Title: Integrated analysis of proteome, phosphotyrosine-proteome, tyrosine-kinome and tyrosine-phosphatome in acute myeloid leukemia

Authors: Jiefei Tong[1,8†], Mohamed Helmy[2†], Florence M.G. Cavalli[3,8†], Lily Jin[1,8], Jonathan St-Germain[4], Robert Karisch[4], Paul Taylor[1,8], Mark D. Minden[4], Michael D. Taylor[3,5,8], Benjamin G. Neel[4,6], Gary D. Bader[2,7], and Michael F. Moran[1,7,8*]

Affiliations:

[1]Program in Cell Biology, Hospital for Sick Children, Toronto.

[2]The Donnelly Centre, University of Toronto.

[3]Program in Developmental & Stem Cell Biology and Arthur and Sonia Labatt Brain Tumour Research Centre, Hospital for Sick Children, Toronto.

[4]Princess Margaret Cancer Centre, Toronto.

[5]Department of Laboratory Medicine and Pathobiology, University of Toronto.

[6]Departmet of Medicine, NYU School of Medicine, New York, NY, USA.

[7]Department of Molecular Genetics, University of Toronto.

[8]Peter Gilgan Centre for Research and Learning, Hospital For Sick Children, 686 Bay Street, Toronto, ON, M5G 0A4, Canada.

Received: 08-26-2016; Revised: 10-21-2016; Accepted: 02-03-2017

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/pmic.201600361.

*To whom correspondence should be addressed: m.moran@utoronto.ca

†Equal contributions were made by these authors.

Abbreviations: AML, acute myeloid leukemia; **pYome,** phosphotyrosine-proteome; **PTPome,** tyrosine-phosphatome

Keywords: AML / kinome/ Proteome/ PTPome/ pYome

Total number of words: 6429

Abstract: Reversible protein-tyrosine phosphorylation is catalyzed by the antagonistic actions of protein-tyrosine kinases (PTKs) and phosphatases (PTPs), and represents a major form of cell regulation. Acute myeloid leukemia (AML) is an aggressive hematological malignancy that results from the acquisition of multiple genetic alterations, which in some instances are associated with deregulated protein-phosphotyrosine (pY)-mediated signaling networks. However, although individual PTKs and PTPs have been linked to AML and other malignancies, analysis of protein-pY networks as a function of activated PTKs and PTPs has not been done. In this study, mass spectrometry was used to characterize AML proteomes, and phospho-proteome-subsets including pY proteins, PTKs, and PTPs. AML proteomes resolved into two groups related to high or low degrees of maturation according to French-American-British (FAB) classification, and reflecting differential expression of cell surface antigens. AML pY proteomes reflect canonical, spatially organized signaling networks, unrelated to maturation, with heterogeneous expression of activated receptor and non-receptor PTKs. We present the first integrated analysis of the pY-proteome, activated PTKs, and PTPs. Every PTP and most PTKs have both positive and negative associations with the pY-proteome. pY proteins resolve into groups with shared PTK and PTP correlations. These findings highlight the importance of pY turnover and the PTP phosphatome in shaping the pY-proteome in AML.

Statement of significance of the study

In this study we used a battery of proteomics methods to characterize the proteomes of primary AML tumors. This included label-free quantification of total proteome, phosphotyrosine-proteome, and comprehensive characterization of classical phosphotyrosine phosphatases (the PTPome). We demonstrate our first integrated analysis of these different kinds of phospho-proteomics datasets. In particular we provide a so-called cluster-of-clusters in which we relate the profile of tumor protein-phosphotyrosine as a function of activated tyrosine kinases and expressed PTP enzymes. To the best of our knowledge, no such integrated analysis has been published. The data argue that the proteome may have utility as a means to stratify tumors according to their protein expression profiles. Importantly, our results illustrate how the PTPome, not just the protein kinases, influence the phospho-proteome.

1 Introduction

Enormous efforts aim to define genomics-based molecular signatures in order to guide the development of precision treatments for individual malignancies. This goal reflects knowledge that tumorigenesis is driven by combinatorial changes in oncogene and tumor suppressor gene [1]. This is exemplified by acute myeloid leukemia (AML), which is a collection of diseases caused by a variety of recurrent and unique mutations[2-5]. A total of 23 genes were significantly mutated, and another 237 were mutated in two or more samples in the genomes of 200 AML samples [5]. Some of mutated genes are well established as being relevant to AML pathogenesis (e.g., *DNMT3A*, *FLT3*, *NPM1*, *IDH1*, *IDH2*, and *CEBPA*) [5]. Gene expression signatures have been suggested for AML [6, 7]. However, the utility of cancer-associated mRNA expression-based signatures has been questioned [2]. To some extent this may reflect the generally poor correlation between mRNA and protein abundances [8-11]. None of the current classification schemes for AML are entirely prognostic. Nearly 50% of AML samples have a normal karyotype, and many of these genomes lack structural abnormalities [5]. These observations provide a rationale for proteomic studies of AML as an alternative source of molecular features as a basis for classification and treatment.

Characterization of AML proteomes and/or phospho-proteomes by various technical platforms including multi-parameter phospho-flow cytometry [12], MS [13], and reverse-phase protein array [8], suggest that patients may stratify into groups defined by distinct phosphorylation networks, which may have prognostic utility. Protein-phosphotyrosine (pY) modifications are a dynamic product of the antagonistic actions of protein-tyrosine kinases (PTKs) and protein-tyrosine phosphatases (PTPs) (Fig. 1A). Both enzyme classes are well

known for their genetic links to AML [14-17]. PTKs are well established potential drug targets in various malignancies, including AML [18]. PTPs are also principal factors in cancer, wherein they are known to function as positive effectors and/or antagonists of pathways that drive cell transformation [19]. MS analysis of the entire complement of classical PTPs, the PTPome, confirmed that variation in PTP expression affects cellular protein tyrosine phosphorylation [20]. However, the extent to which the protein-pY landscape of a cell is a regulated product of the activated tyrosine-kinome and PTPome, as simply depicted in Figure 1A, has not been systematically investigated.

Herein, we report an integrated analysis of AML proteomes and sub-proteomes encompassing tyrosine phosphorylated proteins, activated PTKs, and the PTPome. Our findings reveal new insight into the existence of diverse PTK-PTP relationships associated with pY networks in AML.

2 Materials and Methods

2.1 AML samples and controls

Samples were obtained with REB approval from the Princess Margaret Hospital leukemia repository (Table S1). AML samples are sterile, viable cryopreserved AML cell suspensions, obtained through Ficoll separation of diagnostic bone marrow aspirates; and normal control cells are peripheral blood mononuclear (PBMC) fractions. All cells were stored under liquid nitrogen before use.

2.2 Total peptide profiling and peptide enrichment by pY and oxPTP antibodies

Fig.1B depicts three integrated procedures that were used to analyze AML samples in this study. AML cells were lysed in a urea buffer and then digested by trypsin as described

previously [21]. For total protein analysis (proteome), 5 μg (protein) starting material was digested to peptides and then analyzed by LC-MS/MS. For protein-pY profiling (pY-ome), 5 mg digested protein was subjected to affinity purification by anti-pY antibody (PTMScan, Cell signaling Technology, Danvers, MA) [22]. For comprehensive profiling of classical PTPs (PTP-ome), 3 mg digested protein was oxidized with pervanadate, and oxidized PTP active site motif containing peptides were enriched by anti-oxPTP mouse antibody (R&D systems, cat#MAB2844) as described previously [20]. Detailed protocols for total proteome, pY-ome and PTP-ome are provided in the supplementary protocol.

2.3 LC-MS/MS analysis

Peptides were separated at an operating temperature of 50°C on a 50-cm Easy-Spray column (75-μm inner diameter) packed with 2 μm C18 resin (Thermo Scientific, Odense Denmark). The peptides were eluted over 120 min (250 nl/min) for pY-ome and PTP-ome analyses, and 240 min for whole proteome analysis. The LC was coupled to an Orbitrap Elite mass spectrometer by using a nano-ESI source (Thermo Fisher Scientific, San Jose, CA). Mass spectra were acquired in a data-dependent mode with an automatic switch between a full scan and up to 10 data-dependent MS/MS scans, using HCD fragmentation. Target value for the full scan MS spectra was 3,000,000 with a maximum injection time of 120 ms and a resolution of 70,000 at m/z 400. The ion target value for MS/MS was set to 1,000,000 with a maximum injection time of 120 ms and a resolution of 17,500 at m/z 400. Repeat sequencing of peptides was kept to a minimum by dynamic exclusion of sequenced peptides for 20 s.

Acquired raw files were analyzed by MaxQuant software (version 1.3.0.5) for identification and quantification on Swiss-Prot database (2013.07 version, 20199 entries). For proteome and pY-ome data, the search included cysteine carbamidomethylation as a fixed

modification, N-terminal acetylation, methionine oxidation, phospho-serine, -threonine and -tyrosine (pY-ome data only) as variable modifications. For PTP-ome data, cysteine converting to cysteic acid was added as variable and cysteine carbamidomethylation was changed from fixed to variable modification. The default search parameters in MaxQuant were used. Minimum number of peptides for protein quantification was 2 unique peptides/protein. Localization probabilities for phosphorylation site and cysteic acid for cysteine were required to exceed 75%. The MS spectra of phosphor-peptides discussed in the Results section are shown in Figure S6. MS information related to all detected pY peptides and PTP-ome peptides is shown in Table S5 and Table S8, respectively.

Bioinformatics analysis was completed by using Perseus software tools [23] (perseus-framework.org/) within the MaxQuant environment, R-program, and Cytoscape. For unsupervised clustering and volcano plots, normalized LFQ protein intensities were $\log_2$ transformed, and with imputation of missing values on a per-sample basis using the Perseus default parameters.

Intensities of pY peptides were normalized to peptide amounts in each sample that were measured by using a micro-BCA assay. For clustering analysis of samples based on pY peptides, imputation of missing peptide values was completed in order to replace zero values.

Correlation coefficients between pY-ome and tyrosine kinases or PTPs were calculated by using the correlation function “Corr”, and method “Spearman” in the R-program. The log2 intensity of peptides or proteins, and the correlation coefficient of different pY sites were used for hierarchical clustering by Euclidean distance with average linkage in Perseus.

2.3 Signaling pathway analysis in AML

The AML pY-ome enrichment map was created using g:Profiler with default configurations[24]. We selected the terms with at least 10 genes (Table S5). For each term, we used our pY expression data to calculate total expression score (TES) and average identification frequency (AIF) that are the total expression of all genes associated with the term and the average identification of the term's proteins among our 12 samples, respectively (Table S5).

The pathway analysis and network visualization was carried out by using Cytoscape (2.8.2) and Cytoscape Enrichment Map application [25] with the following parameters: analysis type = generic, p-value cutoff = 1, FDR Q-value cutoff = 1, overlap coefficient = 0.42 and similarity cutoff = Jaccard + overlap combined. The p-value, TES, AIF and number of genes per term were visualized as the node size, node color intense, node-border color intense and node label, respectively. We selected the most significant terms (ten terms) based on Cytoscape sub-networks and g:Profiler enrichment map. The gene-to-gene interaction network with integrated subcellular localization information was built by using the Cytoscape Genemania application [26]. From Genemania, we retrieved the interactions between the identified genes only, by setting the "related genes" option in Genemania to 0. The subcellular localization information was collected using four databases: LOCATE-human [27], LOCATE-mouse [28], the Human Protein Reference Database [29], and UniprotKB. We clustered the subcellular locations into 6 main locations that are Extracellular, Membranes, Cytoplasm, Organelles, Nucleus and Unknown (Table S6).

2.4 Data and materials availability

MS data have been deposited to the ProteomeXchange Consortium [30] via the PRIDE partner repository with the dataset identifier PXD001170.

[Reviewer Access: Project name: AML_profiling; Project accession: PXD001170; Reviewer account username: reviewer41583@ebi.ac.uk; Password: UqPVjMIH; Access the data at http://tinyurl.com/q77c2mu]

3 Results

3.1 Comprehensive analysis of the AML proteome

In order to address relationships between the AML proteome, pY-ome, and PTP-ome, an experimental platform was implemented, as outlined in Fig.1B. Protein extracts were converted to tryptic peptides and then either analyzed by LC-MS/MS directly, or subjected to affinity purification to enrich for pY-containing peptides, or PTPs as indicated. Proteomic datasets were then investigated for relationships by using an integrated approach involving pathway enrichment and protein-protein interaction-based network analyses.

In order to characterize the AML proteome, a set of 12 primary AML samples was collected (Table S1). FLT3-ITD (internal tandem duplication) was detected in one patient sample (#118). Total protein extracts were subjected to quantitative analysis by MS [21]. Four healthy patient-derived peripheral blood mononuclear cell (PBMC) samples were used as a normal blood cell reference. In aggregate, 4485 distinct protein groups were identified (Table S2), and unsupervised hierarchical clustering, based on 3318 proteins observed in two or more AML patients, resolved the samples into 3 groups (Fig. 1C). One group corresponds to the four PBMC samples, which is significant difference

from other two groups (p<0.05, Fig.S1A and S1B). Another group, designated M_{high}, consists of four samples including morphologically mature M5 and M5a samples (FAB, French-American-British classification), and one that was annotated as *prior myelodysplastic syndrome* (designated PM in Fig.1C). The third group, designated M_{low}, consists of eight samples with minimal (M1) or no (M0) maturation, and including one sample with unknown FAB classification (indicated as NA in Fig.1C), and another that was originally scored as M4, but upon relapse was classified as acute lymphoblastic leukemia (ALL). This indicates that monocytic differentiation, which characterizes M5 FAB classification, is associated with a distinctive proteome discernable by MS analysis at the moderate depth of coverage (approx. 3000 proteins) achieved in this study. 462 proteins were identified as differentially expressed between the M_{high} and M_{low} subgroups (Table.S3). Fifty proteins were very highly differentially expressed (|fold change| >10; p<0.01) between the M_{high} and M_{low} subgroups (Table 1). Among this set of proteins, only the actin-binding protein Fascin (FSCN1) was more highly expressed in the M_{low} group, whereas 49 proteins were more highly expressed in the M_{high} subgroup, including 6 hematopoietic cell lineage markers, and 25 predicted extracellular or secreted proteins.

A number of proteins were found to be significantly differentially expressed when the AML and control PBMC proteomes were compared (Table. S4). Of these, 107 were more highly expressed in PBMC, and 269 more highly expressed in AML. Within the 376 differentially expressed proteins are 15 cancer genes according to the Sanger Cancer Gene Census: CD74, CDK6, DDX6, ETV6, FNBP1, HMGA1, MSH2, MSH6, NDRG1, NUP214, NUP98, PSIP1, RPL22, SMARCB1, and TCEA1 [31]. Data on 200 AMLs from The Cancer Genome Atlas (TCGA) Resource [5], accessed and analyzed by using cBioPortal for Cancer Genomics (www.cbioportal.org), indicated that mutation of these cancer genes is infrequent in AML (Table S4). FLT3 and JAK2, which are mutated in some AML [5], were only detected at the protein level in one sample and showed no significant differences in protein

expression between normal and AML. This may reflect low level expression of these signaling proteins, below our limit of detection in total proteome analyses.

3.2 Tyrosine phosphorylation and pathway analysis

Comprehensive protein-pY analysis was completed to quantitatively characterize the AML pY-ome. 219 pY sites, encompassing 159 proteins, were measured (Table S5). In order to determine the cellular processes and pathways represented in the pY dataset, pathway enrichment analyses were conducted. This revealed statistically significant functional groups (Fig. 2A). An AML enrichment map, created by using g:Profiler [24, 32], resulted in over 1,600 GO, KEGG and REAC terms. Terms with at least 10 genes were selected (232 terms) (Table S6). Additionally, pathway enrichment analysis was completed by using the Cytoscape Enrichment Map application [25]. As shown in Fig. 2A, the network includes five disconnected terms. The three most statistically significant groups (Fig. 2A, encircled with dashed lines) were *cell surface receptor signaling pathway*, *response to peptide*, and *peptidyl-tyrosine phosphorylation*. In order to explore additional functional relationships within the AML pY-ome, we further used this set of genes and Cytoscape Genemania [26] to construct an AML gene-to-gene interaction network, shown in Fig.2B (Table S7). Phospho-protein expression level, identification frequency, and known subcellular localization information were used to arrange the interaction network. The resultant schema depicts a network consistent with the transduction of extra cellular signaling cues across the plasma membrane, through membrane-associated signaling components, and leading to cytoplasmic and nuclear effectors (Fig. 2B). Two detected RTKs, FLT3 and KIT, which function atop activated pathways (Fig. 2B) are known to be mutated in AML [5].

3.3 An activated kinome in AML

Unsupervised hierarchical clustering based on the quantified pY-peptides divided AML samples into two groups (Fig.S2A). Each group contains samples from both the M_{low} and M_{high} categories, suggesting the degree of AML cell maturation per se is not associated with gross differences in protein tyrosine phosphorylation. Thirty-three protein kinases were among the identified pY-containing proteins. Activation-loop (A-loop, DFG–pY–APE motif), as shown in Fig. S2B, were detected in eight kinases. Fig. 3 presents a matrix of phosphorylated kinases arranged in a hierarchical (top-to-bottom) manner, with receptor tyrosine kinases (RTKs) followed by non-receptor PTKs, followed by non-tyrosine protein kinases. MS ion currents for pY-peptides can be compared in the horizontal direction. The maximum magnitude of the MS intensity for each pY peptide species is shown in the last column (in shades of blue), as an indicator that some pY peptides may have been present in low levels or have low MS response rates.

Each of the samples contained two or more pY-containing non-receptor PTKs (Fig. 3A), which, according to the computed interaction network (Fig. 2B), are coupled to plasma membrane-associated receptors. Signals derived from SRC-family tyrosine kinases, and the non-receptor tyrosine kinase SYK, were 6-fold and 26-fold higher in Group 2 compared with Group 1, respectively (Fig. 3A). Half the samples, including four in Group 1 and two in Group 2, did not contain a detected RTK. Three RTKs were measured including KIT and FLT3, which are well known to be activated in AML [33, 34], and FGFR3, which is not generally associated with acute leukemia. FLT3 mutation (FTL3-ITD) was only identified in sample #118 (Table S1), which had highest phosphorylation signal at position Y936 (Fig. 3). While FLT3 protein was only detected in sample #228 (Table S2), which had highest

phosphorylation signal at FLT3 Y969. pYome analysis (Fig. 3) revealed tyrosine phosphorylations of FLT3 in some samples which did not have FLT3 mutation or detectable FLT3 protein, indicating pYome analysis is a sensitive and complementary tool for analysis of signaling pathways in patient samples.

3.4 PTP-ome quantification

Anti-oxPTP peptide antibody was used to enrich PTP peptides as described previously [20]. Sixteen classical PTPs were quantified from eight AML samples (Fig. 4A; additional MS information is provided in Table S8). Venn analysis illustrates that three PTPs were themselves subject to tyrosine phosphorylation, and eleven were also measured as part of the total proteome analysis (Fig. 4B). We note that the AP-MS approach for PTP-ome characterization identified more PTPs [35] than total proteome analysis [11], and they were quantified in a greater number of samples. All PTPs identified by total proteome or pYome were quantified by PTPome. Therefore, the AP-MS method provided more thorough data towards the analysis of the impact of the PTP-ome on total cellular tyrosine phosphorylation, as described below.

The influences of the activated tyrosine kinome and PTP-ome on the pY-ome in AML were considered. There was a strong correlation (coefficient of determination R^2 >0.65, p <0.05) between measures of activated tyrosine kinases and the overall level of protein-pY (Fig. 4C). In addition, the PTP-ome may not be simply a negative regulator of cellular protein-pY, since there was a moderate positive correlation, although not significant (p>0.05), between the level of expressed PTPs and cellular protein-pY (R^2 >0.3, Fig. 4D). There was no correlation between PTP expression level and activated tyrosine kinases.

3.5 Integrated analysis: the pY-ome as a product of the activated kinome and PTP-ome

Correlation analyses were performed in order to further reveal relationships between the pY-ome, activated kinome, and PTP-ome. The correlation coefficients relating pY-ome and kinome or PTP-ome are shown in Table S9, and an integrated heat map of correlation coefficients between pY-ome and kinome/PTP-ome is shown in Fig. 4E (see also Fig.S3). Instances where there are a positive correlations between pY-ome and kinome, but negative correlations with the PTP-ome may represent examples of net antagonistic regulation of protein phosphorylation by kinase and dephosphorylation by PTP (Fig. 1A).

The dendrogram on top of the heat map in Fig. 4E largely separated the kinome and PTP-ome. Strikingly, every PTP, as well as most kinases, had both positive and negative associations with the pY-ome. The kinases were largely separated into 2 groups (Group A and B in Fig. 4E). The three measured RTKs (KIT, FGFR3, FLT3) did not cluster together, and only one cluster contained both kinases (DYRK1A, DYRK2) and phosphatases (PTPRB, PTPRG, PTPN13, PTPN18).

Analysis of the horizontal dendrogram revealed 17 clusters of pY sites (Table S9). Five clusters that contain more than 10 pY sites are shown in Figure 4E, numbered 4, 8, 9, 12 and 13. The sequence contexts of the pY sites associated with these five clusters are distinctive, as shown in Fig. 4E (see also Table S9). Four to six representative pY sites from each of these clusters are shown to the right of the determined consensus sequence logo. Another five clusters with more than five, but less than 10 pY sites are shown in Figure S4. In general, pY-ome cluster 4 is positively correlated with group A kinases, including HCK and ABL2, and several PTPs. Cluster 8 and 9 are both highly positively correlated with group B kinases, including FGR, SYK, BTK, but differ in their PTP correlations. Cluster 12 and 13

show strong positive correlation with all of the kinases. Cluster 12 shows strong negative correlations with most of the phosphatases, whereas cluster 13 shows moderate positive correlation with most of the phosphatases.

Although many kinases show both positive and negative correlations with the pY-ome, a subgroup of 5 kinases (FGR, SYK, BTK, PTK2B and SGK223) in Group B has positive correlations with a majority of the pY-ome. PTPN1 and PTPN2 (see asterisks in Fig. 4E) are structurally and functionally related [36], but show distinct relationships with the pY-ome. PTPN1 has modest positive and strong negative correlations in clusters 4 and 9, respectively, whereas PTPN2 has almost opposite relationships in these two regions (Fig. 4E and Table S9). Figure S5 shows a more detailed list of 58 protein-pY sites highly discordant in their correlation with PTPN1 and PTPN2. In general, PTPN1 has negative correlation with most of pY sites (44 sites), whereas PTPN2 is negatively correlated with only 14 sites. Indeed, four reported PTPN1 substrates, FLT3 [37], SYK [38], STAM2 [39] and PXN [40], showed negative correlations with PTPN1 expression. Approximately half of the pY sites that were negatively correlated with PTPN2 expression are annotated for nucleic acid interaction/localization such as RPS13, SRRM2, GSTp1, SF3A3 and RPS10 (Fig.S5).

4 Discussion

Classification of AML according to the FAB system is based on morphologic features, along with flow cytometry analysis of surface markers, cytogenetics, and assessment of recurrent molecular abnormalities. So far, none of the current classification schemes for AML are entirely prognostic. Comprehensive proteome analysis segregated AML into two significantly different groups, designated M_{low} and M_{high}. Among the highly differentially expressed proteins are six known hematopoietic surface antigens, and more than twenty other

secreted/extracellular proteins. This is consistent with the known heterogeneity in AML antigen expression [41]. Our findings illustrate the potential for comprehensive or targeted proteome profiling as an approach to complement FAB classification of AML. FAB classification is commonly for AML, but does not take into account some prognostic factors. The World Health Organization (WHO) has developed a newer system for AML classification that includes some of these factors [42].

Pathway analysis of the AML pY-ome was consistent with the canonical view that these malignancies are dependent on, if not driven by, activated pY-mediated signaling networks generally proceeding from the plasma membrane to the nucleus (Fig. 2). Clustering analysis of the AML pY-omes revealed two groups (Fig. 1D), one of which (Group 2) showed a higher overall level of protein-pY, and a greater complement of activated non-receptor tyrosine kinases compared with the other (Fig. 3). SYK and SRC-family kinases have been identified as therapeutic targets in AML [43, 44]. Both Group 1 and Group 2 contained some samples with activated FLT3 and/or KIT, both implicated as targets in AML [33, 34], but Group 1 on average had a lower level of activated non-receptor tyrosine kinases (Fig. 3). Curiously, one of the Group 2 tumors expressed activated FGFR3, as indicated by its pY-containing A-loop peptide. FGFR3 is a target in t(4;14) multiple myeloma [45, 46] and widely expressed in chronic leukemia [47], but has not been established as a target in AML. These results illustrate the potential for pY-focused phospho-proteomics as a systematic approach for the discovery of candidate tyrosine kinase targets [48, 49], and may be instructive towards testing primary AML tumors *ex vivo* for sensitivity to tyrosine kinase inhibitors (TKIs).

Only recently have proteomics technologies emerged to facilitate comprehensive analysis of the classical PTPs, the PTP-ome [20]. This study represents a primary attempt to integrate cellular protein-pY patterns with the expression of activated kinases and the PTP-ome. It is conceivable that positively and negatively correlated PTP expression with a given pY site reflects the indirect activation of phosphorylation (e.g. dephosphorylation of an inhibitory pY site on an upstream PTK) and a direct role in dephosphorylation, respectively.

Almost half of the measured pY-ome was positively correlated with the PTP-ome (Fig. 4E, cluster 4, 8 and 13). It was reported recently that PTP activity in acute leukemia patients was high compared to the controls [50]. PTPN1 and PTPN2 are structurally and functionally related [36], but in AML their correlation with the pY-ome, particularly with respect to clusters 4 and 9, were contrasting (Fig. 4E). This may reflect differences in their subcellular localization, which has been shown to regulate their access to substrate RTKs [e.g. ref. 51]. Our results indicate that a high level of classical PTP expression in AML neoplasms does not necessarily result in low levels of pY-containing proteins, and supports the notion that protein-pY turnover is elevated in AML.

In conclusion, this study indicates and emphasizes the complexities involved in the biological regulation and technical measurement of protein phosphorylation. The comparison of the relative influences of the activated tyrosine kinome and PTP-ome on the pY-ome in AML indicated a generally stronger contribution by the kinome than the PTP-ome. Our findings illustrate that the expression of PTPs, which are highly variably expressed in cell lines, tissues and tumors [20] will have a strong influence on pY networks. Awareness of this may be of particular importance when modulation or monitoring protein-pY is a therapeutic aim.

5 References

[1] Vogelstein, B., Kinzler, K. W., Cancer genes and the pathways they control. *Nat Med* 2004, *10*, 789-799.

[2] Chapman, M. A., Lawrence, M. S., Keats, J. J., Cibulskis, K.*, et al.*, Initial genome sequencing and analysis of multiple myeloma. *Nature* 2011, *471*, 467-472.

[3] Graubert, T. A., Mardis, E. R., Genomics of acute myeloid leukemia. *Cancer journal (Sudbury, Mass.)* 2011, *17*, 487-491.

[4] Morgan, G. J., Walker, B. A., Davies, F. E., The genetic architecture of multiple myeloma. *Nat Rev Cancer* 2012, *12*, 335-348.

[5] TCGA, Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013, *368*, 2059-2074.

[6] Bullinger, L., Dohner, K., Bair, E., Frohling, S.*, et al.*, Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med* 2004, *350*, 1605-1616.

[7] Valk, P. J., Verhaak, R. G., Beijen, M. A., Erpelinck, C. A.*, et al.*, Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* 2004, *350*, 1617-1628.

[8] Kornblau, S. M., Tibes, R., Qiu, Y. H., Chen, W.*, et al.*, Functional proteomic profiling of AML predicts response and survival. *Blood* 2009, *113*, 154-164.

[9] Nishizuka, S., Charboneau, L., Young, L., Major, S.*, et al.*, Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. *Proc Natl Acad Sci U S A* 2003, *100*, 14229-14234.

[10] Tian, Q., Stepaniants, S. B., Mao, M., Weng, L.*, et al.*, Integrated genomic and proteomic analyses of gene expression in Mammalian cells. *Mol Cell Proteomics* 2004, *3*, 960-969.

[11] Varambally, S., Yu, J., Laxman, B., Rhodes, D. R.*, et al.*, Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell* 2005, *8*, 393-406.

[12] Irish, J. M., Hovland, R., Krutzik, P. O., Perez, O. D.*, et al.*, Single cell profiling of potentiated phospho-protein networks in cancer cells. *Cell* 2004, *118*, 217-228.

[13] Walters, D. K., Goss, V. L., Stoffregen, E. P., Gu, T. L.*, et al.*, Phosphoproteomic analysis of AML cell lines identifies leukemic oncogenes. *Leuk Res* 2006, *30*, 1097-1104.

[14] Blume-Jensen, P., Hunter, T., Oncogenic kinase signalling. *Nature* 2001, *411*, 355-365.

[15] Lemmon, M. A., Schlessinger, J., Cell signaling by receptor tyrosine kinases. *Cell* 2010, *141*, 1117-1134.

[16] Labbe, D. P., Hardy, S., Tremblay, M. L., Protein tyrosine phosphatases in cancer: friends and foes! *Progress in molecular biology and translational science* 2012, *106*, 253-306.

[17] Doepfner, K. T., Boller, D., Arcaro, A., Targeting receptor tyrosine kinase signaling in acute myeloid leukemia. *Critical reviews in oncology/hematology* 2007, *63*, 215-230.

[18] Fabbro, D., Cowan-Jacob, S. W., Mobitz, H., Martiny-Baron, G., Targeting cancer with small-molecular-weight kinase inhibitors. *Methods Mol Biol* 2012, *795*, 1-34.

[19] Julien, S. G., Dube, N., Hardy, S., Tremblay, M. L., Inside the human cancer tyrosine phosphatome. *Nat Rev Cancer* 2011, *11*, 35-49.

[20] Karisch, R., Fernandez, M., Taylor, P., Virtanen, C*., et al.*, Global proteomic assessment of the classical protein-tyrosine phosphatome and "Redoxome". *Cell* 2011, *146*, 826-840.

[21] Wei, Y., Tong, J., Taylor, P., Strumpf, D*., et al.*, Primary tumor xenografts of human lung adeno and squamous cell carcinoma express distinct proteomic signatures. *J Proteome Res* 2011, *10*, 161-174.

[22] Rush, J., Moritz, A., Lee, K. A., Guo, A*., et al.*, Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat Biotechnol* 2005, *23*, 94-101.

[23] Hubner, N. C., Mann, M., Extracting gene function from protein-protein interactions using Quantitative BAC InteraCtomics (QUBIC). *Methods* 2011, *53*, 453-459.

[24] Reimand, J., Arak, T., Vilo, J., g:Profiler--a web server for functional interpretation of gene lists (2011 update). *Nucleic acids research* 2011, *39*, W307-315.

[25] Merico, D., Isserlin, R., Stueker, O., Emili, A., Bader, G. D., Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PloS one* 2010, *5*, e13984.

[26] Montojo, J., Zuberi, K., Rodriguez, H., Kazi, F*., et al.*, GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics* 2010, *26*, 2927-2928.

[27] Sprenger, J., Lynn Fink, J., Karunaratne, S., Hanson, K*., et al.*, LOCATE: a mammalian protein subcellular localization database. *Nucleic acids research* 2008, *36*, D230-233.

[28] Fink, J. L., Aturaliya, R. N., Davis, M. J., Zhang, F*., et al.*, LOCATE: a mouse protein subcellular localization database. *Nucleic acids research* 2006, *34*, D213-217.

[29] Keshava Prasad, T. S., Goel, R., Kandasamy, K., Keerthikumar, S*., et al.*, Human Protein Reference Database--2009 update. *Nucleic acids research* 2009, *37*, D767-772.

[30] Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A.*, et al.*, ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature biotechnology* 2014, *32*, 223-226.

[31] Futreal, P. A., Coin, L., Marshall, M., Down, T.*, et al.*, A census of human cancer genes. *Nat Rev Cancer* 2004, *4*, 177-183.

[32] Reimand, J., Kull, M., Peterson, H., Hansen, J., Vilo, J., g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic acids research* 2007, *35*, W193-200.

[33] Paschka, P., Marcucci, G., Ruppert, A. S., Mrozek, K.*, et al.*, Adverse prognostic significance of KIT mutations in adult acute myeloid leukemia with inv(16) and t(8;21): a Cancer and Leukemia Group B Study. *J Clin Oncol* 2006, *24*, 3904-3911.

[34] Nakao, M., Yokota, S., Iwai, T., Kaneko, H.*, et al.*, Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. *Leukemia* 1996, *10*, 1911-1918.

[35] Cox, J., Mann, M., Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem* 2011, *80*, 273-299.

[36] Stuible, M., Doody, K. M., Tremblay, M. L., PTP1B and TC-PTP: regulators of transformation and tumorigenesis. *Cancer Metastasis Rev* 2008, *27*, 215-230.

[37] Stuible, M., Zhao, L., Aubry, I., Schmidt-Arras, D.*, et al.*, Cellular inhibition of protein tyrosine phosphatase 1B by uncharged thioxothiazolidinone derivatives. *Chembiochem* 2007, *8*, 179-186.

[38] Cheng, A., Bal, G. S., Kennedy, B. P., Tremblay, M. L., Attenuation of adhesion-dependent signaling and cell spreading in transformed fibroblasts lacking protein tyrosine phosphatase-1B. *J Biol Chem* 2001, *276*, 25848-25855.

[39] Stuible, M., Abella, J. V., Feldhammer, M., Nossov, M.*, et al.*, PTP1B targets the endosomal sorting machinery: dephosphorylation of regulatory sites on the endosomal sorting complex required for transport component STAM2. *J Biol Chem* 2010, *285*, 23899-23907.

[40] Takino, T., Tamura, M., Miyamori, H., Araki, M.*, et al.*, Tyrosine phosphorylation of the CrkII adaptor protein modulates cell migration. *J Cell Sci* 2003, *116*, 3145-3155.

[41] Terstappen, L. W., Safford, M., Konemann, S., Loken, M. R.*, et al.*, Flow cytometric characterization of acute myeloid leukemia. Part II. Phenotypic heterogeneity at diagnosis. *Leukemia* 1992, *6*, 70-80.

[42] Falini, B., Tiacci, E., Martelli, M. P., Ascani, S., Pileri, S. A., New classification of acute myeloid leukemia and precursor-related neoplasms: changes and unsolved issues. *Discov Med* 2010, *10*, 281-292.

[43] Dos Santos, C., McDonald, T., Ho, Y. W., Liu, H.*, et al.*, The Src and c-Kit kinase inhibitor dasatinib enhances p53-mediated targeting of human acute myeloid leukemia stem cells by chemotherapeutic agents. *Blood* 2013, *122*, 1900-1913.

[44] Hahn, C. K., Berchuck, J. E., Ross, K. N., Kakoza, R. M.*, et al.*, Proteomic and genetic approaches identify Syk as an AML target. *Cancer Cell* 2009, *16*, 281-294.

[45] Trudel, S., Ely, S., Farooqi, Y., Affer, M.*, et al.*, Inhibition of fibroblast growth factor receptor 3 induces differentiation and apoptosis in t(4;14) myeloma. *Blood* 2004, *103*, 3521-3528.

[46] Trudel, S., Stewart, A. K., Rom, E., Wei, E.*, et al.*, The inhibitory anti-FGFR3 antibody, PRO-001, is cytotoxic to t(4;14) multiple myeloma cells. *Blood* 2006, *107*, 4039-4046.

[47] Wu, J. Y., Huang, L., Xiong, J., Cao, Y.*, et al.*, [Expression of a tumor related gene fgfr3 mRNA in leukemic cells and its clinical significance]. *Zhongguo shi yan xue ye xue za zhi / Zhongguo bing li sheng li xue hui = Journal of experimental hematology / Chinese Association of Pathophysiology* 2008, *16*, 738-741.

[48] Haley, J., White, F. M., Adaptive protein and phosphoprotein networks which promote therapeutic sensitivity or acquired resistance. *Biochem Soc Trans* 2014, *42*, 758-764.

[49] Moran, M. F., Tong, J., Taylor, P., Ewing, R. M., Emerging applications for phospho-proteomics in cancer molecular therapeutics. *Biochim Biophys Acta* 2006, *1766*, 230-241.

[50] Akdogan, E., Cengiz, S., Yilmaz, M., Sonmez, M.*, et al.*, Evaluation of protein tyrosine phosphatase activity in patients with acute leukemia. *Contemp Oncol (Pozn)* 2013, *17*, 83-87.

[51] St-Germain, J. R., Taylor, P., Zhang, W., Li, Z.*, et al.*, Differential regulation of FGFR3 by PTPN1 and PTPN2. *Proteomics* 2015, *15*, 419-433.

6 Acknowledgments:

We thank Princess Margaret Hospital and their generous specimen donors. Support for this research was provided by The Canadian Cancer Society Research Institute, Canadian Institutes for Health Research, and Canada Research Chairs Program (MFM), Genome Canada (MDT), and the Brain Tumour Foundation of Canada (FMGC).

Conflict of interest: The authors declare no conflicts of interest.

Figure Legends

Figure 1. Profile of total protein in AML. (A) The dynamic regulation of protein tyrosine phosphorylation by kinases (kinome) and phosphotyrosine phosphatase (PTP-ome). (B) The Experimental design and proteome analysis of AML tumors. AML cells were treated as illustrated and three MS datasets were obtained from LC-MS/MS analysis: Proteome, pY-ome and PTP-ome. (C) AML proteome analysis. Unsupervised clustering of AML and PBMC cells according to the normalized variation of abundance of 3318 proteins. The French-American-British morphology classification (FAB classification; M0 through M5), when known, is indicated. The samples cluster into three main groups including the control PBMC samples (shaded grey); a group designed M_{high} (shaded blue) comprising AML samples related to M5 classification cases; and a third group, designated M_{low}, mainly comprising AML cases with minimal (M1) or no (M0) maturation. Averages of protein expression in two experiments were used for statistical analysis.

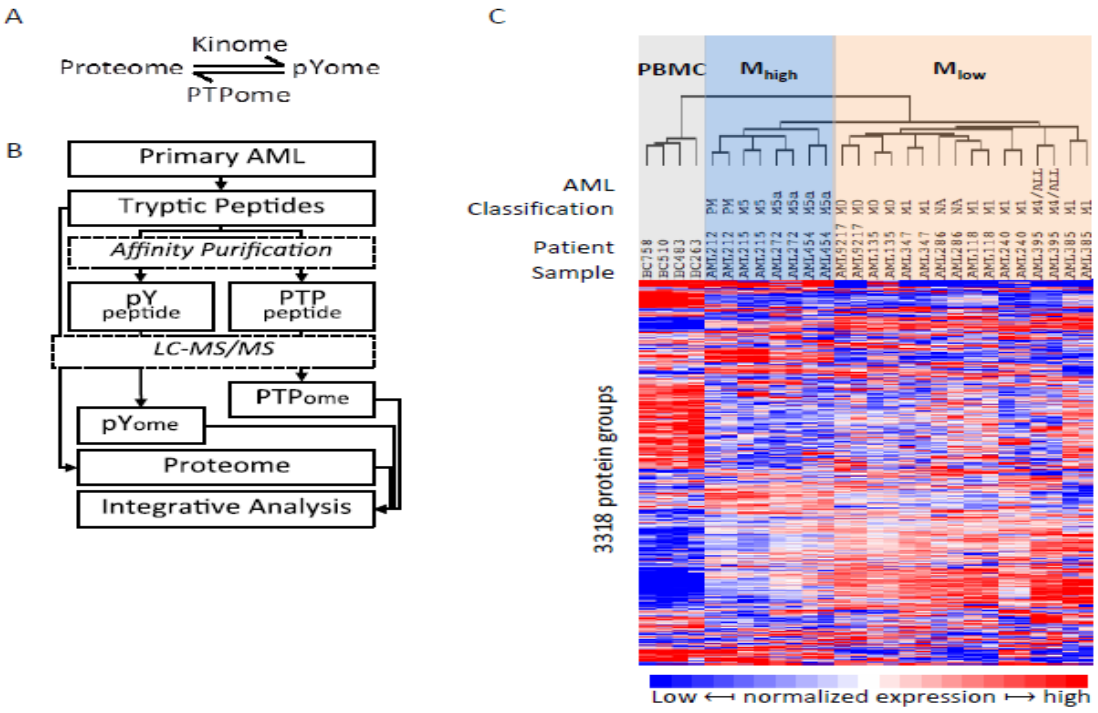

Figure 2. Biological functional network analysis of AML pY-ome**. (**A) Pathway enrichment analysis of AML pY-ome. The network represents the identified GO and KEGG terms (nodes) and the relationship between them (edges) based on similarity of the associated genes/proteins. The node size reflects the significance [-log10(p-value)]. The node label and color are the number of proteins and the total expression score (TES) of each term, respectively. The node border size reflects the average identification frequency (AIF) of the proteins. The edge weight reflects the similarity between terms. The dotted red circles indicate the most significant functional groups. **(**B) Construction and analysis of AML pY-ome interaction network. The TES, AIF and the localization information were added to the network and the proteins were arranged based on the cellular localization. The node color represents the expression level while the node size represents to number of samples where the protein was identified (1 to 12). The node shape represents the cellular localization of the protein. Proteins with more than two locations were attributed as ALL (represented with circle). The interaction network was constructed by using Cytoscape Genemania Application, and with the proteins in the most significant GO terms identified in the pathway enrichment analysis.

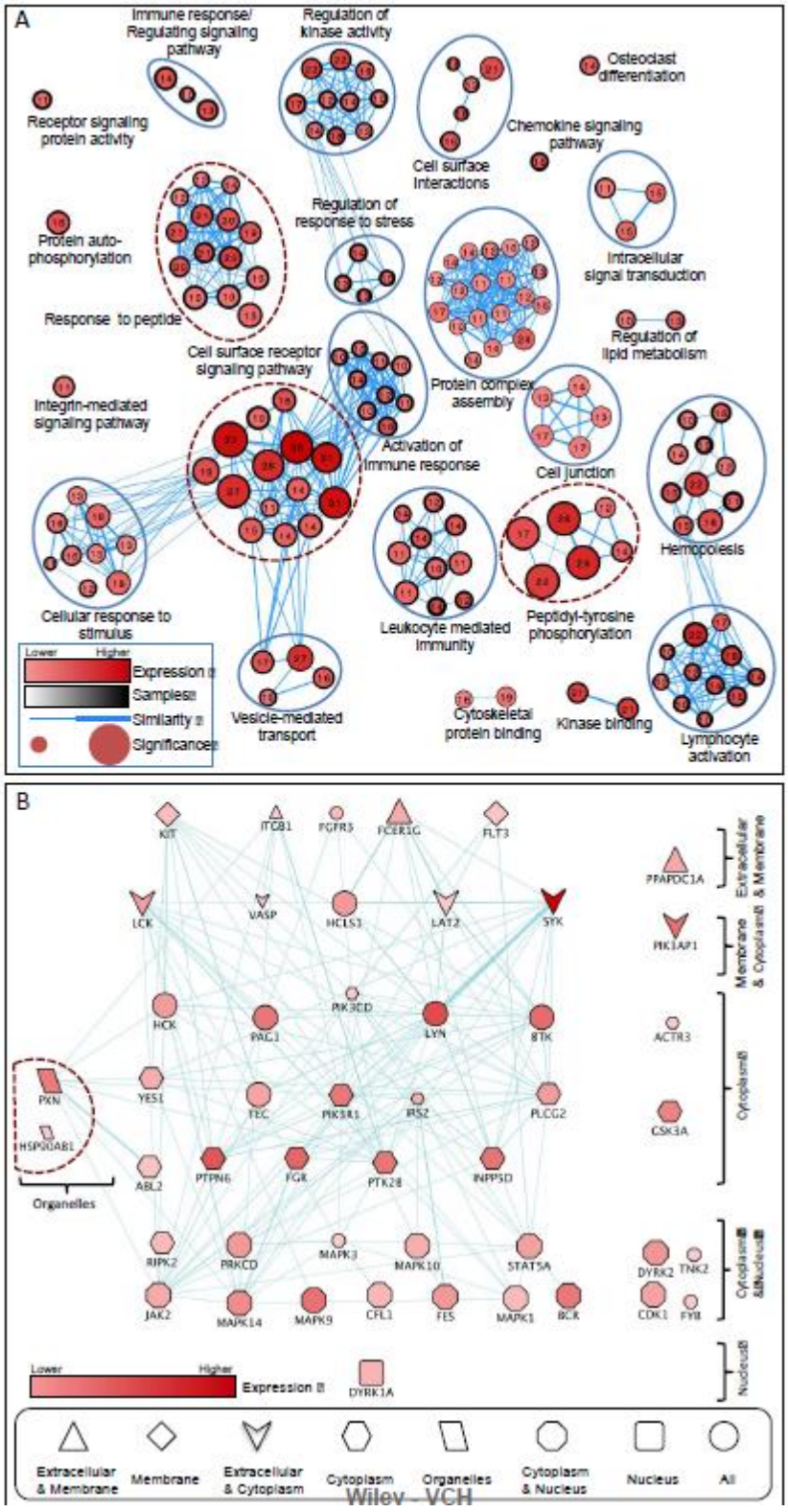

Figure 3. Protein kinase phosphorylation in AML. Heat map showing the relative abundance of pY peptides for the indicated protein kinases that were detected in AML. Phosphopeptides were quantified according to integrated extracted ion currents with MaxQuant software and normalized to sample starting material. The Blue column on the far right represents Log10 maximum intensities for each phosphor-peptide across all AML samples. The ratio of signals from Group 2 and Group 1 samples (fourth column) was calculated as the ratio of mean average of quotients of summed signals from Group 2 peptides vs. summed signals from Group 1 peptides (Table S5). In cases of zero divisors, a value of 10 was used. In cases of only a single peptide with a zero divisor, the ratio was set to >10.

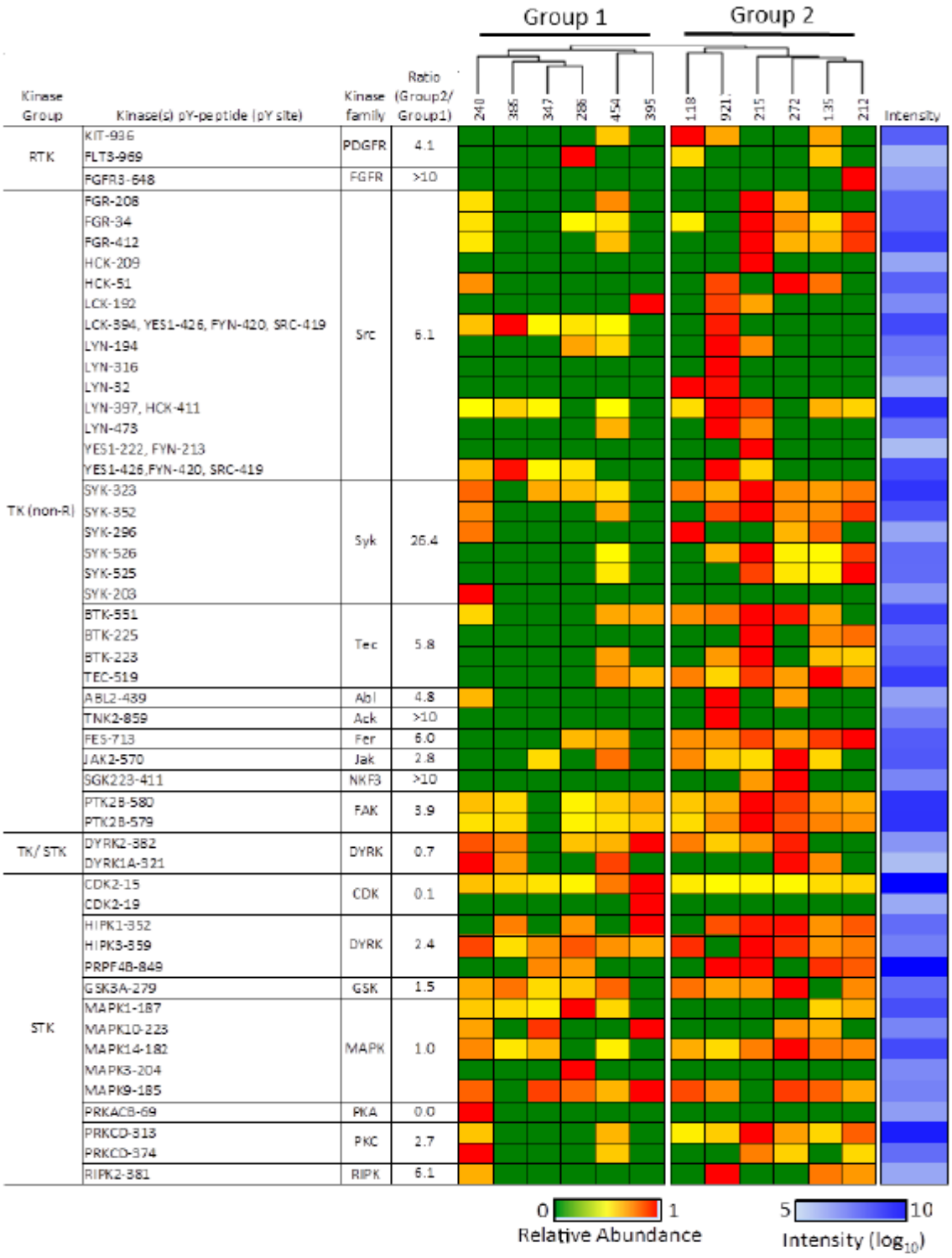

Figure 4. The integrated analysis of protein-tyrosine phosphatases (PTP-ome), relationship among pY-ome, kinome and PTP-ome, and AML proteome. (A) Unsupervised clustering of PTP signature peptides in the indicated AML samples. (B) Venn analysis showing overlap between proteins detected by whole proteome analysis, pY-enrichment (pY-ome), and anti-oxPTP antibody enrichment (PTP-ome). **(**C) and (D) The correlation of MS intensities derived from all pY-containing peptides compared with tyrosine kinase A-loop phospho-peptides (C) or PTP signature peptides (D). (E) Heat map of the correlation coefficients between total pY peptides (pY-ome) and kinase pY-peptides (kinome) or PTP peptides (PTP-ome). Correlations determined by using the Corr() function in R. Sequence frequency analysis is shown for five clusters having more than ten phosphor-tyrosine peptides. Sequence logo plots represent amino acid frequencies for 6 amino acids from both sides of the phosphorylation site (www.weblogo.berkeley.edu). Asterisks indicate PTPN1 and PTPN2. Proteins that contain representative pY sites from the 5 clusters are listed to the right of the sequence logos.

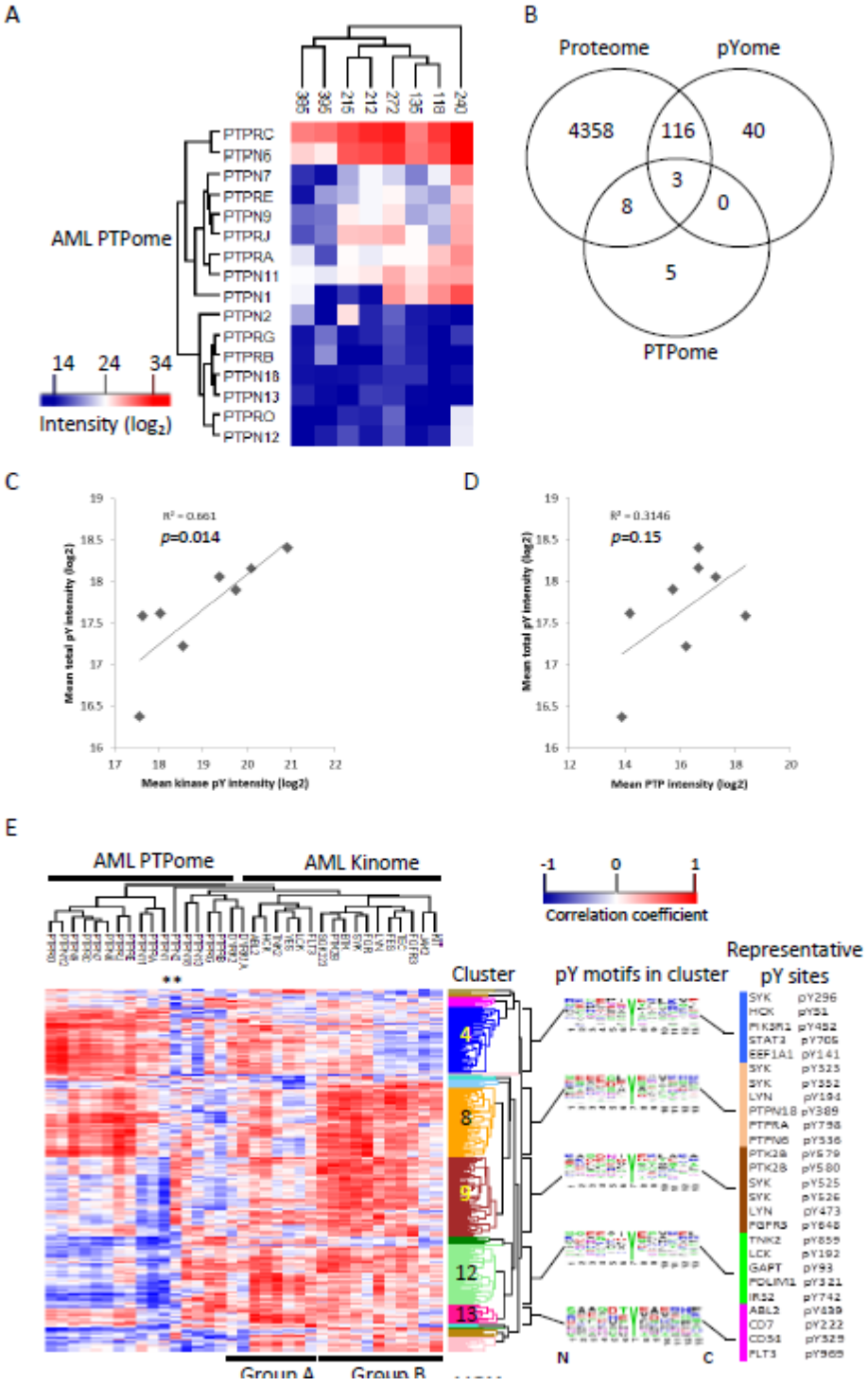

Table 1. Proteins highly differentially expressed in M_{high} vs M_{low} AML (p<0.01 & |Fold Change|>10)

	Protein	Description	Annotation	Unique peptides	M_{high}:M_{low} Log2
1	CD11b	ITGAM, Integrin alpha-M	Hematopoietic cell lineage	10	5.84
2	CD14	Monocyte differentiation antigen	Hematopoietic cell lineage	11	4.35
3	CD36	Thrombospondin receptor	Hematopoietic cell lineage	2	4.15
4	CD41	ITGA2B, Integrin alpha 2b	Hematopoietic cell lineage	28	5.76
5	CD42c	GP1BB, Platelet glycoprotein Ib beta chain	Hematopoietic cell lineage	7	3.77
6	CD61	ITG3B, Integrin beta-3	Hematopoietic cell lineage	2	5.43
7	AZU1	Azurocidin	Extracellular	8	5.00
8	BPI	Bactericidal permeability-increasing protein	Extracellular	19	6.07
9	CTSG	Cathepsin G	Extracellular	22	4.79
10	CTSS	Cathepsin S	Extracellular	14	3.87
11	DEFA1	Neutrophil defensin 1	Extracellular	2	3.60
12	DEFA3	Neutrophil defensin 3	Extracellular	3	5.96

Received: 08-26-2016; Revised: 10-21-2016; Accepted: 02-03-2017

13	ELANE	Neutrophil elastase	Extracellular	12	4.92
14	FCN1	Ficolin-1	Extracellular	8	4.58
15	FGA	Fibrinogen alpha chain	Extracellular	26	3.68
16	FGB	Fibrinogen beta chain	Extracellular	22	3.80
17	FGG	Fibrinogen gamma chain	Extracellular	23	3.51
18	HP	Haptoglobin	Extracellular	19	3.98
19	LGALS3	Galectin-3	Extracellular	7	3.66
20	LTF	Lactotransferrin	Extracellular	63	6.85
21	LYZ	Lysozyme C	Extracellular	13	4.42
22	MMP9	Matrix metalloproteinase-9	Extracellular	20	5.12
23	MMRN1	Multimerin-1	Extracellular	18	3.94
24	PF4	Platelet factor 4	Extracellular	4	4.01
25	PLBD1	Phospholipase B-like 1	Extracellular	16	4.34
26	PPBP	Platelet basic protein	Extracellular	8	4.67
27	S100A8	Protein S100-A8	Extracellular	19	4.93
28	S100A9	Protein S100-A9	Extracellular	13	6.25
29	TUBA4A	Tubulin alpha-4A chain	Extracellular	6	3.81
30	PRTN3	Myeloblastin	Predicted secreted	6	5.05
31	RNASE3	Eosinophil cationic protein	Predicted secreted	9	5.33
32	BASP1	Brain acid soluble protein 1	Membrane associated	14	3.51
33	KCTD12	BTB/POZ domain-containing protein	Membrane associated	13	4.37
34	NCF2	Neutrophil cytosol factor 2	Membrane associated	20	4.39
35	NCF1B	Neutrophil cytosol factor 1B (pseudogene product)	Membrane associated	2	4.30
36	RAB27A	Ras-related protein	Membrane associated	9	3.40
37	CKAP4	Cytoskeleton-associated protein 4	Transmembrane	23	3.54
38	CYBB	Cytochrome b-245 heavy chain	Transmembrane	15	5.87
39	PLP2	Proteolipid protein 2	Transmembrane	2	3.59
40	STOM	Erythrocyte band 7 integral membrane protein	Transmembrane	18	3.65
41	ANXA3	Annexin A3	Calcium binding	27	4.47
42	ANXA5	Annexin A5	Calcium binding	19	3.68

43	EPX	Eosinophil peroxidase	Calcium binding	42	4.71
44	S100A6	Protein S100-A6	Calcium binding	4	5.57
45	HK3	Hexokinase-3	Metabolic enzyme	26	4.89
46	SULT1A1	Sulfotransferase 1A1	Metabolic enzyme	4	3.51
47	TYMP	Thymidine phosphorylase	Metabolic enzyme	17	5.07
48	FSCN1	Fascin	Cytoskeleton	15	-3.65
49	TUBB1	Tubulin beta-1 chain	Cytoskeleton	21	4.72
50	MNDA	Myeloid cell nuclear differentiation antigen	Nuclear antigen	34	6.22

Supplementary information: 6 figures and 9 tables.