathway, the activation signature has to reflect transient changes that occur in the biological system within minutes, not hours.

To establish that endogenous mRNA expression signature can accurately describe an activation of a fast biochemical pathway, three major issues had to be addressed. First, we had to define a robust signature for a particular biochemical system. Next, this signature was used as a read out in a high-throughput screen aimed at finding inhibitors

of the pathway. And finally, we had to validate the hits from the screen in terms of their selectivity and specificity.

We chose the ERK pathway (also known as the p42/p44 MAP kinase pathway) as a target pathway for the screen. ERK pathway plays a major role in the control of cell growth, cell differentiation and cell survival ([23], [24], [25]). The protein kinase cascade Raf>MEK>ERK, also referred to as the mitogen activated protein (MAP) kinase module, is activated in mammalian cells through receptor tyrosine kinases (RTKs), G-protein coupled receptors and integrins ([26]). Activated ERKs can translocate into the nucleus where they phosphorylate transcription factors ([27]; [28]; [29]). ERK pathway is often upregulated in human tumors [30], and as such is an attractive target for anticancer therapy. Furthermore, because the pathway has been extensively studied in the last two decades, one could take advantage of the copious array of tools developed to probe the system.

Results

Identification of expression signature of PDGFR/ERK pathway.

In gene expression-based high throughput screening the expression signature is used as a surrogate of activation state of a system. The expression of genes in the signature should reflect the changes in the system induced by addition of a ligand and should return to a base, "inactive", state upon introduction of inhibitors of the pathway.

To monitor activation status of the ERK pathway we had to select markers that changed their expression upon introduction of a ligand. We chose PDGF-BB as an activator of a pathway in fibroblast and neuroblastoma cells expressing PDGF receptor beta. The activation state of the members of pathway can be traced by increase in their phosphorylation levels shortly after introduction of the growth factor [31]. In particular, ERK phosphorylation peaks at about 15-20 minutes after induction, and then decreases to background levels some 20-30 minutes later ([32], [33]).

In order to select expression markers of activated ERK pathway, we compared serum starved cells incubated with and without PDGF. PDGF-BB can stimulate a number of pathways through SH2-containing proteins binding multiple tyrosine-phosphorylated sites expressed on the activated PDGF receptors beta [34]. To select against compounds that do not inhibit ERK pathway, but affect other PDGF-BB induced pathways, we used known ERK pathway inhibitors, U0126 and PD98059, to refine expression signature

([35], [36]). We compared total expression signatures in cells without PDGF (plus carrier), with PDGF (plus carrier), with PDGF and U0126, and with PDGF and PD98059. For expression signature, we selected only genes which were either highly expressed in cells with PDGF only and low in cells without PDGF or when inhibitors were present, or the genes that were low in cells with PDGF and high in cells without PDGF or when inhibitors were present. To select markers with maximum expression in activated state, and minimum expression without stimulation, we developed and applied rank-pairwise-comparison algorithm as described in Methods.

Top three markers selected were c-fos, early growth response 1 (EGR1), and activity-regulated cytoskeleton-associated protein (ARC). All three genes were previously shown to be upregulated by activation of ERK. ([37], [38], [39]). Two additional genes, ribosomal protein L23a mRNA and ATP synthase were selected as internal controls, because their expression did not depend upon introduction of PDGF-BB or ERK pathway inhibitors U126 and PD98059. The expected pattern of expression of marker and control genes was confirmed by RT-PCR analysis (Figure1).

High-throughput screening to find inhibitors of PDGFR/ERK pathway.

To prove that expression signature of activated PDGFR/ERK pathway can accurately describe the activation status of the system we used the mRNA expression of signature genes to infer a phosphorylation state of PDGFR/ERK pathway in a cell-based screen of a compound library. In this screen we were looking for inhibitors of PDGFR/ERK

pathway, i.e. the compounds that would abate phosphorylation of members of PDGFR/ERK pathway, thus decreasing expression levels of endogenous mRNA of signature genes.

The screen was performed as follows. Adherent cells were serum-starved overnight and compounds were added to the media in the morning (one compound in DMSO per well). After half-hour incubation with compounds, PDGF-BB was added to cells. 45 minutes after addition of growth factor, medium was discarded, cells were lysed, RNA extracted, RT-PCR was performed for the gene expression signatures, and PCR amplicons were quantified by SBE (Figure 1). The library of compounds was prepared in 384 well plates; each plate had 6 replicas in the screen, 3 with average concentration of compounds 10μm, and 3 with average concentration of compounds 50μM. We postulated that compound inhibited the signature if the expression of two marker genes normalized by expression of control genes was significantly (more than one standard deviation) lower than average expression in all wells on a replica plate. Compounds that inhibited signature of activated PDGFR/ERK pathway in 4 out of 6 replicas in primary screen were selected as hits.

Validation of hit compounds.

The three hit compounds selected in the screen were Aurintricarboxylic Acid (Free acid), Aurintricarboxylic Acid triammonium salt (Aluminon), and Quinacrine dihydrochloride (Mepacrine) (Figure 2A).

Western analysis of total cell lysates treated with these compounds along with PDGF demonstrated, that both Aurintricarboxylic Acid and its salt, but not Quinacrine dihydrochloride, abrogated phosphorylation of ERK (Figure 2B). Members of ERK pathway are phosphorylated in response to activation of the pathway [40], thus disruption of phosphorylation proves that Aurintricarboxylic Acid (ATA) is an inhibitor of ERK pathway.

The third hit compound, Quinacrine dihydrochloride, did not inhibit ERK phosphorylation, but it has been previously shown to be a non-specific inhibitor of Phospho Lipase A(2) (PLA(2))([41], [42]). Activated ERK phosphorylates PLA(2) [43], and as a result stimulate translation of c-fos and EGR1 genes [44].

Disruption of phosphorylation of ERK by ATA was an indication that ATA inhibited PDGFR/ERK pathway upstream of ERK. Subsequent analysis has shown that phosphorylation of both MEK (Figure 3A) and PDGF receptor (Figure 3B) was abrogated in response to ATA, thus pointing to the PDGF receptor as a possible ATA target. PDGFR inhibition was shown in both human and mouse fibroblasts, and in neuroblastoma cell line.

ATA withdrawal from cells before introduction of PDGF has not substantially affected inhibition of PDGFR. Figure 4 represents the results of an experiment where TIP5 cells were first incubated with ATA for 30min, then ATA was removed and cells were washed

with ATA-free medium. When PDGF was added to cells, PDGFR phosphorylation was still inhibited, despite the fact that ATA was added to cells prior to washing.

To localize a receptor domain necessary and sufficient for ATA inhibition, we used three chimeras. First chimera, EK-R, consists of the extracellular domain of EGFR and the transmembrane and cytoplasmic domains of c-kit. [45]. Second, TEL/PDGFR-beta protein is a fusion of amino terminus of TEL, containing the helix-loop-helix (HLH) domain, to the transmembrane and cytoplasmic domain of the PDGFR-beta [46]. Third chimera, PER, is comprised of the ectodomain of PDGFR in-frame with the transmembrane and cytoplasmic domains of EGF receptor [47]. ATA did not inhibit phosphorylation of c-kit kinase domain in activated EK-R protein (Figure 5A) and phosphorylation of PDGFR kinase domain in TEL/PDGFR-beta (Figure 5B). These results suggest that extracellular domain of c-kit and PDGFR is necessary for ATA inhibition of the receptors. Moreover, ATA inhibited phosphorylation of PDGF-activated PER chimera in TIP5 cells (Figure 5C), demonstrating that extracellular domain of PDGFR is sufficient for ATA inhibition. On the other hand, PER construct constitutively activated in 501MEL cells was not inhibited by ATA, indicating that ATA did not inhibit dimerization of PDGF receptor (results not shown). Taken together, these results suggest that ATA blocks ligand binding site of PDGFR, preventing binding PDGF to the receptor.

To estimate specificity and selectivity of ATA inhibition, we choose one receptor from PDGFR kinase family, c-Kit, and two tyrosine kinase receptors from other kinase

families, EGFR and IGF-1R [48]. While EGFR and IGF-1R phosphorylation does not seem to be affected by ATA, c-Kit was inhibited almost as strongly as PDGFR (Figure 6).

The selectivity exhibited by ATA parallels sequence similarities between receptors. The alignment of protein sequences of extracellular domains of PDGFR-beta and c-Kit showed that 26% of the residues were identical and 41% highly similar. No significant similarity was found between extracellular domains of PDGFR and EGFR, c-Kit and EGFR, PDGFR and IGFR, c-Kit and IGFR, and EGFR and IGFR. Reciprocally, the scores of sequence identity and similarity between kinase domains of PDGFR and c-Kit, 51% and 66% respectively, were higher than scores for all other tested receptor pairs (Table 1).

Structure-activity relationships in the series of ATA analogues.

To determine whether the inhibitory properties of ATA were unique, the inhibitory activity of ATA was compared with that of a series of ATA analogues. Searching ACD, we identified a diverse set of ATA structural analogs (Figure 7) to test three hypotheses on structure activity relationship in the series. First three compounds, Methylenedisalicylic acid, Salicylic acid and 3-Methylsalicylic acid (Figure 7A), were chosen to examine if the number of rings in the compound affects the inhibitory strengths of ATA. Second group of compounds, Uranine K, Phenolphthalein Sodium Salt, and Aurin (Figure 7B), was used to evaluate importance of ATA carboxyl and hydroxyl

groups. The remaining compounds were utilized to test if group substitutions and ring modifications would change ATA inhibitory properties. No compounds in the series inhibited PDGFR at low micromolar concentration. In the first group, Methylenedisalicylic Acid (Figure 7D), but not Methylsalicylic or Salicylic acids inhibited PDGFR phosphorylation at 50microM, thus suggesting that increasing the number of rings from one to three boosts inhibitory properties. The positions and number of carboxyl and hydroxyl groups were essential for PDGFR inhibition, as indicated by the fact that no compounds in the second group inhibited PDGFR at 100 microM concentration. These results corroborate earlier reports that both Aurin triphenyl methane ring system and the carboxylic acid groups are necessary for ATA inhibitory properties ([49],. [50], [51], [52]).

In the third group, Basic Violet 3, Ethyl Violet and Victoria Pure Blue BO inhibited PDGFR in 5-10 microM range (Figure 7E). Interestingly, these three compounds exhibited less specific pattern of receptor inhibition than ATA, inhibiting not only c-Kit (Figure 7F), but also EGFR (Figure 7G) and IGFR at 10-100 microM (Figure 7H). Moreover, different from ATA, Ethyl Violet and Victoria Pure Blue were readily translocating across cell membrane, as indicated by their inhibition of cytoplasmic TEL/PDGFR in Ba/F3 cells at 10 microM (Figure 7I). Taken together, these results suggest that Basic Violet 3, Ethyl Violet and Victoria Pure Blue BO inhibitory mechanism is different from a mechanism of receptor inhibition of ATA.

Discussion

ATA inhibits interactions on cell surface in vivo.

Aurintricarboxylic acid (ATA), a polymeric carboxylated triphenylmethane derivate with a molecular weight range of 422-6,500 ([53], [54]), has displayed a wide range of biological activity.

In vitro, probably because polyanionic structure of ATA contributes to ATA's nonspecific inhibition of enzymes in cell-free system [55], ATA has been reported to inhibit multiplicity of proteins. ATA has been shown to inhibit enzymes dependent on protein-nucleic acid interactions ([56], [57]), such as DNA and RNA polymerases ([58], [59]), reverse transcriptase ([58], [60] , [61], [62]), nucleases ([63], [64], [65]), primases ([58]), topoisomerases ([66], [67]), ribonucleotide reductases [68], aminoacyl-tRNA synthetase [69], NFkappaB [70] and HIV-1 integration protein [71]. A number of proteins that do not bind nucleic acids has been shown to be inhibited by ATA in vitro, for example phosphatases ([72], [73]), NAD(H)/NADP(H)-requiring enzymes [74], aminopropyltransferases [75], mu- and m-calpain [76], delta aminolevulinic acid dehydratase [77], glucose-6-phosphate dehydrogenase [78], phenylalanine:tRNA ligase [79] and kinases, such as phosphofructokinase [80], ERK, p38 MAPK, IKK [81], IP3Ks and IPMK [82]. In addition, ATA has been reported to inhibit protein synthesis in vitro ([83], [84], [85]).

In vivo, on the other hand, ATA has exhibited a more limited range of activity. ATA has been reported to affect activity of receptors, most in the 10 microM and higher range. ATA blocked binding of gp120, the HIV coat protein, to its receptor, CD4, at 3-10 microM ([51], [86]), obviated binding of interferon-alpha to its receptor in a dose-dependent manner in 12-50 microM range [86], and prevented von Willebrand factor binding to platelet receptor glycoprotein Ib (GPIb) ([87], [88], [54]). ATA was shown to be a N-methyl-D-aspartate (NMDA) receptor antagonist with IC50 of 26.9 microM [89] and was reported to antagonize excitotoxicity at both NMDA and non-NMDA glutamate receptors in 50-100microM range [90]. ATA inhibited projesterone receptor at 100-500microM [91], estradiol receptor at 100-200 microM [92], and glucocorticoid receptor complex at 50-200microM [93]. ATA also was reported to activate IGF-1R (25-100 microM) ([94], [95]) and erbB4 (10 microM) [96].

At 0.5-2.5 micrograms/ml (approximately 1.2-6 microM) ATA was reported to reverse transformed phenotype of cells transfected with basic fibroblast growth factor (bFGF) fused to a signal peptide sequence (spbFGF cells) [97]. It was suggested, on the basis of ATA fluorescence studies, that ATA binds to acidic fibroblast growth factor, altering its physicochemical properties and decreasing its mitogenic activity [98], although these results were not confirmed by more direct biochemical methods.

All the interactions described above take place on the cell surface, which is consistent with the observations that ATA does not readily penetrate cellular membrane. Radioactive ATA was reported not to be taken up by HeLa cells, green monkey kidney

cells, rabbit reticulocytes, or a variety of bacterial cells [99] and ATA did not inhibit proteins in cells at concentrations hundreds of times higher than concentrations sufficient for inhibition in vitro [90]. At high concentration, though, Cho and coworkers report that ATA fluorescence was detected inside the nuclei of cells treated with 500 microM ATA [73].

Here we report that ATA inhibits PDGFR receptor at 1microM and a member of the same RTK family, c-Kit, at 2-5microM, and does not inhibit IGF-R1 and EGFR at comparable concentrations. At concentrations up to 100 microM we did not detect any activation effect of ATA on IGF-1R. As shown in Figure 4, ATA still inhibited PDGF receptor even when the cells were washed to remove compound before introduction of PDGF, indicating that ATA did not inhibit PDGF itself. The washing step, however, has somewhat increased ATA concentration required to inhibit PDGFR phosphorylation, which is consistent with ATA binding extracellular domain of PDGF receptor. Data on inhibition of chimera receptors, presented in Figure 5, shows that extracellular domain of PDGF was necessary and sufficient for ATA inhibition of receptor, and ATA did not inhibit constitutively active PER chimera. Taken together, these results suggest that ATA blocks ligand-binding site of PDGF receptor. As this interaction takes place on the cell surface, ATA does not need to translocate across the cell membrane and a low micromolar concentration of ATA is sufficient to inhibit receptor activation.

As ATA inhibits PDGFR and c-Kit, it is possible that inhibition of other receptors of this family is responsible for previously reported inhibition of members of Januse kinase

(JAK)/STAT pathway in RAW 263.7 cells [100]. Lin and coworkers reported that in RAW 264.7 cells ATA inhibited a number of members of STAT signaling pathways, namely STAT1, STAT3, STAT5, STAT6, Jak1 at 30microM and higher, and Jak2 at 3 microM pathway [100], as well as some cytosolic kinases, namely ERK, p38 MAPK, and IKK at 100 microM [81]. PDGFR, KIT/SCFR, CSF-1R and FLK2/FLT3 are closely related receptor tyrosine kinases, belonging to the same RTK family [48]. GM-CSF and M-CSF receptors in RAW 264.7 macrophages ([101], [102]) have been reported to signal through the JAK/STAT pathway ([103], [104]). By inhibiting these receptors, ATA could affect the downstream members of the JAK/STAT pathway, as well as members of the other common intermediate cascade initiating from cytokine receptors, Ras/Raf/MEK/ERK (MAPK) [105].

Applications of endogenous gene expression-based methods for HTS.

ATA identification as an inhibitor of PDGF receptor, an upstream member of PDGFR/ERK pathway, provides evidence that assay based on endogenous gene expression can be successfully used to find inhibitors of signal transduction cascades in a high throughput screening of compound libraries. The results of this study suggest that in a high throughput setting gene expression signature can be utilized as a reliable, though indirect, read-out of activation status of a pathway, and, moreover, it is possible to select a set of marker genes whose endogenous expression pattern could successfully predict the ability of a compound to modulate a fast biochemical pathway.

This expression-based screening method can be easily adopted for any pathway as long as modulation of the pathway results in some expression changes and the half-life of the transcripts allows mRNA isolation in a high throughput setting. Investigating PDGFR/ERK pathway, we took advantage of the existing commercially available inhibitors of ERK cascade to refine the signature. However, this step is not necessary and can be omitted if no such reagents exist for a pathway of interest. For a novel unexplored pathway one could define activation signature by comparing expression before and after activation of the pathway with the caveat that this method could produce an extra amount of false positives.

Another advantage of the method, that can be especially beneficial for a not very well investigated pathways, is that by utilizing expression of the genes downstream of the pathway as a read-out, we are not limiting ourselves to finding inhibitors to a particular member of the pathway, but to the pathway as a whole, thus maximizing effectiveness of the screen. The downside of the approach, though, is that an additional effort is required for identification of the target of the newly discovered hits, but this is a common problem for cell-based screening methods.

Utilizing endogenous expression of the pathway genes as a read-out can also be advantageous for very unstable systems, where introduction of a reporter or overexpression of proteins of interest might disturb the equilibrium in the system and thus complicate the interpretation of the results of the screen.

The major limitation of the screen as it stands now is a limited sensitivity and resolution of SBE, the current method utilized for detection of gene expression signature in a high throughput setting. To be reliably detected by SBE, the signature has to have a small number of transcripts, thus limiting the signature to only a handful of genes. Moreover, the mRNA has to be amplified by RT-PCR, thus complicating detection by decreasing signal to noise ratio. While, as demonstrated in this study, in its current form gene expression based high throughput screening can still be successfully used for finding inhibitors of fast biochemical pathways, in the future this method will benefit from developing a more sensitive and reliable detection technique that could allow tracing subtle changes in endogenous mRNA expression by utilizing more complex and diverse signatures.

Conclusions

With a rapid progress in finding new distinct proteins and their splice variants, there is a growing need for developing reliable and universally adaptable high throughput biological assays. Here we describe a high throughput gene expression-based screening assay that employs a limited endogenous mRNA expression signature to monitor an activation state of a signal transduction pathway. This flexible and reproducible method is demonstrated by identifying a novel low-micromolar inhibitor of PDGF receptor in a screening of a small chemical compound library.

The results of this study suggest that it is possible to find modulators of pathways by inferring activation state of a biological system through monitoring its endogenous mRNA expression signature, and not measuring the activity of the pathway directly. This can be especially beneficial for finding modulators of novel molecular targets with unknown mechanism of action. Currently, before a protein can be used as a drug target, some sort of function has to be attributed to it. This is often done by assigning the protein of interest to a protein class by sequence homology with known members of that class [106]. Using mRNA expression signature as a read-out in a cell-based assay may help circumvent this step and directly proceed to finding modulators for the novel proteins. Additionally, using internal reference, such as mRNA expression, as a read-out, eliminates necessity for overexpressing target of interest, co-expression of multiple proteins, or expression of "reporter" genes, typical for high-throughput cell-based assays [107].

In conclusion, the versatility and adaptability of assays based on monitoring a limited endogenous mRNA expression signature suggest that these assays should find widespread use in experiments directed toward identification of novel modulators of signaling pathways and will prove a powerful tool in drug discovery and in basic research.

Materials and Methods.

Reagents.

EK-R construct was kindly provided by Dr. Ullrich. PER chimera was a gift of Dr. Tyson and Dr. Bradshaw.

Chemical compounds Apigenin, U0126, Quinacrine Dihydrochloride and Aurintricarboxylic Acid were obtained from Calbiochem (http://www.calbiochem.com); Methyl Violet B base, Rhodamine 6 G tetrafluoroborate, Sulforhodamine, Ethyl Violet, Victoria Pure Blue BO, Rhodamine B, Lissamine Green B, Methyl violet 2B, Rhodamine 6G, (L-Asp)2Rhodamine 110 TFA, Rhodamine 110 chloride, Eosin B, Rhodamine 123 hydrate, Rhodamine 19 perchlorate, Acid Fuchsin calcium salt, p-Rosolic acid, Basic Violet2, Gentian Violet , Pararosaniline hydrochloride, Salicylic acid were purchased from Sigma (http://www.sigmaaldrich.com); and 3-Methylsalicylic acid , 5,5'-Methylenedisalicylic acid, Phenolphthalein sodium salt, and Uranine K were obtained from ABCR (http://www.abcr.de).

Growth factors PDGF, EGF, and SCF were obtained from Cell Signaling (http://www.cellsignal.com), R^3IGF from Sigma, and IL3 from R@D Systems (http://www.rndsystems.com).

Cell culture reagents RPMI 1640, DMEM, HAM's F-10 were purchased from Mediatech (http://cellgro.com), penicillin, streptomycin from Invitrogen (http://invitrogen.com), fetal bovine serum (FBS) from Sigma (http:// www.sigmaaldrich.com).

p44/42 MAP Kinase, Phospho-p44/42 MAP Kinase (Thr202/Tyr204), MEK1/2, Phospho-MEK1/2 (Ser217/221), PDGF-BB, Phospho-PDGFR-beta (Tyr751), Phospho-EGFR (Tyr1068), c-Kit, Phospho-c-Kit (Tyr719), IGF-IαR, and Phospho-IGF-IR (Tyr1131)/Insulin Receptor (Tyr1146) antibodies were obtained from Cell Signaling. EGFR antibodies were purchased from Santa Cruz Biotechnology (http://scbt.com).

Cells.

SH-SY5Y neuroblastoma cells were purchased from American Type Culture Collection (http://www.atcc.org). The IL3-dependent pro-B lymphoid cell line Ba/F3 and Ba/F3 cells expressing TEL/PDGFR-beta [46]were obtained from D. Gary Gilliland. TIP5 primary fibroblasts [108] were a gift from Stephen Lessnick. We thank Dr. Ruth Halaban for 501 MELl human melanoma cells. SH-SY5Y, BaF3, and BaF3 cells expressing TEL/PDGFR-beta were maintained in RPMI 1640 medium, TIP5 cells were cultured in DMEM, and 501 MEL cells were grown in Ham's 10 medium. Medium for IL3 dependent Baf3 cells was supplemented with 0.05 ng/mL IL3. Media for all cell lines contained 10% fetal bovine serum, 10 U/ml penicillin, and 10 μg/ml streptomycin. All cells were grown at 37°C in 5% CO_2.

Characterization of activation signature for ERK/PDGFR pathway.

SH-SY5Y cells were grown to confluence and starved overnight in serum-free medium in order to silence any sustained effects from growth factor signaling. Prior to induction with 50ng/mL PDGF, cells were treated with pathway inhibitors 74 microM Apigenin or 50 microM U0126, or with DMSO (carrier) for 60 minutes. Total RNA was isolated 30 minutes after PDGF addition. Each sample was done twice. The RNA was processed and hybridized with Affymetrix U133A GeneChips as described in [18].

To define ERK/PDGFR activation signature, a pair-ranking algorithm was used. Genes were ranked according to their raw expression values on each chip. Ten genes with maximum change in ranking were selected for each one of three pairs of conditions: cells with PDGF versus cells without PDGF, cells with PDGF versus cells with PDGF and Apigenin, and cells with PDGF versus cells with PDGF and U0126. Three genes common for all three conditions were selected as signature of activated ERK/PDGFR pathway.

These genes were early growth response 1 (EGR1), c-fos, and activity-regulated cytoskeleton associated protein (APC). Control genes independent of PDGF induction were selected among the genes with minimum change in ranking through all conditions. The genes chosen were ATP synthase, nuclear gene encoding mitochondrial protein, and ribosomal protein L23a.

In a high throughput setting the signature of activated ERK/PDGFR pathway was trimmed from three to two genes based on their relative strength of expression in TIP5 cells.

Screening methods.

TIP5 cells were grown to confluence and starved overnight with 20 microL serum-free medium per well of 384-well plates. 20 microL of compounds diluted in media were added so that final concentration of compounds would be approximately 10 microM in three out of six replicas, and 50 microL in three remaining replicas. After 30 minutes of compound treatment, cells were induced with 40 microL of PDGF diluted in media (final PDGF concentration 40ng/mL). After 40 minutes of PDGF induction, media was discarded, cells were lysed and RNA was extracted and quantified as described in [10]. The primers used for multiplex PCR reaction are as follows. Early growth response 1 (EGR1): 5'- AGC GGA TAA CAC CTC ATA CCC ATC CCC TGT -3' and 5'- AGC GGA TAA CTG TCC TGG GAG AAA AGG TTG -3', c-fos: 5'- AGC GGA TAA CGC TTC CCT TGA TCT GAC TGG -3' and 5'- AGC GGA TAA CAT GAT GCT GGG AAC AGG AAG -3', ATP synthase: 5'- AGC GGA TAA CCA AAG CCC ATG GTG GTT ACT -3' and 5'- AGC GGA TAA CGC CCA ATA ATG CAG ACA CCT -3', L23a: 5'- AGC GGA TAA CAA GAA GAA GAT CCG CAC GTC -3' and 5'- AGC GGA TAA CCG AAT CAG GGT GTT GAC CTT -3'.

The following primers were used for SBE reaction. EGR1: 5'- TTC CCC CTG CTT TCC CG -3', c-fos: 5'- TGC CTC TCC TCA ATG ACC CT -3', ATP synthase: 5'- GAC TGT GGC TGA ATA CTT CA -3', L23a: 5'- GTC TGC CAT GAA GAA GAT AGA A -3'.

To select compounds that inhibited expression of pathway signature the following procedure was performed. For each compound on each plate four ratios were determined: expression of EGR1 vs. expression of ATP synthase ($V_{EGR1/ATP\ synthase}$), EGR1 vs. L23a ($V_{EGR1/L23a}$), c-fos vs. ATP synthase ($V_{c\text{-}fos/ATP\ synthase}$), and c-fos vs. L23a ($V_{c\text{-}fos/L23a}$). For each plate a median (μ) and standard deviation (σ) were determined for each of four ratios. A compound was considered a plate hit if each of the four ratios for compounds were at least one standard deviation smaller than median ($V < (\mu-\sigma)$, or ($V_{EGR1/ATP\ synthase} < \mu_{EGR1/ATP\ synthase} - \sigma_{EGR1/ATP\ synthase}$)&($V_{EGR1/L23a} < \mu_{EGR1/L23a} - \sigma_{EGR1/L23a}$)&($V_{c\text{-}fos/ATP\ synthase} < \mu_{c\text{-}fos/ATP\ synthase} - \sigma_{c\text{-}fos/ATP\ synthase}$)&($V_{c\text{-}fos/L23a} < \mu_{c\text{-}fos/L23a} - \sigma_{c\text{-}fos/L23a}$)). Compounds that were plate hits in four out of six replicas were selected for further consideration.

Western blotting.

If cells had to be transfected with expression construct, 501 MEL cells of TIP5 cells were grown overnight to 50% confluence and transfected using Fugene 6 transfection reagent (roche.com) as recommended by manufacturer. 24 hours after transfection, medium was exchanged for a serum free one, and cells were serum starved overnight.

Otherwise, adherent cells (TIP5, MEL 501) were grown to confluence, serum starved overnight, and treated with compounds and growth factors as indicated. Cells growing in suspension (BaF3 cells and BaF3 cells expressing TEL/PDGFR protein), were grown to 10^6 cells/mL and treated with compounds as indicated. After treatment media was removed, adherent cells were scraped with sample buffer from cell signaling, suspension

cells were pelleted and resuspended in sample buffer. The resulting lysates of approximately $1x10^5$ cells were boiled, chilled, run on 4-15% gradient gels from BioRad (biorad.com), transferred to the polyvinylidene difluoride membrane from Millipore (http://millipore.com), blocked, probed and visualized as recommended by antibody manufacturers.

Sequences alignment.

Comparative sequence analysis between PDGF (UniProtKB/Swiss-Prot entry P09619), c-Kit (UniProtKB/Swiss-Prot entry P10721), EGFR (UniProtKB/Swiss-Prot entry P00533), and IGFR (UniProtKB/Swiss-Prot entry P08069) was performed with BLAST 2 SEQUENCES ([109], [110]).

Acknowledgements

We thank members of the laboratory of T.R. Golub, the Broad Institute Cancer Program, and, in particular, K. Stegmaier, D. Peck, C. Ladd-Acosta, S. Gabriel, S. Lessnick, J. Grabarek and M. Reich for their contributions, H. Widlund and V. Igras for cell culture advice and Eric Lander for helpful discussions.

Figure Legends.

Figure 1. A. Genes (in rows) sorted by their expression in samples (columns) with or without U0126, Apigenin, and PDGF. Red indicates high relative expression, blue low expression. B. Screening schema overview. B. RT-PCR of signature genes in 4 sample wells.

Figure 2. A. Hit compounds identified in the screen. B. Western analysis of total TIP5 cell lysates. Cells were serum starved overnight and treated with ATA, Aluminon, and Quinacrine Dihydrochloride in presence of PDGF.

Figure 3. A. Western analysis of total TIP5 cell lysates. Cells were serum starved overnight and treated with ATA and PDGF probed with MEK1/2 antibody. B. Western analysis of total TIP5 cell lysates. Cells were serum starved overnight and treated with ATA and PDGF.

Figure 4. Western analysis of total TIP5 cell lysates.. Cells were serum starved overnight and then incubated with ATA. After ATA was removed by washing, cells were induced with PDGF.

Figure 5. A. Western analysis of total TIP5 cell lysates. Cells were transfected with EK-R plasmid, serum-starved overnight and treated with ATA, EGF and PDGF. B. Western analysis of total lysates of Ba/F3 cells expressing TEL/PDGFR-beta fusion protein. Cells

were treated with ATA.C. Western analysis of total TIP5 cell lysates. Cells were transfected with PER plasmid, serum-starved overnight and treated with ATA and PDGF.

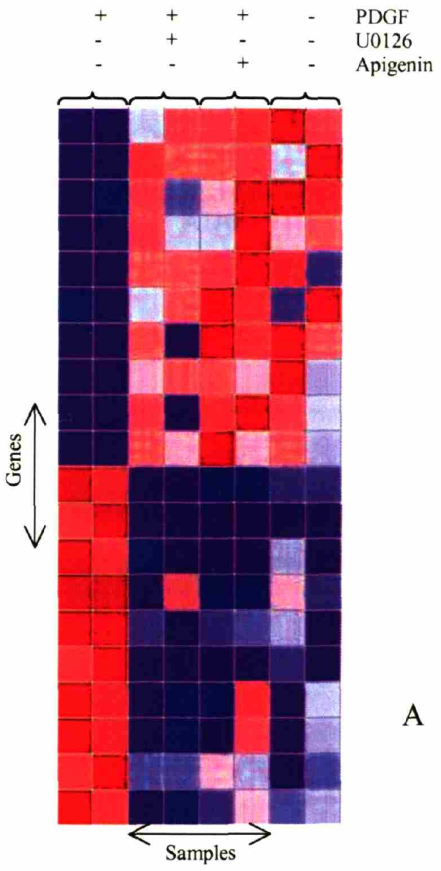
Figure 6. A. Western analysis of total MEL501 cell lysates. Cells were serum starved overnight and treated with ATA and SCF. B. Western analysis of total TIP5 cell lysates. Cells were serum starved overnight and treated with ATA and EGF. C. Western analysis of total TIP5 cell lysates. Cells were serum starved overnight and treated with ATA and IGF.

Figure 7. A.B.C. ATA analogs. D. E. Western analysis of total TIP5 cell lysates. Cells were serum starved overnight and treated with 5,5'-Methylenedisalicylic acid (abbreviated MSD in D), Ethyl Violet, Victoria Pure Blue BO and Basic Violet 3 (abbreviated EV, VPB, and BV3, respectively, in E) and PDGF. F. Western analysis of total MEL501 cell lysates. Cells were serum starved overnight and treated with Ethyl Violet, Victoria Pure Blue BO, and Basic Violet 3 and SCF. G. Western analysis of total TIP5 cell lysates. Cells were serum starved overnight and treated with Ethyl Violet, Victoria Pure Blue BO, and Basic Violet 3 and EGF. H. Western analysis of total TIP5 cell lysates. Cells were serum starved overnight and treated with Ethyl Violet, Victoria Pure Blue BO, and Basic Violet 3 and IGF. I. Western analysis of total lysates of Ba/F3 cells expressing TEL/PDGFR-beta fusion protein. Cells were treated with Ethyl Violet and Victoria Pure Blue BO.

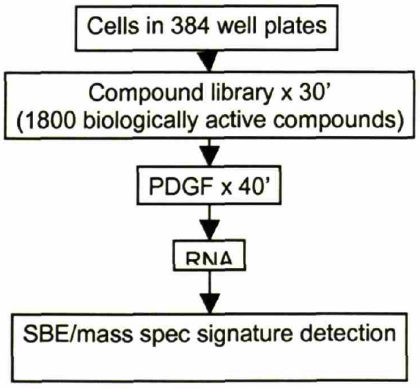
Abbreviations

PDGF, platelet derived growth factor; EGF, epidermal growth factor; IGF, insulin-like growth factor; ERK, extracellular regulated kinase; MEK, mitogen/extracellular signal-regulated kinase; MAP, mitogen activated protein; RTK, receptor tyrosine kinases; ATA, Aurintricarboxylic Acid; EGR1, early growth response 1; ARC, activity-regulated cytoskeleton-associated protein; PLA(2), Phospho Lipase A(2); HTS, high throughput screen.

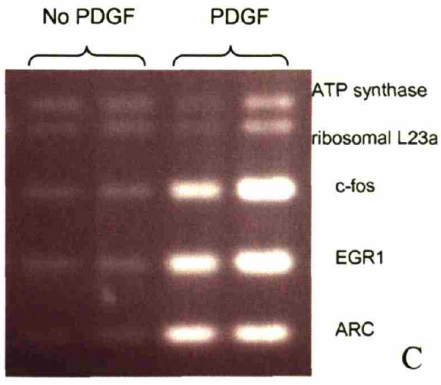
Figure 1.


+ + + - PDGF
- + - - U0126
- - + - Apigenin
Genes
Samples

A

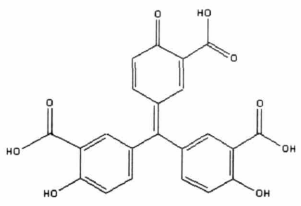

Cells in 384 well plates
Compound library x 30'
(1800 biologically active compounds)
PDGF x 40'
RNA
SBE/mass spec signature detection

B

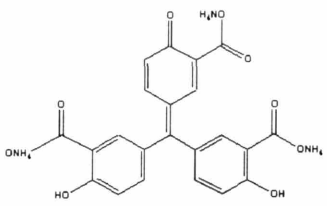

No PDGF
PDGF
ATP synthase
ribosomal L23a
c-fos
EGR1
ARC

C

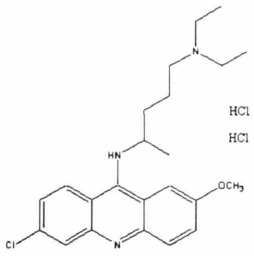
Figure 2



Aurintricarboxylic Acid
(ATA)

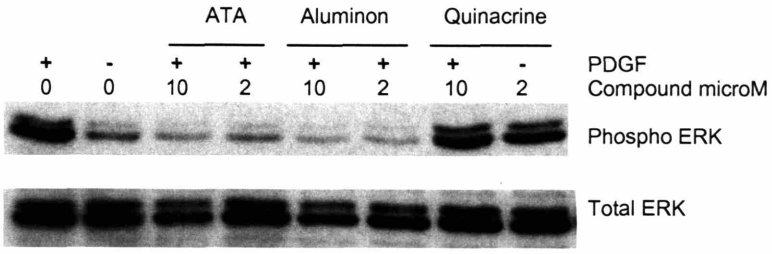


ATA triammonium salt
(Aluminon)



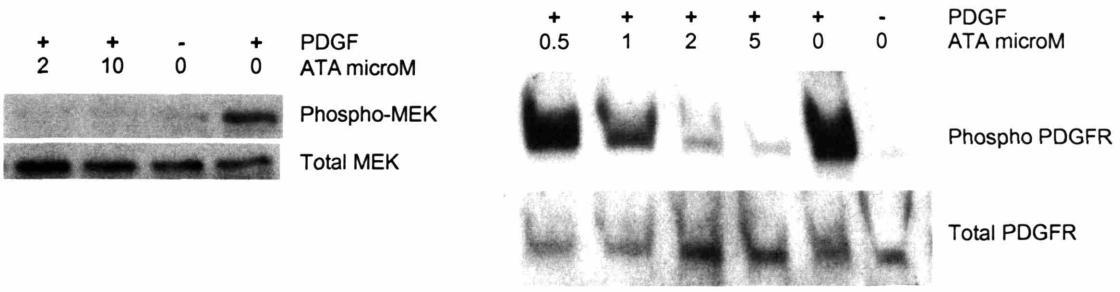
Quinacrin
Dihydrochloride

A



B

Figure 3


+ + - + PDGF
2 10 0 0 ATA microM
Phospho-MEK
Total MEK
+ + + + + - PDGF
0.5 1 2 5 0 0 ATA microM
Phospho PDGFR
Total PDGFR

A B

Figure 4.

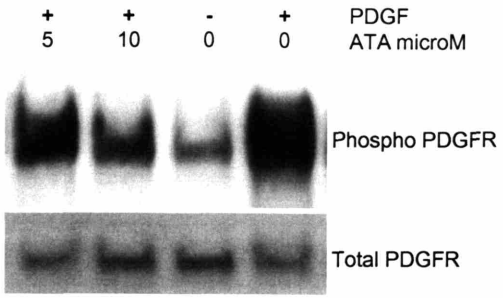
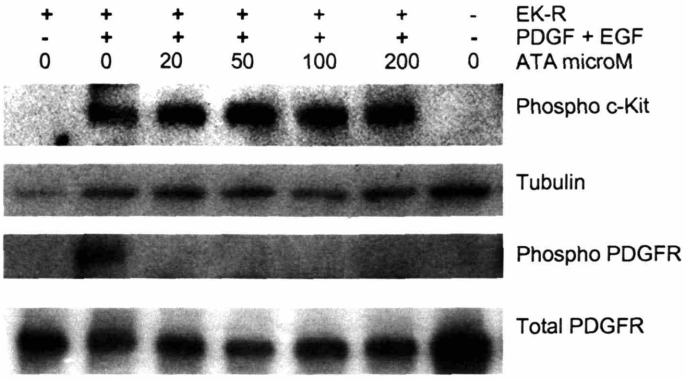
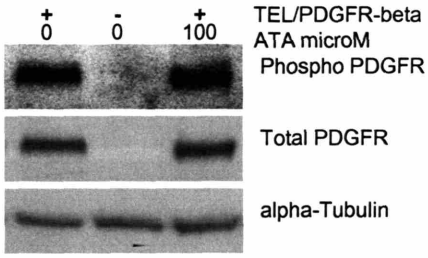


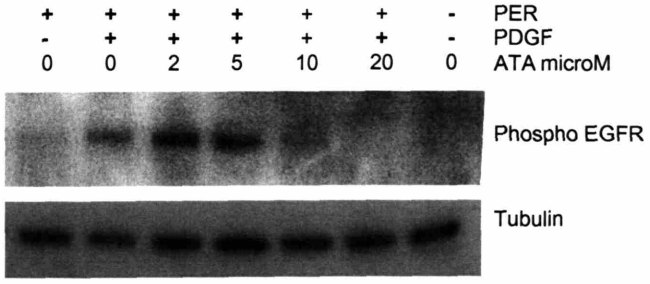
Figure 5.


+ + + + + + - EK-R
- + + + + + - PDGF + EGF
0 0 20 50 100 200 0 ATA microM
Phospho c-Kit
Tubulin
Phospho PDGFR
Total PDGFR

A


+ - + TEL/PDGFR-beta
0 0 100 ATA microM
Phospho PDGFR
Total PDGFR
alpha-Tubulin

B


+ + + + + + - PER
- + + + + + - PDGF
0 0 2 5 10 20 0 ATA microM
Phospho EGFR
Tubulin

C

Figure 6.

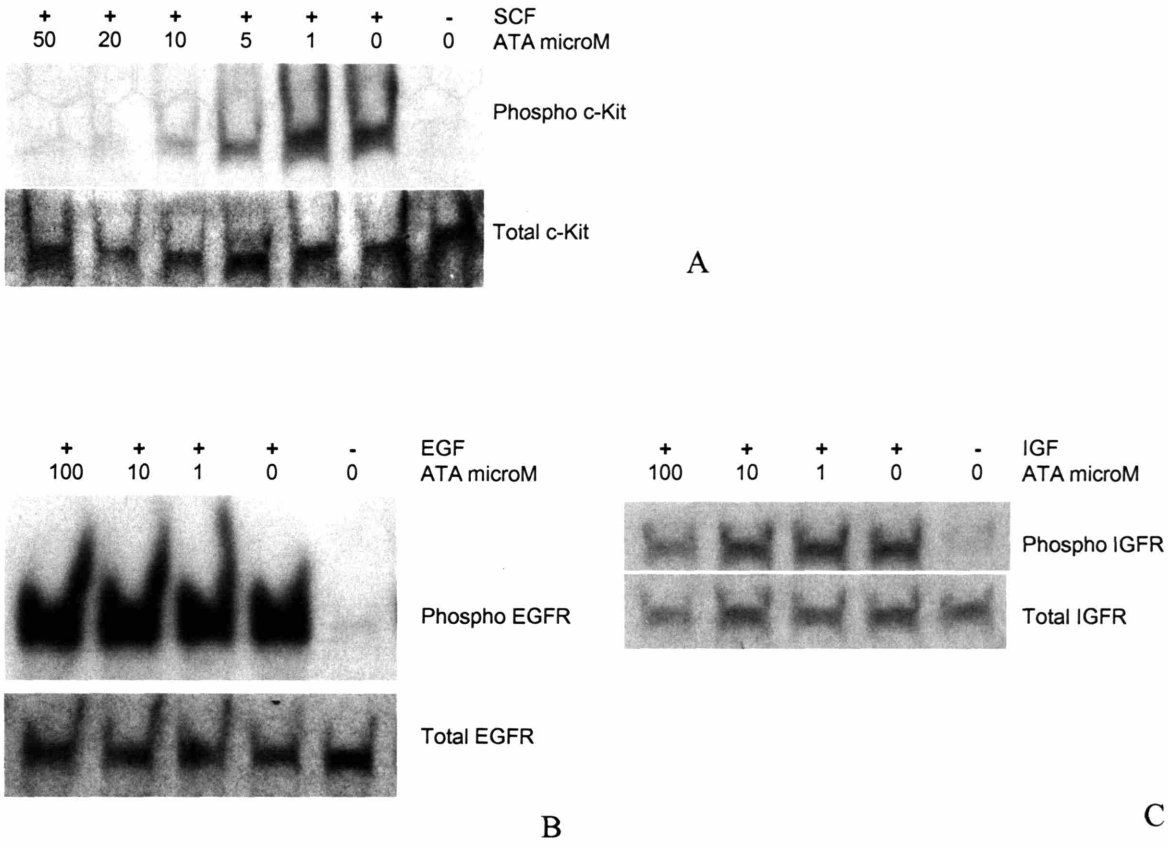
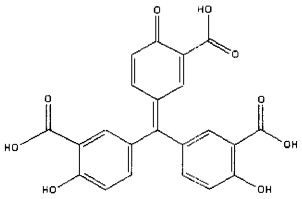
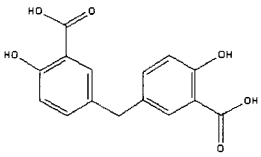

+ + + + + + - SCF
50 20 10 5 1 0 0 ATA microM
Phospho c-Kit
Total c-Kit
A
+ + + + - EGF
100 10 1 0 0 ATA microM
Phospho EGFR
Total EGFR
B
+ + + + - IGF
100 10 1 0 0 ATA microM
Phospho IGFR
Total IGFR
C

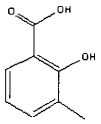
Figure 7.



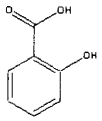
Aurintricarboxylic acid



5,5'-Methylenedisalicylic acid

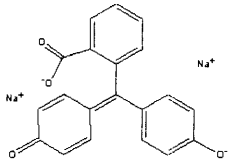


3-Methylsalicylic acid

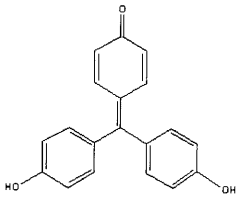


Salicylic acid

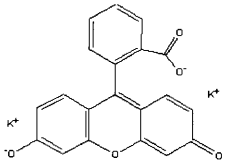
A



Phenolphthalein sodium salt

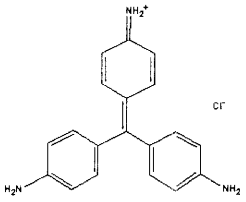


Aurin

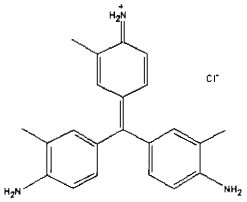


Uranine K

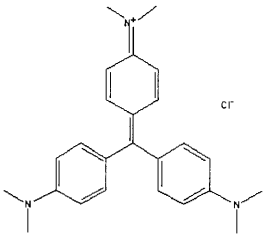
B



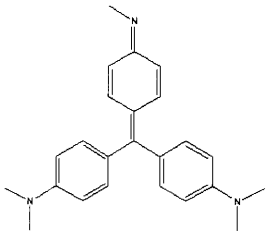
Pararosaniline hydrochloride



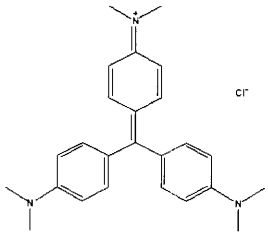
Basic Violet 2



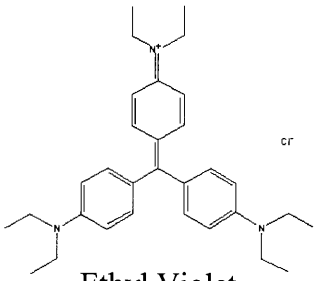
Methyl Violet 2B



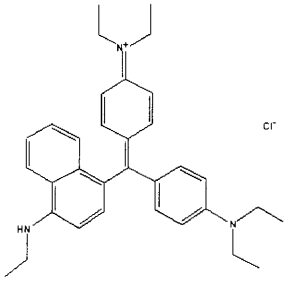
Methyl Violet B base



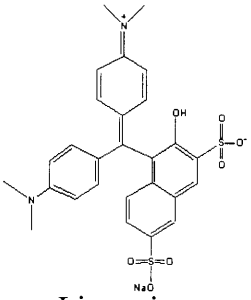
Basic Violet 3



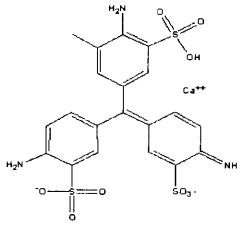
Ethyl Violet



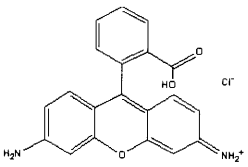
Victoria Pure Blue BO



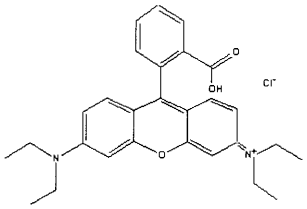
Lissamine Green B



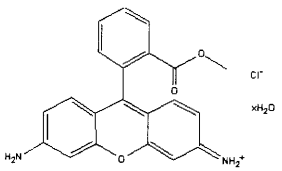
Acid Fuchsin calcium salt



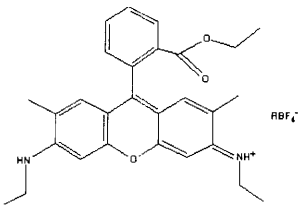
Rhodamine 110 chloride



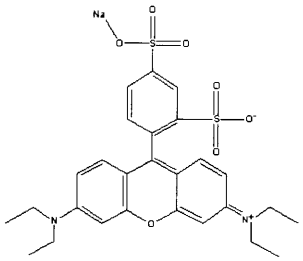
Rhodamine B



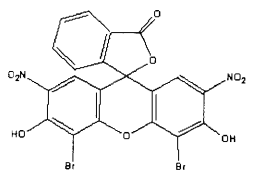
Rhodamine 123 hydrate



Rhodamine 6 G tetrafluoroborate

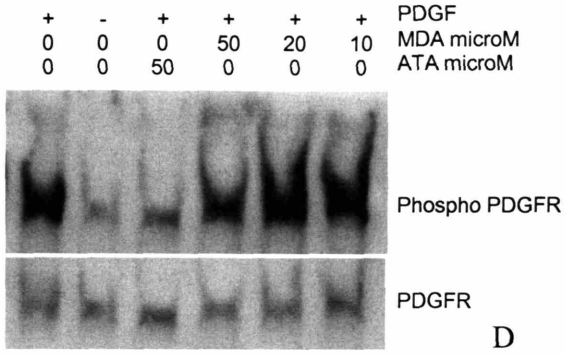


Sulforhodamine B

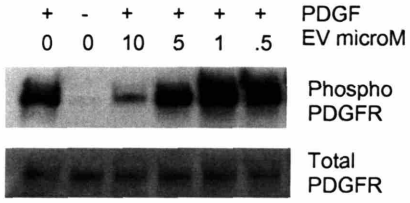


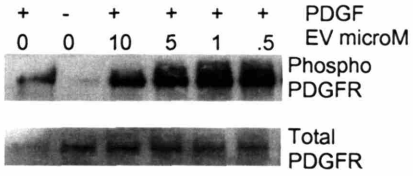
Eosin B

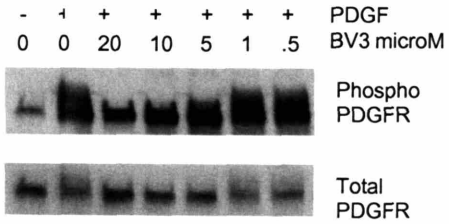
C


+ - + + + + PDGF
0 0 0 50 20 10 MDA microM
0 0 50 0 0 0 ATA microM
Phospho PDGFR
PDGFR

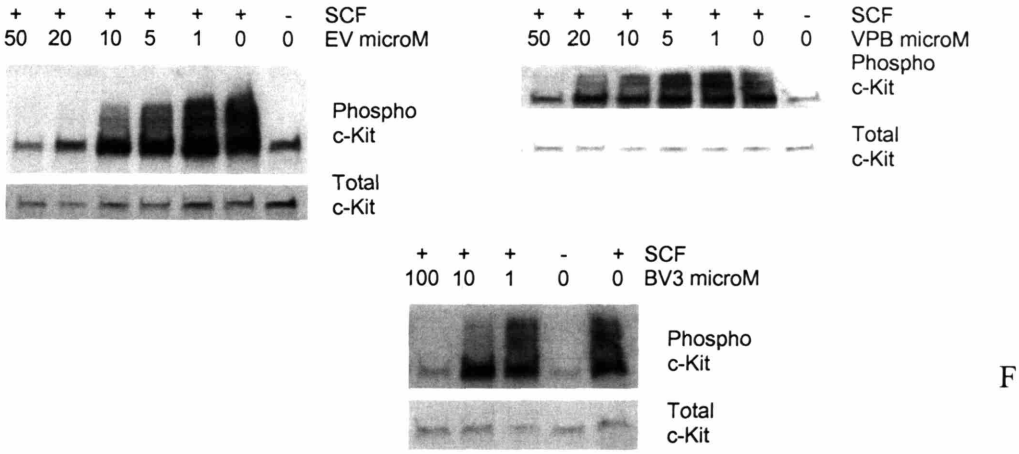
D


+ - + + + + PDGF
0 0 10 5 1 .5 EV microM
Phospho PDGFR
Total PDGFR

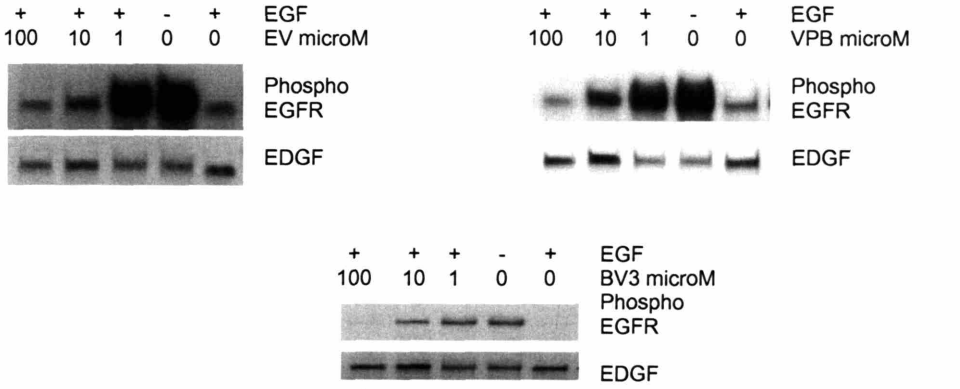

+ - + + + + PDGF
0 0 10 5 1 .5 EV microM
Phospho PDGFR
Total PDGFR


- + + + + + + PDGF
0 0 20 10 5 1 .5 BV3 microM
Phospho PDGFR
Total PDGFR

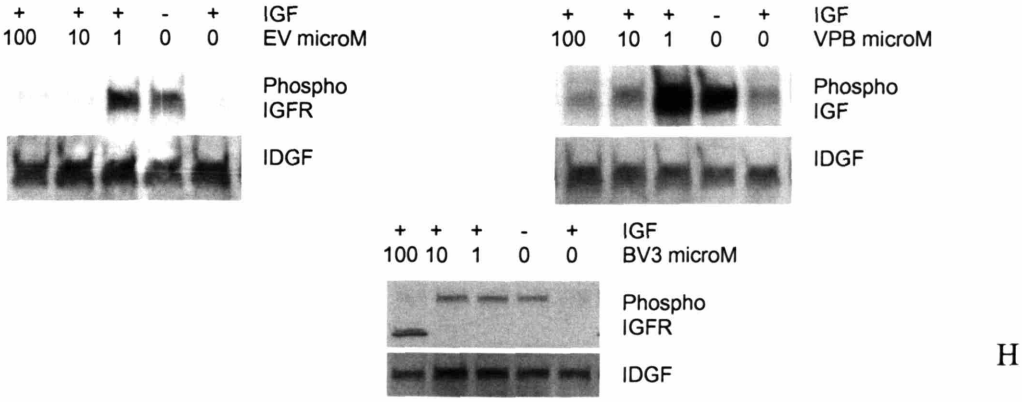
E


+ + + + + + - SCF
50 20 10 5 1 0 0 EV microM
Phospho c-Kit
Total c-Kit
+ + + + + + - SCF
50 20 10 5 1 0 0 VPB microM
Phospho c-Kit
Total c-Kit
+ + + - + SCF
100 10 1 0 0 BV3 microM
Phospho c-Kit
Total c-Kit

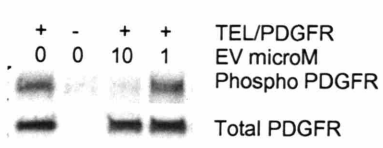
F

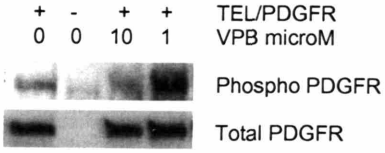

+ + + - + EGF
100 10 1 0 0 EV microM
Phospho EGFR
EDGF
+ + + - + EGF
100 10 1 0 0 VPB microM
Phospho EGFR
EDGF
+ + + - + EGF
100 10 1 0 0 BV3 microM
Phospho EGFR
EDGF

G


+ + + - + IGF
100 10 1 0 0 EV microM
Phospho IGFR
IDGF
+ + + - + IGF
100 10 1 0 0 VPB microM
Phospho IGF
IDGF
+ + + - + IGF
100 10 1 0 0 BV3 microM
Phospho IGFR
IDGF

H


+ - + +
0 0 10 1
TEL/PDGFR
EV microM
Phospho PDGFR
Total PDGFR


+ - + +
0 0 10 1
TEL/PDGFR
VPB microM
Phospho PDGFR
Total PDGFR

I

Table 1.

Extracellular domain	Identities	Positives
PDGFR-beta vs. cKit	26%	41%
PDGFR-beta vs. EGFR	No significant similarity was found	
PDGFR-beta vs. IGFR	No significant similarity was found	
cKit vs. EGFR	No significant similarity was found	
cKit vs. IGFR	No significant similarity was found	
EGFR vs. IGFR	No significant similarity was found	

Kinase domain	Identities	Positives
PDGFR-beta vs. cKit	51%	66%
PDGFR-beta vs. EGFR	40%	62%
PDGFR-beta vs. IGFR	29%	44%
cKit vs. EGFR	38%	59%
cKit vs. IGFR	28%	46%
EGFR vs. IGFR	36%	54%

CONCLUSIONS

Strategies for monitoring endogenous gene expression in a high throughput setting.

Finding small molecules that modulate protein function is of primary importance in drug development and in basic research. In this thesis I describe developing endogenous gene expression-based methods aimed to facilitate the identification of such molecules for cell signaling pathways.

The first approach, described in the first chapter of the thesis, resulted in developing a strategy that reduced array probe redundancy. This allowed us to design and test smaller, but still sufficiently accurate oligonucleotide arrays, which could be used in screening applications. Unfortunately, even in the modified version, the oligonucleotide arrays were still too expensive and laborious for screening compound libraries for modulators of the pathways, as this required thousands of chips. However, the methods and strategies developed to decrease probe redundancy in oligonucleotide arrays can be useful in designing new generations of small, limited-probe, genome-wide arrays for screening applications.

As monitoring a genome-wide gene expression signature appeared impractical for a high-throughput setting, another strategy was used to make expression based high-throughput screening for modulators of signaling pathways possible. A small signature consisting of five genes was developed to monitor activation status of PDGFR/ERK pathway. The

signature was used to screen some 1800 biologically active compounds for inhibitors of the pathway. The screen resulted in identifying novel PDGFR inhibitor, aurintricarboxylic acid (ATA). The mode of action of ATA was investigated to determine that ATA inhibits PDGFR by blocking ligand binding to the receptor.

The success of the proof of concept experiment, which resulted in finding novel inhibitor of a pathway of interest, suggest, that endogenous gene expression-based methods are a viable addition to the toolbox that currently exists to identify modulators of signaling pathways. The flexibility and adaptability of the endogenous expression based methods make them especially attractive to use in finding modulators of novel pathways, pathways that are not well suitable to screening by conventional methods due to absence of reagents such as reporters or overexpressed targets, or novel targets of well-investigated pathways.

Challenges in adaptation of endogenous gene expression signature to a high throughput setting.

While the procedure for defining the signature of the pathway is straightforward and easily adaptable for any pathway that changes cellular expression in response to stimulus, a number of challenges were discovered along the way.

PDGFR/ERK pathway is activated by phosphorylation of proteins in the cascade. If members of the pathway are not constitutively activated, their activation and subsequent

return to inactive state is a relatively fast process, thus making timing of every screening procedure crucial for success of the screen. The time constrains of the biological system should be taken into account while screening libraries with high number of compounds, as the screening should proceed in batches, ensuring that cells in all wells are exposed to compounds, and, more important, to growth factors, for similar amounts of time.

The importance if the timing in endogenous gene expression based screening of a signaling pathway can be illustrated by the following example.

The signature for the screen was defined using SH-SY5Y cells, which were initially selected for the screen. The cell grew well in a high throughput setting, were suitable for existing RNA-isolation protocol, and the PDGFR/ERK pathway was reliably induced in the cells upon introduction of PDGF. To define the signature of activated PDGFR/ERK pathway, growth factor-induced changes in total gene expression in the cells were monitored with oligonucleotide microarrays, and then five genes were selected as a signature suitable for high-throughput screening.

As the project progressed, the protocols and methodology evolved, which resulted in changes in cell-handling procedure. The new protocols became incompatible with the initially selected SH-SY5Y cells, thus dictating a move to a different cell system. TIP5 cells were chosen for the high-throughput screening, being more suitable for the new protocols. The signature, developed for SH-SY5Y cells, was tested for TIP5 cells. The signature appeared to be readily transferable to a different cell type with one exception.

In TIP5 cells one of the three previously synchronized markers had maximum expression at least 20 minutes later than the other two, thus severely limiting optimal detection time. This timing appeared to be unsustainable in a high-throughput setting and for the most part only two out of three markers were effectively detected.

Another unforeseen complication was encountered while defining the signature of the pathway. Total RNA was sampled and gene expression was monitored with oligonucleotide microarrays at 0.5, 1, 2, 3, 4 hours after introduction of PDGF. As subsequent analysis has shown, there was an apparent absence of genes with expression decreasing during the first four hours after activation of the pathway. As a result, all the markers in the signature had increased expression upon activation of the pathway, and introduction of inhibitor of the pathway abolished the expression of the markers.

In this situation I had to rely on stable expression of control genes to weed out false positive wells, such as wells with very low total expression due to lower cell count or toxicity of compounds, and wells with compounds affecting RNA synthesis, such as polymerase inhibitors and DNA/RNA intercalators. Ideally, the signature should contain both genes that increase and decrease their expression upon introduction of modulators of pathway of interest, thus reducing the number of false positives and increasing signal to noise ratio. Still, even with the "one-sided" signature, the screening can be productive, as demonstrated in this study by finding a novel inhibitor of PDGFR/ERK pathway.

Future directions.

While the endogenous gene expression based screening method for cell signaling proved to be robust and practical for a high-throughput setting, in the future the method can benefit from improving the transcript detection system used to monitor expression of the signature gene transcripts. The current method, SBE mass-spectroscopy, due to its limited sensitivity and resolution, currently puts a set of restrictions on the gene signature. To be reliably detected by SBE, the signature has to have a small number of transcripts, thus limiting the signature to only a handful of genes. Moreover, the mRNA has to be amplified by RT-PCR, thus complicating detection by decreasing signal to noise ratio. Developing a more sensitive and reliable detection technique will allow tracing subtle changes in endogenous mRNA expression by utilizing more complex and diverse signatures.

Summary.

To summarize, in this thesis I describe development and subsequent testing of endogenous expression-based methods for high-throughput screening applications. The results of this study suggest that it is possible to find modulators of pathways by inferring activation state of a biological system through monitoring its endogenous mRNA expression signature, and not measuring the activity of the pathway directly. The strategy is validated by a high throughput gene expression-based screening assay that employs a limited endogenous mRNA expression signature to monitor an activation state

of a signal transduction pathway. A novel low-micromolar inhibitor of PDGF receptor is identified in screening a small chemical compound library, demonstrating the utility and applicability of this flexible and reproducible approach for finding modulators of cell-signaling pathways.

Taken together, these results demonstrate the versatility and adaptability of assays based on monitoring endogenous mRNA expression signature. In the future these simple, economical and reliable assays should find widespread use in experiments directed toward identification of novel modulators of signaling pathways and will prove a powerful tool in drug discovery and in basic research.

References

1. von Ahsen O, Bomer U: **High-throughput screening for kinase inhibitors.** *Chembiochem* 2005, **6**(3):481-490.
2. Liu B, Li S, Hu J: **Technological advances in high-throughput screening.** *Am J Pharmacogenomics* 2004, **4**(4):263-276.
3. Sundberg S: **High-throughput and ultra-high-throughput screening: solution- and cell-based approaches.** *Curr Opin Biotechnol* 2000, **11**(1):47-53.
4. Clemons P: **Complex phenotypic assays in high-throughput screening**. *Curr Opin Chem Biol* 2004, **8**(3):334-338.
5. Brown E, Schreiber S: **A signaling pathway to translational control.** *Cell* 1996, **86**(4):517-520.
6. Haggarty S, Mayer T, Miyamoto D, Fathi R, King R, Mitchison T, Schreiber S: **Dissecting cellular processes using small molecules: identification of colchicine-like, taxol-like and other small molecules that perturb mitosis**. *Chem Biol* 2000, **7**(4):275-286.
7. Stockwell BR: **Frontiers in chemical genetics**. *Trends Biotechnol* 2000, **18**(11):449-455.
8. Way JC: **Covalent modification as a strategy to block protein-protein interactions with small-molecule drugs.** *Curr Opin Chem Biol* 2000, **4**(1):40-46.
9. Marton M, DeRisi J, Bennett H, Iyer V, Meyer M, Roberts C, Stoughton R, Burchard J, Slade D, Dai H *et al*: **Drug target validation and identification of**

secondary drug target effects using DNA microarrays. *Nat Med* 1998, **4**(11):1293-1301.

10. Stegmaier K, Ross K, Colavito S, O'Malley S, Stockwell B, Golub T: **Gene expression-based high-throughput screening(GE-HTS) and application to leukemia differentiation.** *Nat Genet* 2004, **36**(3):257-263.

11. Antipova A, Tamayo P, Golub T: **A strategy for oligonucleotide microarray probe reduction.** *Genome Biol* 2002, **3**(12).

12. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray**. *Science* 1995, **270**(5235):467-470.

13. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H *et al*: **Expression monitoring by hybridization to high-density oligonucleotide arrays**. *Nat Biotechnol* 1996, **14**(13):1675-1680.

14. Schadt EE, Li C, Ellis B, Wong WH: **Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data**. *J Cell Biochem Suppl* 2001, **Suppl**(37):120-125.

15. Affymetrix: **Statistical Algorithms Reference Guide [http://www.affymetrix.com/support/technical/technotes/statistical_reference_guide.pdf]**. In. Santa Clara, CA.; 2001.

16. Technology WIMIo: **Center for Genome Research, Cancer Genomics Web Site, Publications [http://www-genome.wi.mit.edu/cgi-bin/cancer/publications/pub_menu.cgi]**. In. Edited by Research CfG.

17. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA *et al*: **Molecular classification of cancer: class discovery and class prediction by gene expression monitoring**. *Science* 1999, **286**(5439):531-537.

18. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, Gaasenbeek M, Angelo M, Reich M, Pinkus GS *et al*: **Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning**. *Nat Med* 2002, **8**(1):68-74.

19. Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME, Kim JY, Goumnerova LC, Black PM, Lau C *et al*: **Prediction of central nervous system embryonal tumour outcome based on gene expression**. *Nature* 2002, **415**(6870):436-442.

20. Dasarathy BV: **Nearest neighbor (NN) norms : nn pattern classification techniques**. Los Alamitos, Calif. Washington: IEEE Computer Society Press; 1991.

21. Technology WIMIo: **Center for Genome Research, Cancer Genomics Web Site, Software [www-genome.wi.mit.edu/cancer/software/software.html]**. In.

22. Johnston P: **Cellular assays in HTS**. *Methods Mol Biol* 2002, **190**:107-116.

23. Pages G, Lenormand P, L'Allemain G, Chambard JC, Meloche S, Pouyssegur J: **Mitogen-activated protein kinases p42mapk and p44mapk are required for fibroblast proliferation.** *Proc Natl Acad Sci U S A* 1993, **90**(18):8319-8323.

24. Cowley S, Paterson H, Kemp P, Marshall C: **Activation of MAP kinase kinase is necessary and sufficient for PC12 differentiation and for transformation of NIH 3T3 cells**. *Cell* 1994, **77**(6):841-852.

25. Le Gall M, Chambard J, Breittmayer J, Grall D, Pouyssegur J, Van Obberghen-Schilling E: **The p42/p44 MAP kinase pathway prevents apoptosis induced by anchorage and serum removal.** *Mol Biol Cell* 2000, **11**(3):1103-1112.

26. Widmann C, Gibson S, Jarpe M, Johnson G: **Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human.** *Physiol Rev* 1999, **79**(1):143-180.

27. Chen R, Sarnecki C, Blenis J: **Nuclear localization and regulation of erk- and rsk-encoded protein kinases.** *Mol Cell Biol* 1992, **12**(3):915-927.

28. Traverse S, Gomez N, Paterson H, Marshall C, Cohen P: **Sustained activation of the mitogen-activated protein (MAP) kinase cascade may be required for differentiation of PC12 cells. Comparison of the effects of nerve growth factor and epidermal growth factor.** *Biochem J* 1992, **288**(2):351-355.

29. Brunet A, Roux D, Lenormand P, Dowd S, S K, J P: **Nuclear translocation of p42/p44 mitogen-activated protein kinase is required for growth factor-induced gene expression and cell cycle entry.** *EMBO J* 1999, **18**(3):664-674.

30. Kim S, Hahn J, Min Y, Yoo N, Ko Y, Lee W: **Constitutive activation of extracellular signal-regulated kinase in human acute leukemias: combined role of activation of MEK, hyperexpression of extracellular signal-regulated kinase, and downregulation of a phosphatase, PAC1.** *Blood ;(11):* 1999, **93**(11):3893-3899.

31. Heldin C, Westermark B: **Mechanism of action and in vivo role of platelet-derived growth factor.** *Physiol Rev :* 1999, **79**(4):1283-1316.

32. Shaw J, Chou I, Anand B: **Rapid phosphorylation of microtubule-associated proteins through distinct mitogenic pathways.** *J Biol Chem* 1988, **263**(3):1459-1466.

33. Kolch W: **Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions**. *Biochem J* 2000, **351**(2):289-305.

34. Pawson T: **Regulation and targets of receptor tyrosine kinases.** *Eur J Cancer* 2002, **38**(Suppl 5):S3-10.

35. Favata M, Horiuchi K, Manos E, Daulerio A, Stradley D, Feeser W, Van Dyk D, Pitts W, Earl R, Hobbs F *et al*: **Identification of a novel inhibitor of mitogen-activated protein kinase kinase.** *J Biol Chem* 1998, **273**(29):18623-18632.

36. Dudley D, Pang L, Decker S, Bridges A, Saltiel A: **A synthetic inhibitor of the mitogen-activated protein kinase cascade.** *Proc Natl Acad Sci U S A* 1995, **92**(17):7686-7689.

37. Westwick J, Cox A, Der C, Cobb M, Hibi M, Karin M, Brenner D: **Oncogenic Ras activates c-Jun via a separate pathway from the activation of extracellular signal-regulated kinases.** *Proc Natl Acad Sci U S A* 1994, **91**(13):6030-6034.

38. Harada T, Morooka T, Ogawa S, Nishida E: **ERK induces p35, a neuron-specific activator of Cdk5, through induction of Egr1.** *Nat Cell Biol* 2001, **3**(5):453-459.

39. Ying S, Futter M, Rosenblum K, Webber M, Hunt S, Bliss T, Bramham C: **Brain-derived neurotrophic factor induces long-term potentiation in intact adult hippocampus: requirement for ERK activation coupled to CREB and upregulation of Arc synthesis.** *J Neurosci* 2002, **22**(5):1532-1540.

40. Lewis TS, Shapiro PS, Ahn NG: **Signal transduction through MAP kinase cascades.** *Adv Cancer Res* 1998, **74**:49-139.

41. Wenzel D, Innis J: **Arrhythmogenic and antiarrhythmic effects of lipolytic factors on cultured heart cells.** *Res Commun Chem Pathol Pharmacol* 1983, **41**(3):383-396.

42. Martin T, Wysolmerski R, Lagunoff D: **Phosphatidylcholine metabolism in endothelial cells: evidence for phospholipase A and a novel Ca2+-independent phospholipase C.** *Biochim Biophys Acta :* 1987, **917**(2):296-307.

43. Croxtall J, Choudhury Q, Flower R: **Glucocorticoids act within minutes to inhibit recruitment of signalling factors to activated EGF receptors through a receptor-dependent, transcription-independent mechanism.** *Br J Pharmacol* 2000, **130**(2):289-298.

44. Lerea L, Carlson N, McNamara J: **N-methyl-D-aspartate receptors activate transcription of c-fos and NGFI-A by distinct phospholipase A2-requiring intracellular signaling pathways.** *Mol Pharmacol* 1995, **47**(6):1119-1125.

45. Kubota Y, Angelotti T, Niederfellner G, Herbst R, Ullrich A: **Activation of phosphatidylinositol 3-kinase is necessary for differentiation of FDC-P1 cells following stimulation of type III receptor tyrosine kinases.** *Cell Growth Differ 1998 Mar;9(3):247-56* 1998, **9**(3):247-256.

46. Carroll M, Tomasson M, Barker G, Golub T, Gilliland D: **The TEL/platelet-derived growth factor beta receptor (PDGF beta R) fusion in chronic myelomonocytic leukemia is a transforming protein that self-associates and activates PDGF beta R kinase-dependent signaling pathways.** *Proc Natl Acad Sci U S A* 1996, **93**(25):14845.

47. Tyson D, Larkin S, Hamai Y, Bradshaw R: **PC12 cell activation by epidermal growth factor receptor: role of autophosphorylation sites.** *Int J Dev Neurosci* 2003, **21**(2):63-74.

48. Blume-Jensen P, Hunter T: **Oncogenic kinase signalling.** *Nature* 2001, **411**(6835):355-365.

49. Kreamer B, Anderson J, Liu D, Sparks M, Richardson A: **Inhibition of cell-free protein synthesis by analogues of aurintricarboxylic acid.** *Can J Biochem* 1978, **56**(12):1162-1167.

50. Igarashi S, Zmean J: **Effect of triphenylmethane derivatives on cell-free macromolecular synthesis. II. mRNA-ribosome binding.** *Can J Biochem* 1975, **53**(2):124-127.

51. Schols D, Baba M, Pauwels R, Desmyter J, De Clercq E: **Specific interaction of aurintricarboxylic acid with the human immunodeficiency virus/CD4 cell receptor.** *Proc Natl Acad Sci U S A* 1989, **86**(9):3322-3326.

52. Gagliardi A, Collins D: **Inhibition of angiogenesis by aurintricarboxylic acid.** *Anticancer Res* 1994, **14**(2A):475-479.

53. Wang P, Kozlowski J, Cushman M: **Isolation and structure elucidation of low molecular weight components of aurintricarboxylic acid (ATA).** *J Org Chem* 1992, **57**:3861-3866.

54. Weinstein M, Vosburgh E, Philips M, Turner N, Chute-Rose L, Moake J: **Isolation from commercialaurintricarboxylic acid of the most effective polymeric inhibitorsof von Willebrand factor interaction with platelet glycoprotein Ib. Comparison with other polyanionic and polyaromatic polymers.** *Blood* 1991, **78**:2291.

55. Bina-Stein M, Tritton TR: **Aurintricarboxylic acid is a nonspecific enzyme inhibitor.** *Mol Pharmacol* 1976, **12**:191.

56. Blumenthal T, Landers T: **The inhibition of nucleic acid-binding proteins by aurintricarboxylic acid.** *Biochem Biophys Res Commun* 1973, **55**(3):680-688.

57. Gonzalez R, Haxo R, Schleich T: **Mechanism of action of polymeric aurintricarboxylic acid, a potent inhibitor of protein--nucleic acid interactions.** *Biochemistry* 1980, **19**(18):4299-4303.

58. Nakane H, Balzarini J, De Clercq E, Ono K: **Differential inhibition of various deoxyribonucleic acid polymerases by Evans blue and aurintricarboxylic acid.** *Eur J Biochem Biophys* 1988, **177**(91).

59. Seki S, Tsutsui K, Oda T: **Effect of aurintricarboxylic acid on in vitro DNA synthetic activity of mouse ascites sarcoma cells.** *Biochem Biophys Res Commun* 1977, **79**:179.

60. Baba M, Schols D, Pauwels R, Balzarini J, De Clercq E: **Fuchsin acid selectively inhibits human immunodeficiency virus (HIV) replication in vitro.** *Biochem Biophys Res Commun* 1988, **155**:1404.

61. Cushman M, Wang P, Chang S, Wild C, De Clercq E, Schols D, Goldman M, Bowen J: **Preparation and anti-HIV activities of aurintricarboxylic acid fractions and analogues: direct correlation of antiviral potency with molecular weight.** *J Med Chem* 1991, **34**(1):329.

62. Hunt D, Wagner R: **Inhibition by aurintricarboxylic acid and polyethylene sulfonate of RNA transcription of vesicular stomatitis virus.** *J Virol* 1975, **16**(5):1146-1153.

63. Hallick RB, Chelm BK, Gray PW, Orozco EM: **Use of aurintricarboxylic acid as an inhibitor of nucleases during nucleic acid isolation.** *Nucleic Acids Res* 1977, **4**:3055.

64. Schwartzman R, Cidlowski J: **Mechanism of tissue-specific induction of internucleosomal deoxyribonucleic acid cleavage activity and apoptosis by glucocorticoids.** *Endocrinology* 1993, **133**(2):591.

65. Schulz-Harder B, Tata J: **Inhibition of nuclear ribonuclease activity during transcription in vitro by aurintricarboxylic acid and vanadyl ribonucleoside complexes.** *Biochem Biophys Res Commun* 1982, **104**(3):903-910.

66. Catchpoole DR, Stewart BW: **Inhibition of topoisomerase II by aurintricarboxylic acid: implications for mechanisms of apoptosis.** *Anticancer Res* 1994, **14**:853.

67. Benchokroun Y, Couprie J, Larsen AK: **Aurintricarboxylic acid, a putative inhibitor of apoptosis, is a potent inhibitor of DNA topoisomerase II in vitro and in Chinese hamster fibrosarcoma cells.** *Biochem Phram* 1995, **49**:305.

68. Baumann H, Hofmann R, Lammers M, Schimpff-Weiland G, Follmann H: **Aurintricarboxylic acid and polynucleotides as novel inhibitors of ribonucleotide reductases.** *Z Naturforsch* 1984, **39**(3-4):276-281.

69. Igarashi S, Zmean J: **Effect of triphenylmethane derivatives on cell-free macromolecular synthesis. I. Aminoacyl-tRNA synthetase.** *Can J Biochem* 1975, **53**(2):120-123.

70. Sharma R, Garg B, Kurosaki H, Goto M, Otsuka M, Yamamoto T, Inoue J: **Aurine tricarboxylic acid, a potent metal-chelating inhibitor of NFkappaB-DNA binding.** *Bioorg Med Chem* 2000, **8**(7):1819-1823.

71. Cushman M, Sherman P: **Inhibition of HIV-1 integration protein by aurintricarboxylic acid monomers, monomer analogs, and polymer fractions.** *Biochem Biophys Res Commun* 1992, **185**(1):85-90.

72. Liang F, Huang Z, Lee S, Liang J: **Aurintricarboxylic acid blocks in vitro and in vivo activity of YopH, an essential virulent factor of Yersinia pestis, the agent of plague**. *J Biol Chem* 2003, **278**(43):41734-41741.

73. Cho H, Lee D, Shrestha S, Shim Y, Kim K, Kim M, Lee K, Won J, Kang J: **Aurintricarboxylic Acid Translocates across the Plasma Membrane, Inhibits Protein Tyrosine Phosphatase and Prevents Apoptosis in PC12 Cells**. *Mol Cells* 2004, **18**(1):46.

74. Thompson D, Reed M: **Inhibition of NAD(H)/NADP(H)-requiring enzymes by aurintricarboxylic acid.** *Toxicol Lett* 1995, **81**:141.

75. Hibasami H, Tanaka M, Nagai J: **Kinetic studies on inhibition of aminopropyltransferases by aurintricarboxylic acid in vitro.** *Chem Biol Interact* 1982, **40**:319.

76. Posner A, Raser K, Hajimohammadreza I, Yuen P, Wang K: **Aurintricarboxylic acid is an inhibitor of mu- and m-calpain.** *Biochem Mol Biol Int* 1995, **36**(2):291-299.

77. Allain P, Leblondel G, Streichenberger G: **Anti-inflammatory effect of aluminum or aurin tricarboxylic acid.** ***C R Seances Soc Biol Fil* 1980, 174**(1):**68-73**.

78. Hsu C, Joshi J: **Aurinetricarboxylic acid: a potent inhibitor of glucose-6-phosphate dehydrogenase.** *Chem Biol Interact* 1977, **17**(1):61-67.

79. Carias JR, Julien R: **Competitive inhibition of Phe: tRNA ligase of sheep embryos by aurintricarboxylic acid.** *FEBS Lett* 1975, **56**:303.

80. McCune SA, Foe LG, Kemp RG, Jurin RR: **Aurintricarboxylic acid is a potent inhibitor of phosphofructokinase.** *Biochem J* 1989, **259**:925.

81. Tsi C, Chao Y, Chen C, Lin W: **Aurintricarboxylic acid protects against cell death caused by lipopolysaccharide in macrophages by decreasing inducible nitric-oxide synthase induction via IkappaB kinase, extracellular signal-regulated kinase, and p38 mitogen-activated protein kinase inhibition.** *Mol Pharmacol* 2002, **62**(1):90-101.

82. Mayr G, Windhorst S, Hillemeier K: **Antiproliferative plant and synthetic polyphenolics are specific inhibitors of vertebrate inositol-1,4,5-trisphosphate 3-kinases and inositol polyphosphate multikinase.** *J Biol Chem* 2005, **280**(14):13229-13240.

83. Woodley C, Roychowdhury M, MacRae T, Olsen K, Wahba A: **Protein synthesis in brine shrimp embryos. Regulation of the formation of the ternary complex (Met-tRNAf X eIF-2 X GTP) by two purified protein factors and phosphorylation of Artemia eIF-2.** *Eur J Biochem* 1981, **117**(3):543-551.

84. Stewart M, Grollman A, Huang M: **Aurintricarboxylic acid: inhibitor of initiation of protein synthesis.** *Proc Natl Acad Sci U S A* 1971, **68**(1):97-101.

85. Leader D: **Aurintricarboxylic acid inhibition of the binding of phenylalanyl-tRNAa to rat liver ribosomal subunits.** *FEBS Lett* 1972, **22**(2):245-248.

86. Gan Y, Weaver J, Pine P, Zoon K, A. A: **Aurin tricarboxylic acid, the anti-AIDS compound, prevents the binding of interferon-alpha to its receptor.** *Biochem Biophys Res Commun* 1990, **172**(3):1298-1303.

87. Phillips MD, Moake JL, Nolasco L, Turner N: **Aurin tricarboxylic acid: a novel inhibitor of the association of von Willebrand factor and platelets.** *Blood* 1988, **72**: 1898.

88. Girma J, Fressinaud E, Christophe O, Rouault C, Obert B, Takahashi Y, Meyer D: **Aurin tricarboxylic acid inhibits platelet adhesion to collagen by binding to the 509-695 disulphide loop of von Willebrand factor and competing with glycoprotein Ib.** *Thromb Haemost* 1992, **68**(6):707-713.

89. Zeevalk G, Schoepp D, Nicklas W: **Aurintricarboxylic acid prevents NMDA-mediated excitotoxicity: evidence for its action as an NMDA receptor antagonist.** *J Neurochem* 1993, **61**(1):386-389.

90. Zeevalk G, Schoepp D, Nicklas W: **Excitotoxicity at both NMDA and non-NMDA glutamate receptors is antagonized by aurintricarboxylic acid: evidence for differing mechanisms of action.** *J Neurochem* 1995, **64**(4):1749-1758.

91. Moudgil V, Eessalu T: **Sensitivity of progesterone receptors to aurintricarboxylic acid. Inhibition of nuclear uptake, ATP and DNA binding.** *Biochim Biophys Acta* 1980, **627**(3):301-312.

92. Moudgil V, Weekes G: **Inhibition of hen oviduct estradiol receptor by aurintricarboxylic acid.** *FEBS Lett* 1978, **94**(2):324-326.

93. Moudgil V, Caradonna V: **Inhibition of nuclear uptake of glucocorticoid-receptor complex by aurintricarboxylic acid.** *J Steroid Biochem* 1982, **17**(6):585-589.

94. Beery R, Haimsohn M, Wertheim N, Hemi R, Nir U, Karasik A, Kanety H, Geiger A: **Activation of the insulin-like growth factor 1 signaling pathway by the antiapoptotic agents aurintricarboxylic acid and Evans Blue.** *Endocrinology* 2001, **142**:3098.

95. Haimsohn M, Beery R, Karasik A, Kanety H, Geiger A: **Aurintricarboxylic acid induces a distinct activation of the IGF-I receptor signaling within MDA-231 cells.** *Endocrinology* 2002, **143**:837.

96. Okada N, Koizumi S: **Tyrosine phosphorylation of ErbB4 is stimulated by aurintricarboxylic acid in human neuroblastoma SH-SY5Y cells.** *Biochem Biophys Res Commun* 1997, **230**(2):266-269.

97. Benezra M, Vlodavsky I, Yayon A, Bar-Shavit R, Regan J, Chang M, Ben-Sasson S: **Reversal of basic fibroblast growth factor-mediated autocrine cell transformation by aromatic anionic compounds.** *Cancer Res* 1992, **52**(20):5656-5662.

98. Lozano R, Rivas G, Gimenez-Gallego G: **Destabilization, oligomerization and inhibition of the mitogenic activity of acidic fibroblast-growth factor by aurintricarboxylic acid.** *Eur J Biochem* 1997, **248**(1):30-36.

99. Grollman AP: **Inhibition of messenger ribonucleic acid attachment to ribosomes. II. Proposed mechanism for the design of novel antiviral agents**. In: *Antimicrobial Agents and Chemotherapy.* Edited by Hobby GL. New York.: American Society of Microbiology; 1968: 36-40.

100. Chen C, Chao Y, Chang Y, Hsu M, Lin W: **Inhibition of cytokine-induced JAK-STAT signalling pathways by an endonuclease inhibitor aurintricarboxylic acid.** *Br J Pharmacol* 2002, **137**(7):1011-1020.

101. Novak U, Harpur A, Paradiso L, Kanagasundaram V, Jaworowski A, Wilks A, Hamilton J: **Colony-stimulating factor 1-induced STAT1 and STAT3 activation is accompanied by phosphorylation of Tyk2 in macrophages and Tyk2 and JAK1 in fibroblasts.[Abstract/Free Full Text]**. *Blood* 1995, **86**:2948.

102. Mui A, Wakao H, O'Farrell A, Harada N, Miyajima A: **Interleukin-3, granulocyte-macrophage colony stimulating factor and interleukin-5 transduce signals through two STAT5 homologs.** *EMBO J* 1995, **14**:1166.

103. Stark G, Kerr I, Williams B, Silverman R, Schreiber R: **How cells respond to interferons.** *Annu Rev Biochem* 1998, **67**:227-264.

104. Schindler C, Darnell JE: **Transcriptional responses to polypeptide ligands: the Jak-Stat pathway.** *Annu Rev Biochem* 1995, **64**:621.

105. McCubrey J, May W, Duronio V, Mufson A: **Serine/threonine phosphorylation in cytokine signal transduction.** *Leukemia* 2000, **14**(1):9-21.

106. Bridges A: **Chemical inhibitors of protein kinases.** *Chem Rev* 2001, **101**(8):2541-2572.

107. Hemmila I, Hurskainen P: **Novel detection strategies for drug discovery.** *Drug Discov Today* 2002, **7**(18 Suppl):S150-156.

108. Lessnick S, Dacwag C, Golub T: **The Ewing's sarcoma oncoprotein EWS/FLI induces a p53-dependent growth arrest in primary human fibroblasts.** *Cancer Cell* 2002, **1**(4):393-401.

109. Tatusova TA, Madden TL: **BLAST 2 SEQUENCES**. In.; 1999.

110. Tatusova TA, Madden TL: **Blast 2 sequences - a new tool for comparing protein and nucleotide sequences**. *FEMS Microbiol Lett* 1999, **174**:247-250.

CURRICULUM VITAE

Alena A. Antipova

ADDRESS

24 Center St., Woburn, MA, 01801
alena@mit.edu

EDUCATION

Massachusetts Institute of Technology, Ph.D. in Chemistry, expected September 2005

Moscow Institute of Physics and Technology, M.S. in Molecular Biology, 1996

Moscow Institute of Physics and Technology, B.S. in Mathematics and Physics, 1994

PROFESSIONAL EXPERIENCE

Broad Institute, Massachusetts Institute of Technology, Research Assistant, 2000 to present

Harvard Medical School, Research Assistant, 1997-1999

USC School of Medicine, Research Assistant, 1996-1997

Institute of Gene Biology, Research Assistant, 1994-1996

TEACHING EXPERIENCE

Teaching Assistant, Massachusetts Institute of Technology, 1999-2000

AWARDS

Advanced Fellowship, Moscow Institute of Physics and Technology, 1991-1996
FEBS Youth Grant, 1995

PUBLICATIONS

Antipova A.A., Stockwell B.R., Golub T.R. GE-HTS for Signal Transduction, EMBO/EMBL Symposium on "Functional Genomics; Exploring the Edges of Omics", 2004. (Talk)

Antipova A.A., Tamayo P., Golub T.R. A strategy for oligonucleotide microarray probe reduction. Genome Biol. 2002;3(12)

Itoh K., Antipova A., Ratcliffe M.J., Sokol S. Interaction of dishevelled and Xenopus axin-related protein is required for Wnt signal transduction. Mol Cell Biol. 2000 Mar;20(6):2228-38