

# Sensitive Quantitative Proteomics of Human Hematopoietic Stem and Progenitor Cells by Data-independent Acquisition Mass Spectrometry

## Authors

Sabine Amon, Fabienne Meier-Abt, Ludovic C. Gillet, Slavica Dimitrieva, Alexandre P. A. Theocharides, Markus G. Manz, and Ruedi Aebersold

## Correspondence

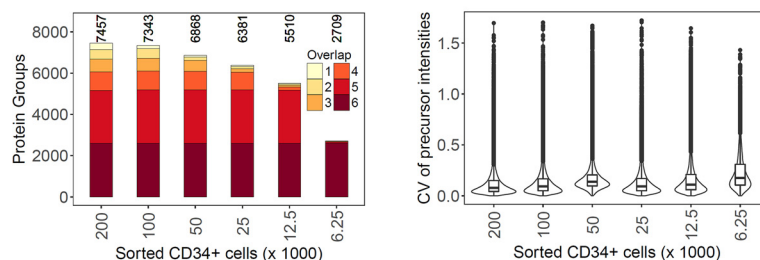
aebersold@imsb.biol.ethz.ch

## In Brief

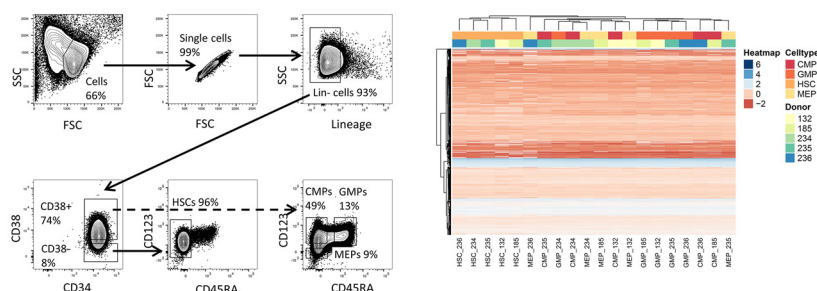
To date, technical limitations have precluded the robust quantitative proteomic analysis of rare cell types. We describe a highly sensitive mass spectrometry-based proteomic workflow for the analysis of human hematopoietic stem cells and three progenitor cell types. More than 5,000 protein groups could be consistently quantified from 25,000 sorted hematopoietic stem and progenitor cells. The data reproducibly identified characteristic patterns of differentially expressed proteins in the tested populations that indicated biochemical differences not apparent by transcriptomic analyses on equivalent samples.

## Graphical Abstract

Sensitive DIA-MS proteomics workflow for analysis of small numbers of FACS-isolated stem/progenitor cells:



Application to sets of 25,000 FACS-isolated human hematopoietic stem and progenitor cells (HSCs/MPPs, CMPs, MEPs, GMPs):



## Highlights

- Sensitive DIA-MS proteomics workflow for analysis of small cell numbers.
- Application to healthy, FACS-isolated, human hematopoietic stem and progenitor cells.
- Good correlation between proteome and transcriptome for committed progenitors.
- Discrepant regulation of protein and mRNA in stem/multipotent progenitor cells.

# Sensitive Quantitative Proteomics of Human Hematopoietic Stem and Progenitor Cells by Data-independent Acquisition Mass Spectrometry\*<sup>§</sup>

Sabine Amon<sup>‡‡</sup>, Fabienne Meier-Abt<sup>‡‡</sup>,  Ludovic C. Gillet<sup>‡</sup>, Slavica Dimitrieva<sup>¶</sup>, Alexandre P. A. Theocharides<sup>§</sup>, Markus G. Manz<sup>§</sup>, and Ruedi Aebersold<sup>‡||\*\*</sup>

Physiological processes in multicellular organisms depend on the function and interactions of specialized cell types operating in context. Some of these cell types are rare and thus obtainable only in minute quantities. For example, tissue-specific stem and progenitor cells are numerically scarce, but functionally highly relevant, and fulfill critical roles in development, tissue maintenance, and disease. Whereas low numbers of cells are routinely analyzed by genomics and transcriptomics, corresponding proteomic analyses have so far not been possible due to methodological limitations. Here we describe a sensitive and robust quantitative technique based on data-independent acquisition mass spectrometry. We quantified the proteome of sets of 25,000 human hematopoietic stem/multipotent progenitor cells (HSC/MPP) and three committed progenitor cell subpopulations of the myeloid differentiation pathway (common myeloid progenitors, megakaryocyte-erythrocyte progenitors, and granulocyte-macrophage progenitors), isolated by fluorescence-activated cell sorting from five healthy donors. On average, 5,851 protein groups were identified per sample. A subset of 4,131 stringently filtered protein groups was quantitatively compared across the 20 samples, defining unique signatures for each subpopulation. A comparison of proteomic and transcriptomic profiles indicated HSC/MPP-specific divergent regulation of biochemical functions such as telomerase maintenance and quiescence-inducing enzymes, including isocitrate dehydrogenases. These are essential for maintaining stemness and were detected at proteome, but not transcriptome, level. The method is equally applicable to almost any rare cell type, including healthy and cancer stem cells or physiologically and pathologically infiltrating cell populations. It thus provides essential new information toward the detailed biochemical understanding of cell development

and functionality in health and disease. *Molecular & Cellular Proteomics* 18: 1454–1467, 2019. DOI: 10.1074/mcp.TIR119.001431.

In multicellular organisms, normal physiological functions and pathophysiological mechanisms are the result of the interplay of multiple cell types at various stages of differentiation. A prototypic example is the mammalian hematopoietic system where hematopoietic stem/multipotent progenitor cells (HSCs/MPPs) can differentiate into various functionally divergent cell lineages, including the downstream formation of common myeloid progenitors (CMPs)<sup>1</sup>, megakaryocyte-erythrocyte progenitors (MEPs), or granulocyte-macrophage progenitors (GMPs) (1, 2). When this process is altered, e.g. upon genetic or epigenetic changes in HSCs, abnormal, (pre)leukemic stem cell subpopulations may form, eventually resulting in clonal hematopoiesis and the onset of acute myeloid leukemia (3–5). To gain insight into the biochemical changes underlying cellular differentiation and to unravel factors involved in the early development of malignant hematopoietic diseases, highly refined analysis of the different cell subpopulations of the hematopoietic cell system is crucially needed (6).

Hematopoietic stem cells are critically rare compared with other hematopoietic cell types (7). Other numerically scarce, but functionally relevant, cell subpopulations include preleukemic stem cells (3–5, 8), leukemic stem cells (9), cancer stem cells in solid tumors (10, 11), circulating tumor cells (12, 13), and infiltrating T cells in solid tumors (14). Although the isolation of such rare cell types is supported by specific surface expression of cluster of differentiation (CD) markers such as CD34, CD38, CD123, CD45RA, and CD10 (15–17), normally no more than a few thousand cells per subpopulation can be isolated by fluorescence-activated cell sorting (FACS) from a

From the <sup>‡</sup>Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, 8093 Zurich, Switzerland; <sup>§</sup>Hematology, University and University Hospital Zurich, 8091 Zurich, Switzerland; <sup>¶</sup>Functional Genomics Center Zurich, ETH Zurich and University of Zurich, 8057 Zurich, Switzerland; <sup>||</sup>Faculty of Science, University of Zurich, 8057 Zurich, Switzerland

\* Author's Choice—Final version open access under the terms of the Creative Commons CC-BY license.

Received October 26, 2018, and in revised form, March 8, 2019

Published, MCP Papers in Press, April 11, 2019, DOI 10.1074/mcp.TIR119.001431

single person. For example, the preparation of 25,000 sorted human HSCs requires up to 4 l of steady-state blood or a leukapheresis procedure following hematopoietic stem and progenitor cell (HSPC) mobilization, making further upscaling difficult. Whereas a few thousand cells can be routinely analyzed by modern imaging and genomic profiling technologies (1, 2, 16–19), proteome-level measurements, particularly the reproducible quantification of thousands of proteins across sample cohorts, has remained technically challenging for minute samples. Indeed, highly enriched human HSPC subpopulations have, to our knowledge, not been analyzed by unbiased large-scale proteomic analysis, even though global protein expression determines cellular functionality and provides critical information on the cellular differentiation process. Proteomic analysis of FACS-isolated cells has in general been reported only in studies focused on optimizing specific technical parts of the workflow, such as the cell sorting step itself (20), sample preparation (21, 22), or sample fractionation (23). Others used 400,000 cells as starting material, which restricted the scope of the analyses to large pools of murine samples (24) or *in vitro* model systems. Furthermore, no systematic assessment of the reproducibility or consistency of the proteomic results of small numbers of sorted cells has been performed, other than comparing protein identification numbers. It is therefore evident that the robust, reproducible, and quantitative proteomic analysis of minute samples, such as for example highly enriched HSPC, represents a significant technical and scientific advance.

Here, we present and apply an integrated workflow for the high-coverage, quantitative proteome profiling of minute amounts of sorted cells. It is based on data-independent acquisition (DIA)-MS on the Orbitrap Lumos platform and peptide centric signal extraction and analysis. DIA-MS is a massively parallel-in-time acquisition method of fragment ion mass spectra of all detectable precursors in a sample. It provides a complete, yet convoluted, quantitative fragment ion map record of a sample (25). Peptide-centric analysis (26, 27) of DIA datasets results in quantitative peptide matrices (25) of sufficient consistency and reproducibility to support label-free comparisons of large sample cohorts. To date, DIA studies on hybrid quadrupole-time-of-flight (QqTOF) (26, 28, 29) or Orbitrap (30, 31) platforms typically used microgram amounts of total peptide mass for analysis (and even larger amounts of actually processed starting material), a quantity that is one to two orders of magnitude above the quantity achievable by FACS isolation of rare hematopoietic cell types.

To overcome limitations of working with small amounts of proteins, we established a method to reproducibly identify and quantify nearly 6,000 protein groups with a median coefficient of variance (CV) of 9% for 125 ng of HEK293 tryptic peptides (see “Results”). This unprecedented performance was then used to profile minute amounts of highly enriched human HSCs/MPPs, CMPs, MEPs, and GMPs. The resulting protein *versus* sample data matrix revealed factors and biochemical pathways involved in quiescence, stemness maintenance, and cell differentiation. Comparison with RNAseq analyses demonstrated proteome-specific regulation of “stemness maintaining networks” in HSCs/MPPs.

#### EXPERIMENTAL PROCEDURES

*Experimental Design and Statistical Rationale*—Sample numbers: For method development, samples derived from cultured cells (HEK293 cell line) and human CD34+ hematopoietic stem/progenitor cells (isolated by FACS) were used in varying amounts as described. For final analysis of donor samples, material from five individuals (four cell types per donor) was analyzed, resulting in a total of 20 samples.

Replicates: HEK293 samples were analyzed in triplicates as technical replicates; CD34+ samples were processed in parallel as triplicates (process replicates). Due to the limited amount of sample material, donor samples were analyzed in single replicates, and different individuals were considered biological replicates. The number of samples/donors was limited by the material available from the hospital for this study.

Controls: Because this study did not involve a case/control design, no control samples were necessary.

Randomization: For LC-MS/MS analysis, the run order of donor samples (different individuals, different cell types) was randomized.

Statistical tests: Statistical procedures built into the software mapDIA (32) were used.

*HEK293 Cell Culture*—HEK293 cells (ATCC) were grown in Dulbecco's Modified Eagle's Medium (10% fetal bovine serum, 50  $\mu\text{g}/\mu\text{l}$  penicillin, 50  $\mu\text{g}/\mu\text{l}$  streptomycin) until confluence. Harvested cells were washed twice with 1x phosphate buffered saline and counted using an Invitrogen Countess automated cell counter (ThermoFisher Scientific, Waltham, MA). For the HEK293 peptide dilution series,  $5 \times 10^6$  cells were lyophilized and processed in bulk according to the protocol described below.

*Human Hematopoietic Stem/Progenitor Cell Samples*—Fresh human HSPCs were obtained from healthy stem cell donors (Clinical Hematology, University Hospital Zurich, Zurich, Switzerland) with informed consent and approval of the local ethics committee (KEK-ZH-Nr: 2015–0564). After mobilization with granulocyte-colony stimulating factor, stem cell-enriched mononuclear leukocytes were obtained via leukapheresis from peripheral blood for clinical stem cell transplantation. Any leftover material after clinical transplantation was collected for further cell preparation.

*Cell Preparation, Flow-cytometric Analysis, and Cell Sorting*—Human CD34+ hematopoietic stem/progenitor cells were enriched from mononuclear cells using immunomagnetic beads according to the manufacturer's instructions (CD34 MicroBead Kit; Miltenyi Biotec, Bergisch Gladbach, Germany). Following enrichment, CD34+ cells were frozen in liquid nitrogen. For analysis and sorting of hematopoietic stem cell-enriched cells and myeloid progenitors, CD34+ cells were thawed and stained with Tricolor/phycoerythrin (PE)-Cy5-conjugated antibodies specific for lineage markers: CD2, S5.5; CD3, 7D6; CD4, S3.5; CD7, CD7–6B7; CD8, 3B5; CD14, TuK4; CD19, SJ25-C1; CD56, MEM-188 (ThermoFisher Scientific-Invitrogen); CD10, HI10a; CD11b, ICRF44; CD20, 2H7 (BioLegend, San Diego, CA); CD235a,

<sup>1</sup> The abbreviations used are: CMP, common myeloid progenitors; CV, coefficient of variation; DDA, data-dependent acquisition; DIA, data-independent acquisition; FDR, false discovery rate; GMP, granulocyte-macrophage progenitors; GO, gene ontology; HSC, hematopoietic stem cells; HSPC, hematopoietic stem and progenitor cells; MEP, megakaryocyte-erythrocyte progenitors; MPP, multipotent progenitor cells.

GA-R2 (BD Biosciences, Allschwil, Switzerland); and PE-Cy7-conjugated anti-CD34, 8G12 (BD Biosciences); fluorescein isothiocyanate-conjugated anti-CD38, HIT2 (BD Biosciences); allophycocyanin (APC)-conjugated anti-CD123, 6H6 (ThermoFisher Scientific-Invitrogen); and APC780-conjugated anti-CD45RA, HI100 (ThermoFisher Scientific-Invitrogen). Apoptotic cells were excluded in the analysis by Hoechst staining. Because less than 1% of all cells were apoptotic (see [supplemental Fig. S1](#)) and because of concerns regarding interference of the Hoechst stain with the mass spectrometric analysis, this step was omitted for further cell sorting. For compensation of the fluorochromes, cells and/or beads (anti-mouse Ig, kappa/negative control compensation particles; BD Biosciences) were singly stained with the antibodies. Compensation was performed automatically using the DIVA software (BD Biosciences) and checked manually.

For FACS analysis of intracellular isocitrate dehydrogenase 1 (IDH1) expression, the IntraPrep Leukocytic Permeabilization Reagent Kit (Beckman Coulter, Brea, CA) was used together with phycoerythrin (PE)-conjugated anti-IDH1, D2H1 (Cell Signaling Technology, Danvers, MA) and PE-conjugated isotype control, DA1E (Cell Signaling Technology).

All analyses were performed on a four laser-equipped LSR Fortessa machine (BD Biosciences), all sorting was performed on a five laser-equipped FACS Aria III machine (BD Biosciences).

Gates were set using fluorescence minus one and unstained controls according to (17). The degree of purification of cell populations was determined by back-gating and double sorting (see [supplemental Fig. S2](#)) and determined to be >95% in all cases. FACS data were analyzed with FlowJo software (FlowJo LLC, Ashland, OR).

Highly enriched human HSCs/MPPs, CMPs, GMPs, and MEPs were isolated from five healthy HSPC donors ([supplemental Table S1](#)) using fluorescence-activated cell sorting and the cell surface markers CD34, CD38, CD123, and CD45RA. Lymphoid progenitors were excluded by adding the marker CD10 to the lineage mixture. For CD34+ cell dilution, purification control, fluorescence-minus-one, and validation experiments, CD34+ cells from an additional six healthy HSPC donors were used (ID 273, ID 227, ID 223, ID 183, ID 151, ID 340).

For MS analysis, 25,000 cells each were collected in 300  $\mu$ l phosphate buffered saline solution in protein low-binding micro-centrifuge tubes (Eppendorf, Hamburg, Germany). All remaining material was sorted into separate micro-centrifuge tubes for the purpose of library generation. This resulted, depending on the donor sample and the cell type, in 13 additional samples (4x CMP, 4x HSC, 2x GMP, 3x MEP) with 40,000 to 270,000 cells. Samples were pelleted by centrifugation at standard force for viable HSPCs of 400 g for 15 min. We found that a critical step for reproducibility of material recovery was to very carefully remove the supernatant above the cells, leaving 50  $\mu$ l on top of the pellet to avoid losing the nonadherent pelleted cells. The tubes were then snap frozen in liquid nitrogen, and the remaining FACS buffer was lyophilized.

**Sample Preparation for Mass Spectrometry**—The lyophilized cell pellets were resuspended in 10  $\mu$ l (200  $\mu$ l for bulk HEK293 preparation for the peptide dilution series) of 8 M urea in 100 mM ammonium hydrogen carbonate and lysed aided by sonication with a VialTweeter (Hielscher, Teltow, Germany) at an amplitude of 60%, a cycle of 60% and a duration of 20 s for three times with intermediate cooling on ice. One aliquot containing 128,000 HEK293 cells and one sample with 100,000 CD34+ hematopoietic stem/progenitor cells were submitted to a BCA protein assay following the manufacturer's guidelines (ThermoFisher Scientific-Pierce) to determine the protein content and the required amount of protease to be added.

Samples were diluted to 4 M urea with 100 mM ammonium hydrogen carbonate and treated with 1.25 U Benzonase Nuclease (Sigma-Aldrich, Darmstadt, Germany) per 25,000 cells for 30 min at 37 °C. Reduction of disulfide bonds was carried out by the addition of

tris-(2-carboxyethyl)-phosphine (Sigma-Aldrich) to 5 mM and incubation at 37 °C for 30 min with shaking in a Thermomixer (Eppendorf), followed by alkylation of free thiol groups by the addition of iodoacetamide (Sigma-Aldrich) to 10 mM in the dark for 30 min at room temperature. Samples were diluted to 1 M urea, and sequencing grade trypsin (Promega, Madison, WI) was added at an enzyme-to-substrate ratio of 1:50 (extrapolated from the BCA assay results) for overnight digestion at 37 °C. After adjusting to 2% formic acid, samples were desalted with Empore Disks C18 (3 M, Saint Paul, MN) self-packed into StageTips format (33). Samples were dried using a vacuum centrifuge and resuspended in 11  $\mu$ l of 2% acetonitrile and 0.1% formic acid with the addition of indexed Retention Time (iRT) peptides (34) (Biognosys, Schlieren, Switzerland) for the following MS analysis. An aliquot corresponding to 1,500,000 cells (285  $\mu$ g protein) of the HEK293 sample was desalted with a 50-mg Sep-Pak tC18 cartridge (Waters, Milford, MA). After drying by vacuum centrifugation, the HEK293 sample was dissolved at 1  $\mu$ g/ $\mu$ l in 2% acetonitrile and 0.1% formic acid with added iRT peptides. The peptide concentration was determined using the quantitative fluorimetric peptide assay (ThermoFisher Scientific-Pierce) following the manufacturer's instructions. The leftover material of the CD34+ samples after MS-injection was also submitted to peptide concentration determination.

For the generation of the HEK293 spectral library, an aliquot corresponding to 80  $\mu$ g peptides was subjected to high pH reversed-phase peptide fractionation using a dedicated spin column kit (ThermoFisher Scientific-Pierce). Fractionation was carried out according to the manufacturer's recommendation. Briefly, eight fractions were collected by eluting with 300  $\mu$ l each of increasing acetonitrile content (from 5% to 50% in 0.1% trimethylamine). The fractions were dried by vacuum centrifugation and redissolved in 25  $\mu$ l of 2% acetonitrile, 0.1% formic acid with iRT peptides, of which 2  $\mu$ l were injected per measurement. Additionally, unfractionated whole lysate of HEK293 was measured by LC-MS for the purpose of library generation.

**Mass Spectrometry Analysis**—Nanoflow LC-MS/MS measurements were carried out on an EASY-nLC 1200 system (ThermoFisher Scientific) coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (ThermoFisher Scientific) equipped with a Nanospray Flex ion source.

Peptides were separated on an Acclaim PepMap 100 C<sub>18</sub> column (ThermoFisher Scientific) with an inner diameter of 75  $\mu$ m, a length of 25 cm, and a particle size of 2  $\mu$ m. The column was operated at room temperature and at a flow rate of 300 nL/min. LC solvent A consisted of 98% water, 2% acetonitrile, and 0.1% formic acid, LC solvent B was composed of 80% acetonitrile, 20% water and 0.1% formic acid. Peptides were separated by a linear gradient from 5 to 37% B over 120 min, except for the purpose of library generation (240 min for unfractionated samples and 180 min for high pH reversed-phase fractions of HEK293, respectively).

The instrument was operated either in the DDA or DIA mode. In both cases, fragmentation was accomplished by higher energy collision dissociation at a normalized collision energy setting of 27%. The resolution of the Orbitrap analyzer was set to 120,000 and 30,000 for MS1 and MS2, with a maximum injection time of 100 ms and 50 ms, respectively. The mass range monitored in MS1 was 350–1,500 *m/z* and in MS2 200–1,800 *m/z* for DIA and the auto *m/z* normal scan range mode for DDA. The automatic gain control (AGC) target was set to 2e5 in MS1 and 5e5 or 8e4 in MS2 for DIA or DDA, respectively.

DDA measurements utilized the top speed setting, where one MS1 survey scan was followed by the acquisition of MS/MS spectra for a cycle time of up to a maximum of 3 s. In DIA mode, one MS1 scan was followed by 40 MS2 windows of equal width (15 *m/z*) with an overlap of 1 *m/z*, covering precursors in the range of 399.5–1,000.5 *m/z*. This resulted in a cycle time of 3.4 s. Data were acquired with Xcalibur 4.0.27.10 and Tune Plus version 2.1.

**Data Analysis**—DDA data were searched accordingly by Mascot (35) (Matrix Science, version 2.5.1) as well as Comet (36) version 2016.01 rev. 2 against the Swissprot reviewed subset of the human UniProt database (downloaded on 2016.07.06) containing 20,199 protein entries plus one additional protein entry for the concatenated iRT peptide sequences, plus as many appended decoy sequences generated by sequence reversal (keeping C-terminal K and R residues). The settings were as following: enzyme = trypsin, missed cleavages =  $\leq 2$ , peptide tolerance =  $\pm 10$  ppm, MS/MS tolerance = 0.02 Da, fixed modification = carbamidomethylation on cysteine, variable modification = oxidation on methionine.

Spectral libraries from DDA runs were generated as previously described (37). Briefly, the peptide-level false discovery rates (FDR) for the search results was independently adjusted to 1% for Mascot and Comet results using PeptideProphet (38) (TPP v4.7 rev 0 (39)). Mascot and Comet results were then combined by iProphet (40) and filtered to 1% protein FDR by Mayu (41) (v 1.08). Compilation of a consensus library used for the query of the DIA measurements was carried out by SpectraST (42, 43) (v 5.0) with the following filter criteria applied: include peptides at 1% peptide FDR for proteins at 1% protein FDR, six fragments per peptide, and fragment  $m/z$  range of 350–1,800  $m/z$ . The HEK293 library included entries of 112,227 peptide precursors from 9,127 protein groups, the combined library for HEK293 cells and HSPC library samples 146,610 peptide precursors from 10,057 protein groups.

DIA data were evaluated by Spectronaut 11 (31) (Biognosys) querying the above-mentioned library created from DDA runs with the following settings: data extraction with a tolerance of 10 ppm and 25 ppm for MS1 and MS2 level, with a dynamic retention time extraction window and automatic nonlinear iRT retention time calibration, identification at a precursor Q-value cutoff of 0.01 and protein Q-value of 0.01, quantification based on MS2 level area, and without cross-run normalization. The Spectronaut report (information on precursor, not fragment level) was exported for further processing in R. Protein groups were counted as defined by Spectronaut. For benchmarking datasets (HEK peptide dilution series and different collections of FACS-isolated CD34+ cell numbers), reports were used as such (without additional filtering, normalization or imputation) to generate the results for Fig. 1 and Fig. 2. CV calculations were computed in R on the per-triplicate dilution set level by dividing the standard deviation of the raw precursor intensities as reported by Spectronaut by the mean of the raw precursor intensities.

DDA data for the HEK293 dilution series were processed identical to the files for the library generation up to the step of adjusting the peptide FDR to 1% with PeptideProphet and compared with the corresponding DIA data at this level.

**Proteomic Data Processing for HSPC Samples**—The entries in the Spectronaut report were filtered to a protein Q-value cutoff of 0.01. We then filtered the data to keep only peptide precursors that were either (i) present in all donors for at least one cell type or (ii) in all cell types for at least one donor. Protein groups covered by only a single peptide were excluded from the differential analysis. The precursor quantitation matrix contained 785,280 entries, thereof 50,382 missing values (6.4%), for an average of 1.28 missing value per precursor across the 20 samples. Missing quantity values were imputed with random values in the range of 0.7 to 0.9 of the minimum value observed for the corresponding precursor. Normalization was performed based on the total ion current (normalization factor is calculated as mean of the sum of the precursor quantities across all measurements divided by sum of precursor quantities of the corresponding measurement).

The filtered, imputed, and normalized precursor data matrix was submitted to mapDIA v3.0.2 (32) for differential analysis. mapDIA calculates protein fold changes in a condition-based pairwise fashion by

using the most stable peptides to perform the fold-change calculation. The replicate design set up was used, with standard deviation factor 2 and minimum correlation (median intraprotein correlation cutoff) 0.2.

For GO enrichment and search tool for recurring instances of neighboring genes (STRING) analysis, only unambiguous, single-protein entries from the results were considered and protein groups with multiple members ignored. This reduced the number of candidate proteins for these analysis steps to 3,364.

**RNA Isolation, Quality Determination, and Sequencing**—10,000 cells from each population were sorted into RNeasy lysis buffer containing beta-mercaptoethanol. Total RNA was purified according to manufacturer instructions using the RNeasy Plus Micro Kit (Qiagen, Hilden, Germany). RNA quality was determined using the Agilent 2100 Bioanalyzer System (Agilent Technologies, Santa Clara, CA). The samples demonstrated RNA integrity numbers  $\geq 8.0$  and were of high enough quality for a poly-A tail-based approach for RNA sequencing. RNA sequencing was performed as described in (44) using the Illumina HiSeq 4000 sequencing platform.

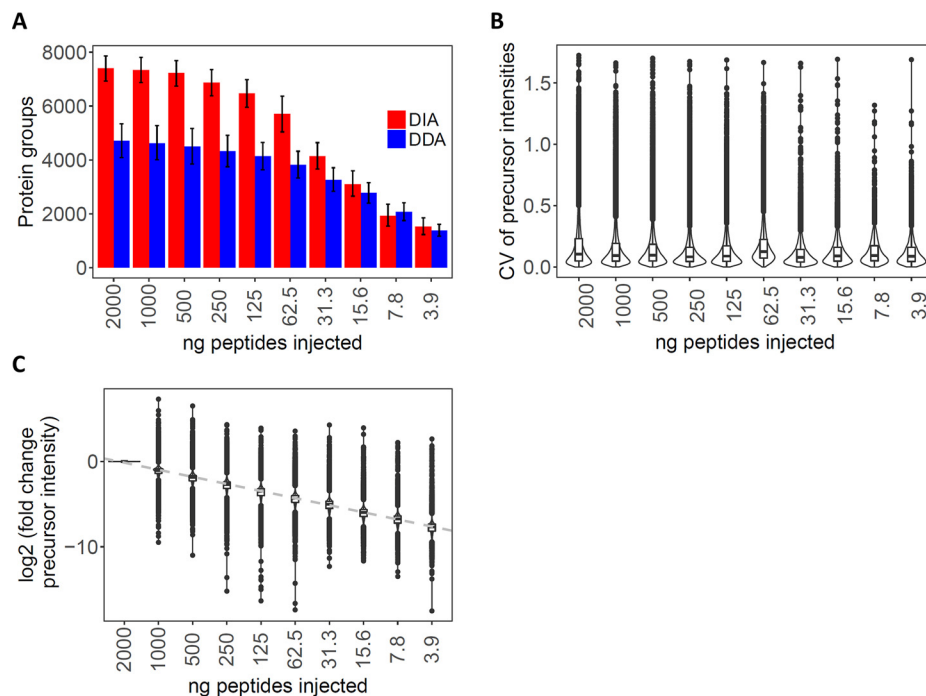
**RNA Data Normalization and Processing**—RNA sequencing data reads were quality-checked with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were trimmed with Trimmomatic version 0.33 (4 and 3 bases were hard-trimmed from the start and end respectively; adapter trimming was done at the end). Trimmed reads were aligned to the Ensembl GRCh38 human reference genome and transcriptome with STAR version 2.5.1b (45). The average number of high-quality reads, reads aligned, and reads uniquely aligned per sample were 43.2 million, 42.6 million, and 35.9 million, respectively. Gene expression was quantified using the R/Bioconductor package Rsubread version 1.26.0 (46). Differentially expressed genes between cell types were identified using the R/Bioconductor package DESeq2 version 1.16.1 (47).

**Gene Ontology Enrichment Analysis**—For gene set enrichment analysis (GSEA), gene sets were retrieved from the Gene Ontology Consortium database ([www.geneontology.org](http://www.geneontology.org)) on 2017–05–25 and 2017–05–27 (see supplemental Table S5). Ranked lists were created from the normalized and filtered proteome and transcriptome data using  $\log_2(\text{fold change})$  or  $\log_2(\text{fold change}) \times (-\log_{10}(\text{adjusted } p \text{ value}))$  as ranking criterion. GSEA was performed on the preranked lists using the GSEA software (v3.0, <http://www.broadinstitute.org/gsea>) with the minimum gene set size set to five and the remaining settings as defaults (19). Enrichments were deemed significant when  $\text{FDR} < 0.25$  as suggested by (48). The number of genes used to calculate the enrichment scores depicted in Fig. 4B are shown in supplemental Table S6.

**Network Analysis**—Proteins significantly and differentially regulated compared with the corresponding mRNA were analyzed with STRING (<http://www.string-db.org>, v10.5) for their network connections. Significance was defined as  $\text{FDR} < 0.01$  for both protein and mRNA data. Protein-protein interactions were computed based on the experimental and database evidence channels (49). The network was then loaded into Cytoscape using the STRING interaction score as distance and width for the edges. When available, the GO terms found enriched in Fig. 4B were mapped in color to the corresponding protein nodes.

**Quantitative Polymerase Chain Reaction (qPCR)**—Isolated mRNA (see above) was reverse transcribed according to manufacturer's instructions using the SuperScript IV VIL0 Master Mix with ezDNase enzyme (SuperScript IV Vilo Master Mix with ezDNase Enzyme, ThermoFisher Scientific). Quantitative analysis of cDNA was performed using Taqman probes and master mix (TaqMan Gene Expression Master Mix, ThermoFisher Scientific). Individual probes included Hs00271858\_m1 (IDH1), Hs00255867\_m1 (GAR1), Hs00950764\_g1 (NHP2), and Hs99999903\_m1 (ACTB) (ThermoFisher Scientific). ACTB was used as housekeeping control gene. Expression values

**FIG. 1. HEK293 tryptic digests dilution series.** (A) Number of protein groups identified in DDA (blue) and DIA (red) mode, respectively, for decreasing loads of HEK293 tryptic peptides. The bars in the negative and positive directions represent the number of protein groups identified in common (intersection) or in total (union) for the technical triplicate injections at the indicated peptide loads, respectively. (B) Distribution of the CV for the peptide precursor intensities for the technical (process) triplicate injections for each sample load. (C) Distribution of the fold change (log<sub>2</sub> scale) of the average precursor intensities between a given sample load and that at 2,000 ng sample load.



were calculated using a Delta CT approach. Results were based on technical duplicates and biological triplicates.

## RESULTS

**Optimization of DIA-MS for Small Sample Loads**—Most DIA-MS applications reported so far minimally used 0.5–2  $\mu$ g of total peptide mass per injection (26, 28–31). To make the method compatible with the protein amounts extractable from FACS-isolated rare cells, we first optimized the acquisition scheme on an Orbitrap Fusion Lumos mass spectrometer to extend the application of DIA-MS to lower sample quantities, while minimizing attrition of quantified proteins and quantitative accuracy.

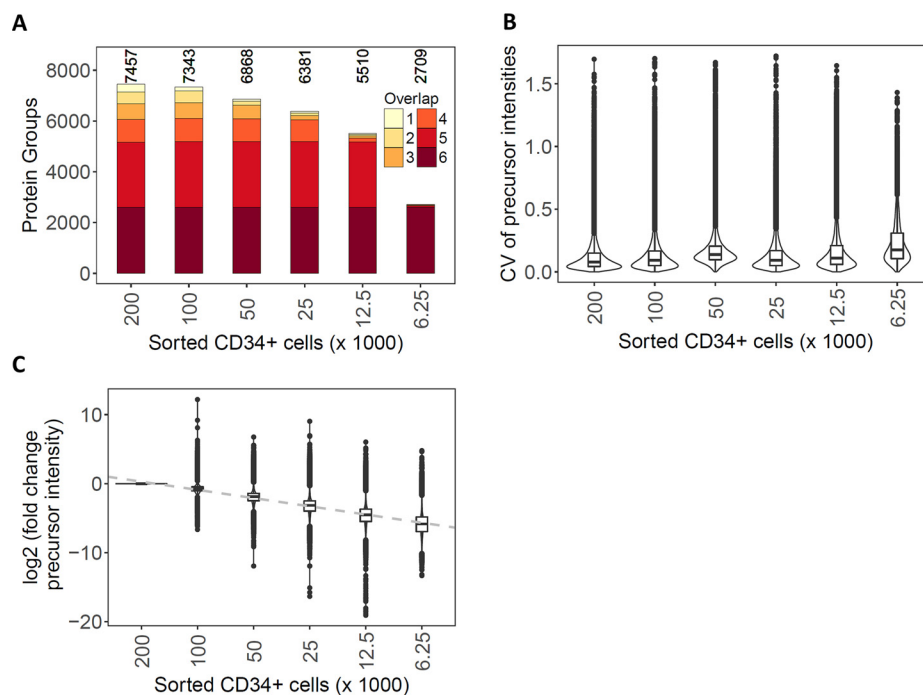
We developed and benchmarked the procedure with a peptide dilution series of HEK293 cells. Specifically, we refined sample handling as well as DIA-MS-specific measures (Fig. 1; see also supplemental text and Figs. S3–S9). In brief, the sample handling procedure minimized sample losses *e.g.* through surface adsorption, which are unavoidable when minute sample amounts are processed (supplemental Fig. S3). The mass spectrometric method maximized the use of ions injected into the mass spectrometer. Initially we monitored the ion density distribution in dependence of mass to charge ( $m/z$ ) and retention time as a function of sample load (supplemental Figs. S4A–4C). We then optimized the DIA acquisition windows by adjusting the fill times. The benchmarking results obtained via the optimized MS method are summarized in Fig. 1. Fig. 1A shows a comparison of the number of protein groups identified by standard DDA and the optimized DIA method as a function of sample load. The results indicate that for sample loads above 30 ng the DIA mode consistently identified a higher number of peptide precursors and protein

groups than DDA. Fig. 1B shows an assessment of the quantitative reproducibility and accuracy of data generated in DIA mode. The measurements achieved an average peptide quantification CV of less than 10% for triplicate injections across the whole dilution series. Further, the peptide quantification values obtained for the consecutive dilution steps retained linearity throughout the entire dilution range (Fig. 1C). We also show that, at 30 ng sample loads, longer fill time increases further the number of identifications (supplemental Figs. S4D–S4F), demonstrating that DIA performance is optimal at very low sample loads by leveraging the fill time capabilities of ion-trap instruments. In summary, the results indicate that an optimized economy of the available precursor ions on the Orbitrap Lumos significantly extends the performance of DIA proteomics toward sample amounts in the low nanogram range with minimum attrition in terms of identified proteins and quantitative accuracy.

To assess the reproducibility and accuracy of the peptide/protein identification and quantification for the whole procedure, we collected from the FACS instrument triplicate samples of 200,000; 100,000; 50,000; 25,000; 12,500; and 6,250 human CD34<sup>+</sup> hematopoietic stem/progenitor cells and processed them with our sample preparation procedure in parallel. We determined that ca. 3.2  $\mu$ g of total peptide mass could be recovered from 200,000 sorted cells after the entire process, of which ~73% (8/11  $\mu$ l) could be injected by the autosampler for DIA measurements. By extrapolation, we therefore estimated that we injected ~2,327; 1,164; 582; 291; 145; and 73 ng of peptides, respectively, from the lower numbers of sorted cells. In comparison, 26.8  $\mu$ g of peptide mass was obtained from 200,000 HEK293 cells when pro-

### FIG. 2. Dilution series of human CD34+ hematopoietic cells isolated by FACS.

(A) Number of protein groups cumulatively identified across the technical replicates for decreasing numbers of FACS-isolated human CD34+ hematopoietic cells. The color scale represents the consistency of protein group identifications across the runs. (B) Distribution of the CV for the peptide precursor intensities for the technical (process) triplicate injections of processed FACS-isolated cells. (C) Distribution of the fold change (in log<sub>2</sub> scale) of the average precursor intensities between a given sample load and that at 200,000 FACS-isolated human CD34+ hematopoietic cells.



cessed in bulk. This difference in the amount of peptides obtained from the same number of HEK293 and hematopoietic stem/progenitor cells is roughly in agreement with the 4–5 times smaller cell volume expected for the human CD34+ hematopoietic cells compared with HEK293 cells (50, 51) and the expected increase in loss from processing low sample amounts.

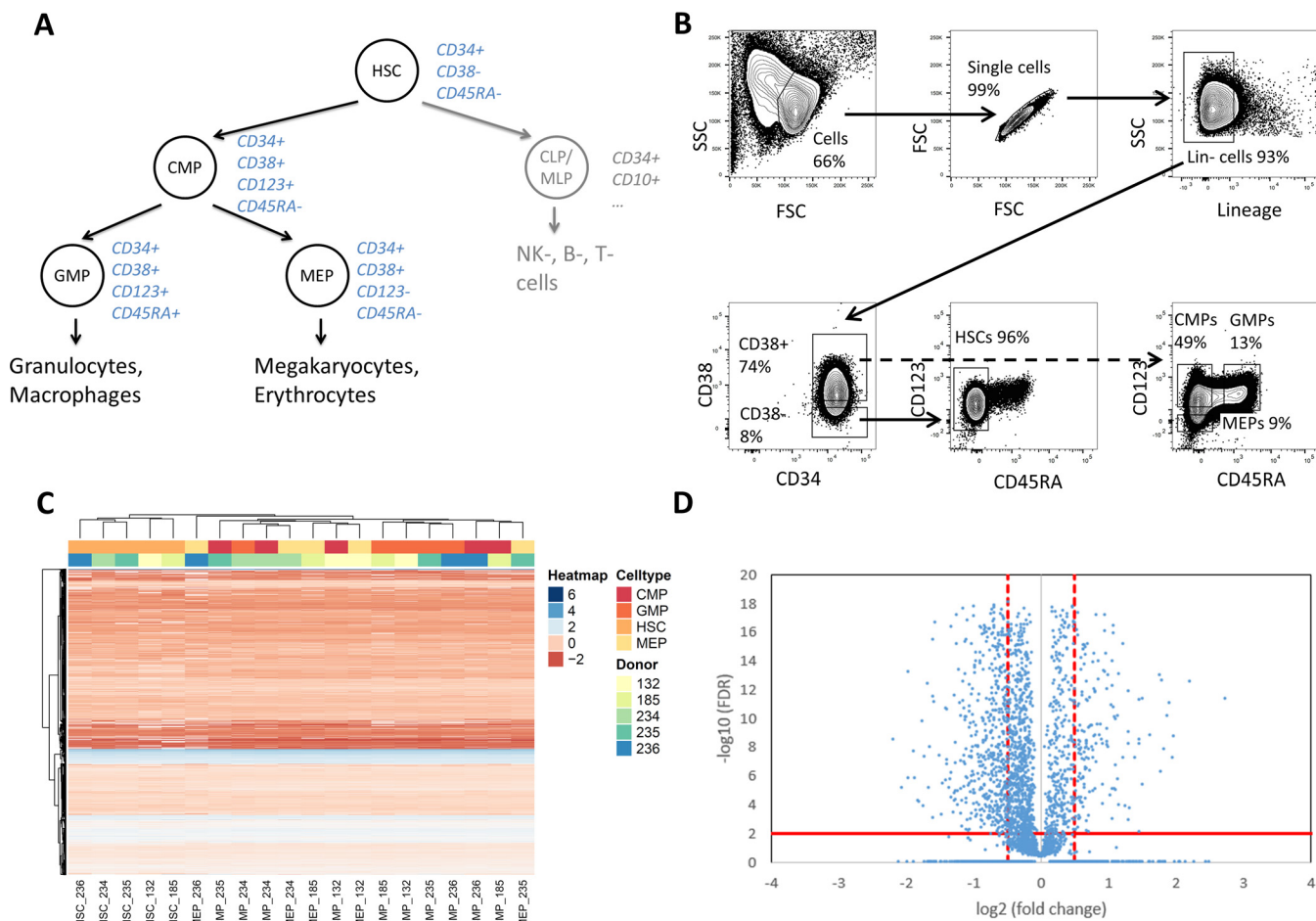
Along the dilution series of FACS-isolated cells the average number of identified protein groups decreased from 6,955 for 200,000 to 4,833 for 12,500 sorted human CD34+ hematopoietic cells (Fig. 2A). A median CV below 14% (Fig. 2B, supplemental Table S2) and a good linearity of quantification was maintained for these measurements (Fig. 2C). At 6,250 cells, the number of identified protein groups decreased to 2,248 and the CV of quantification increased to above 17%. However, the overall peptide quantification remained well correlated throughout the entire CD34+ series (supplemental Fig. S9), indicating that protein quantification remained quite accurate even for samples containing as few as 6,250 cells. The 2,709 protein groups identified in these samples were almost entirely subsumed in the set of proteins identified from higher cell numbers (supplemental Fig. S10), indicating that after normalization of signal intensities, meaningful comparisons across sample cohorts are feasible, even in cases in which one or several samples are only available in minute quantities.

Relating the results from sorted human CD34+ hematopoietic cells to those of the HEK293 peptide dilution series, we noted an attrition in the number of identified proteins and their quantification accuracy for cell numbers below 25,000, corresponding to <300 ng of peptide mass on column. Overall, the optimized DIA and sample preparation method provided repro-

ducible identification and quantification results (in all three replicates) for more than 5,100 protein groups from as little as 25,000 FACS-isolated human CD34+ hematopoietic cells.

*Proteomic Analysis Underscores Ontogenetic Distance Between Individual Human Hematopoietic Cell Types*—We applied the newly developed method to profile the proteome of human CD34+ hematopoietic cell subpopulations isolated from the peripheral blood of five HSPC donors (age 28–57 y, see supplemental Table S1). Four highly enriched subpopulations, including rare HSCs/MPPs, CMPs, MEPs, and GMPs, respectively, were isolated by FACS (Figs. 3A and 3B), processed and analyzed by DIA-MS. Guided by the cell dilution experiment described above, 25,000 cells were collected for each subpopulation. To support peptide-centric analysis of the DIA data, we generated a spectral library specific for the cell types of this study (see “Experimental Procedures”). We identified on average 5,851 protein groups for the different human HSPC populations (supplemental Fig. S11, supplemental Table S3). To increase the robustness of the following differential comparison, we applied additional stringent filtering criteria that resulted in a final list of 4,131 protein groups (from 39,264 peptide precursors) that were quantified consistently with at least two peptides across the samples.

Because the cell samples were derived from unrelated donors of different age and the analyzed cell populations are relatively close in the cell differentiation tree, we first tested whether the quantitative protein measurements were sufficiently accurate and reproducible (see supplemental Figs. S12 and S13) to confidently detect cell subtype-specific differences despite the expected interperson variability. The



**FIG. 3. Proteome profiles of human hematopoietic stem and progenitor cell subpopulations.** (A) Human hematopoietic cell hierarchy with respective cell surface markers depicted in blue (15–17). (B) FACS strategy, depicted on magnetic-activated cell sorting-preselected  $CD34^+$  hematopoietic cells isolated from healthy HSPC donors. Shown are the analysis gates. Highly enriched HSCs/MPPs (referred to as HSCs) are  $CD34^+CD38^-CD45RA^-$ , highly enriched CMPs are  $CD34^+CD38^+CD123^+CD45RA^-$ , highly enriched GMPs are  $CD34^+CD38^+CD123^+CD45RA^+$ , and highly enriched MEPs are  $CD34^+CD38^+CD123^-CD45RA^-$ . (C) Nonsupervised hierarchical clustering (Euclidean distance) heatmap (78) of intensities for the peptides identified in HSCs, CMPs, GMPs, and MEPs (shades of red) isolated from five different donors (shades of blue). The peptide intensities are centered and scaled and depicted in color shades from red to blue. The missing peptide intensity values are shown in white. (D) Volcano plot of differential analysis of proteins. Comparison of HSCs to the average of the three other cell types. Abbreviations: HSPC, hematopoietic stem and progenitor cell; HSC, hematopoietic stem/multipotent progenitor cell; CMP, common myeloid progenitor; CLP/MLP, common/multipotent lymphoid progenitor; GMP, granulocyte-macrophage progenitor; MEP, megakaryocyte-erythrocyte progenitor; SSC, side scatter; FSC, forward scatter.

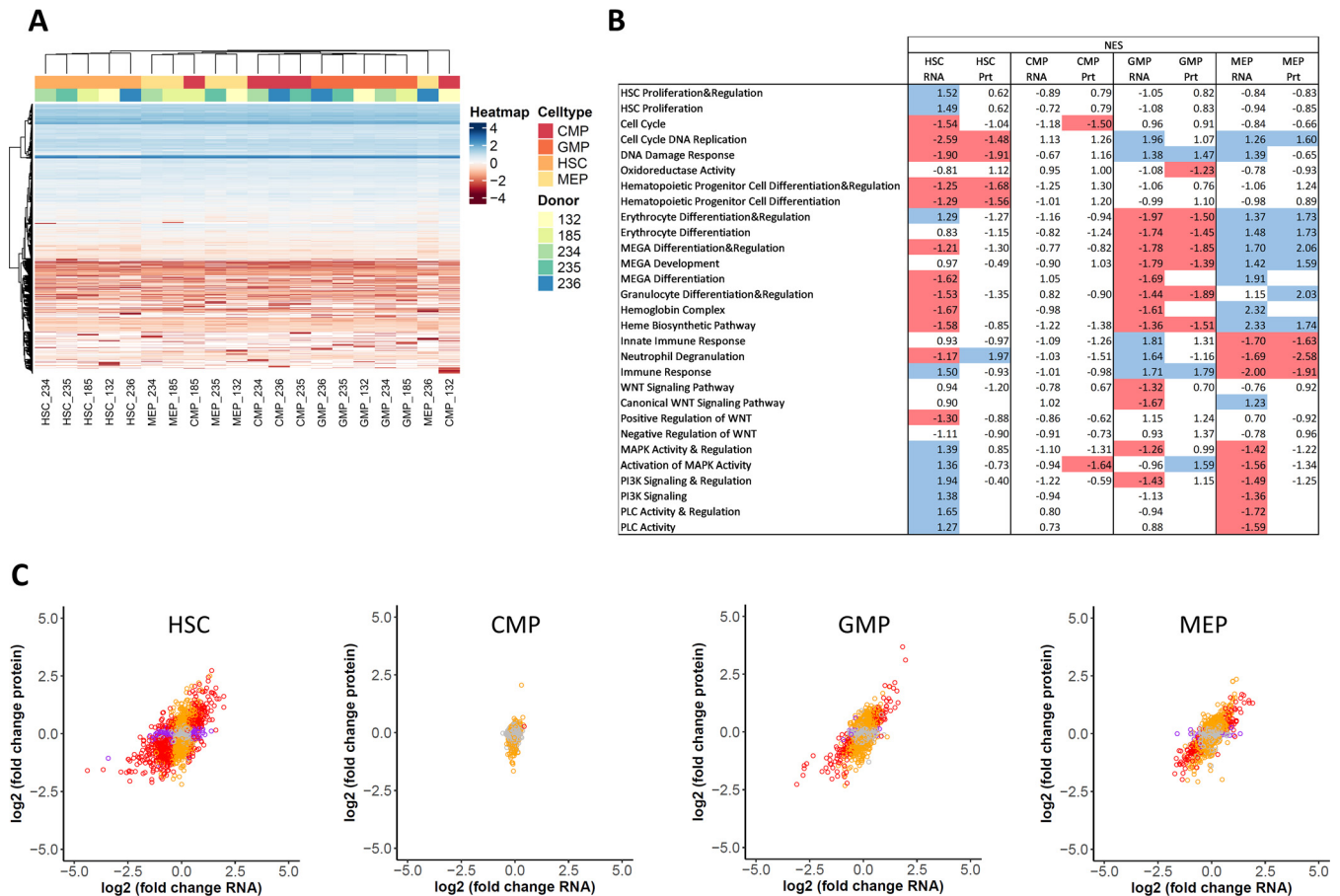
summary heatmap of these comparisons (Fig. 3C) indicates that the proteome profiles clustered according to cell subtype rather than donor. HSCs/MPPs clustered the furthest away from the other cell subpopulations, while CMPs were found to be more similar to GMPs and MEPs for some donors, in agreement with the ontogenetic distances expected between the different cell lineages (Fig. 3A).

Differentially expressed proteins were detected in the various cell subpopulations with roughly similar numbers of proteins being significantly up- or down-regulated, whereby significance was defined by the cutoffs of  $\text{FDR} < 0.01$  and  $\log_2(\text{fold change}) > 0.5$  (Fig. 3D and supplemental Fig. S14). Cutoffs were chosen based on the protein fold-changes observed in the volcano plots (supplemental Fig. S14A). Using these cutoffs and

comparing HSCs/MPPs to the average of all three remaining subpopulations (CMPs, GMPs, MEPs) (52), 1,008 proteins were determined to be differentially abundant in HSCs/MPPs. In accordance with the close ontogenetic distance of CMPs to GMPs and MEPs, the number of proteins with significantly changed expression between these cell types was somewhat lower: for GMPs 489, for MEPs 370, and for CMPs 64.

The availability of DIA data for each cell type from five donors allowed us to confidently identify proteins that were consistently detectable in some cell types but expressed below the detection limit of the measurement in other cell types and thus provided particularly important biological information (see supplemental Fig. S15). For example, the enzyme myeloperoxidase was consistently detectable at a low level in





**FIG. 4. Proteome-transcriptome correlation in human hematopoietic stem and progenitor cell subpopulations.** (A) Nonsupervised hierarchical clustering (Euclidean distance) heatmap (78) of the intensity for the transcripts identified in HSCs/MPPs (referred to as HSCs), CMPs, GMPs, and MEPs (shades of red) isolated from five different HSPC donors (shades of blue). The transcript intensities are centered and scaled and depicted in color shades from red to blue. The transcripts with missing transcript intensity values in all samples were removed because they could not be handled by the clustering algorithm. Remaining missing transcript intensity values are shown in white. Clustering was observed mostly according to cell type, not according to donor. (B) GO enrichment analysis showed good alignment of protein and mRNA data. GSEA was performed for ranked mRNA and protein lists using GO processes from the Gene Ontology Consortium database as gene sets. Shown are normalized enrichment scores for the individual gene sets. Significantly up-regulated gene sets are marked in blue color; significantly downregulated gene sets are marked in red color. Significance was defined as FDR < 0.25, specific cell subpopulations were compared with the average of the remaining three cell types, and log<sub>2</sub>(fold change) was used as ranking criterion. Empty fields mean that no enrichment could be calculated. Abbreviations: MEGA, megakaryocyte; MAPK, mitogen-activated protein kinase; PI3K, phosphoinositide-3-kinase; PLC, phospholipase C. (C) Correlation between proteomics and transcriptomics data for HSCs/MPPs (referred to as HSCs), CMPs, GMPs, and MEPs. Dots are depicted in red when the FDR was below 0.01 both for protein and RNA data, orange when FDR < 0.01 for protein data, purple when FDR < 0.01 for RNA data, and gray when FDR ≥ 0.01 for both protein and RNA data. (D) Network analysis of significantly up-regulated proteins with concomitant significantly downregulated mRNA in HSCs/MPPs. Two clusters were especially prominent, including the snoRNPs and telomerase complex proteins GAR1, DKC1, NOP10, NHP2, and the quiescence-inducing NAD(P)H-producing proteins IDH1, IDH3A, and IDH3B. HSCs/MPPs were compared with the average of the other three subpopulations; cutoffs were set at FDR < 0.01 for protein and RNA data. Colors depict GO terms found enriched in Fig 4B.

CMPs and at a much higher level in GMPs but was below the limit of detection in MEPs and HSCs/MPPs. The proliferation marker protein Ki-67 was not detected in HSCs/MPPs but was consistently detected in CMPs, GMPs, and MEPs. Similarly, all five members of the condensin-1 complex (SMC2, SMC4, NCAPD2, NCAPG, and NCAPH), a complex responsible for chromatin condensation, were present below the detection limit in HSCs/MPPs but could be clearly detected in the other cell types studied.

Overall, these results show the consistent detection of quantitative protein patterns that are characteristic for ontogenetically close cell types.

*Gene Ontology Enrichment Analyses for Proteomics and Transcriptomics Data*—From the same sorting experiments, a further 10,000 cells were isolated for RNAseq. Similar to the proteomic results, the transcriptomic data also revealed clustering mostly by cell type, rather than by donor (Fig. 4A). As for proteins, the HSCs/MPPs transcript profiles clustered furthest

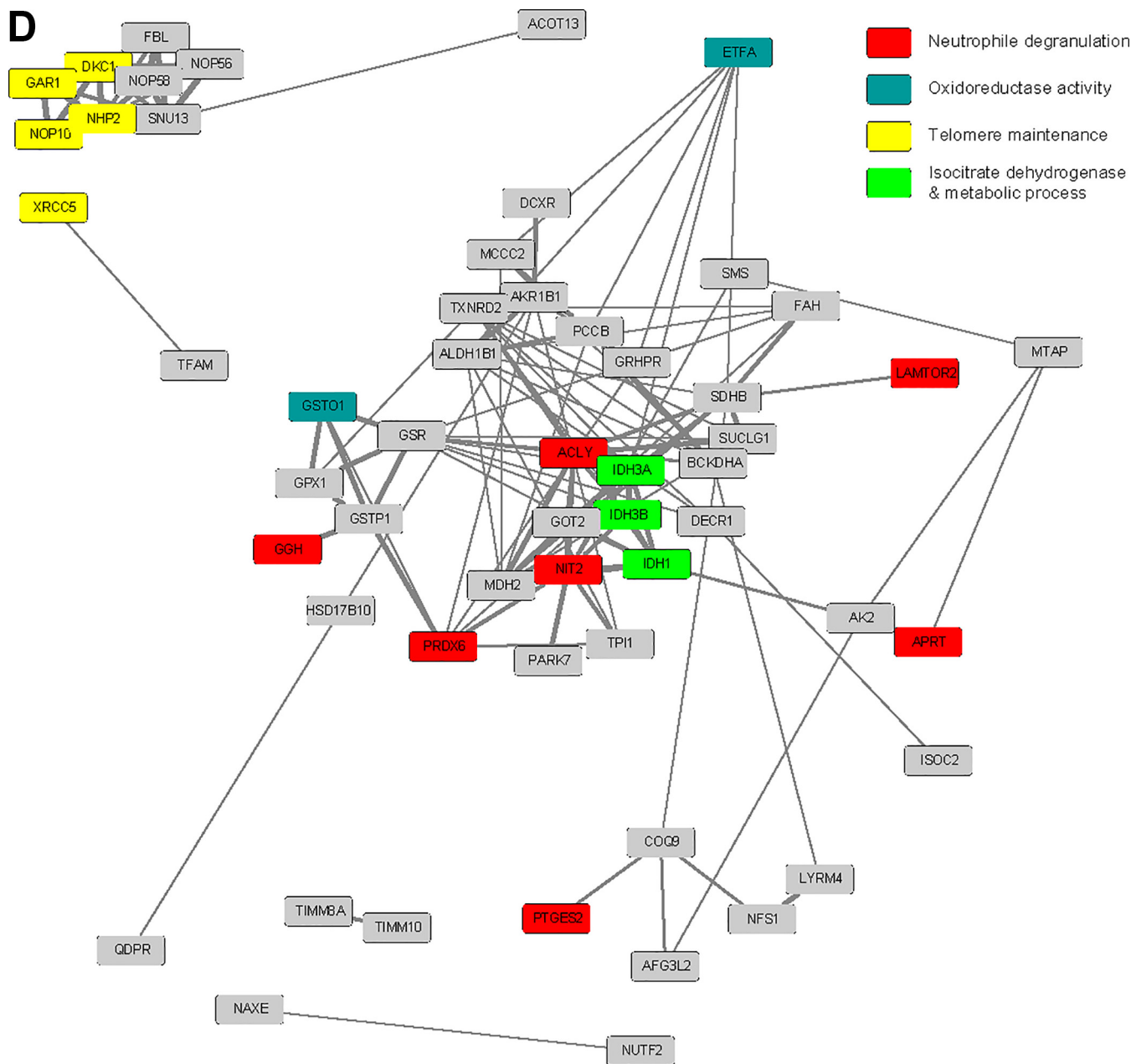


Fig. 4—continued

apart from the other cell subpopulations, whereas CMPs were more similar to GMPs and MEPs.

To further assess the proteomic and transcriptomic results, we performed gene set enrichment analyses for specific GO processes involved in hematopoietic stem cell differentiation according to previous studies and functional annotations (1, 2, 18, 19, 24, 48, 52, 53). The proteomic and transcriptomic data closely recapitulated most of the expected changes in GO processes (Fig. 4B, supplemental Figs. S16 and S17). Cell cycle/DNA replication/DNA damage response was found downregulated for HSCs/MPPs, which are more quiescent than progenitor cells, both at the protein and mRNA level.

Erythrocyte differentiation/megakaryocyte development/heme biosynthesis were observed to be up-regulated in MEPs at the protein and mRNA level. (Innate) immune responses were found up-regulated in GMPs at the protein and mRNA level. The canonical WNT pathway was observed to be up-regulated in MEPs at the mRNA level. Mitogen-activated protein kinase, phosphoinositide-3-kinase and phospholipase C pathways were all shown to be up-regulated in HSCs at the mRNA level (1, 2, 18, 19, 52, 53). Transcription factors Gata1 and Gata2 were also manually validated to be present in the CMP and MEP cell types only (supplemental Fig. S15), as expected.

Of note, the proteomic and transcriptomic results showed highest agreement of the GO enrichments for the GMPs and MEPs, both in terms of directionality (up or down) and significance of the pathway enrichments. For the CMPs, only a few significantly enriched pathways were observed, probably due to the position of CMPs between GMPs and MEPs in the developmental system. Interestingly, the HSCs/MPPs showed fewer significantly up- or downregulated GO processes for the proteomics data compared with the transcriptomics results (e.g. HSC proliferation, mitogen-activated protein kinase (MAPK) activity and regulation, phosphoinositide-3-kinase (PI3K) signaling, phospholipase C (PLC) activity (Fig. 4B)). This was in part due to the higher number of transcripts observed (17,355) compared with the number of detected protein groups (4,131).

Overall, the GO pathway analysis was in agreement with the expected properties of the respective cell types, thereby validating our protein and mRNA results against former studies.

**Discrepant Protein and mRNA Regulation in Highly Enriched HSCs/MPPs**—To investigate the complementary value of the protein quantitative information, we decided to examine proteins that were found differentially regulated compared with their mRNAs (Fig. 4C and supplemental Figs. S14B–14E).

The protein *versus* mRNA fold-change plots (Fig. 4C) showed good correlation for MEPs ( $R^2$  of 0.50) and GMPs ( $R^2$  of 0.41) but somewhat lower values for the HSCs/MPPs ( $R^2$  of 0.32) and CMPs ( $R^2$  of 0.06), in line with the results from GO enrichment analysis (Fig. 4B).

STRING analysis of proteins differentially regulated at the protein and transcript level revealed two major protein-protein association networks in HSCs/MPPs, the first including the small nucleolar ribonucleoproteins (snRNPs) and telomerase maintenance proteins GAR1, DKC1, NOP10, and NHP2 and the second, the quiescence-inducing NAD(P)H-producing isocitrate dehydrogenase proteins IDH1, IDH3A, and IDH3B (Fig. 4D, Experimental Procedures). Both association networks were up-regulated on protein (see also supplemental Figs. S15 and S18B) and downregulated on mRNA level in HSCs/MPPs (see supplemental Fig. S18A).

#### DISCUSSION

HSCs are mostly quiescent cells, *i.e.* a major fraction of cells is expected to be in G0 of the cell cycle. These cells have been shown to contain very low levels of mRNA, while still maintaining a relatively constant protein mass overall. Proteins may therefore provide a better readout of the cellular state of quiescent cells than the mRNAs. The same may be postulated for relatively quiescent (pre)leukemic and cancer stem cells. Until now, extensive global proteome data could not be obtained in critically rare cell types such as human HSPC subpopulations or (pre)leukemic and cancer stem cells. To achieve the most comprehensive and accurate protein quantification presently possible for minute, clinically relevant samples, we developed and applied an integrated sample

preparation/DIA-MS method, using an Orbitrap instrument. We report several critical observations for sample preparation to avoid loss of cell pellets and of hydrophobic peptides (supplemental Fig. S3) and regarding optimization of ion trap fill times in DIA (supplemental Fig. S4).

We applied our newly developed DIA acquisition scheme to the analysis of 25,000 FACS-isolated human HSPC subpopulations and could quantify more than 5,800 protein groups per cell subtype. Of these, we used a stringently filtered subset (4,131 protein groups) for further analysis. Importantly, our results also demonstrate that even though the number of protein identifications decreased when lowering the number of cells further to 12,500 or even 6,250 cells, quantification retained a very high level of accuracy (Figs. 1C and 2C, supplemental Figs. S7 and S9). In practice, this means that, even if fewer than 25,000 cells are available from FACS, protein quantification results can still be confidently compared with those from higher cell counts, if the proteins are detectable in both samples.

Analysis of key marker proteins with strong differences in expression between individual HSPC subpopulations demonstrated the expected results for the proteome data. In line with its role in neutrophil immune reactions and its localization in the alfa granules of granulocytes (54), the enzyme myeloperoxidase was not detected in HSCs or MEPs, seen at low protein levels in CMPs and at high levels in GMPs. Myeloperoxidase is a marker for granulocytes, and its presence can be used to distinguish between myeloid and lymphatic origins of acute leukemias (54). The presence of myeloperoxidase mainly in GMPs is thus in agreement with the differentiation potential of GMPs to granulocytes (55). Furthermore, in line with the quiescent state of HSCs/MPPs, proliferation and mitotic markers such as Ki-67 and condensin-1 complex members were not detected in this cell type (56, 57). These results provided validation of the proteome data in HSPC samples with critically low cell numbers, thereby allowing for comparisons of proteome and respective transcriptome data.

Proteome and corresponding transcriptome data showed the same clustering pattern. Moreover, cell subpopulations were more decisive for the clustering than genomic variability, indicating that the data quality achieved allowed us to detect biologically relevant protein patterns in a noisy background. Furthermore, enrichment analyses for GO terms in the different HSPC subpopulations reinforced the good alignment of protein and transcript data, thereby further validating the quality and information content of the proteomic results. Whereas GMPs and MEPs had very similar GO enrichment results for proteins and transcripts, HSCs/MPPs—though overall well aligned for proteomics and transcriptomics—showed fewer significantly up- or downregulated GO processes for the proteomics data, compared with the transcriptomics results (e.g. HSC proliferation, MAPK activity and regulation, PI3K signaling, PLC activity). This is in part due to the lower coverage achieved for proteins compared with tran-

scripts. In addition, this could also indicate the presence of alternative processes regulating the abundance of transcripts and proteins (58). Given the similar data quality (number of protein groups identified and quantified, CV) for the various analyzed cell types, the different behavior of HSCs/MPPs is likely to be biologically significant and may at least in part reflect the noncycling state of HSCs/MPPs as opposed to the cycling state of hematopoietic progenitor cells.

In line with these findings, proteins that showed discrepant regulation between their proteomic and transcriptomic data (59) were observed almost exclusively in HSCs/MPPs. Buffering of mRNA alterations at the level of protein concentrations is a well-known phenomenon, illustrating that transcript levels by themselves are not sufficient to predict protein levels in many scenarios (60). In HSCs/MPPs two clusters were identified, which are up-regulated on the protein level while downregulated on the mRNA level in HSCs/MPPs. The first is a strongly interconnected protein module that includes several snoRNPs and telomerase maintenance proteins and is deemed essential for long-lived stem cells (61–63). Telomerase activity in hematopoietic cells is associated with self-renewal potential and has been shown to decrease upon myeloid differentiation (62). Mutations in these telomerase maintenance proteins result in dyskeratosis congenita, a syndrome characterized by bone marrow failure and an increased risk for acute myeloid leukemia and myelodysplastic syndromes (64). Differential protein and mRNA regulation has been reported for these proteins and was attributed to post-translational mechanisms (63, 65). The second cluster consisted of the IDH proteins IDH1, IDH3A, and IDH3B that have previously been shown to maintain quiescence in hair follicle stem cells (66). IDHs are also thought to play a key role in hematopoietic stem cell homeostasis and were reported to be mutated in ~20% of acute myeloid leukemias (67, 68). IDH proteins catalyze the oxidative decarboxylation of isocitrate to alpha-ketoglutarate and are involved in adaptation to hypoxia, histone demethylation, and DNA modification (69). IDH1 is a cytosolic/peroxisomal homodimer whereas IDH3 is a mitochondrial heterotetramer composed of two alpha, one beta, and one gamma subunits (70–72). Mutant IDH enzymes have neomorphic activity, leading to the formation of the (R) enantiomer of 2-hydroxyglutarate and causing DNA and histone hypermethylation, altered gene expression, and blocked differentiation of hematopoietic progenitor cells (69). Acute myeloid leukemia treatments targeted at mutant IDH proteins have entered clinical routine, such as the IDH2 inhibitor enasidenib that received FDA approval for the treatment of relapsed or refractory acute myeloid leukemia on August 1, 2017 ([www.fda.gov](http://www.fda.gov)). Other IDH inhibitors are currently being evaluated in clinical trials (69).

qPCR validation experiments supported down-regulation of these targets in HSCs relative to the other HSPC subpopulations, whereas flow cytometry validation experiments con-

firmed high IDH1 protein expression in HSCs ([supplemental Figs. S18A, S18B, and S19](#)). In contrast to the proteomics data, equally high IDH1 flow cytometry intensity was seen in HSCs relative to CMPs, which could be due to differences in subcellular distribution of IDH1 in for example peroxisomes (71, 72).

These examples illustrate the relevance of generating high-quality proteomic data for well-defined cell subpopulations for the identification of biological processes that cannot be detected by genomic or transcriptomic analysis. Though this seems particularly evident for quiescent cells, we expect that proteomic data will bring an invaluable layer of biological information complementary to that of the transcriptomic data for many other cell subtypes. The presented application of DIA to acquire robust protein quantification data on low sample loads can thus be expected to increase our understanding of the dynamics of cell type-specific networks and to complete our knowledge on differentiation processes at play in healthy and pathological numerically scarce cells (19, 52).

In future research, it will be important to further refine the different cell subpopulations. CD34+CD38-CD45RA- HSCs/MPPs, CD34+CD38+CD123+CD45RA- CMPs, CD34+CD38+CD123+CD45RA+ GMPs, and CD34+CD38+CD123-CD45RA-MEPs could indeed be further divided into biologically even more refined subpopulations (2, 16). Also, proteomic data from cell types isolated directly from bone marrow or from cord blood rather than from mobilized HSPCs obtained from donors after artificial stimulation will need to be obtained. Those new samples will enable to increase the number of peptide assays in the spectral library, potentially allowing to identify additional proteins in the present and future DIA datasets. Further fine tuning of the sample handling steps may be necessary to cope with possibly even lower sample amounts than those analyzed in this study. The dilution series in HEK cells demonstrated a faster than linear signal drop-off for hydrophobic peptides suggesting adsorptive losses to surfaces. Possible avenues to explore would be to miniaturize cell lysis and digestion with the help of novel microfluidic devices (73) and to use LC columns with further reduced inner diameters (74). Thus, by combining improvements in sample processing and instrument design, it can be expected that the method will allow to robustly quantify proteins from even lower sample loads. Furthermore, transcription factors are of low abundance and are therefore difficult to detect, especially by large-scale proteomic analysis without fractionation. Specifically concerning the Hox and Gata genes, we checked our spectral library and confirmed that it contained peptide assays for 12 HOX proteins and 4 GATA proteins, almost all exclusively originating from the human HEK fractionation library. Future developments in DIA-MS analysis are required for reliable detection of such key low abundance proteins.

In summary, we describe a sensitive mass spectrometric method that allows generating highly accurate and reproduc-

ible protein quantification data from minute amounts of highly homogeneous cell subpopulations enriched by FACS. This technology allows dissecting the biochemical processes in play in specific cell subpopulations of interest with unprecedented sensitivity, depth of coverage, and reproducibility. Thereby, it paves the way for global proteomic analyses in clinically highly relevant but numerically scarce cell populations such as (pre)leukemic stem cells in hematopoietic malignancies as well as cancer stem cells from solid tumors.

**Acknowledgments**—We are grateful to Audrey van Droogen (ETH Zurich) for providing an aliquot of HEK293 cells; Ian Lienert, Lukas Reiter, and Oliver Bernhardt (Biognosys AG, Schlieren) for the Fill Time Reporter software to extract the fill times from the raw data files; and Alexander Leitner (ETH Zurich) for support with MS method development and manuscript preparation. We thank Patrizia Belleda and Asuka Fry (University Hospital Zurich) for the initial processing of donor samples and Claudia Dumrese (University of Zurich) for assistance with FACS. We acknowledge Witold Wolski (Functional Genomics Center Zurich) for assistance with the statistical analysis of the omics data and data interpretation. We thank Patrick Pedrioli (ETH Zurich) for providing the script to export MSMS spectra for the single-protein hits.

This study was supported by two grants to Fabienne Meier-Abt (Filling the Gap Grant of the University of Zurich; Krebsliga Zurich Grant) and the Clinical Research Priority Program Human Hematolymphatic Diseases of the University of Zurich (Markus G. Manz). Alexandre P.A. Theocharides was supported by the Clôëtta Foundation. Sabine Amon was supported by the ERC grant HLA-DR15 in MS (FP7-IDEAS-340733); Ludovic Gillet was supported by the grants Proteomics 4D (ERC-2014-AdG 670821) and PrECISE (H2020-EU.3.1 668858). Access to the instrumentation was supported by the following grants to Ruedi Aebersold: Proteomics 4D (ERC-2014-AdG 670821), ULTRA-DD (FP7-JTI 115766), and ETH Scientific Equipment.

#### DATA AVAILABILITY

The mass spectrometric DDA and DIA raw data files and associated search results of the HEK peptide dilution series have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository (75, 76) with the dataset identifier PXD009246. The mass spectrometric DIA raw data files and associated search results of the CD34+ FACS-isolated dilution series and FACS-isolated stem and progenitor cells have been deposited with the dataset identifier PXD009255. RNA sequencing data have been deposited in NCBI's Gene Expression Omnibus (GEO) (77) (<https://www.ncbi.nlm.nih.gov/geo/>) with the accession number GSE113182.

\* R.A. holds shares of Biognosys AG, which operates in the field covered by the article. The remaining authors declare no competing financial interests.

☐ This article contains supplemental material Tables S1–S6 and Figs. S1–S19.

\*\* To whom correspondence should be addressed. E-mail: aebersold@imbsb.biol.ethz.ch.

‡‡ These authors contributed equally to this work.

Author contributions: S.A., F.M.-A., and L.C.G. performed research; S.A., F.M.-A., L.C.G., and S.D. analyzed data; S.A., F.M.-A., L.C.G.,

and R.A. wrote the paper; and A.P.A.T., M.G.M., and R.A. designed research.

#### REFERENCES

- Laurenti, E., Doulatov, S., Zandi, S., Plumb, I., Chen, J., April, C., Fan, J. B., and Dick, J. E. (2013) The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nat Immunol* **14**, 756–763
- Notta, F., Zandi, S., Takayama, N., Dobson, S., Gan, O. I., Wilson, G., Kaufmann, K. B., McLeod, J., Laurenti, E., Dunant, C. F., McPherson, J. D., Stein, L. D., Dror, Y., and Dick, J. E. (2016) Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351**, aab2116
- Jan, M., Snyder, T. M., Corces-Zimmerman, M. R., Vyas, P., Weissman, I. L., Quake, S. R., and Majeti, R. (2012) Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci Transl Med* **4**, 149ra118
- Corces-Zimmerman, M. R., Hong, W. J., Weissman, I. L., Medeiros, B. C., and Majeti, R. (2014) Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proc Natl Acad Sci U.S.A.* **111**, 2548–2553
- Shlush, L. I., Zandi, S., Mitchell, A., Chen, W. C., Brandwein, J. M., Gupta, V., Kennedy, J. A., Schimmer, A. D., Schuh, A. C., Yee, K. W., McLeod, J. L., Doedens, M., Medeiros, J. J., Marke, R., Kim, H. J., Lee, K., McPherson, J. D., Hudson, T. J., HALT Pan-Leukemia Gene Panel Consortium, Brown, A. M., Yousif, F., Trinh, Q. M., Stein, L. D., Minden, M. D., Wang, J. C., and Dick, J. E. (2014) Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* **506**, 328–333
- Vedi, A., Santoro, A., Dunant, C. F., Dick, J. E., and Laurenti, E. (2016) Molecular landscapes of human hematopoietic stem cells in health and leukemia. *Ann. N.Y. Acad. Sci.* **1370**, 5–14
- Notta, F., Doulatov, S., Laurenti, E., Poepl, A., Jurisica, I., and Dick, J. E. (2011) Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science* **333**, 218–221
- Shlush, L. I., Mitchell, A., Heisler, L., Abelson, S., Ng, S. W. K., Trotman-Grant, A., Medeiros, J. J. F., Rao-Bhatia, A., Jaciw-Zurakowsky, I., Marke, R., McLeod, J. L., Doedens, M., Bader, G., Voisin, V., Xu, C., McPherson, J. D., Hudson, T. J., Wang, J. C. Y., Minden, M. D., and Dick, J. E. (2017) Tracing the origins of relapse in acute myeloid leukaemia to stem cells. *Nature* **547**, 104–108
- Behbehani, G. K., Samusik, N., Bjornson, Z. B., Fantl, W. J., Medeiros, B. C., and Nolan, G. P. (2015) Mass cytometric functional profiling of acute myeloid leukemia defines cell-cycle and immunophenotypic properties that correlate with known responses to therapy. *Cancer Discov* **5**, 988–1003
- Tallerico, R., Conti, L., Lanzardo, S., Sottile, R., Garofalo, C., Wagner, A. K., Johansson, M. H., Cristiani, C. M., Karre, K., Carbone, E., and Cavallo, F. (2017) NK cells control breast cancer and related cancer stem cell hematological spread. *Oncoimmunology* **6**, e1284718
- Luo, M., Clouthier, S. G., Deol, Y., Liu, S., Nagrath, S., Azizi, E., and Wicha, M. S. (2015) Breast cancer stem cells: Current advances and clinical implications. *Methods Mol Biol.* **1293**, 1–49
- Ni, X., Zhuo, M., Su, Z., Duan, J., Gao, Y., Wang, Z., Zong, C., Bai, H., Chapman, A. R., Zhao, J., Xu, L., An, T., Ma, Q., Wang, Y., Wu, M., Sun, Y., Wang, S., Li, Z., Yang, X., Yong, J., Su, X. D., Lu, Y., Bai, F., Xie, X. S., and Wang, J. (2013) Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. *Proc Natl Acad Sci U.S.A.* **110**, 21083–21088
- Alva, A., Friedlander, T., Clark, M., Huebner, T., Daignault, S., Hussain, M., Lee, C., Hafez, K., Hollenbeck, B., Weizer, A., Premasekharan, G., Tran, T., Fu, C., Ionescu-Zanetti, C., Schwartz, M., Fan, A., and Paris, P. (2015) Circulating tumor cells as potential biomarkers in bladder cancer. *J. Urol.* **194**, 790–798
- Azizi, E., Carr, A. J., Plitas, G., Cornish, A. E., Konopacki, C., Prabhakaran, S., Nainys, J., Wu, K., Kiseliovas, V., Setty, M., Choi, K., Fromme, R. M., Dao, P., McKenney, P. T., Wasti, R. C., Kadaveru, K., Mazutis, L., Rudensky, A. Y., and Pe'er, D. (2018) Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174**, 1293–1308
- Manz, M. G., Miyamoto, T., Akashi, K., and Weissman, I. L. (2002) Prospective isolation of human clonogenic common myeloid progenitors. *Proc Natl Acad Sci U.S.A.* **99**, 11872–11877

16. Doulatov, S., Notta, F., Eppert, K., Nguyen, L. T., Ohashi, P. S., and Dick, J. E. (2010) Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat Immunol.* **11**, 585–593
17. Mori, Y., Chen, J. Y., Pluvinage, J. V., Seita, J., and Weissman, I. L. (2015) Prospective isolation of human erythroid lineage-committed progenitors. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 9638–9643
18. Terskikh, A. V., Miyamoto, T., Chang, C., Diatchenko, L., and Weissman, I. L. (2003) Gene expression analysis of purified hematopoietic stem cells and committed progenitors. *Blood* **102**, 94–101
19. Qiao, W., Wang, W., Laurenti, E., Turinsky, A. L., Wodak, S. J., Bader, G. D., Dick, J. E., and Zandstra, P. W. (2014) Intercellular network structure and regulatory motifs in the human hematopoietic system. *Mol. Syst. Biol.* **10**, 741
20. Kasuga, K., Katoh, Y., Nagase, K., and Igarashi, K. (2017) Microproteomics with microfluidic-based cell sorting: Application to 1000 and 100 immune cells. *Proteomics* **17**, 13–14
21. Martin, J. G., Rejtar, T., and Martin, S. A. (2013) Integrated microscale analysis system for targeted liquid chromatography mass spectrometry proteomics on limited amounts of enriched cell populations. *Anal. Chem.* **85**, 10680–10685
22. Li, S., Plouffe, B. D., Belov, A. M., Ray, S., Wang, X., Murthy, S. K., Karger, B. L., and Ivanov, A. R. (2015) An integrated platform for isolation, processing, and mass spectrometry-based proteomic profiling of rare cells in whole blood. *Mol. Cell. Proteomics* **14**, 1672–1683
23. Di Palma, S., Stange, D., van de Wetering, M., Clevers, H., Heck, A. J., and Mohammed, S. (2011) Highly sensitive proteome analysis of FACS-sorted adult colon stem cells. *J. Proteome Res.* **10**, 3814–3819
24. Cabezas-Wallscheid, N., Klimmeck, D., Hansson, J., Lipka, D. B., Reyes, A., Wang, Q., Weichenhan, D., Lier, A., von Paleske, L., Renders, S., Wunsche, P., Zeisberger, P., Brocks, D., Gu, L., Herrmann, C., Haas, S., Essers, M. A. G., Brors, B., Eils, R., Huber, W., Milsom, M. D., Plass, C., Krijgsvelde, J., and Trumpp, A. (2014) Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell* **15**, 507–522
25. Gillet, L. C., Leitner, A., and Aebersold, R. (2016) Mass spectrometry applied to bottom-up proteomics: Entering the high-throughput era for hypothesis testing. *Annu. Rev. Anal. Chem. (Palo Alto Calif)* **9**, 449–472
26. Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, O111.016717
27. Ting, Y. S., Egerton, J. D., Payne, S. H., Kim, S., MacLean, B., Kall, L., Aebersold, R., Smith, R. D., Noble, W. S., and MacCoss, M. J. (2015) Peptide-centric proteome analysis: An alternative strategy for the analysis of tandem mass spectrometry data. *Mol. Cell. Proteomics* **14**, 2301–2307
28. Collins, B. C., Gillet, L. C., Rosenberger, G., Röst, H. L., Vichalkovski, A., Gstaiger, M., and Aebersold, R. (2013) Quantifying protein interaction dynamics by SWATH mass spectrometry: Application to the 14-3-3 system. *Nat. Methods* **10**, 1246–1253
29. Liu, Y., Buil, A., Collins, B. C., Gillet, L. C., Blum, L. C., Cheng, L. Y., Vitek, O., Mouritsen, J., Lachance, G., Spector, T. D., Dermitzakis, E. T., and Aebersold, R. (2015) Quantitative variability of 342 plasma proteins in a human twin population. *Mol. Syst. Biol.* **11**, 786
30. Egerton, J. D., Kuehn, A., Merrihew, G. E., Bateman, N. W., MacLean, B. X., Ting, Y. S., Canterbury, J. D., Marsh, D. M., Kellmann, M., Zabrouskov, V., Wu, C. C., and MacCoss, M. J. (2013) Multiplexed MS/MS for improved data-independent acquisition. *Nat. Methods* **10**, 744–746
31. Bruderer, R., Bernhardt, O. M., Gandhi, T., Miladinovic, S. M., Cheng, L. Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C., Vitek, O., Rinner, O., and Reiter, L. (2015) Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell. Proteomics* **14**, 1400–1410
32. Teo, G., Kim, S., Tsou, C. C., Collins, B., Gingras, A. C., Nesvizhskii, A. I., and Choi, H. (2015) mapDIA: Preprocessing and statistical analysis of quantitative proteomics data from data independent acquisition mass spectrometry. *J. Proteomics* **129**, 108–120
33. Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670
34. Escher, C., Reiter, L., MacLean, B., Ossola, R., Herzog, F., Chilton, J., MacCoss, M. J., and Rinner, O. (2012) Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **12**, 1111–1121
35. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567
36. Eng, J. K., Jahan, T. A., and Hoopmann, M. R. (2013) Comet: An open-source MS/MS sequence database search tool. *Proteomics* **13**, 22–24
37. Schubert, O. T., Gillet, L. C., Collins, B. C., Navarro, P., Rosenberger, G., Wolski, W. E., Lam, H., Amodè, D., Mallick, P., MacLean, B., and Aebersold, R. (2015) Building high-quality assay libraries for targeted analysis of SWATH MS data. *Nat. Protoc.* **10**, 426–441
38. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392
39. Deutsch, E. W., Mendoza, L., Shteynberg, D., Farrah, T., Lam, H., Tasman, N., Sun, Z., Nilsson, E., Pratt, B., Prazen, B., Eng, J. K., Martin, D. B., Nesvizhskii, A. I., and Aebersold, R. (2010) A guided tour of the Trans-Proteomic Pipeline. *Proteomics* **10**, 1150–1159
40. Shteynberg, D., Deutsch, E. W., Lam, H., Eng, J. K., Sun, Z., Tasman, N., Mendoza, L., Moritz, R. L., Aebersold, R., and Nesvizhskii, A. I. (2011) iProphet: Multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol. Cell. Proteomics* **10**, M111.007690
41. Reiter, L., Claassen, M., Schrimpf, S. P., Jovanovic, M., Schmidt, A., Buhmann, J. M., Hengartner, M. O., and Aebersold, R. (2009) Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. *Mol. Cell. Proteomics* **8**, 2405–2417
42. Lam, H., Deutsch, E. W., Eddes, J. S., Eng, J. K., King, N., Stein, S. E., and Aebersold, R. (2007) Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* **7**, 655–667
43. Lam, H., Deutsch, E. W., Eddes, J. S., Eng, J. K., Stein, S. E., and Aebersold, R. (2008) Building consensus spectral libraries for peptide identification in proteomics. *Nat. Methods* **5**, 873–875
44. Picelli, S., Faridani, O. R., Bjorklund, A. K., Winberg, G., Sagasser, S., and Sandberg, R. (2014) Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181
45. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
46. Liao, Y., Smyth, G. K., and Shi, W. (2013) The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108
47. Love, M. I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550
48. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550
49. Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N. T., Roth, A., Bork, P., Jensen, L. J., and von Mering, C. (2017) The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368
50. Shariatmadar, S., Sharma, S., Cabana, R., Powell, S., Ruiz, P., and Krishan, A. (2008) Electronic volume of CD34 positive cells from peripheral blood apheresis samples. *Cytometry B Clin. Cytom.* **74**, 182–188
51. Mateus, A., Matsson, P., and Artursson, P. (2013) Rapid measurement of intracellular unbound drug concentrations. *Mol. Pharm.* **10**, 2467–2478
52. Novershtern, N., Subramanian, A., Lawton, L. N., Mak, R. H., Haining, W. N., McConkey, M. E., Habib, N., Yosef, N., Chang, C. Y., Shay, T., Frampton, G. M., Drake, A. C., Leskov, I., Nilsson, B., Preffer, F., Domb-

- kowski, D., Evans, J. W., Liefeld, T., Smutko, J. S., Chen, J., Friedman, N., Young, R. A., Golub, T. R., Regev, A., and Ebert, B. L. (2011) Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296–309
53. Tarafdar, A., Dobbin, E., Corrigan, P., Freeburn, R., and Wheadon, H. (2013) Canonical Wnt signaling promotes early hematopoietic progenitor formation and erythroid specification during embryonic stem cell differentiation. *PLoS ONE* **8**, e81030
54. Ravnsborg, T., Houen, G., and Hojrup, P. (2010) The glycosylation of myeloperoxidase. *Biochim. Biophys. Acta* **1804**, 2046–2053
55. Berliner, N., Hsing, A., Graubert, T., Sigurdsson, F., Zain, M., Bruno, E., and Hoffman, R. (1995) Granulocyte colony-stimulating factor induction of normal human bone marrow progenitors results in neutrophil-specific gene expression. *Blood* **85**, 799–803
56. Zhu, L., Xing, S., Zhang, L., Yu, J. M., Lin, C., and Yang, W. J. (2017) Involvement of Polo-like kinase 1 (Plk1) in quiescence regulation of cancer stem-like cells of the gastric cancer cell lines. *Oncotarget* **8**, 37633–37645
57. Hirota, T., Gerlich, D., Koch, B., Ellenberg, J., and Peters, J. M. (2004) Distinct functions of condensin I and II in mitotic chromosome assembly. *J. Cell Sci.* **117**, 6435–6445
58. Gunaratne, J., Schmidt, A., Quandt, A., Neo, S. P., Saraç, O. S., Gracia, T., Loguerio, S., Ahrné, E., Xia, R. L., Tan, K. H., Lossner, C., Bähler, J., Beyer, A., Blackstock, W., and Aebersold, R. (2013) Extensive mass spectrometry-based analysis of the fission yeast proteome: The *Schizosaccharomyces pombe* PeptideAtlas. *Mol. Cell. Proteomics* **12**, 1741–1751
59. Haas, S., Hansson, J., Klimmeck, D., Loeffler, D., Velten, L., Uckelmann, H., Wurzer, S., Prendergast, A. M., Schnell, A., Hexel, K., Santarella-Mellwig, R., Blaszkiewicz, S., Kuck, A., Geiger, H., Millsom, M. D., Steinmetz, L. M., Schroeder, T., Trumpp, A., Krijgsveld, J., and Essers, M. A. (2015) Inflammation-induced emergency megakaryopoiesis driven by hematopoietic stem cell-like megakaryocyte progenitors. *Cell Stem Cell* **17**, 422–434
60. Liu, Y., Beyer, A., and Aebersold, R. (2016) On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550
61. Allsopp, R. C., Morin, G. B., DePinho, R., Harley, C. B., and Weissman, I. L. (2003) Telomerase is required to slow telomere shortening and extend replicative lifespan of HSCs during serial transplantation. *Blood* **102**, 517–520
62. Morrison, S. J., Prowse, K. R., Ho, P., and Weissman, I. L. (1996) Telomerase activity in hematopoietic cells is associated with self-renewal potential. *Immunity* **5**, 207–216
63. Lin, P., Mobasher, M. E., Hakakian, Y., Kakarla, V., Naseem, A. F., Ziai, H., and Alawi, F. (2015) Differential requirements for H/ACA ribonucleoprotein components in cell proliferation and response to DNA damage. *Histochem. Cell Biol.* **144**, 543–558
64. Vulliamy, T., Beswick, R., Kirwan, M., Marrone, A., Digweed, M., Walne, A., and Dokal, I. (2008) Mutations in the telomerase component NHP2 cause the premature ageing syndrome dyskeratosis congenita. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 8073–8078
65. Westman, B. J., and Lamond, A. I. (2011) A role for SUMOylation in snoRNP biogenesis revealed by quantitative proteomics. *Nucleus* **2**, 30–37
66. Ambrus, A. M., Lemons, J., Raitman, I., Krishnan, N., Jelinek, D., Suh, E., Remillard, M., Shilo, N., Haley, E., Sun, L., Xiao, R., Flores, A., Wang, D., White, A., Lowry, W., and Collier, H. (2016) The NADPH-production enzyme isocitrate dehydrogenase maintains quiescence in hair follicle stem cells. *FASEB J.* **30**, 1260.5
67. Figueroa, M. E., Abdel-Wahab, O., Lu, C., Ward, P. S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H. F., Tallman, M. S., Sun, Z., Wolniak, K., Peeters, J. K., Liu, W., Choe, S. E., Fantin, V. R., Paietta, E., Lowenberg, B., Licht, J. D., Godley, L. A., Delwel, R., Valk, P. J., Thompson, C. B., Levine, R. L., and Melnick, A. (2010) Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567
68. Montalban-Bravo, G., and DiNardo, C. D. (2018) The role of IDH mutations in acute myeloid leukemia. *Future Oncol.* **14**, 979–993
69. Medeiros, B. C., Fathi, A. T., DiNardo, C. D., Pollyea, D. A., Chan, S. M., and Swords, R. (2017) Isocitrate dehydrogenase mutations in myeloid malignancies. *Leukemia* **31**, 272–281
70. Bzymek, K. P., and Colman, R. F. (2007) Role of alpha-Asp181, beta-Asp192, and gamma-Asp190 in the distinctive subunits of human NAD-specific isocitrate dehydrogenase. *Biochemistry* **46**, 5391–5397
71. Antonenkov, V. D., and Hiltunen, J. K. (2012) Transfer of metabolites across the peroxisomal membrane. *Biochim. Biophys. Acta* **1822**, 1374–1386
72. Farr, R. L., Lismont, C., Terlecky, S. R., and Fransen, M. (2016) Peroxisome biogenesis in mammalian cells: The impact of genes and environment. *Biochim. Biophys. Acta* **1863**, 1049–1060
73. Zhu, Y., Piehowski, P. D., Zhao, R., Chen, J., Shen, Y., Moore, R. J., Shukla, A. K., Petyuk, V. A., Campbell-Thompson, M., Mathews, C. E., Smith, R. D., Qian, W. J., and Kelly, R. T. (2018) Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nat. Commun.* **9**, 882
74. Zhu, Y., Zhao, R., Piehowski, P. D., Moore, R. J., Lim, S., Orphan, V. J., Pasa-Tolic, L., Qian, W. J., Smith, R. D., and Kelly, R. T. (2018) Subnanogram proteomics: Impact of LC column selection, MS instrumentation and data analysis strategy on proteome coverage for trace samples. *Int. J. Mass Spectrom.* **427**, 4–10
75. Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Rios, D., Dianes, J. A., Sun, Z., Farrah, T., Bandeira, N., Binz, P. A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R. J., Kraus, H. J., Albar, J. P., Martinez-Bartolome, S., Apweiler, R., Omenn, G. S., Martens, L., Jones, A. R., and Hermjakob, H. (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32**, 223–226
76. Vizcaino, J. A., Csordas, A., Del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q. W., Wang, R., and Hermjakob, H. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, 11033
77. Edgar, R., Domrachev, M., and Lash, A. E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210
78. Gu, Z., Eils, R., and Schlesner, M. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849