Eight-channel iTRAQ enables comparison of the activity of six leukemogenic tyrosine kinases

Pierce, A; Unwin, R D; Evans, C A; Griffiths, S; Carney, L; Zhang, L; Jaworska, E; Lee, C F; Blinco, D; Okoniewski, M J; Miller, C J; Bitton, D A; Spooncer, E; Whetton, A D

Pierce, A; Unwin, R D; Evans, C A; Griffiths, S; Carney, L; Zhang, L; Jaworska, E; Lee, C F; Blinco, D; Okoniewski, M J; Miller, C J; Bitton, D A; Spooncer, E; Whetton, A D (2008). Eight-channel iTRAQ enables comparison of the activity of six leukemogenic tyrosine kinases. Molecular & Cellular Proteomics, 7(5):853-63.

Postprint available at:
http://www.zora.uzh.ch

Posted at the Zurich Open Repository and Archive, University of Zurich.
http://www.zora.uzh.ch

Originally published at:
Molecular & Cellular Proteomics 2008, 7(5):853-63.
Eight-channel iTRAQ enables comparison of the activity of six leukemogenic tyrosine kinases

Abstract

There are a number of leukemogenic protein-tyrosine kinases (PTKs) associated with leukemic transformation. Although each is linked with a specific disease their functional activity poses the question whether they have a degree of commonality in their effects upon target cells. Exon array analysis of the effects of six leukemogenic PTKs (BCR/ABL, TEL/PDGFRbeta, FIP1/PDGFRalpha, D816V KIT, NPM/ALK, and FLT3ITD) revealed few common effects on the transcriptome. It is apparent, however, that proteome changes are not directly governed by transcriptome changes. Therefore, we assessed and used a new generation of iTRAQ tagging, enabling eight-channel relative quantification discovery proteomics, to analyze the effects of these six leukemogenic PTKs. Again these were found to have disparate effects on the proteome with few common targets. BCR/ABL had the greatest effect on the proteome and had more effects in common with FIP1/PDGFRalpha. The proteomic effects of the four type III receptor kinases were relatively remotely related. The only protein commonly affected was eosinophil-associated ribonuclease 7. Five of six PTKs affected the motility-related proteins CAPG and vimentin, although this did not correspond to changes in motility. However, correlation of the proteomics data with that from the exon microarray not only showed poor levels of correlation between transcript and protein levels but also revealed alternative patterns of regulation of the CAPG protein by different oncogenes, illustrating the utility of such a combined approach.
Eight-channel iTRAQ enables comparison of the activity of 6 leukaemogenic tyrosine kinases.

*Andrew Pierce¹, *Richard D. Unwin¹, Caroline A. Evans¹, Stephen Griffiths¹, Louise Carney¹, Liqun Zhang¹, Ewa Jaworska¹, Chia-Fang Lee¹, David Blinco¹, Michal J. Okoniewski², Crispin J. Miller², Danny A. Bitton², Elaine Spooncer¹, Anthony D. Whetton¹

¹Stem Cell and Leukaemia Proteomics Laboratory, University of Manchester, Christie Hospital, Kinnaird House, Kinnaird Road, Manchester, M20 4QL, UK.

²Bioinformatics Group, Paterson Institute for Cancer Research, University of Manchester, Wilmslow Road, Manchester, M20 4BX, UK.

*These authors contributed equally to this work

Running Title: Proteomic analysis of leukaemic oncogenes by 8-plex iTRAQ

Corresponding author: Anthony D. Whetton,
Email anthony.whetton@manchester.ac.uk
Tel: +44 (0)161 446 8247
Fax: +44 (0)161 446 8203
Summary

There are a number of leukemogenic protein tyrosine kinases (PTKs) associated with leukaemic transformation. Whilst each is linked with a specific disease their functional activity poses the question whether they have a degree of commonality in their effects upon target cells. Exon array analysis of the effects of 6 leukemogenic PTKs (BCR/ABL, TEL/PDGFRβ, FIP1/PDGFRα, D816V KIT, NPM/ALK and FLT3ITD) revealed few common effects on the transcriptome. It is apparent, however, that proteome changes are not directly governed by transcriptome changes. Therefore, we have assessed and used a new generation of iTRAQ tagging, enabling eight-channel relative quantification discovery proteomics, to analyse the effects of these 6 leukemogenic PTKs. Again these were found to have disparate effects on the proteome with few common targets. BCR/ABL had the greatest effect on the proteome and had more effects in common with FIP1/PDGFRα. The proteomic effects of the 4 type III receptor kinases were relatively remotely related. The only protein commonly affected was eosinophil-associated ribonuclease 7. Five out of 6 PTKs affected the motility-related proteins CapG and vimentin, though this did not correspond to changes in motility. However, correlation of the proteomics data with that from the exon microarray not only showed poor levels of correlation between transcript and protein levels, but also revealed alternative patterns of regulation of the CapG protein by different oncogenes, illustrating the utility of such a combined approach.
Introduction

Human leukaemias are often associated with chromosome translocations which generate a fusion transcript that can be implicated in disease pathogenesis (for a review see (1)). Many of these translocation gene products encode for constitutively activated protein tyrosine kinases (PTKs). These include BCR/ABL (2), a PTK associated with Chronic Myeloid Leukaemia and TEL/PDGFRβ (3), observed in some cases of Chronic Myelomonocytic Leukaemia. FIP1L1-PDGFRα fusion gene is associated with the pathogenesis of about 50% of patients with the hypereosinophilic syndrome (4). A t(2;5) translocation results in the fusion kinase NPM/ALK, found in cases of anaplastic large cell lymphoma (5). Recently other forms of PTK dysregulation have been observed in myeloproliferative disorders and the leukaemias: The Kit receptor PTK undergoes activating mutations (D816V) in acute myeloid leukaemia (AML) (6) and in solid tumors; internal tandem duplication at the juxtamembrane region of the flt3 gene is a somatic change detected in 20% of cases of AML that leads to enhanced ligand independent PTK activity in the FLT3 Internal Tandem Duplication (ITD) form of the protein (7).

Targeted inhibition of the PTKs (such as KIT, BCR/ABL, TEL/PDGFRβ) via inhibitors that bind the ATP pocket of the kinase domain has been a successful strategy to achieve haematologic or molecular remission (8). In CML the use of Imatinib, a BCR/ABL kinase inhibitor, has profoundly altered approaches to treatment (9). However, after 5 years a significant proportion of patients develop
a resistance to the drug, although other molecular therapies remain effective (10). The requirement for further developments in treatment of this and other leukaemias is apparent and further detail of mechanisms for oncogenic tyrosine kinase-mediated leukaemic transformation are required.

The development of systems biology offers new opportunities for the study of oncogenic effects of genes deregulated in primary cells from patients with leukaemia. Investigation of protein levels in such cells is important as, in hematopoietic cells, changes in the levels of mRNA do not necessarily act as a predictor for changes in the proteome (11). The effects of the TEL/PDGFRβ oncogene on a haematopoietic cell line proteome have been shown to be driven by altered transcription plus post-translational regulation of protein levels (12). However, sensitive protein expression analysis has been problematic as amplification of protein, unlike DNA or RNA, cannot be achieved. Relative quantification proteomics using higher sensitivity mass spectrometry techniques now offers the potential to compare and contrast the effects of leukemogenic PTKs. Currently an isobaric tag-based methodology for peptide relative quantification (iTRAQ) coupled to multidimensional liquid chromatography and tandem mass spectrometry enables the assessment of protein levels where 4 samples can be compared for their common effects (13). Here we report the use of a novel 8 channel relative quantification mass spectrometry technique to compare the relative effects of 6 oncogenic tyrosine kinases, BCR/ABL, TEL/PDGFRβ, FIP1/PDGFRα, D816V KIT, NPM/ALK and FLT3ITD. The effects
of the oncogenes on the proteome are distinctive and do not follow normal oncogene classification patterns.

Experimental Procedures

Cell lines

Ba/F3 cells were transfected with either an empty MSCV retroviral vector, or MSCV containing BCR/ABL, TEL/PDGFRβ, FIP1/PDGFRα, D816V KIT, NPM/ALK or FLT3ITD. The resultant cell lines were maintained in culture in Fischers (Invitrogen) with 10% (v/v) horse serum (Biowest, Nuaillé, France). Ba/F3-MSCV cells were grown in Fischers with 10% (v/v) horse serum supplemented with 5% mIL-3 (conditioned media from X63-Ag-653 cells).

Microarray analysis

RNA was prepared using TRIzol (Invitrogen), DNase treated then cleaned using MinElute RNeasy Clean up Kit (Qiagen) per manufacturer's instructions. Transcriptome analysis was undertaken using murine GeneChip® Mouse Exon 1.0 ST Arrays by the CR-UK Affymetrix microarray facility (Paterson Institute, Manchester, UK; http://bioinformatics.picr.man.ac.uk/vice/Protocols.vice). All data were analysed in R using Bioconductor (14). Expression summaries were generated using RMA (15) with a custom Chip Definition File (CDF) file produced using the makecdfenvs library (Irizarry et al, unpublished).
grained gene level annotation was provided using the X:MAP database (http://xmap.picr.man.ac.uk) and the exonmap BioConductor package (16). It is unreasonable to assume that a given gene should have the same baseline level of expression in each of the different cell lines or that the same level of response should be observed across all cell lines. Consequently, it is unreasonable to expect either the expression levels or fold changes to come from a single distribution. Thus each cell line was analysed separately, and the set of probesets exceeding a 2 fold change selected. Up- and down-regulated probesets were treated separately. For each probeset, the number of cell lines in which it was found to be up or down regulated was determined, and used to provide a score. Data were permuted in order to create a null distribution of these scores, from which the False Discovery Rate (FDR) (17) was estimated. The set of probesets found to be consistently differentially expressed in 6, 5 or 4 oncogene-expressing cell lines were identified and then filtered using exonmap to exclude those that did not hit within an Ensembl (18) annotated gene, or that hit the genome at more than one location. Remaining probesets were then mapped via exonmap to their target genes.

Liquid chromatography and mass spectrometry

Cell pellets were produced and processed as previously described (12) with minor modifications. In brief, cells were lysed in 1M triethylammonium bicarbonate (TEAB) with 0.1% (w/v) SDS. 50 µg of protein from each cell line
was reduced, alkylated, and subjected to tryptic hydrolysis prior to labelling with 8 channel iTRAQ reagent (Applied Biosystems, Framingham, MA) in 1M TEAB, according to the manufacturer’s instructions (in all experiments labelling exceeded 98% of total identified peptides). Two biological replicates of MSCV Ba/F3 cells were labelled to provide an internal control for the experiment and assess biological variation. The entire experiment (including generation of cell pellets) was performed twice. Peptides were fractionated off-line using an SCX cation exchange column (40 fractions) prior to RP-LC-MS/MS. Data were acquired using an independent data acquisition protocol where an MS scan was taken then the two highest abundance (ion current) ions were selected for fragmentation, followed by dynamic exclusion for 1 minute.

**Data Analysis**

Data was processed by a ‘Thorough’ search against a mouse CDS database (mouse_KBMS5_0_20050302, 115,660 entries) using the Paragon algorithm (19) within ProteinPilot v2.0 software with trypsin as the digest agent and default settings (Applied Biosystems, Warrington, UK). This software calculates a %age Confidence which reflects the probability that the hit is a false positive, so that at the 95% confidence level, there is a false positive identification rate of around 5% (19). While this software automatically accepts all peptides with a confidence of identification >1%, only proteins which had at least one peptide with >95% were initially recorded. These low confidence peptides therefore do not identify a protein by themselves, but may support the presence of a protein identified using
other peptides. Performing the search against a concatenated database containing both forward and reversed sequences (therefore 231,320 entries) allowed estimation of the false discovery level.

For protein relative quantification, only MSMS spectra which were unique to a particular protein, and where the sum of the signal-to-noise ratio for all of the peak pairs is >9 were used for quantification (default software settings). The accuracy of each protein ratio is given by a calculated ‘error factor’ in the software, and a p-value to assess whether the protein is significantly differentially expressed. Error factor is calculated as $10^{(95\% \text{ confidence error})}$, where this 95% confidence error is the weighted standard deviation of the weighted average of log ratios multiplied by the Student t-factor for n-1 degrees of freedom, where n is the number of peptides contributing to protein relative quantification. P-value is determined by calculating Student t-factor by dividing (Weighted Average of Log Ratios – log Bias) by the Weighted Standard Deviation, allowing determination of the p-value, with n-1 degrees of freedom, again where n is the number of peptides contributing to protein relative quantification. For the identification of expression differences, each experimental run was initially considered separately. To be identified as being significantly differentially expressed, a protein had to be quantified with at least three spectra (allowing generation of a p-value), have a p-value < 0.05, and have a ratio fold change >1.2 or <0.8 in both experimental replicates. These fold change limits were selected on the basis of our previous work with the four-plex iTRAQ reagent.
(12), and were shown here to be equally applicable to the eight-plex iTRAQ relative quantification by the inclusion of internal controls (Figure 2D).

Data was further analysed using Genespring GX microarray analysis package (Agilent Technologies).

**Western blotting and activity assays**

Western blotting was performed using standard protocols. Antibodies used were actin (1:500; Sigma, Poole, UK), CapG, (0.25µg/mL; Abcam, Cambridge UK), Abl, Flt3, Kit, PDGFRα (1µg/mL, 0.1µg/mL, 1µg/mL, 1µg/mL; Santa Cruz, Calne, UK), phosphotyrosine 4G10 (50ng/mL; Upstate, Milton Keynes, UK), PDGFRβ (1µg/mL; BD Pharmingen, Oxford, UK), NPM (neat hybridoma supernatent, LRF Diagnostics Unit, Oxford, UK).

**Quantitative real-time PCR analysis**

Primers and probes were designed using the Universal Probe Library Assay Design Centre from Roche Applied Sciences (https://www.roche-applied-science.com/sis/rtpcr/upl/adc.jsp). Q-RT-PCR assays were performed in triplicate on a 7900HT Fast Real-Time PCR System. ∆∆Ct values were calculated for each sample against the average of the two housekeeping genes which were employed to calculate fold change using the $2^{-\Delta\Delta C_t}$ method (20).

**Chemotaxis assays**
Chemotaxis assays were performed as previously described (21) except that serum free StemSpan medium (Stem Cell Technologies, Vancouver, BC) was employed to preclude oncogene-mediated responses to serum associated agonists, and assays were carried out in 24 transwell plates with 5µm inserts (Sigma).

**Results**

**A common cell population for relative analysis of PTK effects.**

The role of tyrosine kinases in leukaemogenesis is undisputed. The panoply of kinases involved suggests there could be common downstream targets and our aim was to systematically analyse several leukaemogenic protein tyrosine kinases. Six different leukaemogenic PTKs were expressed in a common Ba/F3 background for this purpose (Figure 1A). Ba/F3 cells were chosen because of their widespread usage in studies on leukemogenesis. Each transfected cell line was shown to be independent of IL-3 for growth. Furthermore clones of each cell line that had the same doubling time as empty vector-transfected cells (MSCV Ba/F3) cultured in IL-3 were chosen (data not shown). Expression of the PTKs led to differential effects on protein tyrosine phosphorylation (Figure 1B).

**Comparative analysis of the effect of expression of leukaemogenic tyrosine kinases on the transcriptome**
Genechip microarrays have been used extensively in the study of oncogene effects on the transcriptome. Exon arrays now allow a genome-wide, exon-level analysis on a single array to investigate alternative splicing and gene expression.

RNA was produced from each cell line and the effects of the oncogene on the transcriptome assessed. The exon microarray data indicated that there were relatively few common changes (data can be found at http://bioinformatics.picr.man.ac.uk/vice/) at the gene level imposed on the Ba/F3 cells by the 6 PTKs. Where up-regulation has occurred, 86 probesets (matching 44 genes) showed changed levels (>2fold) in 6 out of 6 (6/6) oncogenes, with a further 311 probesets (194 genes) in 5/6 and 949 probesets (598 genes) in 4/6. Given the extensive nature of these arrays (i.e. the large number of probesets they contain), calculated false discovery rates (the proportion of probesets which would appear in each category by chance) are 75.4%, 55.5% and 48.1% respectively. Clustering exons to genes enables higher confidence of an effect. However, only 5 genes (comprising 33 probesets) show up-regulation of at least half of their exons in 6/6 oncogenes (Table 1).

Analysis of the down-regulated genes shows a similar pattern. While down-regulation appears to be more prevalent than up-regulation, still only 396 probesets (from 117 genes) were significantly altered by 6/6 oncogenes, with 1232 (391 genes) and 2742 (1110 genes) being down-regulated by 5/6 and 4/6 oncogenes respectively. Again the nature of these experiments means a
calculated false discovery rates for these probesets of 17.5%, 22.2% and 22.7% respectively. Further analysis at gene level showed that only 12 genes (comprising 122 probesets) have two-fold (or greater) down-regulation of at least half of their exons in 6/6 oncogenes (Table 1).

While a more in-depth analysis is in progress to detect variation in gene splicing, it remains clear from these data that the mode(s) of action of the six oncogenes being studied have very few common features at the transcriptome level. The lack of commonality at the transcriptome level does not exclude the possibility that these oncogenes do in fact share a common pathway for transformation, potentially via splicing. Further analysis of this dataset is underway to identify splicing effects of PTKs.

We then considered other means for PTKs to elicit common effects. It is now well established that oncogenic tyrosine kinases, not only post-translationally modify proteins but also initiate post-translational regulation of the proteome via processes such as altered rates of translation, or stability via post-translational modification e.g. ubiquitination (22). Only analysis of proteomic changes can reveal these targets as mRNA levels will remain unperturbed. We have previously shown that oncogene-induced changes in the proteome do not always correlate to changes in the transcriptome (12). In addition changes seen in protein levels in primitive hematopoietic cells undergoing development do not
correlate precisely to changes in the transcriptome (11). We therefore investigated the effects of these oncogenes on the proteome.

**Isobaric tagging for relative quantification from 8 samples in a single mass spectrometry experiment.**

We have previously shown that the iTRAQ isobaric tagging approach gives highly reproducible data in the sense that peptides from the same protein show similar relative quantification patterns and also biological replicates show markedly similar trends (11, 12). A new generation of iTRAQ isobaric tagging reagent was employed for 8 channel peptide relative quantification proteomics. This employs the same chemistry to label peptides via free amine groups as previously described (13) but are constructed to enable 8 different samples to be relatively quantified in a single tandem mass spectrometry experiment (Figures 2A and 2B). In the 8-plex reagent, reporter ion masses are now 113.1-119.1, and 121.1. The mass at 120.1 is omitted, to avoid contamination from phenylalanine immonium ion (m/z 120.08). The extra channels are added by increasing the overall mass of the balance group. In addition, all mass differences are now encoded using $^{13}$C and $^{15}$N around a ring structure in the reporter group (Figure 2A). A representative MS/MS spectrum showing the new reporter ion region is shown in Figure 2B.

To ensure appropriate relative quantification using the new reagent, a defined protein mix was digested with trypsin and labelled with the 8 channel iTRAQ
reagent (Figure 2C) as previously described with the 4 channel iTRAQ reagent (12). The appropriate protein relative quantification levels were found to be achieved with the approach employed with the mixture of proteins (see legend to Figure 2). Analysis of a standard 8-protein mix shows that in all cases this method is capable of protein relative quantification (all proteins except ovotransferrin and carbonic anhydrase were present in a 1:1 ratio in the isobarically tagged tryptic digest samples, see figure legend for detail). It should be noted that, in our hands, the approach slightly underestimated the degree of fold change anticipated, shown by the diamond symbols (Figure 2C), a feature we have previously observed with the 4-plex reagent. These data show that the 8-plex iTRAQ can be employed for assessment of relative PTK effects on the proteome.

**Comparative analysis of the effect of expression of leukaemogenic tyrosine kinases on the proteome**

To analyse potential common proteome differences elicited by the 6 PTKs, eight channel isobaric tagging with tandem mass spectrometry was employed. Proteomics data was obtained from two experiments where 1380 protein and 1614 protein identifications with a confidence score greater than 95% were made respectively. Searching against a reversed database allowed calculation of false discovery rates for these two experiments as 0.146% and 0.806% at the protein level respectively. Relative quantitation analysis was performed on these
datasets separately and then averaged together to give a list of 1886 proteins (See supplementary Table 1).

In each case, labelling efficiency was estimated to be >98%, by comparing the total number of identified potential reactive sites (i.e. N-termini and lysine side chains) with the number of iTRAQ modifications at these sites. In addition, there was some concern that, although the total amount of protein analysed in the 8-plex experiment was the same as in a standard 4-plex (i.e. 400µg), there is now only half of the material in each channel (50µg per 8-plex, compared to 100µg per 4-plex). However, we compared the number of peptides identified with confidence >90% in this dataset and a 4-plex dataset from similar samples (i.e. whole cell lysates), and showed that in each case around 8% of these peptides were not included in the final dataset due to low iTRAQ reporter ion intensity. Therefore there appears to be no deleterious effects of halving the amount of sample in each channel to allow expansion from a 4-plex to an 8-plex relative quantification experiment.

Further analysis was performed to select proteins where there was a high confidence of protein identification and relative quantification using criteria defined in the analysis software (see Methods and legend to Supplementary Data Table 2). Relative quantification of the effects of 6 PTK oncogenes on 829 proteins is therefore shown where two channels were biological replicates of mock-transfected Ba/F3 cells, included to confirm reproducibility of the approach
as was seen in our previous work (11, 12). Figure 2D shows the ratio for the majority of proteins lies between 0.8 and 1.2, the window previously described (12) beyond which a protein can be considered to be potentially changed in expression level. For comparison the relative protein expression levels of oncogene-transfected Ba/F3 cells compared to MSCV Ba/F3 are shown (Figure 2E-2J). These demonstrate the quantitative variability in the effect of each oncogene on the proteome as a whole, BCR/ABL having the largest gross effect.

The changes noted for the oncogenes in both experimental replicates were: BCR/ABL (168 proteins changed, 77 down, 91 up), TEL/PDGFRβ (67 proteins changed, 22 down and 45 up), NPM/ALK (87 proteins changes, 59 down, 28 up), D816V KIT (106 proteins changes, 54 down, 52 up), FLT3ITD (73 proteins changes, 38 down, 35 up), and FIP1/PDGFRα (124 proteins changes, 43 down, 81 up). The identities of these proteins are shown in Supplementary Data Tables 3-8.

Protein changes were examined further using a dendrogram approach to identify any clustering of oncogene effects. The analysis showed neither clustering of type 3 leukaemogenic receptor tyrosine kinases (FLT3ITD, D816V KIT, FIP1/PDGFRα, TEL/PDGFRβ) nor oncogenes associated with myeloproliferative disorders (Figure 3A). D816V KIT did cluster with FLT3-ITD but this was in a group of 3 with NPM/ALK, the only PTK in the set associated with lymphoma. Construction of an Edwards-Venn diagram containing proteins with a change in
expression between control cells and each oncogene-transfected cell population also revealed the lack of major overlap in effect between any of the oncogenes (Figure 3B). Surprisingly, FIP1/PDGFRα and TEL/PDGFRβ oncogenes showed very little similarity. We therefore conclude that leukaemogenic tyrosine kinases have pleiotropic effects on global protein expression. This is underscored by analysis of the gene ontogeny for proteins affected (in expression level) by the oncogenic PTKs in that effects on many protein groups were recorded and the effects of each oncogene were broadly similar (data not shown).

The number of proteins within the dataset affected by >4 of the oncogenic PTKs was small (Table 2) and covers a wide variety of protein functions. Eosinophil associated ribonuclease 7 is a member of the RNase A superfamily involved in inflammatory processes mediated by eosinophils. CapG binds the end of actin filaments and regulates cellular structure (23). Scinderin (adseverin) cleaves actin filaments and has previously been shown to be changed in expression in the hematopoietic stem cell to progenitor cell transition (24). Vimentin is a member of the intermediate filament family of proteins that is increased in expression with all PTKs except FIP1/PDGFRα. Two related protein phosphatase inhibitors (SET proteins) are decreased in 4/6 and 5/6 PTK transfected cell lines respectively, D816V KIT cells being the only cells to show no change in either. These proteins have been implicated in BCR/ABL mediated transformation (25, 26). The two 40S ribosomal proteins S2 and S15 both display an increase except with the TEL/PDGFRβ and NPM/ALK transfected Ba/F3 cells.
This relative lack of common effects demonstrates diversity in mechanism of action for leukaemogenic PTKs.

**Correlation of peptide expression data to exon expression data**

In the past, comparison of protein expression data to transcript expression has been achieved, but is always undermined by the inability of both standard expression array platforms and bottom-up proteomics analyses to accurately resolve different isoforms. We have established novel bioinformatic approaches to link this proteomic data to the exon array data to check the correlation between protein/peptide abundance and mRNA expression. This enables analysis of oncogenic PTK effects on splice isoform generation. We can now assess the correlation of proteome to transcriptome at three levels: peptide to exon probesets; >1 peptide to exon data; protein expression (summated peptide data) versus gene specific expression (summed exon data). Pearson correlation between exon arrays and peptide data, at the exon vs. peptide level, for MSCV versus oncogene was: 0.47 for BCR/ABL; 0.22 for NPM/ALK; 0.28 for TEL/PDGFRβ; 0.30 for FLT3ITD; 0.25 for D816V KIT and 0.40 for FIP1/PDGFRα. The data set will be the subject of a much more in depth analysis (manuscript in preparation) but the main conclusion relevant here is that there is relatively poor correlation between changes at the exon and peptide levels. Thus a majority of protein changes cannot be predicted by mRNA analysis. Selected examples of this correlation and its anomalies are illustrated in Figure 4 where protein expression data for proteins shown in Table 2 has been compared to
exon array data. The transcriptome and proteome expression levels for Ribose-5-phosphate isomerase and eosinophil major basic protein show a high degree of correlation. The data for CapG however clearly demonstrates that the data obtained from RNA expression levels does not always correlate to protein expression levels. The fidelity of the exon array and iTRAQ data was confirmed by use of QRT PCR and western blot analysis. The data for CapG also illustrates that different oncogenes can affect a protein/pathway by different mechanisms. Whilst CapG is down-regulated in both NPM/ALK and BCR/ABL transfected cells this occurred via different mechanisms: NPM/ALK cells showed a decrease in transcript level resulting in a similar decrease in protein level; BCR/ABL cells display no decrease in transcript yet protein levels are decreased.

**Correlation of proteomic changes to motility effects.**

Systematic proteomic analysis enables oncogenic effects on specific functional pathways to be assessed and further analysis performed. The oncogene-transfected cell lines were shown to display differential cell motility in serum free conditions (Figure 5). In cells expressing BCR/ABL motility is decreased, as observed in previous studies (27, 28). However, expression of TEL/PDGRFβ in these conditions enhanced motile response as did D816VKIT. Vimentin, scinderin and CapG are proteins associated with the cytoskeleton and motility that are affected by almost all of the 6 PTKs in this analysis. Thus their
modulation by oncogenic PTKs can be seen to be inconsistent with the differential effects observed on motile response.

We therefore further analysed the data for all proteins affected by each PTK whose gene ontology suggests association with motility. Within this set were proteins that were perturbed (increased or decreased) in expression levels by BCR/ABL and perturbed in the opposite direction by TEL/PDGFRβ and D816VKIT. Two candidates were found that could affect motile responses. These were High mobility Group protein B2 (HMGB2) and Septin 1. HMGB2 has previously been associated with altered motility and is therefore a candidate target protein for altered motile responses instigated by these oncogenes.

Discussion

The discovery of PTKs that are associated with leukaemogenic transformation have led to investigation of the specific roles for these kinases in transformation (29-31). Often a singular oncogenic PTK is the subject of a focussed study seeking to identify downstream effectors modulated by that specific kinase, such as motility effects or genomic stability mediated by BCR/ABL (32, 33). The molecular components of any oncogene-mediated change are then targets for analysis in respect of the effects of other PTKs: much information exists describing the effects of BCR/ABL on a variety of pathways and processes (34, 35) but a much less substantial scientific literature has developed on the effects
of other oncogenic kinases on these pathways. In this fashion the determination of common and uncommon events promoted by specific kinases will progress at a relatively slow rate. Given the growing number of kinases associated with the leukaemias we have sought to establish whether novel systems biology approaches can determine the effects of PTKs on the proteome and transcriptome.

To achieve a meaningful foundation dataset the requirement is for a common cellular background on which to examine the effects of PTKs. Ba/F3 cells are commonly employed in studies on leukaemogenesis and are therefore suitable for initial analyses (36). Whilst this approach demands follow up studies on other cell lines and primary material, a successful primary screening strategy on a common background will inevitably save time in understanding common events seen in transformation.

The analysis of the dataset we have constructed has shown very few common leukaemogenic PTK mediated events. At the exon/transcript level, these include increases in hepatocellular carcinoma associated antigen, LR8 protein and CD55 antigen. CD55, an inhibitor of the complement system, is widely expressed in malignant tumours and reported to have a role in decreasing complement mediated tumour cell lysis, cell motility and metastasis (for a review see (37)). Other probesets also showed changes in 6/6 oncogenes, but these were accompanied by no changes in other exons from that gene. This may represent
alternative splicing events. However, given the large number of probesets on each array (over 1.4 million probesets) it is almost inevitable that many false positives will be called ‘changed’ and that a subset of these will be called in all oncogenes by random chance. The false discovery rates for identification of an up-regulation in 6/6 oncogenes suggests that 75.4% of those identified could be expected by chance. We reasoned that, since this analysis is done at the probeset level, those probesets which are falsely called ‘changed’ in 6/6 oncogenes will be unlikely to map to the same gene locus by random chance. So, in this background only genes which had a large proportion of its probesets shown to be altered in all oncogenes can be reported as potential common targets. Unfortunately, the protein products of these genes were not identified in the proteomic analysis, presumably due to their low abundance. Data generated using exon arrays offers the opportunity to consider the possible effects of transformation by individual oncogenes on splice variants and gene expression. Further analysis on exon specific effects to consider oncogene-mediated splice variant generation is underway.

It is now becoming more accepted that changes in the proteome do not necessarily occur as a result of changes in transcription. To look for common oncogenic targets in the proteome, we utilised the newly developed 8 channel iTRAQ reagent to simultaneously assess protein expression in cells expressing six different oncogenes, with the remaining two channels taken up by biological replicates of empty vector-transfected cells. The 8 channel iTRAQ system is a
redesigned version of the 4-plex reagent to allow incorporation of the extra isotopes required to extend the reporter ion mass range. This requires a new ‘balance’ group and minor modifications to the reporter, while retaining the same labelling chemistry via an NHS-ester. In our hands, this new reagent retains many of the features of the 4-plex reagent in that labelling efficiency remains high, we see an apparent increase in signal in a UV detector (at 214nm) during SCX fractionation, and peptides appear to have slightly increased retention on SCX. There is no evidence that peptide elution is different for peptides labelled with different iTRAQ reagents (on either SCX or RP media), peptide fragmentation is good and the iTRAQ reporter ion signal(s) are generally among the most abundant peaks in the majority of spectra. Using a defined mixture, we have shown that, like the 4-plex, the 8-plex is capable of detecting differences in peptide levels between two samples but the degree of change is underestimated. This is unimportant in a study where a low false positive rate is required, as in our studies. We have therefore demonstrated that 8 channel relative quantification mass spectrometry can compare protein profiles (and in the future phosphopeptide profiles (38) in hematopoietic cells).

Once again, however, the proteomics analysis supported the hypothesis that the leukaemogenic properties of the six oncogenes shared little in common. Only one protein was shown to be affected by all six oncogenes, namely eosinophil-associated ribonuclease 7 (also known as ribonuclease 7). Other proteins found to be altered in 5/6 or 4/6 oncogenes include two members of the SET protein
phosphatase family (I2PP2A and Acidic nuclear phosphoprotein pp32). This protein is known to be involved in BCR/ABL-positive chronic myeloid leukaemias (25). Also, several proteins identified are related to the actin cytoskeleton structure. This would suggest a common effect on cell shape or cell motility. Functional assays for motility suggests that, while BCR/ABL appears to decrease the motile response the SDF1, and both TEL/PDGRFβ and D816VKIT increase motility, this pattern does not correlate with the change in expression of these proteins. Likewise, further analysis of the data for proteins whose expression follows this pattern yields two candidates, Septin 1 and HMGB2. These are both therefore candidates for proteins which play a role in oncogene-mediated alterations in motility. However, it is clear that, since other motility associated proteins remain unaltered, or do not follow this pattern, that the oncogenes which do affect motility probably do so using distinct mechanisms.

The exon array approach has been coupled to proteome analysis to match relative quantification at the peptide and exon probeset level. This analysis is potentially more powerful that a gene versus protein analysis due to the methods of data generation. Standard proteomics and microarray technologies each may sample different parts of a protein transcript, so it could be that a comparison is actually between probesets at the beginning of a transcript, versus peptides from the end of the protein. This may be reason why proteome and transcriptome changes correlate poorly in previous studies. Using exon arrays allows a much higher resolution mapping, matching peptide expression to the levels of the
mRNA from the exon which codes for it. However, this analysis also shows a
general lack of correlation between changes in expression at the transcript and
proteome level. Interestingly, the correlations between protein and gene, were
better than those between peptide and exon (data not shown), presumably
because averaging peptides/exons reduces the effect of outlying data points. To
confirm these data for a case where there was no correlation, CapG, the
proteome and transcriptome data were checked using western blotting in relative
quantitative PCR respectively. This confirmed the accuracy of the systematic
analyses, and also revealed further differences between the actions of the
oncogenes. Where both NPM/ALK and D816VKIT downregulate CapG by
decreasing the rate of mRNA transcription, BCR/ABL, FLT3-ITD and
FIP1/PDGFRα all show no change in CapG mRNA levels, and therefore
downregulate the protein either by slower rate of translation or by increased
degradation. These findings promote the idea that proteome analysis (at both
the protein and peptide and phosphopeptide level) should be a key element of
systematic analyses of oncogene effects.

Here we have shown that simple extension of effects seen with one PTK to other
leukaemogenic oncogene tyrosine kinases is inappropriate. There is little
correlation between 6 PTKs on the proteome seen using the approach we have
validated here. Rather than deflecting the search for commonalities the study
suggests a focus on specific organelles (to identify common effects on lower
abundance proteins) and enrichment of post-translationally modified proteins as
a target for further analyses. The present analysis shows no correlative links between type III receptor kinase effects, few common effects elicited by 6 PTKs, a raft of effects across many gene ontology groups and a group of proteins commonly affected that therefore enable development of ideas on common events required in PTK action that contribute to transformation.

Acknowledgements

This work was supported by Leukaemia Research Fund UK. We wish to thank Dr. DG Gilliland and Dr. DW Sternberg for provision of Ba/F3 cell lines. We also wish to thank Lynda Capper for technical support.

References

1. Mitelman, F., Johansson, B., and Mertens, F. (2007) The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer.* 7, 233-245.

2. Konopka, J. B., Watanabe, S. M., and Witte, O. N. (1984) An alteration of the human c-Abl protein in K562 leukemia-cells un masks associated tyrosine kinase-activity. *Cell.* 37, 1035-1042.

3. Golub, T. R., Barker, G. F., Lovett, M., and Gilliland, D. G. (1994) Fusion of PDGF Receptor-Beta to a novel Ets-like gene, Tel, in Chronic Myelomonocytic Leukemia with t(5;12) chromosomal translocation. *Cell.* 77, 307-316.

4. Cools, J., DeAngelo, D. J., Gotlib, J., Stover, E. H., Legare, R. D., Cortes, J., Kutok, J., Clark, J., Galinsky, I., Griffin, J. D., Cross, N. C. P., Tefferi, A.,
Malone, J., Alam, R., Schrier, S. L., Schmid, J., Rose, M., Vandenberghhe, P., Verhoef, G., Boogaerts, M., Wlodarska, I., Kantarjian, H., Marynen, P., Coutre, S. E., Stone, R., and Gilliland, D. G. (2003) A tyrosine kinase created by fusion of the PDGFRA and FIP1L1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. *N. Engl. J. Med.* 348, 1201-1214.

5. Morris, S. W., Kirstein, M. N., Valentine, M. B., Dittmer, K. G., Shapiro, D. N., Saltman, D. L., and Look, A. T. (1994) Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science.* 263, 1281-1284.

6. Nagata, H., Worobec, A. S., Oh, C. K., Chowdhury, B. A., Tannenbaum, S., Suzuki, Y., and Metcalfe, D. D. (1995) Identification of a point mutation in the catalytic domain of the protooncogene c-Kit in peripheral blood mononuclear cells of patients who have mastocytosis with an associated hematologic disorder. *Proc. Natl. Acad. Sci. USA.* 92, 10560-10564.

7. Nakao, M., Yokota, S., Iwai, T., Kaneko, H., Horiike, S., Kashima, K., Sonoda, Y., Fujimoto, T., and Misawa, S. (1996) Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. *Leukemia.* 10, 1911-1918.

8. Baselga, J. (2006) Targeting tyrosine kinases in cancer: The Second Wave. *Science.* 312, 1175-1178.

9. Deininger, M., Buchdunger, E., and Druker, B. J. (2005) The development of imatinib as a therapeutic agent for chronic myeloid leukemia. *Blood.* 105, 2640-2653.

10. Hochhaus, A., Kantarjian, H. M., Baccarani, M., Lipton, J. H., Apperley, J. F., Druker, B. J., Facon, T., Goldberg, S. L., Cervantes, F., Niederwieser, D., Silver, R. T., Stone, R. M., Hughes, T. P., Muller, M. C., Ezzeddine, R., Countouriotis, A. M., and Shah, N. P. (2007) Dasatinib induces notable hematologic and cytogenetic responses in chronic-phase chronic myeloid leukemia after failure of imatinib therapy. *Blood.* 109, 2303-2309.
11. Unwin, R. D., Smith, D. L., Blinco, D., Wilson, C. L., Miller, C. J., Evans, C. A., Jaworska, E., Baldwin, S. A., Barnes, K., Pierce, A., Spooncer, E., and Whetton, A. D. (2006) Quantitative proteomics reveals posttranslational control as a regulatory factor in primary hematopoietic stem cells. *Blood*. 107, 4687-4694.

12. Unwin, R. D., Pierce, A., Watson, R. B., Sternberg, D. W., and Whetton, A. D. (2005) Quantitative proteomic analysis using isobaric protein tags enables rapid comparison of changes in transcript and protein levels in transformed cells. *Mol. Cell. Proteomics*. 4, 924-935.

13. Ross, P. L., Huang, Y. L. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics*. 3, 1154-1169.

14. Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y. C., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y. H., and Zhang, J. H. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*. 5, R80.

15. Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B., and Speed, T. P. (2003) Summaries of affymetrix GeneChip probe level data. *Nucleic Acids Res.*. 31, e15.

16. Okoniewski, M. J., Yates, T., Dibben, S., and Miller, C. J. (2007) An annotation infrastructure for the analysis and interpretation of Affymetrix exon array data. *Genome Biol* 8, R79.

17. Benjamini, Y., and Hochberg, Y. (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J. Roy. Stat. Soc. B* 57, 289-300.
18. Hubbard, T. J. P., Aken, B. L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., Down, T., Dyer, S. C., Fitzgerald, S., Fernandez-Banet, J., Graf, S., Haider, S., Hammond, M., Herrero, J., Holland, R., Howe, K., Howe, K., Johnson, N., Kahari, A., Keefe, D., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Melsopp, C., Megy, K., Meidl, P., Overduin, B., Parker, A., Rlic, A., Rice, S., Rios, D., Schuster, M., Sealy, I., Severin, J., Slater, G., Smedley, D., Spudich, G., Trepanion, S., Vilella, A., Vogel, J., White, S., Wood, M., Cox, T., Curwen, V., Durbin, R., Fernandez-Suarez, X. M., Flick, P., Kasprzyk, A., Proctor, G., Searle, S., Smith, J., Ureta-Vidal, A., and Birney, E. (2007) Ensembl 2007. *Nucleic Acids Res.* 35, D610-D617.

19. Shilov, I. V., Seymour, S. L., Patel, A. A., Loboda, A., Tang, W. H., Keating, S. P., Hunter, C. L., Nuwaysir, L. M., and Schaeffer, D. A. (2007) The Paragon Algorithm: A next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell. Proteomics*, T600050-MCP200.

20. Livak, K. J., and Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods*. 25, 402-408.

21. Whetton, A. D., Lu, Y. N., Pierce, A., Carney, L., and Spooner, E. (2003) Lysophospholipids synergistically promote primitive hematopoietic cell chemotaxis via a mechanism involving Vav 1. *Blood*. 102, 2798-2802.

22. Dai, Z. H., Quackenbush, R. C., Courtney, K. D., Grove, M., Cortez, D., Reuther, G. W., and Pendergast, A. M. (1998) Oncogenic Abl and Src tyrosine kinases elicit the ubiquitin-dependent degradation of target proteins through a Ras-independent pathway. *Genes & Dev.* 12, 1415-1424.

23. Witke, W., Li, W., Kwiatkowski, D. J., and Southwick, F. S. (2001) Comparisons of CapG and gelsolin-null macrophages: demonstration of a unique role for CapG in receptor-mediated ruffling, phagocytosis, and vesicle rocketing. *J. Cell Biol.* 154, 775-784.
24. Evans, C. A., Tonge, R., Blinco, D., Pierce, A., Shaw, J., Lu, Y. N., Hamzah, H. G., Gray, A., Downes, C. P., Gaskell, S. J., Spooncer, E., and Whetton, A. D. (2004) Comparative proteomics of primitive hematopoietic cell populations reveals differences in expression of proteins regulating motility. *Blood.* 103, 3751-3759.

25. Neviani, P., Santhanam, R., Trotta, R., Notari, M., Blaser, B. W., Liu, S. J., Mao, H., Chang, J. S., Galietta, A., Uttam, A., Roy, D. C., Valtieri, M., Bruner-Klisovic, R., Caligiuri, M. A., Bloomfield, C. D., Marcucci, G., and Perrotti, D. (2005) The tumor suppressor PP2A is functionally inactivated in blast crisis CML through the inhibitory activity of the BCR/ABL-regulated SET protein. *Cancer Cell* 8, 355-368.

26. Perrotti, D., and Neviani, P. (2006) ReSETting PP2A tumour suppressor activity in blast crisis and imatinib-resistant chronic myelogenous leukaemia. *Br. J. Cancer.* 95, 775-781.

27. Hamzah, H. G., Pierce, A., Stewart, W. A., Downes, C. P., Gray, A., Irvine, A., Spooncer, E., and Whetton, A. D. (2005) Chronic myeloid leukemia CD34+cells have elevated levels of phosphatidylinositol 3,4,5 trisphosphate (PtdIns(3,4,5) P-3 and lack a PtdIns(3,4,5)P-3 response to cytokines and chemotactic factors; effects reversed by imatinib. *Leukemia.* 19, 1851-1853.

28. Pierce, A., Lu, Y. M., Hamzah, H. G., Thompson, S., Owen-Lynch, P. J., Whetton, A. D., and Spooncer, E. (2006) Differential effect of leukaemogenic tyrosine kinases on cell motility is governed by subcellular localisation. *Br. J. Haematol.* 133, 345-352.

29. Correll, P. H., Paulson, R. F., and Wei, X. (2006) Molecular regulation of receptor tyrosine kinases in hematopoietic malignancies. *Gene.* 374, 26-38.

30. Cilloni, D., Messa, E., Messa, F., Carturan, S., Defilippi, I., Arruga, F., Rosso, V., Catalano, R., Bracco, E., Nicoli, P., and Saglio, G. (2006) Genetic abnormalities as targets for molecular therapies in myelodysplastic syndromes. *Ann. N. Y. Acad. Sci.* 1089, 411-23.
31. Delhommeau, F., Pisani, D. F., James, C., Casadevall, N., Constantinescu, S., and Vainchenker, W. (2006) Oncogenic mechanisms in myeloproliferative disorders. *Cell. Mol. Life. Sci.* 63, 2939-53.

32. Penserga, E. T. P., and Skorski, T. (2007) Fusion tyrosine kinases: a result and cause of genomic instability. *Oncogene.* 26, 11-20.

33. Salgia, R., Li, J. L., Ewaniuk, D. S., Pear, W., Pisick, E., Burky, S. A., Ernst, T., Sattler, M., Chen, L. B., and Griffin, J. D. (1997) BCR/ABL induces multiple abnormalities of cytoskeletal function. *J. Clin. Invest.* 100, 46-57.

34. Perrotti, D., and Calabretta, B. (2004) Translational regulation by the p210 BCR/ABL oncoprotein. *Oncogene.* 23, 3222-9.

35. Calabretta, B., and Perrotti, D. (2004) The biology of CML blast crisis. *Blood.* 103, 4010-22.

36. Warmuth, M., Kim, S., Gu, X. J., Xia, G., and Adrian, F. (2007) Ba/F3 cells and their use in kinase drug discovery. *Curr. Op. Oncology.* 19, 55-60.

37. Mikesch, J. H., Schier, K., Roetger, A., Simon, R., Buerger, H., and Brandt, B. (2006) The expression and action of decay-accelerating factor (CD55) in human malignancies and cancer therapy. *Cellular Oncology.* 28, 223-232.

38. Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A., and White, F. M. (2007) Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc Natl. Acad. Sci. USA.* 104, 5860-5865.
### Table One.
Genes identified as changing in all 6 oncogenes on analysis of the exon array data. To be included the genes must change greater than 2 fold in over half of the probe sets. *Denotes where probesets associated with this gene locus but not mapped to an exon are also changed by >2-fold.

| Id                        | Gene description                        | Probesets changing (total number of exonic probesets) |
|---------------------------|-----------------------------------------|--------------------------------------------------------|
| ENSMUSG00000023367        | Hepatocellular carcinoma-associated antigen 112 | 5 (8)                                                  |
| ENSMUSG00000029810        | LR8 protein                             | 8 (10)                                                 |
| ENSMUSG00000031773        | cDNA sequence AK220210                  | 17 (28)                                                |
| ENSMUSG00000026399        | CD55 antigen                            | 7 (10)                                                 |
| ENSMUSG00000057596        | Expressed sequence AI451617             | 4 (6)                                                  |

| Id                        | Gene description                        | Probesets changing (total number of exonic probesets) |
|---------------------------|-----------------------------------------|--------------------------------------------------------|
| ENSMUSG00000018924        | Arachidonate 15-lipoxygenase             | 14 (15)                                                |
| ENSMUSG00000034652        | CD300A antigen                          | 11 (11)                                                |
| ENSMUSG00000019122        | Chemokine (C-C motif) ligand 9           | 8 (9)                                                  |
| ENSMUSG00000030483        | Cytochrome P450 2b13                     | 6 (10)                                                 |
| ENSMUSG00000032089        | Interleukin 10 receptor, alpha           | 13 (11)*                                               |
| ENSMUSG00000029446        | Phosphoserine phosphatase               | 7 (9)                                                  |
| ENSMUSG00000021108        | Protein kinase C, eta                    | 16 (17)                                                |
| ENSMUSG00000027073        | Proteoglycan 2                           | 5 (7)                                                  |
| ENSMUSG00000054366        | RIKEN cDNA A330102H22                    | 5 (10)                                                 |
| ENSMUSG00000040302        | RIKEN cDNA C030048B08                    | 6 (10)                                                 |
| ENSMUSG00000010342        | Tex14                                    | 21 (37)                                                |
| ENSMUSG00000040828        | Tmem146                                  | 10 (13)                                                |
Table 2. Proteins whose expression is altered by the expression of four or more oncogenes.
Proteins shown are those where a confident assessment of a common decrease or increase in 4 or more of the 6 oncogenes has been found. To be included the data must contain more than 2 peptides and have a fold change (control versus oncogene) <0.8 or >1.2 with a p-value of less than 0.05. The significant changes are shown in bold, the underlined values represent ‘changes’ where the confidence levels have not been met.
FIGURE LEGENDS

Figure 1. **Retroviral expression of oncogenes in Ba/F3 cells.**

Ba/F3 cells were transfected with MSCV retroviral vector containing the different oncogenes as indicated. A) following drug selection the successful transfection was confirmed by assessing oncogene expression by western blot analysis on whole cell lysates. B) the effect of oncogene expression on the phosphotyrosine protein content of the cell was assessed by western blot analysis with anti-phosphotyrosine antibody.

Figure 2. **Analysis of protein expression in 6 oncogene transfected cell lines by 8plex isobaric tagging**

A) A schematic of the new 8-channel (8-plex) iTRAQ reagent, including the isotope coding in the new reporter groups. Reporter masses are sequential, but omit 120Da as this is also the mass of a phenylalanine immonium ion. B) an example of the data generated is shown in with sequence determined by the pattern of fragment ions (lower panel) and quantification derived for the intensities of the 8 reporter ions (upper panel; sample relative abundance 1:1:1:1:1:1:1:1). C) relative quantification of a mixture of 8 proteins mixed in predetermined quantities prior to isobaric tagging. The relative ratios in which the proteins were mixed is shown on the figure using diamond symbols. The bars shown represent 95% confidence interval levels. D) further validation of the
method using biological sample replicates is shown where the ratios of the 
MSCV:MSCV Ba/F3 control cells (mixed 1:1 ratio) were labelled and compared 
for protein ratios assessed using the iTRAQ approach. The x-axis indicates the 
protein identification score (Unused score from ProteinPilot software, see 
Methods) which represents the sum of the $\log_{10}(\%$age confidence of peptide 
identification) of each unique peptide associated with only that protein. Each spot 
represents a protein, and only proteins with a score of >4 are included. E-J) 
show the MSCV Ba/F3 cells isobaric tag relative quantification plotted against 
ocogene PTK transfected Ba/F3 cells.

**Figure 3. Analysis of potential oncogene relationships**

A) Dendrogram analysis was performed using the condition tree option in the 
Genespring GX microarray analysis package (Agilent Technologies). Oncogenes 
are clustered according to similar protein expression profiles observed in iTRAQ 
ratios. Green indicates a decrease, red an increase and black indicates no 
change in iTRAQ ratios. The dendrogram tree structure illustrates the relationship 
between the oncogenes. The varying length of the horizontal “branch” indicates 
degree of similarity between the oncogenes-a shorter branch indicates a greater 
degree of similarity. B) An Edwards-Venn diagram constructed to illustrate the 
protein changes commonly caused by the oncogenes.
Figure 4. Correlation of iTRAQ and exon array data
Representative examples of proteins found to change in greater than 4 out of 6 oncogenes are shown. Exon array, iTRAQ, western blot and QRT PCR data is displayed as a fold change compared to MSCV control.

Figure 5. The effect of oncogenic PTK expression on the chemotactic response to SDF-1.
The SDF-1 induced chemotactic response of Ba/F3 cells transfected with the oncogenes shown was assessed in Boyden chamber assays. $10^5$ cells were added to the top well and 100ng/ml SDF-1 to either the top well, bottom well or both wells. The number of cells in the bottom well was counted after 6 hours. Results are the mean ± SEM of three experiments.
SUPPLEMENTARY DATA

Figure legend for Supplementary Table 1 – Table showing all 1886 proteins identified across two replicate analyses of Ba/F3 cells transfected with 6 oncogenic PTKs. The first column denotes the database from which the protein sequence was obtained, along with its accession number. Columns 2 and 3 show the Unused protein score for that protein from each of the two experimental runs. The fourth column contains that name of the protein in that database, and the remaining columns show the (average) ratio of that protein in the oncogene transfected cells compared to an empty-vector control. Since the empty vector control was included twice in the experiment, the control:control ratio is always given.

Figure legend for Supplementary Table 2 - Table showing all 829 high quality relatively quantified proteins from two replicate analyses of Ba/F3 cells transfected with 6 oncogenic PTKs. To be included in this table, a subset of the data shown in Supplementary Table 1, the protein had to be identified with an Unused protein Score of >2, quantification calculated from at least three spectra, and has a calculated error factor of <2. Column headings are as Supplementary Table 1

Figure legend for Supplementary Table 3 - The high quality proteins identified to be differentially expressed between a control cell line and a BCR/ABL
expressing cell line. Column headings are as described in the legend to Supplementary Data Table 1.

**Figure legend for Supplementary Table 4** - The high quality proteins identified to be differentially expressed between a control cell line and a TEL/PDGFRβ expressing cell line. Column headings are as described in the legend to Supplementary Data Table 1.

**Figure legend for Supplementary Table 5** - The high quality proteins identified to be differentially expressed between a control cell line and a NPM/ALK expressing cell line. Column headings are as described in the legend to Supplementary Data Table 1.

**Figure legend for Supplementary Table 6** - The high quality proteins identified to be differentially expressed between a control cell line and a D816V KIT expressing cell line. Column headings are as described in the legend to Supplementary Data Table 1.

**Figure legend for Supplementary Table 7** - The high quality proteins identified to be differentially expressed between a control cell line and a FLT3ITD expressing cell line. Column headings are as described in the legend to Supplementary Data Table 1.
Figure legend for Supplementary Table 8 - The high quality proteins identified to be differentially expressed between a control cell line and a FIP1/PDGFRα expressing cell line. Column headings are as described in the legend to Supplementary Data Table 1.
Figure 1

A

| Molecule            | Molecular Weight |
|---------------------|------------------|
| BCR/ABL             | 250 kDa          |
| TEL/PDGFRβ          | 105 kDa          |
| NPM/ALK             | 75 kDa           |
| FIP1/PDGFRα         | 105 kDa          |
| D816V KIT           | 160 kDa          |
| FLT3ITD             | 160 kDa          |

B

| Lane | Description      |
|------|------------------|
| 1    | MSCV             |
| 2    | FLT3ITD          |
| 3    | D816V KIT        |
| 4    | NPM/ALK          |
| 5    | FIP1/PDGFRα      |
| 6    | TEL/PDGFRβ       |
| 7    | BCR/ABL          |
Figure 2A

Isobaric Tag
Total mass = 305

Reporter Group
Mass 113 – 121
(Retains Charge)
Skip 120

Balance Group
Mass 192-184
Skip 185

Amine specific peptide reactive group (NHS)
Figure 2B

Aldolase 1 – E L S D I A H R

m/z 120 left ‘blank’ for Phe immonium ion

R H A I D S L E
Figure 2F

Unused ProtScore

Log2 (TEL/PDGFb:MSCV)

unusedProtScore.png
Figure 2H
Figure 2I

![Graph showing Log2 (FLT3ITD:MSCV) vs. Unused ProtScore with horizontal lines indicating ratios 1:1.5, 1:2, 1:2:1, and 1:5:1.](image)
Figure 3A

BCR/ABL

FIP1/PDGFRα

FLT3-ITD

D816V KIT

NPM/ALK

TEL/PDGFRβ
Figure 4

Ribose-5-phosphate isomerase

Eosinophil major basic protein

CapG

iTRAQ data
Exon array data
QTPCR data
Western blot data
Figure 5

Number of migrating cells

MSCV
BCR/ABL
TEL/PDGFRb
FLT3ITD
D816V KIT
FIP1/PDGFRa
NPM/ALK

Conditions

+/-SDF  +/-SDF  -/+SDF  -/-SDF