Proteomic and Phosphoproteomic Landscapes of Acute Myeloid Leukemia

Michael H. Kramer[1], Qiang Zhang[2], Robert Sprung[2], Ryan B. Day[1], Petra Erdmann-Gilmore[2], Yang Li[1], Ziheng Xu[1], Nichole M. Helton[1], Daniel R. George[1], Yiling Mi[2], Peter Westervelt[1], Jacqueline E. Payton[3], Sai M. Ramakrishnan[1], Christopher A. Miller[1], Daniel C. Link[1], John F. DiPersio[1], Matthew J. Walter[1], R. Reid Townsend[2], Timothy J. Ley[1].

Affiliations:

[1]Division of Oncology, Department of Internal Medicine, Washington University School of Medicine, Saint Louis, MO

[2]Division of Endocrinology, Metabolism and Lipid Research, Washington University School of Medicine, Saint Louis, MO

[3]Department of Pathology and Immunology, Washington University School of Medicine, Saint Louis, MO

Corresponding author:
Timothy J. Ley, M.D.
timley@wustl.edu

Supplemental Materials

Table of Contents:

Supplemental Methods

Supplemental Figure 1: Correlation of deep-scale proteomic process replicates for human peptides and human phosphopeptides.

Supplemental Figure 2: Effect of endogenous myeloid serine proteases on protein detection.

Supplemental Figure 3: Data quality measures.

Supplemental Figure 4: Unsupervised clustering of proteomic profiles from LFQ data.

Supplemental Figure 5: Principal Component Analysis (PCA) and t-SNE analysis of TMT and LFQ data.

Supplemental Figure 6: Common AML mutations supervise clusters of patients in t-SNE analysis.

Supplemental Figure 7: Unsupervised hierarchical clustering of protein abundance profiles with lineage-associated genes removed.

Supplemental Figure 8: Global post-transcriptional regulation of the H/ACA box snoRNP core complex and THO complex in AML.

Supplemental Figure 9: Immunofluorescence of cells transduced with TurboID constructs.

Supplemental Figure 10. Differentially abundant proteins for DNMT3A mutant samples.

Supplemental Figure 11: Differentially abundant proteins for *TP53*, *WT1* and *SMC3* mutant samples.

Supplemental Figure 12: Differentially abundant proteins for *RUNX1*, *NRAS* and *TET2* mutant samples.

Supplemental Figure 13. Differentially abundant proteins for CEBPA mutant samples.

Supplemental Figure 14. Differentially abundant proteins for FLT3 mutant samples.

Supplemental Figure 15: CD180 and MRC1/CD206 Protein vs. RNA abundance.

Supplemental Figure 16. Combined hierarchical clustering of proteomic and phosphoproteomic data.

Supplemental Table 1: Sample identifiers and peptide quantities.

Supplemental Table 2: Clinical annotation and mutations for patients included in this dataset.

Supplemental Table 3: TMT and LFQ protein abundance measurements for all patient and control samples used in this dataset.

Supplemental Table 4: mRNA abundance for all patient and control samples used in this dataset.

Supplemental Table 5: Spearman correlations of protein (TMT) vs. mRNA abundance across the AML patients included in this study.

Supplemental Table 6: Proteins identified using NPM1 and NPMc TurboID constructs for proximity-labeling of proteins with biotin.

Supplemental Table 7: Phosphosite abundance measurements for all patient and control samples used in this dataset (normalized Mascot Delta Score ≥ 0.5)

Supplemental Methods

Preparation of lysates from banked AML cells

Cryopreserved AML cells were thawed and processed in the presence of diisopropyl fluorophosphate (DFP), a membrane permeable, irreversible inhibitor of serine proteases. Prior to the thawing of cells, a waste beaker with 2 N NaOH was prepared for discard of all pipette tips and other plastic that would come in contact with DFP. The beaker and waste remained in the fume hood for 48 hours, then the plastic waste was bagged and the liquid waste was decanted into a bottle for collection by Washington University Environmental Health Services. Cryovials in sets of two were pulled from liquid nitrogen. The vials were warmed briefly in a 37° C water bath until the frost was off the outside of the tube and the cells had warmed without thawing. Fetal bovine serum 0.5mL (FBS) at room temperature (RT), containing DFP, was added to each vial containing ~ 1 ml of cell suspension. The final concentration of DFP was 2 mM. The cell suspension was pipetted up and down until the cells had thawed (~ 3 min). After thawing the cells were immediately diluted into 10 mL of FBS in a 15 ml-conical test tube. The resuspension procedure was repeated with the second cryovial. The cells were centrifuged at 1500 rpm at 10 °C for 5 min. Supernatants were discarded into the beaker with NaOH. The cells were resuspended in PBS. Cells were washed with ~10 ml of PBS, followed by an additional centrifugation and resuspension and discarding all waste into the beaker of 2 M NaOH. Cell pellets were solubilized in 200 µl of Tris urea lysis buffer (pH 8.0). Samples were transferred, using lysis buffer rinse (50 µl), to a Covaris MilliTUBE with AFA fiber for focused ultrasonication. Lysates were sonicated for 12 min (Peak Incident Power: 70 W, Duty Factor: 50 %, cycles/burst: 200, time: 12 min, temp: 5-8˚ C), placed on ice, and transferred to 1.5 ml tubes. Lysates were spun at 16,000 x g in an Eppendorf centrifuge for 30 min at 4 ºC.

The supernatants were removed and protein concentration determined using a Pierce™ BCA Protein Assay Kit and protein aliquoted (330 µg) into 0.5 ml tubes and stored at -80º C. After thawing the

protein concentrations were made similar (~ 5 mg/ml) by precipitating approximately equal quantities of protein from all lysates using the 2D cleanup kit (GE Healthcare) according to manufacturer's instructions. Protein pellets were solubilized in 40 µl of lysis buffer, and protein concentrations were measured as described above. Lysates were aliquoted for the preparation of labeled peptides and phosphopeptides, as previously described.[1]

Preparation of Healthy Control Bone Marrow Samples

Bone marrow samples from healthy donors were collected as part of an IRB-approved protocol at Washington University in St. Louis. Bone marrow buffy coat cells were collected in the same way as as previously described for AML samples.[2] Samples were not cryopreserved, but were instead treated with 2 mM DFP as described above, and then lineage-depleted (Miltenyi Biotec 130-092-211) or CD34-selected (Miltenyi Biotec 130-100-453) using autoMACS separator, per the manufacturer's instructions. The lineage-depletion kit enriches for hematopoietic stem and progenitor cells by removing mature hematopoietic cells including T cells, B cells, NK cells, dendritic cells, monocytes, granulocytes, and erythroid cells. The lineage depletion kit contains an antibody cocktail against CD2, CD3, CD11b, CD14, CD15, CD16, CD19, CD56, CD123 and CD235a to remove mature cells expressing any of these markers.

PDX samples

Cryopulverized patient-derived xenograft (PDX) tissue from previously described batches[1] of basal and luminal breast cancer was used to validate the deep-scale proteomics/phosphoproteomics protocol for this study. The procedures for generation of PDX tissue were reviewed and approved by the institutional animal care and use committee at Washington University in St. Louis.

Preparation of TMT-labeled peptides

The lyophilized peptides were dissolved in 40 µl of HEPES buffer (100 mM, pH 8.5) and labeled according to the vendor protocol using the TMT-11 reagent kit. The labeled samples were combined into groups of nine samples and two reference pools (TMT-11 plexes), dried, and dissolved in 120 µl of 1% (vol/vol) FA.

The combined TMT-11 labeled samples were desalted as described above for the unlabeled peptides. Eluants were collected into the 1.7 ml Eppendorf tubes, frozen and lyophilized. The efficiency of labeling was >99% as determined by UPLC-Orbitrap-MS.

Preparation of stock solutions

All stock solutions were prepared with HPLC grade water. A stock solution for lysis buffer (10X): A 1M solution of Tris buffer (pH 8.0) was prepared by weighing 6.06 g of Tris HCl and dissolving it in 40 ml water. The pH was adjusted with 6 M HCl and the final volume brought to 50 ml. A stock solution of 1M NaCl was prepared by weighing and dissolving NaCl (5.84 g) in 100 ml water. A stock solution of EDTA was prepared by dissolving EDTA 9.3g in 50 ml water. Sodium fluoride was prepared by weighing and dissolving NaF (0.208g) in 50 ml water. A solution of PMSF 100ml was prepared by weighing and dissolving 0.87g in 50 ml MeOH. It was stored at -20 °C for up to 6 months. Aprotinin (1 mg) was dissolved in 1 ml water. Leupeptin (5 mg) was dissolved in 2.5 ml water. Preparation of digest enzymes: dry Lys C (100 AU) was dissolved in 20 ml of water and stored at −80 °C. Trypsin solution (0.5 µg/µl) was used as supplied by the vendor.

Preparation of lysis buffer

The urea lysis buffer was prepared by diluting and mixing stock solutions to the following final concentrations: 50 mM Tris; 75mM NaCl; 1mM EDTA; 10 mM NaF. Phosphatase inhibitor cocktail 2

(1:100), cocktail 3 (1:100), aprotinin (2 μg/mL), leupeptin (10 μg/ml) and 1 mM PMSF were added to the chilled buffer on ice. It was prepared immediately before use.

Preparation of standards for protein and peptide concentrations

Standards for the calibration curve for determination of protein concentrations were prepared by serially diluting the Kit Albumin standard solution (2 mg/ml) (Pierce BCA Protein Assay Kit) with HPLC grade water: 2 mg/ml, 1.5 mg/ml, 1.0 mg/ml, 0.75 mg/ml, 0.5 mg/ml, 0.25 mg/ml, 0.125 mg/ml, and 0.0625 mg/ml. All standard measurements were performed in triplicate. Standard concentrations were used for the fluorescent peptide assay. These were prepared by serial dilution of the assay kit standard: 3.9 ng/µl, 7.8 ng/ µl, 15.6 ng/ µl, 31.3ng/ µl, 62.5 ng/ µl, 125 ng/ µl and 250 ng/µl. The diluent was water containing 2% (vol/vol) MeCN. Standards were measured in triplicate.

Preparation of solvents for peptide purification

Solid-phase extraction using the SepPak cartridge required 2 ml of MeCN for each sample. The solvents for equilibrating and washing the cartridges were 0.1% (vol/vol) TFA and 1% (vol/vol) FA, respectively. The peptide elution solvent was 50% (vol/vol) MeCN/0.1% (vol/vol) FA. The solvents for SepPak solid-phase extraction were made in advance.

Stage-tip desalting solvents: The solvent for equilibrating and washing of stage-tips for phosphopeptide desalting was 0.1% (vol/vol) FA. The solvent for peptide elution was 50% (vol/vol) MeCN/0.1% (vol/vol) FA.

Basic reverse-phase HPLC solvents: Ammonium formate stock solution (180 mM, pH 10) was prepared by adding 12.5 ml of 28% (wt/vol) ammonium hydroxide (density 0.9 g/ml) to ~300 ml of HPLC-grade water. Formic acid (10% (vol/vol) was added and the pH was adjusted to 10. A liter of

Solvent A (4.5 mM ammonium formate (pH 10) in 2% (vol/vol) MeCN) was prepared by adding 25 ml of ammonium formate stock solution and 20 ml of MeCN to 900 ml of water. The pH was adjusted to 10.0 and the solvent was filtered through a 0.2 µ nylon filter. Solvent B (4.5 mM ammonium formate pH 10 in 90% MeCN) was prepared by adding 25 ml of ammonium formate stock solution to 37.5 ml of water and 400 ml of 100% MeCN. The pH was adjusted to 10 and the solvent was filtered through a 0.2 µ nylon filter.

Preparation of labeled peptides for basic reverse phase chromatography

Lyophilized TMT labeled peptides were dissolved by vortexing in 75 µl of 4.5 mM ammonium formate (pH 10) containing 10% (vol/vol) MeCN. The MeCN concentration was decreased by serial addition of 4.5 mM ammonium formate (pH10) as follows: 150 µL of 4.5mM ammonium formate (pH 10): 75 µl of the 4.5 mM ammonium formate (pH10) buffer. The sample was vortexed after each addition. A final addition of 90 µl of Solvent A brought the sample to the injection volume (465 µl). It was centrifuged at 18,000 x g at RT for 5 min and transferred to an autosampler vial.

Preparation of reagents for phosphopeptide enrichment

The following solvents were prepared for phosphopeptide enrichment and purification: 100% (vol/vol) MeOH, 50% (vol/vol) MeCN/0.1% (vol/vol) FA, and 1% (vol/vol) FA. Solvent for binding and washing agarose beads was 80% (vol/vol) MeCN/0.1% (vol/vol) TFA. The stage-tip elution buffer was 50% (vol/vol) MeCN/0.1% (vol/vol) FA. Agarose beads were equilibrated with a solution made by dissolving ferrous (III) chloride to 10mM in HPLC water. This solution was made immediately prior to use. The IMAC agarose-bead elution buffer was prepared by combining 96.3 ml of 1 M monobasic potassium phosphate with 153.75 ml of 1M dibasic potassium phosphate in 250 ml of HPLC water, yielding 500 ml of a 500 mM solution. The agarose-bead slurry solution consisted of MeCN, methanol, and 0.01% (vol/vol) acetic acid in a 1:1:1 ratio by volume.

UPLC solvents

For the EASY NLC system, solvent A was 1% (vol/vol) FA and solvent B was 100% (vol/vol) MeCN/1% (vol/ vol) FA. The nanoElute (Bruker Daltonics) solvent A was 0.1% (vol/vol) FA and solvent B was 100% (vol/vol) MeCN / 0.1% (vol/vol) FA. UPLC solvents were replaced every 2 weeks.

Preparation of standard peptides for basic RP-HPLC.

Purified tryptic peptides from bovine serum albumin were prepared to benchmark the system basic-pH RP chromatography. High purity bovine serum albumin (0.5 g) was dissolved in 100 mM Tris-HCl buffer (pH 8.0) containing 8 M urea. Reduction with dithiothreitol (DTT; 10mM) was performed at room temperature for 30 minutes. Protein alkylation was performed at room temperature with IAM (55 mM) for 45 min in the dark.

After dilution of the protein solution to 2M urea with Tris buffer (100mM, pH8) trypsin was added at a 1:100 (wt/wt) enzyme to protein. After two hours at room temperature an additional aliquot of trypsin was added and digestion proceeded overnight.

Peptides were desalted using the SepPak method as described. The purified peptides were lyophilized then dissolved in MeCN 2% (vol/vol). The peptide concentration was determined and peptides were divided into 500 µg aliquots, lyophilized, and stored at -20° C.

Equipment setup

Basic-pH reversed phase chromatography:

The HPLC system was prepared by purging solvent lines A and B with their respective buffers. The flow rate for equilibration was 1 ml/min. The column was equilibrated with 100% of Solvent B for 20

min, then 100% of Solvent A for 20 min. Flow through the column was reversed for 10 min to remove any particulate accumulation on the inlet frit. The system peptide benchmarking consisted of two gradient methods: a gradient HPLC run without sample injection followed by chromatography of the albumin peptide standard (500 µg). System performance was evaluated by comparison with peak resolution and retention times from previous optimized chromatography studies using standard peptides. The gradient for equilibrating the reversed-phase column at a flow rate of 1 ml/min was as follows (time in min.),%B: 0,0; 4,0; 19,100; 21,100; 22,0; 24,0; 39,100; 41,100; 42,0; 70,0. The gradient method for basic pH reversed-phase chromatography at a flow rate of 1 ml/min was as follows (time in min.),%B: 0.0; 7,0; 13,16; 73,40; 77,44; 82,60; 98,60; 100,0; 120,0.

UPLC-Orbitrap MS

The labeled peptides were analyzed using high-resolution nano-liquid chromatography tandem mass spectrometry (LC-MS). Chromatography was performed with an Acclaim PepMap 1000 C18 RSLC column (75 µm i.d. × 50 cm) (Thermo-Fisher Scientific™) on an EASY *nano*LC (Thermo Fisher Scientific™). The column was equilibrated with 11 µl of solvent A: 1% (vol/vol) formic acid (FA) at 700 bar pressure. The samples in 1% (vol/vol) FA were loaded (2.5 µl) onto the column with 1% (vol/vol) FA at 700 bar. Peptide chromatography was initiated with the following :solvent A 1% (vol/vol) FA and 5% solvent B (100%vol/vol) MeCN, 1% (vol/vol) FA for 5 min at a flow rate of 250 nl/min. The UPLC gradient method for reversed-phase chromatography of peptides is given in the table below.

Time interval (min)	Gradient (%B)	Flow rate (nL/min)
0	5	250
5	5	250
100	23	250
20	35	250
1	95	250
39	95	250

Spectral acquisition was performed in data-dependent acquisition mode. The full-scan mass spectra were acquired with the Orbitrap mass analyzer with a scan range of *m/z* = 350/1500 and a resolving power set to 70,000. Twelve data-dependent high-energy collision dissociations were performed with a mass resolving power set to 35,000, a fixed lower value of *m/z* 100, an isolation width of 1.2 Da, and a normalized collision energy setting of 32. Maximum injection time was 60 ms for parent-ion analysis and 120 ms for product-ion analysis. The target ions that were selected for MS2 were dynamically excluded for 40 sec. Automatic gain control (AGC) was set at a target value of 3e6 ions for MS1 scans and 1e5 ions for MS2 analysis. Peptide ions with charge states of +1 or of +7or great were excluded for HCD acquisition.

A summary of all Instrument parameters for the analysis of the TMT-labeled peptides and phosphopeptides is given in the table below.

Method Parameters **TMT global**	**Value**
Polarity	Positive
Full MS microscans	1
Orbitrap Resolution	70,000
AGC Target	3×10^6 ion counts
Maximum Ion Time	60ms
Scan Range	350-1500
Number of Dependent Scans	12
Isolation Window	1.2 m/z
Fixed First Mass	100.0 m/z
Activation type	HCD
Collision Energy (%)	32
dd-MS^2 Settings	
Microscans	1
Detector type	Orbitrap
Orbitrap Resolution	35,000
AGC	1×10^5
Maximum Ion time	120
Isolation width	1.2 m/z
Fixed first mass	100.0
Activation type	HCD
HCD collision Energy (%)	32
Charge Exclusion	Include 2-6 charge states

Exlude Isotopes	on
Dynamic Exclusion	40.0s

Method Parameters TMT phosphopeptides	**Value**
Polarity	Positive
Full MS microscans	1
Orbitrap Resolution	70,000
AGC Target	1×10^6 ion counts
Maximum Ion Time	60ms
Scan Range	350-1800
Number of Dependent Scans	12
Isolation Window	0.7 m/z
Fixed First Mass	110.0 m/z
Activation type	HCD
Collision Energy (%)	32
dd-MS^2 Settings	
Microscans	1
Detector type	Orbitrap
Orbitrap Resolution	35,000
AGC	1×10^5
Maximum Ion time	105
Isolation width	1.2 m/z
Fixed first mass	100.0
Activation type	HCD
HCD collision Energy (%)	32
Charge Exclusion	Include 2-6 charge states
Exclude Isotopes	on
Dynamic Exclusion	20.0s

UPLC-tims-TOF LC-MS

The unlabeled, unfractionated peptides from individual samples were analyzed using trapped ion mobility time-of-flight (tims-TOF) mass spectrometry.[3] The peptides were separated using a *nano-ELUTE* chromatograph (Bruker Daltonics,Bremen Germany) interfaced to a timsTOF Pro mass spectrometer (Bruker Daltonics) with a modified nano-electrospray source (CaptiveSpray, Bruker Daltonics).

The samples in 2 µl of 1% (vol/vol) FA were injected onto a 75 µm i.d. × 25 cm Aurora Series column with a CSI emitter (Ionopticks). The column temperature was set to 50° C. The column was

equilibrated using constant pressure (800 bar) with 8 column volumes of solvent A (0.1% (vol/vol) FA). Sample loading was performed at constant pressure (800 bar) at a volume of 1 sample pick-up volume plus 2 µl. The peptides were eluted using one column separation mode with a flow rate of 400 nl/min and using solvents A (0.1% vol/vol) FA and B (0.1% vol/vol FA/MeCN) as given below.

Time (min)	Gradient %B	Flow rate (µL/min)
0	2	0.4
60	15	0.4
30	25	0.4
10	35	0.4
10	80	0.4
10	80	0.4

The mass spectrometer was operated in PASEF mode.[3] A summary of the parameters is given below.

MS settings	**Value**
Scan range	100-1700 m/z
Ion Polarity	Positive
Capillary voltage	1700V
Dry Gas	3.0L/min
Dry temperature	200 °C
Scan mode	PASEF
Number of PASEF ramps	10
Total cycle time	1.15 sec
Charge state	0-5
Precursor Repetitions	linear
Target Intensity	20000
Intensity Threshold	500
Active Exclusion after	0.40 min
TIMS settings	
1/K0 range	0.60-1.60 V-s/cm^2
Collision Energy	20.0-63.0 eV
Ramp time	100 ms
Accumulation time	100 ms
Duty Cycle	100%
Ramp Rate	9.52 Hz
MS averaging	1x
Quad isolation	2 Th for m/z < 700 and 3 Th for m/z > 700

The MS1 and MS2 spectra were recorded from *m/z* 100 to 1700.

Suitable precursor ions for PASEF-MS/MS were selected in real time from tims-TOF MS survey scans by a PASEF scheduling algorithm.[3] A polygon filter was applied to the *m/z* and ion mobility plane to select features most likely representing peptide precursors rather than singly charged background ions. The quadrupole isolation width was set to 2 Th for *m*/*z* < 700 and 3 Th for *m*/*z* > 700, and the collision energy was ramped stepwise as a function of increasing ion mobility: 52 eV for 0–19% of the ramp time; 47 eV for 19–38%; 42 eV for 38–57%; 37 eV for 57–76%; and 32 eV for the remainder. The TIMS elution voltage was calibrated linearly using the Agilent ESI-L Tuning Mix (*m/z* 622, 922, 1222). The scan range for MS1 and MS2 spectral acquisition was set *m/z* 100 to 1700.

LC-MS system performance.

The LC-MS systems were assessed with standards after analysis of every eight samples. The number of identified peptides from a standard tryptic digest of a cell line (Hela) was determined routinely. The lyophilized Hela digest (20 µg) was solubilized by vortexing in 200 µl of 1% (vol/vol) FA in water and aliquoted into autosampler vials for storage at -20° C. The LC-MS systems were monitored daily using the retention times and intensities of standard peptides. Pierce Retention Time Calibration mix peptides were purchased as a 5 pmol/µl stock solution. The stock solution was diluted in a 2 ml volumetric flask to 50 fmol/µL in 1% (vol/vol) FA in water.

The number of unique peptides identified was 17-19,000 for the Orbitrap and 38-43,000 for tims-TOF instruments. The EASYnLC interfaced to the Orbitrap mass spectrometer was monitored with PRTC peptide standards. PRTC retention times are plotted over time to monitor LC performance. The resolution was set to 35000 and the collision energy optimized for optimal definition of reporter ions for the TMT-11 plex.

Protein and phosphopeptide identification and quantification.

The machine data from the LC-MS analysis of isobarically-labeled peptides, using the Q-Exactive mass spectrometer, were converted to peak lists using Proteome Discoverer (version 2.1.0.81, ThermoScientific). MS2 spectra with parent ion charge states of +2, +3 and +4 were analyzed using Mascot software[4] (Matrix Science, London, UK; version 2.7.0). MS2 spectra were searched against a RefSeq (downloaded July 2018) database of human proteins (41,734 entries) and common contaminant proteins (cRAP version 1.0 Jan. 1st, 2012; 116 entries). LC-MS data from off-line-fractionated labeled peptides was searched separately by TMT plexes from the IMAC-enriched LC-MS data. The peptide spectral matches from the contaminant database were removed from final results with the exception of human proteins. The database searches were performed with a fragment ion mass tolerance of 0.02 Da and a parent ion tolerance of 20 ppm. Enzyme was specified as trypsin/P with allowance of a maximum of 4 missed cleavages. A fixed modification of carbamidomethylation of cysteine residues was specified. The following were selected as variable modifications: deamidation of asparagine, deamidation of glutamine, formation of pyro-glutamic acid from N-terminal glutamine, acetylation of protein N-terminal residues, oxidation of methionine, and pyro-carbamidomethylation of N-terminal cysteine. For the IMAC-enriched samples, additional variable modifications of phosphate moieties at serine, threonine and tyrosine residues was specified. Peptide spectral matches (PSM) were filtered at 1% false-discovery rate (FDR) by searching against a reversed database and the ascribed peptide and protein identities were accepted. A minimum of two peptides with unique sequences, not resulting from missed cleavages, was required to accept a protein identification. Only phosphopeptides with a high-probability phosphosite localization normalized Mascot Delta Score[5] of >0.5 were retained for downstream analysis. When multiple phosphopeptides represented the same phosphosite, abundances were averaged to estimate relative phosphosite abundance prior to further analysis.

Label-free “single-shot” LC-MS data from the timsTOF mass spectrometer were converted to peak lists using DataAnalysis (version 5.2, Bruker Daltonics). Spectra were analyzed using MaxQuant software [6] (version 1.6.17). MS2 spectra were search against a RefSeq (ver July 2018) database of human proteins (41,734 entries) and common contaminant proteins (cRAP, version 1.0 Jan. 1st, 2012; 116 entries). Entries of human proteins that were matched to both RefSeq and cRAP were assigned to RefSeq. The searches were performed with a fragment ion mass tolerance of 40 ppm and a parent ion tolerance of 20 ppm. The enzyme search specificity was selected as tryptic/P, allowing for a maximum of 4 missed cleavages. Carbamidomethylation of cysteine residues was specified as a fixed modification. The searches were conducted with the following variable modifications: deamidation of asparagine, deamidation of glutamine, formation of pyro-glutamic acid from N-terminal glutamine, acetylation of protein N-terminus, oxidation of methionine, and pyro-carbamidomethylation of N-terminal cysteine residues. Peptides and protein results were filtered at 1% false-discovery rate (FDR) by searching against a reversed protein sequence database. A minimum of two peptides with unique sequences, not resulting from missed cleavages, was required for identification of a protein. MS1 precursor intensities were used for relative protein quantification and were normalized relative to the average intensity from all samples.

Stastical analysis, normalization and quality assessment of multiplex tandem-mass-tag data

The processing, quality assurance and analysis of isobaric-labeled peptide LC-MS data were performed with proteoQ (version 1.5.0.0, https://github.com/qzhang503/proteoQ), software developed with the tidyverse approach[7] (tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.3.1. https://CRAN.R-project.org/package=tidyverse) with open source software for statistical computing and graphics, R (https://www.R-project.org/) and RStudio (http://www.rstudio.com/). The reporter-ion intensities from each 11-plex or tandem mass tag (TMT) *m/z* values (channels) were converted to logarithmic ratios (base 2), relative to the average reporter-ion intensity of reference

samples within each 11-plex. Within each sample, Dixon's outlier removals were carried out recursively for peptides with greater than two identifying PSM's. The median of the ratios of PSM that could be assigned to the same peptide was first taken to represent the ratios of the incumbent peptide. The median of the ratios of peptides was then taken to represent the ratios of the inferred protein. To align protein ratios across samples, likelihood functions were first estimated for the log-ratios of proteins using finite mixture modeling, assuming two-component Gaussian mixtures.[8] The ratio distributions were then aligned so that the maximum likelihood of log-ratios was centered at zero for each sample. Scaling normalization was performed to standardize the log-ratios of proteins across all samples. To reduce the influence of outliers from either log-ratios or reporter-ion intensities, the values between the 5th and 95th percentile of log-ratios and 5th and 95th percentile of intensity were used in the calculations of standard deviations.

Western Blotting

Lysates were made using 1x NuPAGE LDS Sample Buffer (Invitrogen). The following antibodies were used on the Jess Protein Simple Western Blotting system: anti-FLAG (Abcam ab1162), anti-KDM4A (Abcam ab191433), anti-KDM4B (Cell Signaling #8639), anti-KDM4C (Abcam ab226480), anti-KPNB1 (Abcam ab45938), anti-KPNA3 (Thermo Scientific PA5117127), HRP-conjugated anti-mouse (ProteinSimple), and HRP-conjugated anti-Rabbit (ProteinSimple).

TurboID experiments

C57BL6/J bone marrow was harvested using standard techniques and grown overnight in 4-factor cytokine media (RPMI-1640, 10% FBS, penicillin-streptomycin, 100 ng/ml kit ligand, 10 ng/ml TPO, 50 ng/ml FLT3L, 6 ng/ml IL3). GP2 293T cells (Takara Bio) were transduced with MSCV-IRES-GFP plasmids containing TurboID cDNA alone or fused to the indicated genes using TrasnIT-LT1 (Mirus Bio). The following day lineage depletion was performed using the Miltenyi Lineage Cell Depletion Kit,

and lineage-negative cells were incubated overnight at 37 C. Retrovirus was collected two and three days after transduction and concentrated onto wells of a 6 well plates coated with Retronectin (Takara Bio, 5 ug/ml) by spinning at 2500g for 90 minutes at 32° C. Supernatant was aspirated from well, and lineage-depleted murine bone marrow cells were added to the wells and spun at 280g for 7 minutes at 32° C to infect. Infection was repeated the following day. Two days after the second infection, transduced cells were enriched by sorting for GFP and allowed to recover in culture for 2 days. 1-5 million cells were resuspended at 10^6 cells/ml and incubated with 50 uM biotin (Millipore-Sigma) for 4 hours at 37° C. Cells were harvested, lysed in 1500 ul lysis buffer (25mM Tris-HCl, 150mM NaCl, 1% TritonX-100, pH 7.2) with protease inhibitor cocktail (Millipore-Sigma), and sonicated. Lysate was cleared by spinning 16,000g x 10 minutes at 4° C. High capacity Streptavidin Resin (ThermoFisher) was washed with lysis buffer 500g x 1 minute (80 ul per sample) and aspirated, and 1500 ul cleared cell lysate was added to washed beads. Lysate and beads were incubated at 4° C overnight on a rotator. Beads were pelleted at 500g x 1 minute, supernatant was aspirated, and beads were washed with 1 ml 1% SDS in PBS. Wash was repeated with lysis buffer twice, followed by 1 wash in 50mM Na2HPO4, 500mM NaCl, 1% TritonX100, pH 7.4. Supernatant was aspirated and beads were submitted for mass spectrometry. Samples for Western blot analysis were taken from lysate prior to bead addition.

TurboID mass spectrometry and data analysis

The peptides were prepared using a previously described method for on-bead tryptic digestion.[9] The beads were washed four times with 1 mL of 50 mM ammonium bicarbonate buffer (pH = 8.0) (ABC). The washed beads were resuspended in 40 µL of ABC buffer containing 8 M urea. The proteins were reduced by the addition of tris(2-carboxyethyl)phosphine (TCEP; 2 µL of 0.5 M) and incubation for 60 min at 30 ºC. The reduced proteins were alkylated using iodoacetamide (4 µL of 0.5 M) and incubation for 30 min at room temperature in the dark. The urea was diluted to 1.5 M by adding 167

µL of 50 mM ABC buffer prior to the addition of LysC (1mAU). Samples were digested for 2 h at 30 °C in a Thermomixer with gyration at 750 rpm. Trypsin (1 µg) was added, and the samples were incubated overnight at 30 °C in the Thermomixer gyrating at 750 rpm. The peptides were then transferred to a new tube, the bead samples were washed with an additional 50 µL of ABC buffer and the wash was combined with the peptides. Residual detergent was removed by ethyl acetate extraction.[10] In preparation for desalting, peptides were acidified to pH=2 with 1% TFA final concentration. The peptides were desalted using two micro-tips (porous graphite carbon, BIOMETNT3CAR) (Glygen) on a Beckman robot (Biomek NX), as previously described.[11] The peptides were eluted with 60% MeCN in 0.1% TFA and dried in a Speed-Vac (Thermo Scientific, Model No. Savant DNA 120 concentrator) after adding TFA to 5%. The peptides were dissolved in 20 µl of 1% MeCN in water. An aliquot (10%) was removed for quantification using the Pierce Quantitative Fluorometric Peptide Assay kit (Thermo Scientific, Cat. No. 23290). The remaining peptides were transferred to autosampler vials (Sun-Sri, Cat. No. 200046), dried and stored at -80ºC for LC-MS analysis.

The LC-MS and the raw MS data was processed as described above. The proteins were then quantified with a minimum of two unique peptides, 1% peptide FDR, and 95% protein threshold. Spectral counts were used to quantify protein abundance. The data was subsequently processed using R. All the NA values were first replaced with zero and 0.01 was added. Decoy peptides were then removed. Each sample was then normalized to its total spectral counts, and the data was log-transformed to approximate normality. A filter of a minimum of five spectral counts across either all wildtype or mutant NPM1 samples was applied. Fold changes were calculated with the mean values, and classic t-tests and Benjamini-Hochberg p value adjustment were performed. Significant protein labelling enrichment was called if the protein had a larger than 1.5-fold increase and an adjusted p value less than 0.05. Comparison between both the mutant against wildtype NPM1 as well as mutant

against empty vector TurboID dataset was made, and the overlap proteins were selected. A heatmap of the overlapping enriched proteins was then made with per-row Z-score using the package ComplexHeatmap.[12]

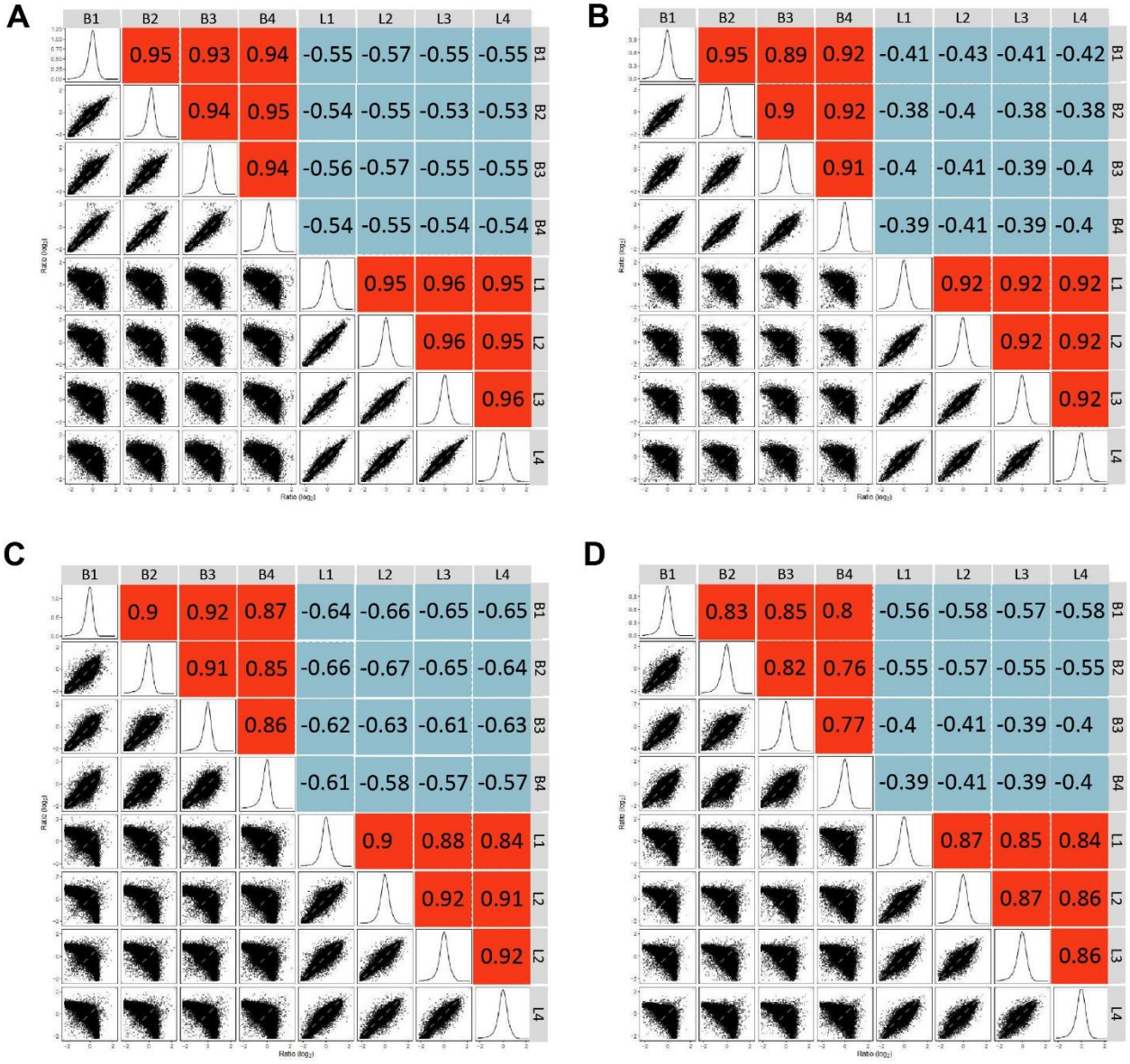
Immunofluorescence

Cytospins were performed using 150,000 sorted, GFP+ cells. Cells were spun on to pre-wet UltraFrost or UltraStick slides (ThermoFisher), 350 rpm x 7 minutes. Cells were ringed using a Pap pen and fixed in 4% paraformaldehyde (PFA) for 10 minutes at room temperature. Fixed cells were wash three times in PBS, then blocked and permeabilized 30 minutes at room temperature in 1% BSA, 0.1% saponin, and 10% goat serum in PBS. Cells were washed with PBS and incubated with anti-BirA antibody (Novus NBP2-59939) at 1:1000 in 1% BSA and 0.1% saponin in PBS overnight at 4 C in a humidified chamber. Slides were washed 3 times in PBS and incubated in 1:500 goat anti-mouse Alexa 633 (Invitrogen) and 1 ug/ml DAPI in 1% BSA and 0.1% saponin in PBS for 45 minutes at room temperature. Slides were washed 3 times with PBS, mounted using Prolong Glass Antifade, and imaged using confocal microscopy and Zen imaging software(Zeiss).

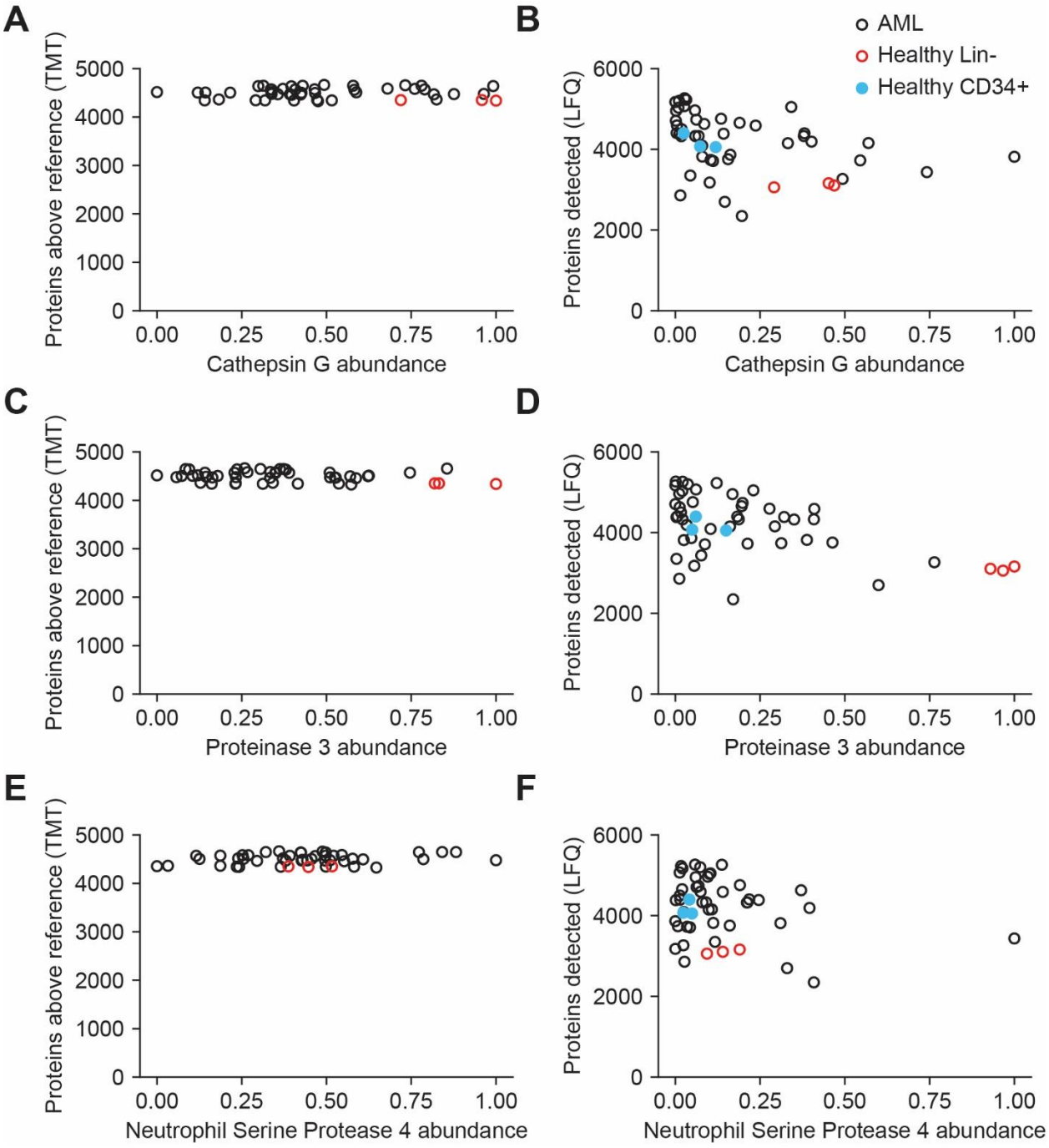
Supplemental References

1. Mertins P, Tang LC, Krug K, et al. Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography-mass spectrometry. *Nat Protoc*. 2018;13(7):1632-1661.
2. Cancer Genome Atlas Research N, Ley TJ, Miller C, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 2013;368(22):2059-2074.
3. Meier F, Brunner AD, Koch S, et al. Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol Cell Proteomics*. 2018;17(12):2534-2545.
4. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 1999;20(18):3551-3567.
5. Savitski MM, Lemeer S, Boesche M, et al. Confident phosphorylation site localization using the Mascot Delta Score. *Mol Cell Proteomics*. 2011;10(2):M110 003830.
6. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008;26(12):1367-1372.

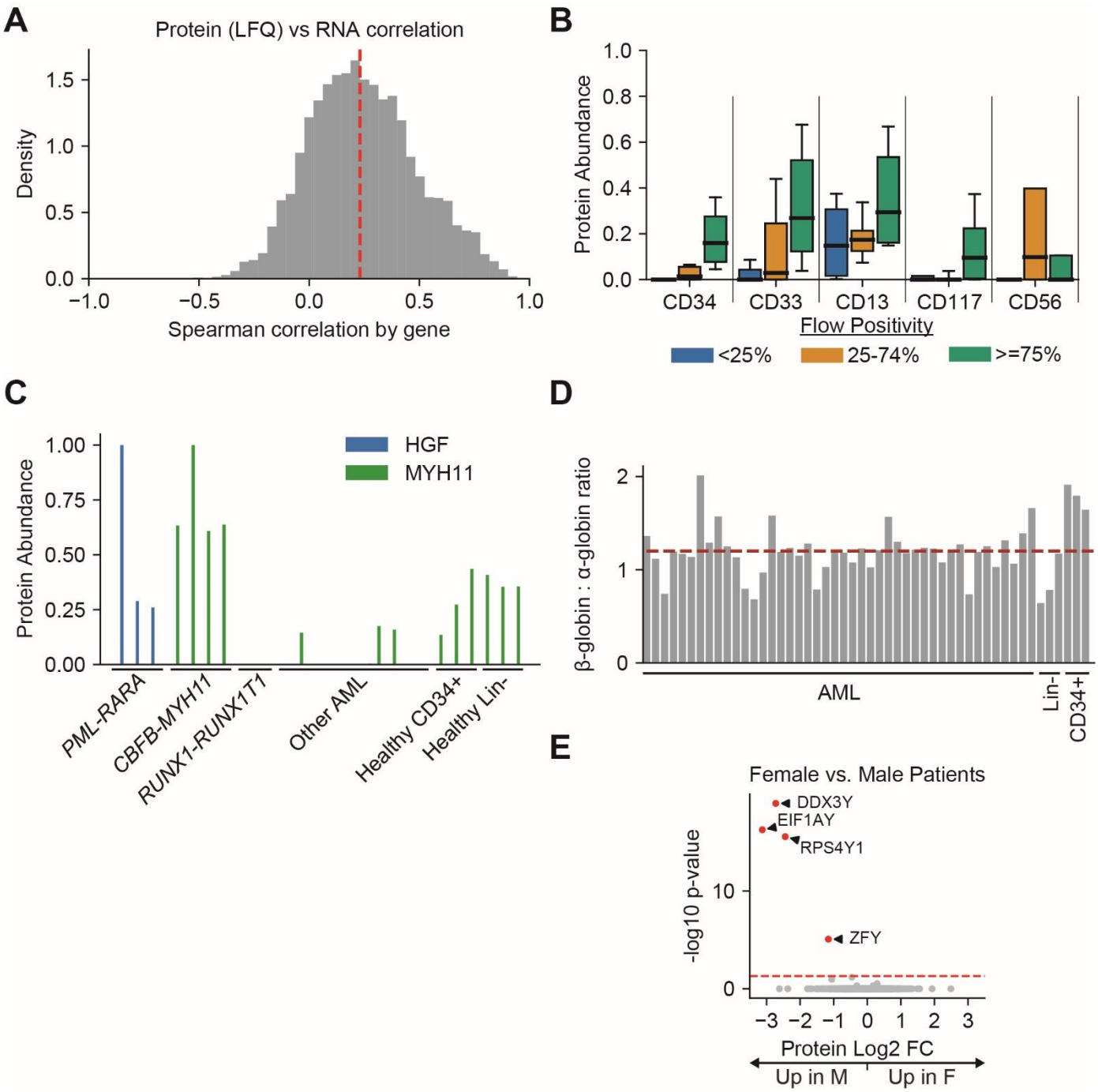
7. Wickham H. Advanced R (ed Second edition.). Boca Raton: CRC Press/Taylor & Francis Group; 2019.
8. Benaglia T, Chauveau D, Hunter DR, Young DS. mixtools: An R Package for Analyzing Finite Mixture Models. *Journal of Statistical Software*. 2009;32(6):1-29.
9. Kim DI, Jensen SC, Noble KA, et al. An improved smaller biotin ligase for BioID proximity labeling. *Mol Biol Cell*. 2016;27(8):1188-1196.
10. Erde J, Loo RR, Loo JA. Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J Proteome Res*. 2014;13(4):1885-1895.
11. Chen ZW, Fuchs K, Sieghart W, Townsend RR, Evers AS. Deep amino acid sequencing of native brain GABAA receptors using high-resolution mass spectrometry. *Mol Cell Proteomics*. 2012;11(1):M111 011445.
12. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32(18):2847-2849.
13. Abbas HA, Mohanty V, Wang R, et al. Decoupling Lineage-Associated Genes in Acute Myeloid Leukemia Reveals Inflammatory and Metabolic Signatures Associated With Outcomes. *Front Oncol*. 2021;11:705627.



Supplementary Figure 1. Correlation of deep-scale proteomic process replicates for human peptides and human phosphopeptides. Pairwise Pearson correlations for peptide (**A,C**) and phosphopeptide (**B,D**) relative abundance in deep-scale analysis of standard basal (B1-B4) and luminal (L1-L4) breast cancer PDX models for this study (**A,B**) and published from the Proteome Characterization Center 1[1] (PCC1; **C,D**). Using the same standard reference tumor tissue, a high degree of reproducibility was achieved with the deep scale proteomic protocol used in this study. For the same PDX model, the Pearson correlations of relative abundances of peptide and phosphopeptides for repeat performance of the protocol were 0.95 (min 0.93, max 0.96) and 0.92 (min 0.89, max 0.95), respectively. Using the reported unprocessed LC-MS data[1], we determined correlations for peptides and phosphopeptides from PCC1 for the same tissue (**C,D**). As expected, relative abundance between the different breast cancer tumor types were minimally correlated, as shown.

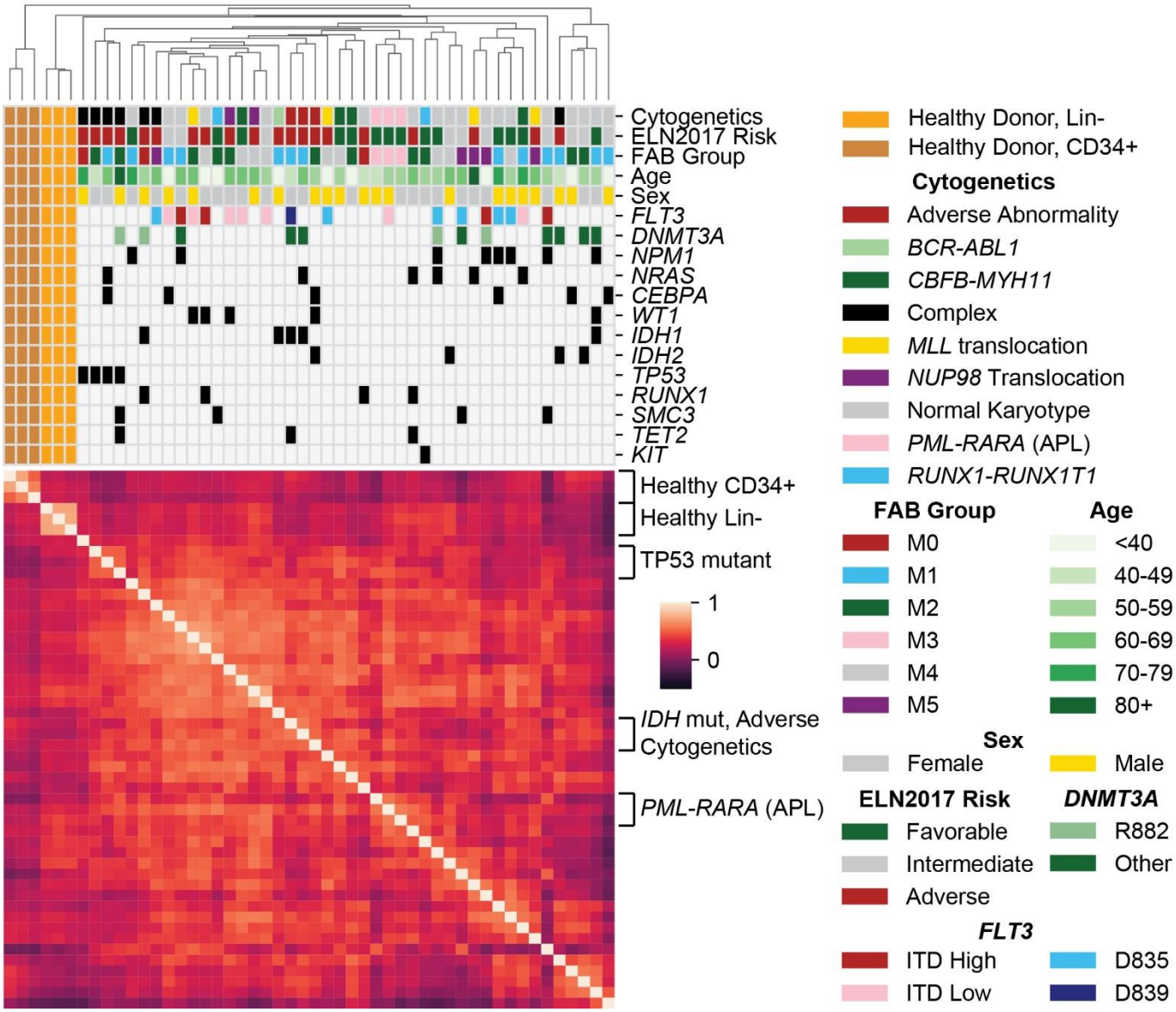


Supplemental Figure 2. Endogenous myeloid-serine proteases do not exert a major effect on detection of proteins. (**A, C, E**) Number of proteins detected at above-average abundance (>reference pool) for each patient or healthy donor sample, compared to the normalized protein abundance of the specified serine protease in each TMT sample (**B, D, F**) Number of proteins detected for each sample, compared to the normalized protein abundance of the specified serine protease, using label-free-quantification (LFQ).

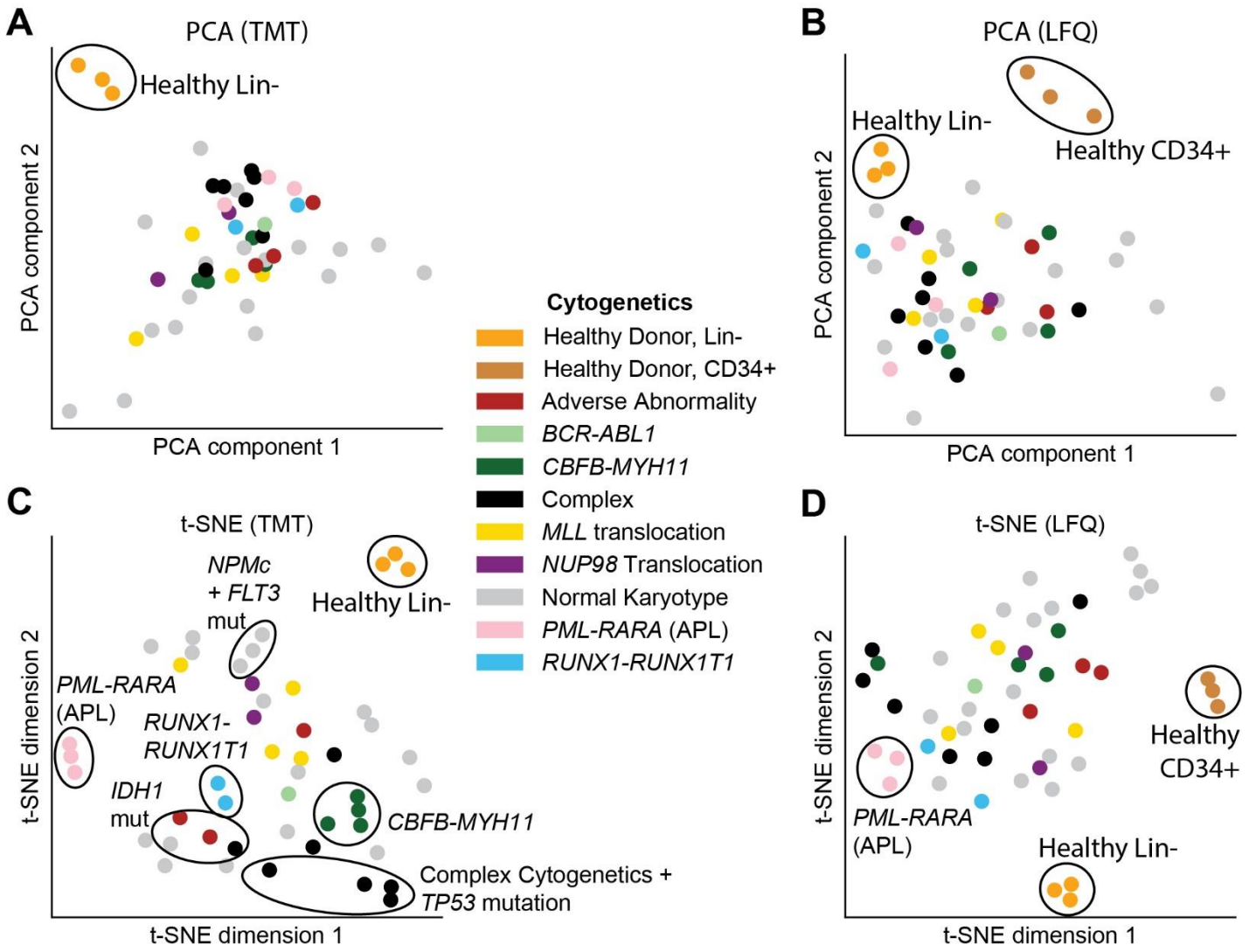


Supplemental Figure 3. Data quality measures. (**A**) Distribution of gene-wise Spearman correlation between proteomic (LFQ) and bulk RNA-seq data across AML samples. Only genes quantifiable by both technologies in at least 20% of AML samples were included in the analysis. Dashed red line represents the median value. (**B**) Measured protein abundance (LFQ) for each of the identified cell-surface proteins in patients grouped by the percentage of cells from their presentation marrow sample that displayed the same respective marker using clinical flow cytometry. Protein abundance is calculated based on MS1 precursor intensities normalized relative to the average intensity from all samples. Each protein expression value is then scaled to have a maximum value across all measured samples of 1, and a minimum value of 0, for this display. (**C**) Measured protein abundance (LFQ) for each of two proteins known to be expressed in only one AML subtype. Shown are 3 AMLs with *PML-RARA* fusions, 4 with *CBFB-MYH11* fusions, 2 with *RUNX1-RUNX1T1* fusions, 11 other

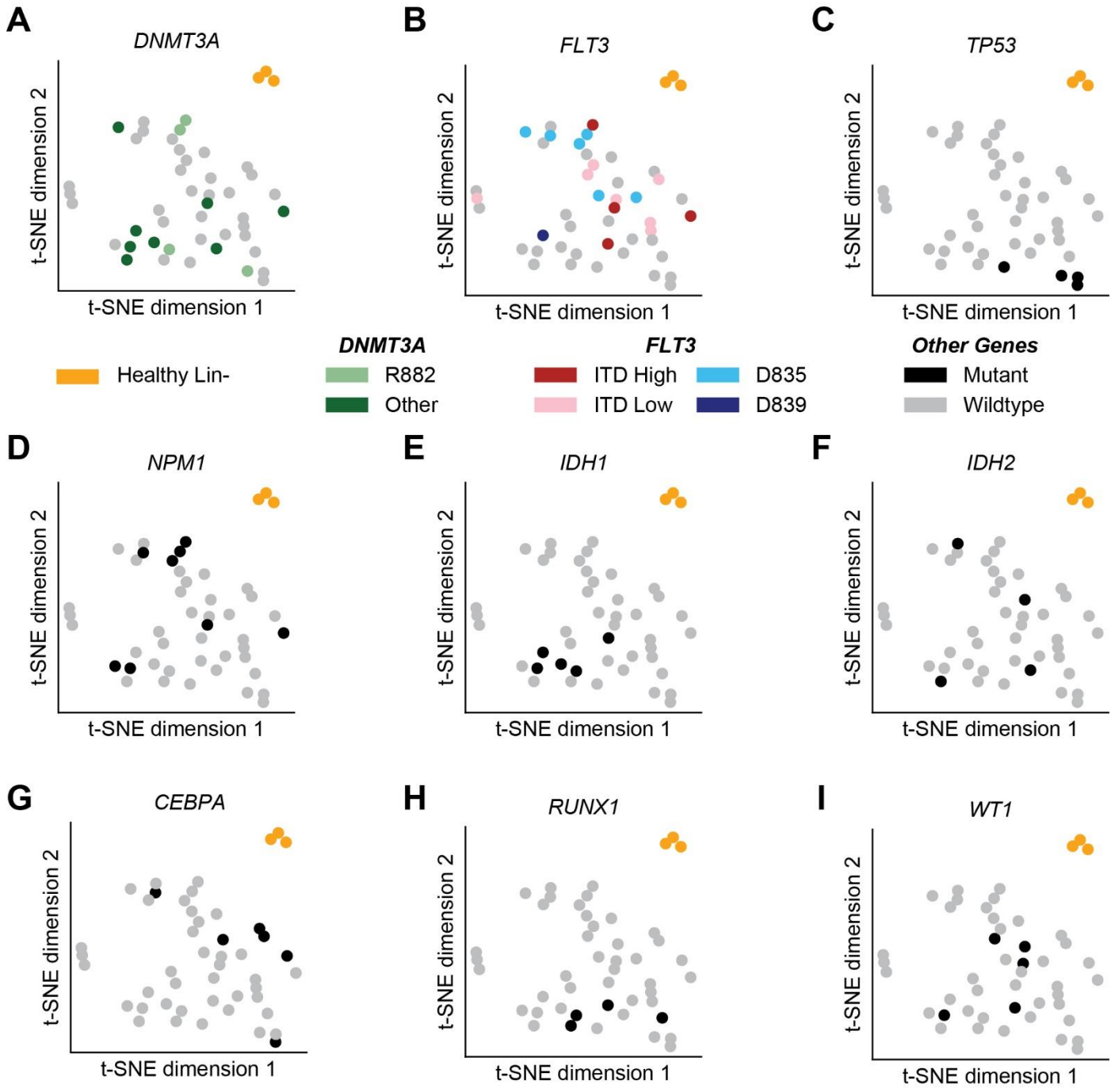
representative AML samples, 3 healthy adult donor bone marrow samples which were CD34 selected (CD34+) and 3 healthy adult donor bone marrow samples which were lineage depleted (Lin-). HGF is overexpressed only in *PML-RARA* initiated AML and MYH11 is overexpressed only in *CBFB-MYH11* initiated AML. RARA and RUNX1T1, which are displayed in Figure 1G based on the TMT method, were not detected using LFQ. (**D**) Ratio of beta globin (HBB) to alpha globin proteins (HBA1+2) in LFQ data for all AML samples, 3 healthy adult donor marrows after lineage depletion (Lin-) and 3 samples of CD34 purified bone marrow cells from healthy donors (CD34+). Note that the average ratio is 1.2, reflecting the stoichiometric abundance of these two proteins in adult hemoglobin. (**E**) Volcano plot showing differential expression in TMT data of female vs. male AML patients. All significantly different proteins are Y-linked. Note that in TMT mass spectrometry, presence of protein in any sample leads to recording of low-level background in all samples, so complete absence of Y-linked proteins in female patients is not measured.



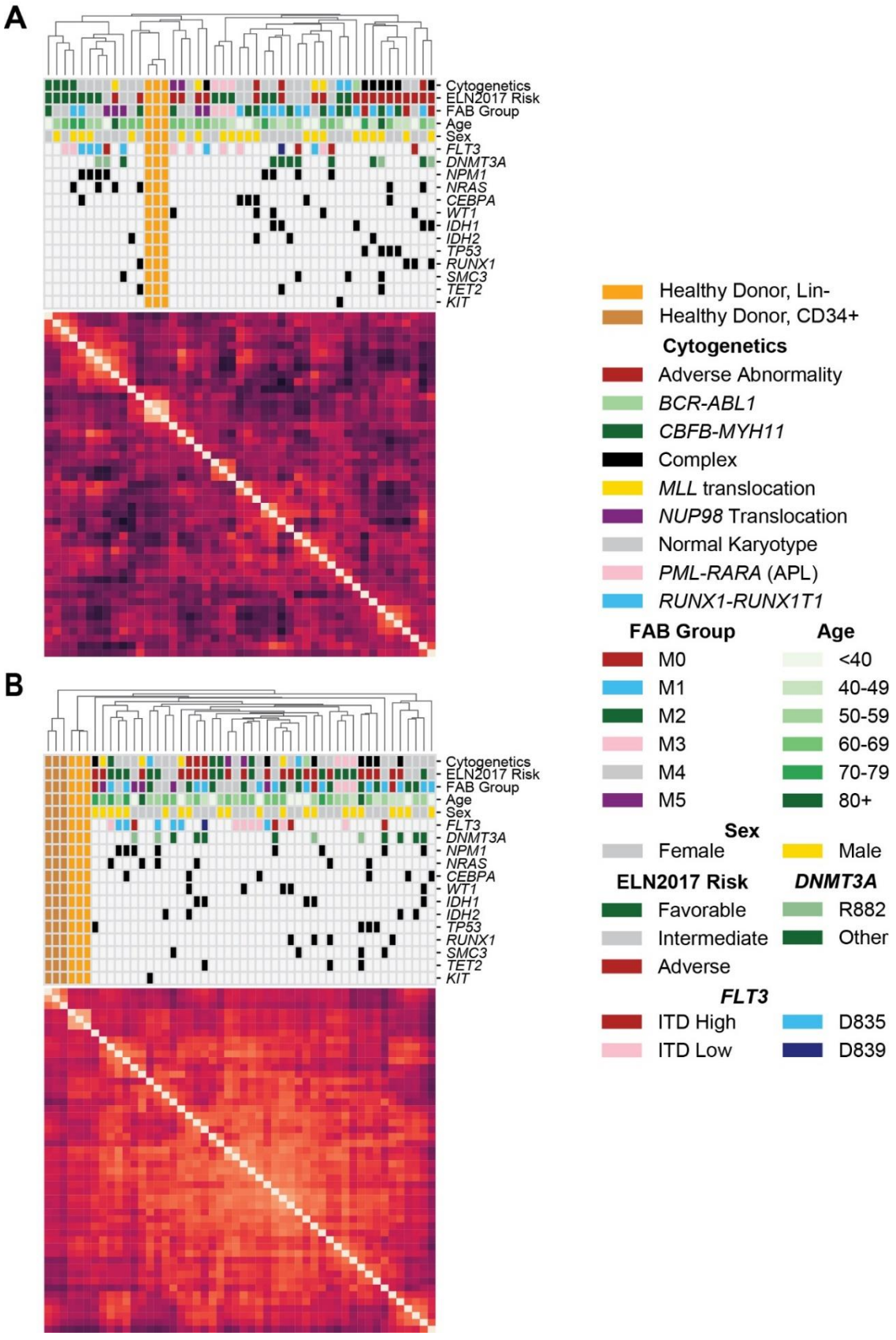
Supplemental Figure 4. Unsupervised clustering of proteomic profiles from LFQ data. Unsupervised clustering of proteomic profiles from LFQ data, revealing distinct clusters of samples, some of which correlate with known molecular covariates, including cytogenetic alterations, FAB subgroups, and recurrent mutations. The heatmap shows a Pearson correlation of protein expression levels among all patients, using LFQ proteomic abundance measurements which were linearly scaled to values between 0 and 1 prior to calculation of Pearson correlation. Clustering was based on the UPGMA algorithm, with similarity scores as shown in the heatmap.



Supplemental Figure 5. Principal Component Analysis (PCA) and t-SNE analysis of TMT and LFQ data. (**A,B**) AML patients and healthy donor bone marrows plotted by first and second principal components based on protein abundance as measured using (**A**) TMT and (**B**) LFQ platforms. (**C,D**) t-SNE plots from protein abundance of AML patients and healthy donor bone marrows as measured using (**C**) TMT and (**D**) LFQ platforms. Each dot represents one patient or healthy donor sample and each sample is colored based on cytogenetics as shown in the legend. Only proteins without missing values were utilized for both PCA and t-SNE.

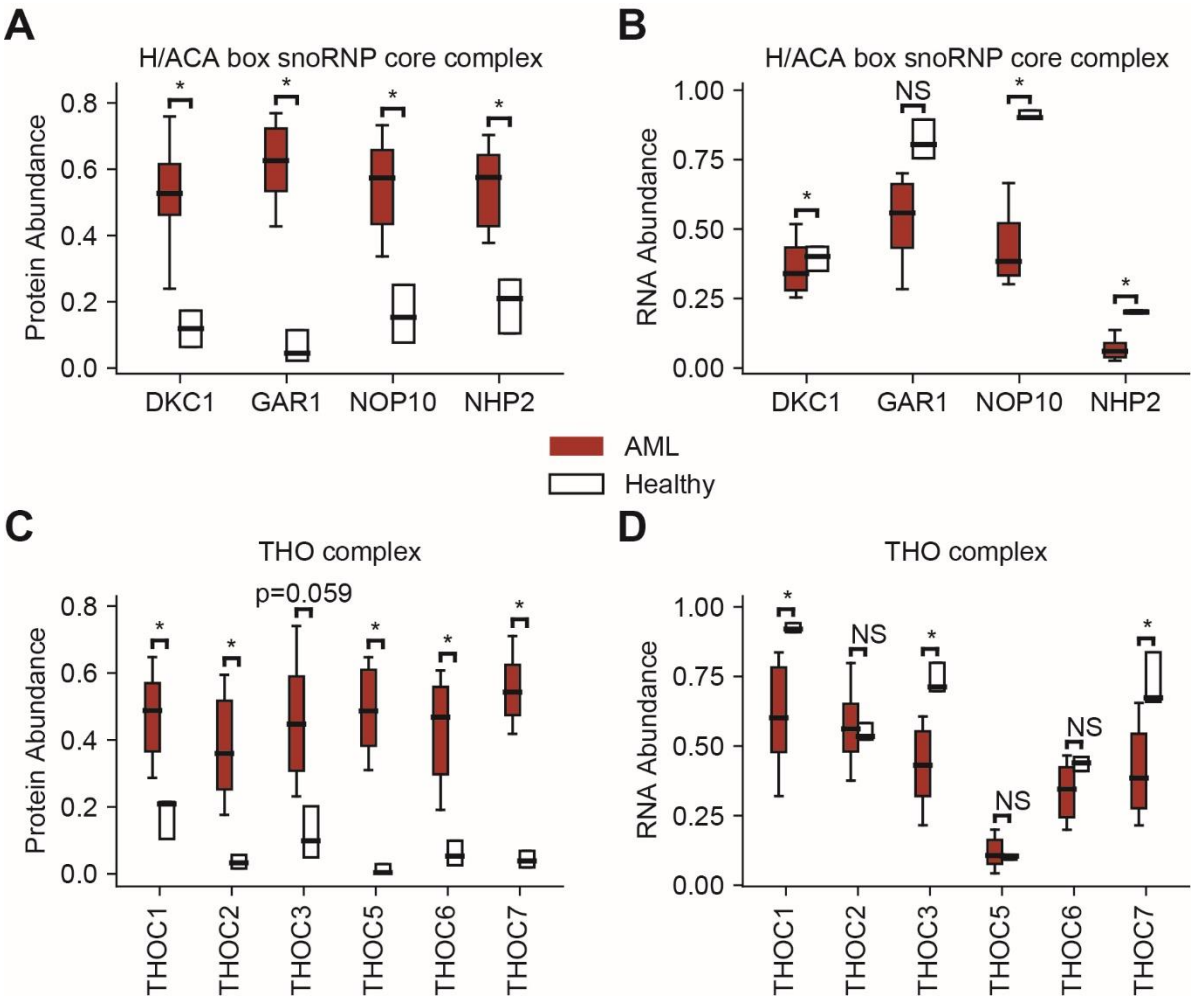


Supplemental Figure 6. Common AML mutations supervise clusters of patients in t-SNE analysis. (**A-I**) t-SNE plot from protein abundance of AML patients and healthy donor bone marrows as measured using the TMT platforms. Each plot has an identical layout of patients based on t-SNE analysis. Plots differ in the coloring of samples, as patient samples are colored as indicated by the mutation status of the gene in the plot title.

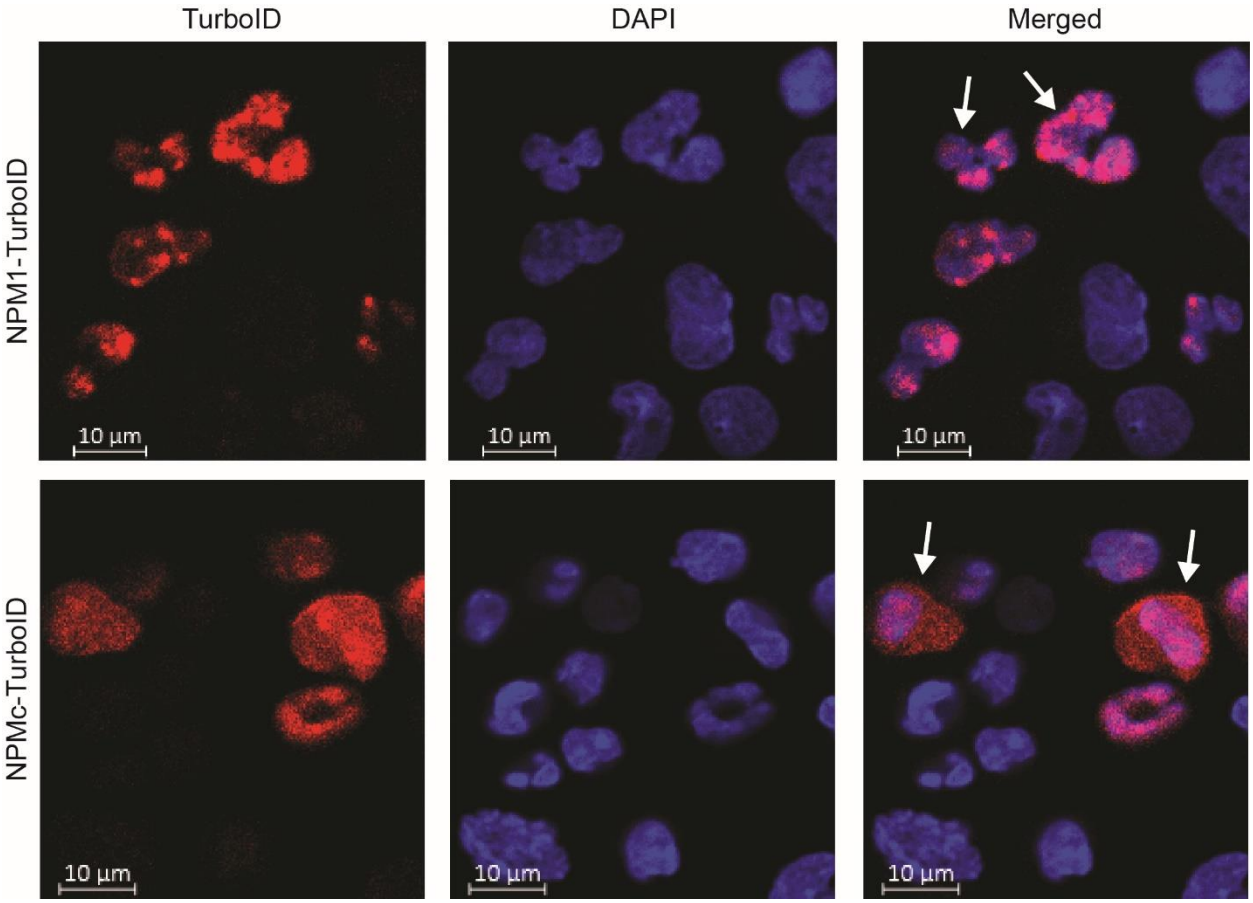


Supplemental Figure 7. Unsupervised hierarchical clustering of protein abundance profiles with lineage-associated genes removed. Unsupervised clustering of proteomic profiles from (**A**) TMT and (**B**) LFQ data, revealing distinct clusters of samples, some of which correlate with known

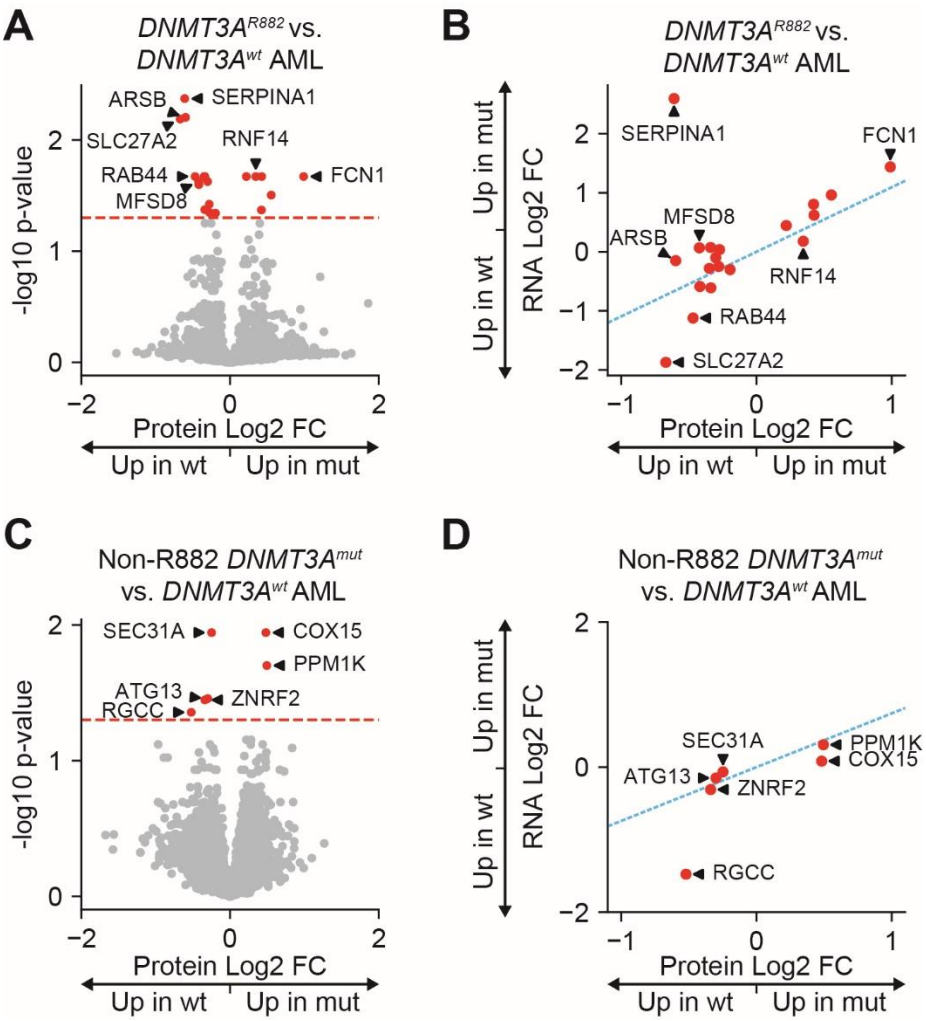
molecular covariates, including cytogenetic alterations, FAB subgroups, and recurrent mutations. The heatmap shows a Pearson correlation of protein expression levels among all patients. For LFQ, proteomic abundance measurements which were linearly scaled to values between 0 and 1 prior to calculation of Pearson correlation. Clustering was based on the UPGMA algorithm, with similarity scores as shown in the heatmap. FAB lineage-associated proteins were removed from this analysis as previously published.[13]



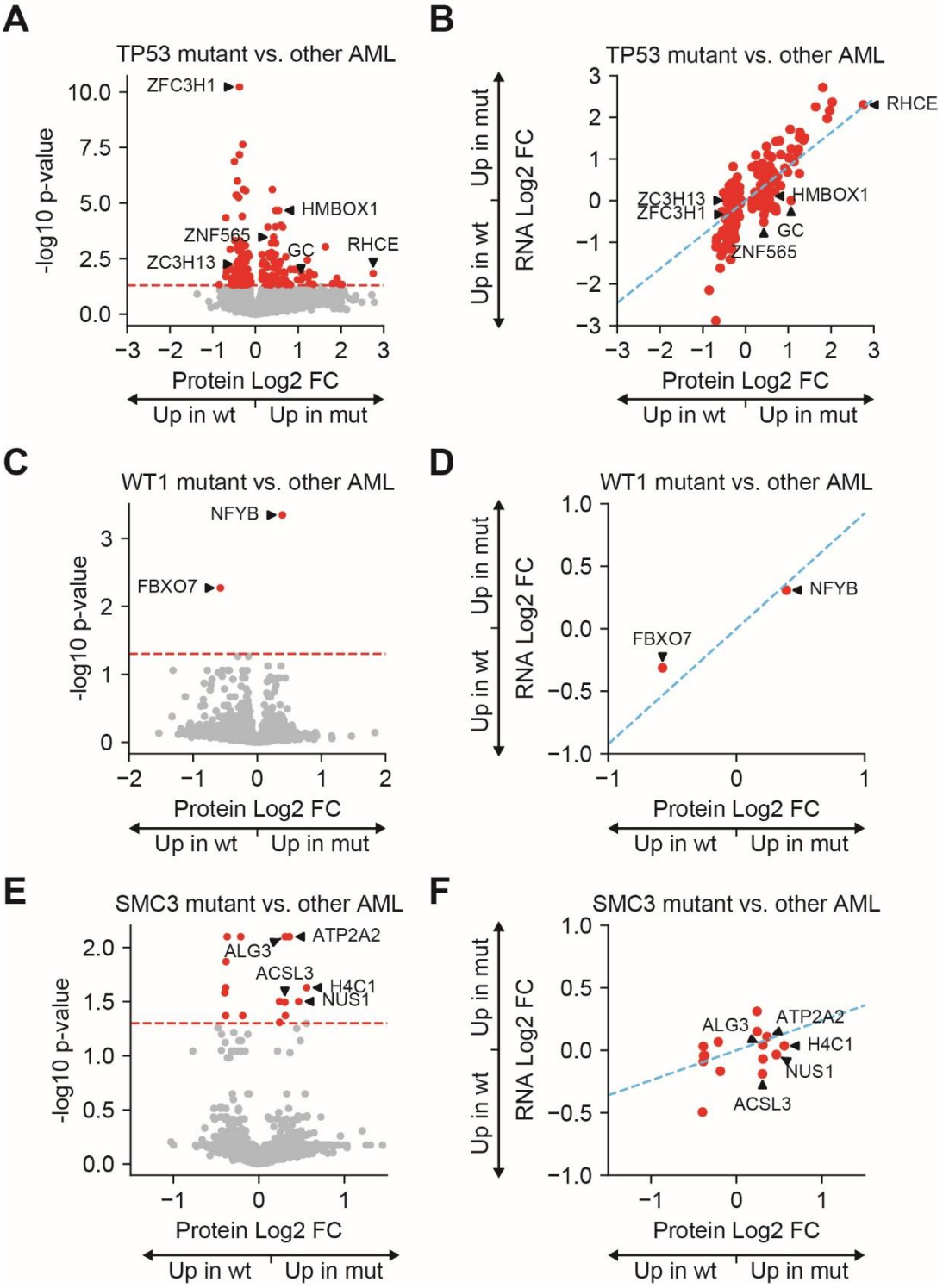
Supplemental Figure 8. Post-transcriptional regulation of the H/ACA box snoRNP core complex and THO complex in AML samples globally. (**A,B**) Normalized abundance of the components of the H/ACA box snoRNP core complex in AML and lineage-depleted healthy control bone marrow in TMT protein abundance data (**A**) and bulk RNA-seq (**B**). Asterisks (*) indicate significant differences ($p<0.05$) by Mann-Whitney U test with multiple-hypothesis correction with Benjamini-Hochberg method across all detected proteins or RNA in our dataset. (**C,D**) Normalized abundance of the components of the Tho complex in AML and lineage-depleted healthy control bone marrow in TMT protein abundance data (**C**) and bulk RNA-seq (**D**). Asterisks (*) as above.



Supplemental Figure 9. Immunofluorescence of cells transduced with TurboID constructs. Mouse bone marrow was harvested and lineage-depleted. These enriched progenitor cells were transduced with an MSCV-IRES-GFP construct expressing TurboID cDNA fused to wildtype NPM1 (top panels) or mutant NPMc protein (bottom panels). Flow sorted GFP+ cells were stained with an antibody specific for the TurboID moiety of the fusion proteins and immunofluorescent imaging was performed. TurboID was detected using anti-BirA antibody to biotin ligase and a secondary Alexa633 (red) antibody. The DAPI stain identifies the nucleus. The NPM1-TurboID protein shows localization in nuclear bodies that correspond to the nucleolus, while protein from NPMc-TurboID construct shows more diffuse localization in the nucleus, and a cytoplasmic blush outside the nucleus. Representative cells are designated with white arrows.

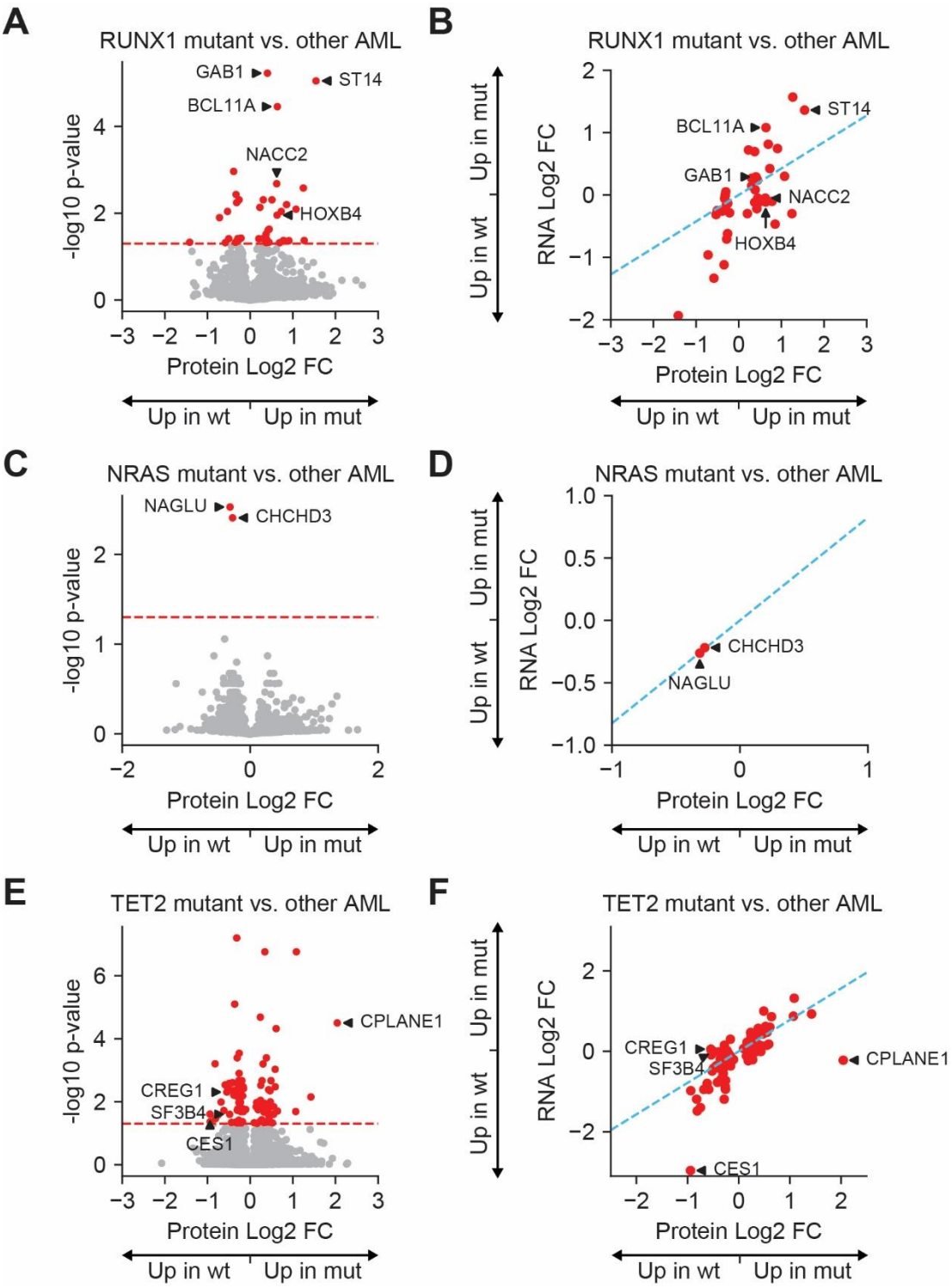


Supplemental Figure 10. Differentially abundant proteins for DNMT3A mutant samples. (**A,C**) Volcano plot showing mean fold change of protein abundance in *DNMT3A*^{R882} mutant (n=4) (**A**) or non-R882 *DNMT3A* mutant (n=8) (**C**) vs. *DNMT3A* wildtype (n=32) AML samples. Red dots represent significantly different proteins ($p < 0.05$ after multiple-hypothesis correction). P-values are calculated using the t-test and corrected for multiple hypothesis testing with Benjamini-Hochberg method. Dashed red line shows $p = 0.05$. (**B, D**) Protein fold change vs. RNA fold change for *DNMT3A*^{R882} mutated (**B**) or non-R882 *DNMT3A* mutant (**D**) vs. *DNMT3A* wildtype AML samples. Dashed blue line shows a line of best fit for the relationship between protein and RNA fold change for all proteins. Only points with a significantly different protein abundance between the specified mutant and wildtype samples are shown. Increased distance from this line of best fit suggests a higher probability of post-transcriptional regulation of protein abundance.



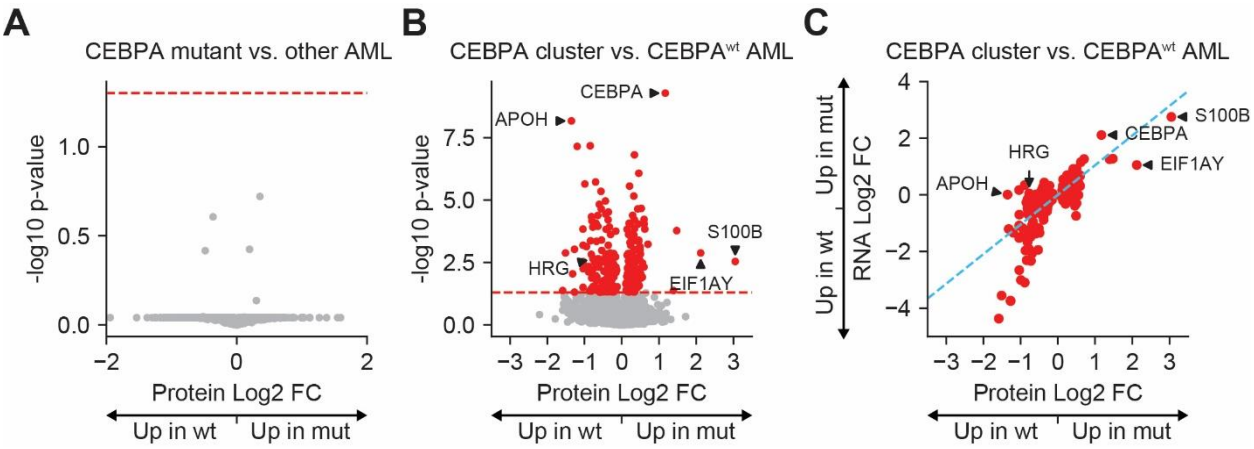
Supplemental Figure 11. Differentially abundant proteins for *TP53*, *WT1* and *SMC3* mutant samples. (**A,C,E**) Volcano plot showing mean fold change of protein abundance in *TP53*^{mut} (n=4) (**A**), *WT1*^{mut} (n=5) (**C**) or *SMC3*^{mut} (n=4) (**E**) vs. wildtype AML samples (for the specified gene). Red dots represent significantly different proteins ($p < 0.05$ after multiple-hypothesis correction). P-values are calculated using the t-test and corrected for multiple hypothesis testing with Benjamini-Hochberg method. Dashed red line shows $p = 0.05$. (**B, D, F**) Protein fold change vs. RNA fold change for *TP53*^{mut} (**B**), *WT1*^{mut} (**D**) or *SMC3*^{mut} (**F**) vs. wildtype AML samples (for the specified gene). Dashed blue line shows a line of best fit for the relationship between protein and RNA fold change for all

proteins. Only points with a significantly different protein abundance between the specified mutant and wildtype samples are shown. Increased distance from this line of best fit suggests a higher probability of post-transcriptional regulation of protein abundance.

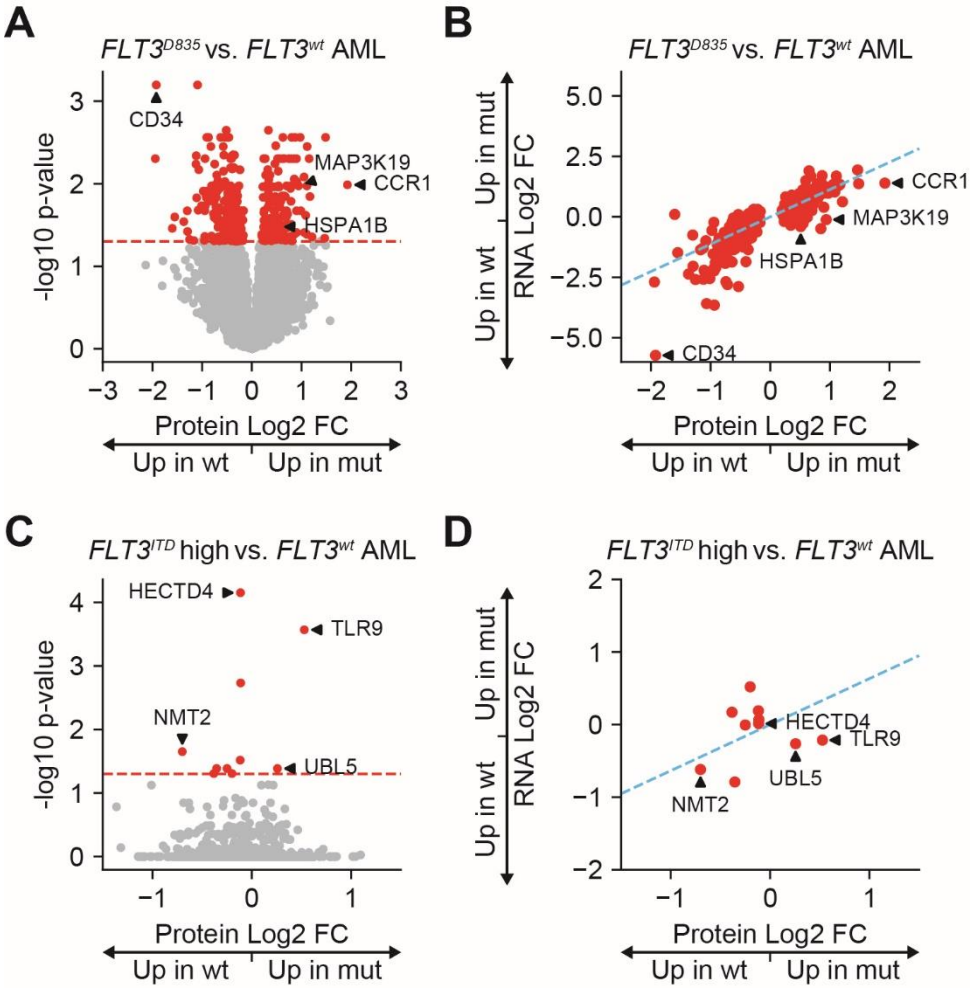


Supplemental Figure 12. Differentially abundant proteins for *RUNX1*, *NRAS* and *TET2* mutant samples. (**A,C,E**) Volcano plot showing mean fold change of protein abundance in *RUNX1*^{mut} (n=4) (**A**), *NRAS*^{mut} (n=6) (**C**) or *TET2*^{mut} (n=3) (**E**) vs. wildtype AML samples (for the specified gene). Red dots represent significantly different proteins ($p < 0.05$ after multiple-hypothesis correction). P-values are calculated using the t-test and corrected for multiple hypothesis testing with Benjamini-Hochberg method. Dashed red line shows $p = 0.05$. (**B, D, F**) Protein fold change vs. RNA fold change for *RUNX1*^{mut} (**B**), *NRAS*^{mut} (**D**) or *TET2*^{mut} (**F**) vs. wildtype AML samples (for the specified gene). Dashed blue line shows a line of best fit for the relationship between protein and RNA fold change for

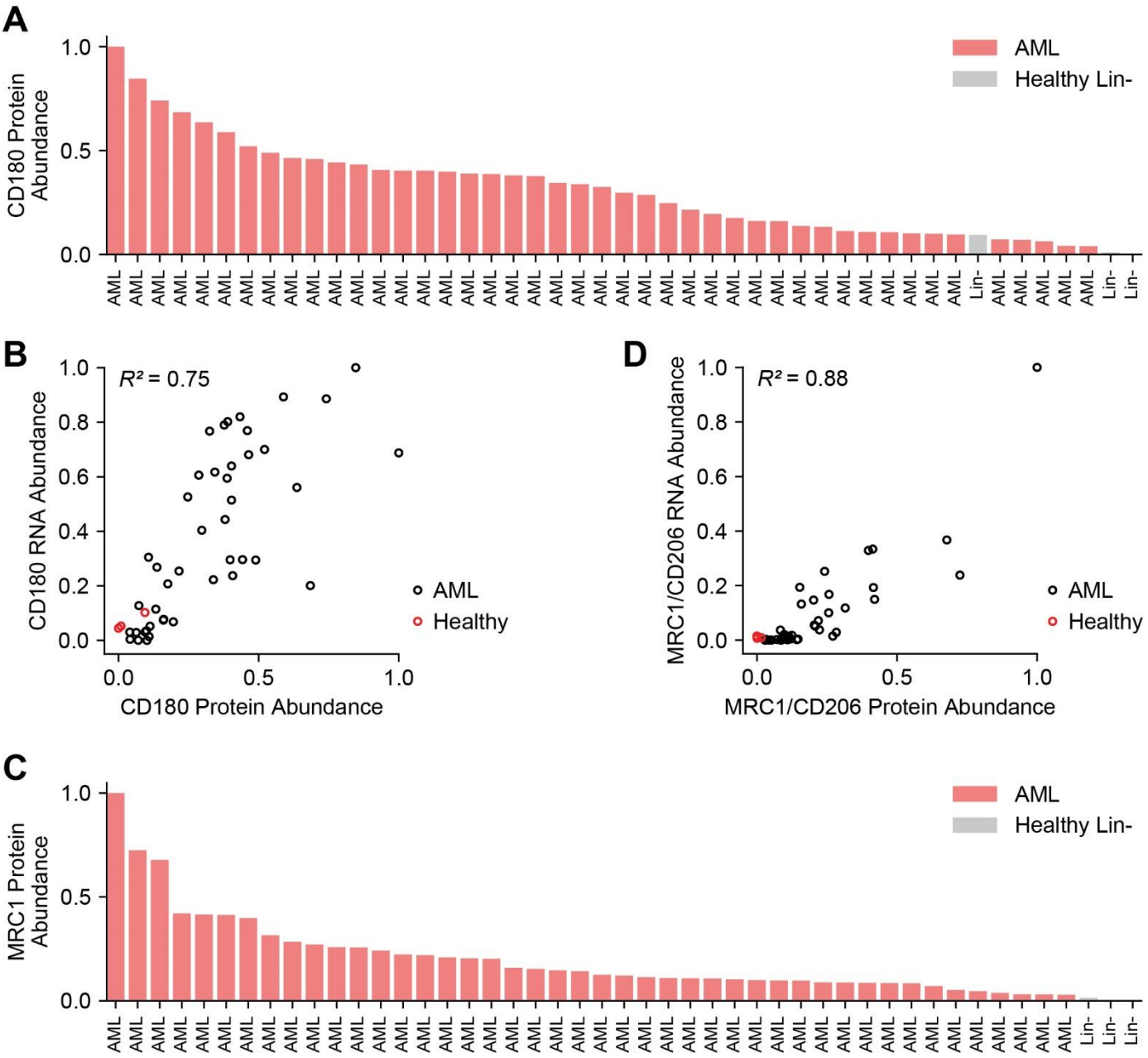
all proteins. Only points with a significantly different protein abundance between the specified mutant and wildtype samples are shown. Increased distance from this line of best fit suggests a higher probability of post-transcriptional regulation of protein abundance.



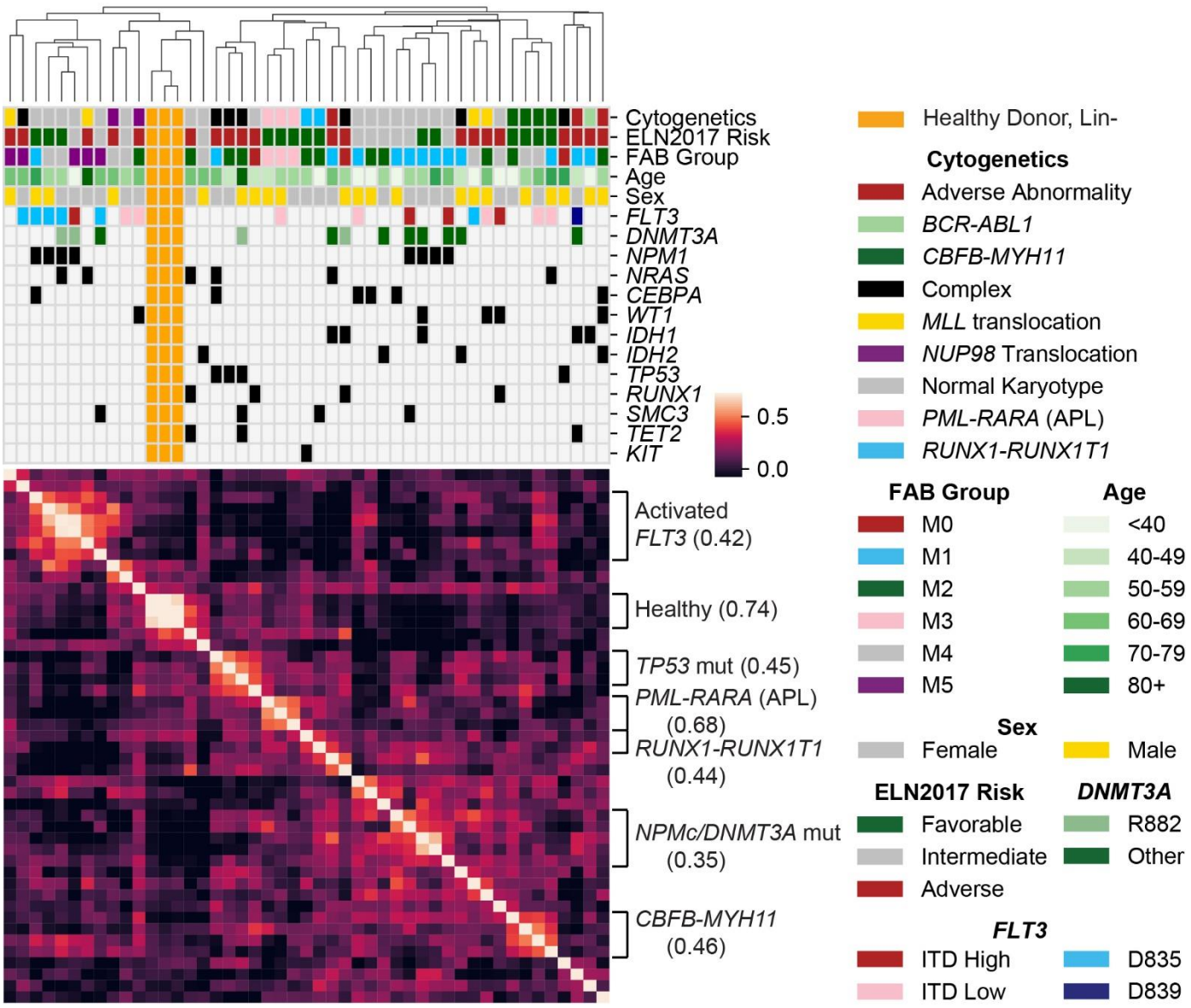
Supplemental Figure 13. Differentially abundant proteins for CEBPA mutant samples. (**A**) Volcano plot showing mean fold change of protein abundance in *CEBPA*^{mut} (n=6) vs. wildtype AML (n=38) samples. P-values are calculated using the t-test and corrected for multiple hypothesis testing with Benjamini-Hochberg method. Dashed red line shows p = 0.05. There were no significantly different proteins in all 6 *CEBPA* mutant samples when pooled (1 biallelic and 5 monoallelic), but there was a group of 3 *CEBPA* mutant samples by unsupervised clustering (**Figure 2,** now referred to as the *CEBPA* cluster) that we further analyzed. (**B**) Volcano plot showing mean fold change of protein abundance in the *CEBPA* cluster (n=3) vs. *CEBPA* wildtype (n=38) AML samples. P-values are calculated using the t-test and corrected for multiple hypothesis testing with Benjamini-Hochberg method. Dashed red line shows p = 0.05. (**C**) Protein fold change vs. RNA fold change for the *CEBPA* cluster vs. *CEBPA* wildtype AML samples. Dashed blue line shows a line of best fit for the relationship between protein and RNA fold change for all proteins. Only points with a significantly different protein abundance between the specified mutant and wildtype samples are shown. Increased distance from this line of best fit suggests a higher probability of post-transcriptional regulation of protein abundance.



Supplemental Figure 14. Differentially abundant proteins for FLT3 mutant samples. (**A,C**) Volcano plot showing mean fold change of protein abundance in *FLT3*^{D835} mutant (n=6) (**A**) or *FLT3*-ITD high mutant (n=4) (**C**) vs. *FLT3* wildtype (n=26) AML samples. Given that many *FLT3*-ITD mutations are subclonal, only the 4 *FLT3*-ITD samples with highest VAF were utilized for this analysis. Red dots represent significantly different proteins ($p < 0.05$ after multiple-hypothesis correction). P-values are calculated using the t-test and corrected for multiple hypothesis testing with Benjamini-Hochberg method. Dashed red line shows $p = 0.05$. (**B, D**) Protein fold change vs. RNA fold change for *FLT3*^{D835} mutant (**B**) or *FLT3*-ITD mutant (**D**) vs. *FLT3* wildtype AML samples. Dashed blue line shows a line of best fit for the relationship between protein and RNA fold change for all proteins. Only points with a significantly different protein abundance between the specified mutant and wildtype samples are shown. Increased distance from this line of best fit suggests a higher probability of post-transcriptional regulation of protein abundance.



Supplemental Figure 15. CD180 and MRC1/CD206 Protein and RNA abundance. (**A, C**) Normalized protein abundance of CD180 (**A**) and MRC1 (**C**) in TMT data for AML patient bone marrow samples, lineage-depleted bone marrow from healthy donors (Healthy Lin-), and CD34-selected bone marrow from healthy donors (Healthy CD34+). (**B, D**) Normalized protein abundance (TMT) vs. normalized RNA abundance for AML samples and healthy, lineage-depleted bone marrow samples (protein) or CD34+ cells (RNA). Data shown for both CD180 (**B**) and MRC1/CD206 (**D**). Note the high concordance of mRNA and protein abundance for most samples.



Supplemental Figure 16. Combined hierarchical clustering of proteomic and phosphoproteomic data. Unsupervised clustering of both proteomic and phosphoproteomic data from 44 AML samples and 3 healthy control bone marrow samples using the UPGMA algorithm. Pearson correlations were calculated separately for protein abundance and total phosphosite abundance between each patient and then averaged to provide a single similarity score, which is shown in the heatmap. Clinical correlates are noted, as in Figure 1.

Supplemental Table Legends (separate files)

Supplemental Table 1: Sample identifiers and peptide quantities.

Supplemental Table 2: Clinical annotation and mutations for patients included in this dataset.

Supplemental Table 3: TMT and LFQ protein abundance measurements for all patient and control samples used in this dataset.

Supplemental Table 4: mRNA abundance for all patient and control samples used in this dataset.

Supplemental Table 5: Spearman correlations of protein (TMT) vs. mRNA abundance across the AML patients included in this study.

Supplemental Table 6: Proteins identified using NPM1 and NPMc TurboID constructs for proximity-labeling of proteins with biotin. Spectral counts of all recurrently identified proteins for NPM1, NPMc and TurboID only constructs provided.

Supplemental Table 7: Phosphosite abundance measurements for all patient and control samples used in this dataset (normalized Mascot Delta Score ≥ 0.5)