Cell Reports, Volume *16*

Supplemental Information

Comprehensive Proteomic Analysis of Human Erythropoiesis

Emilie-Fleur Gautier, Sarah Ducamp, Marjorie Leduc, Virginie Salnot, François Guillonneau, Michael Dussiot, John Hale, Marie-Catherine Giarratana, Anna Raimbault, Luc Douay, Catherine Lacombe, Narla Mohandas, Frédérique Verdier, Yael Zermati, and Patrick Mayeux

Supplemental information

Supplemental figures

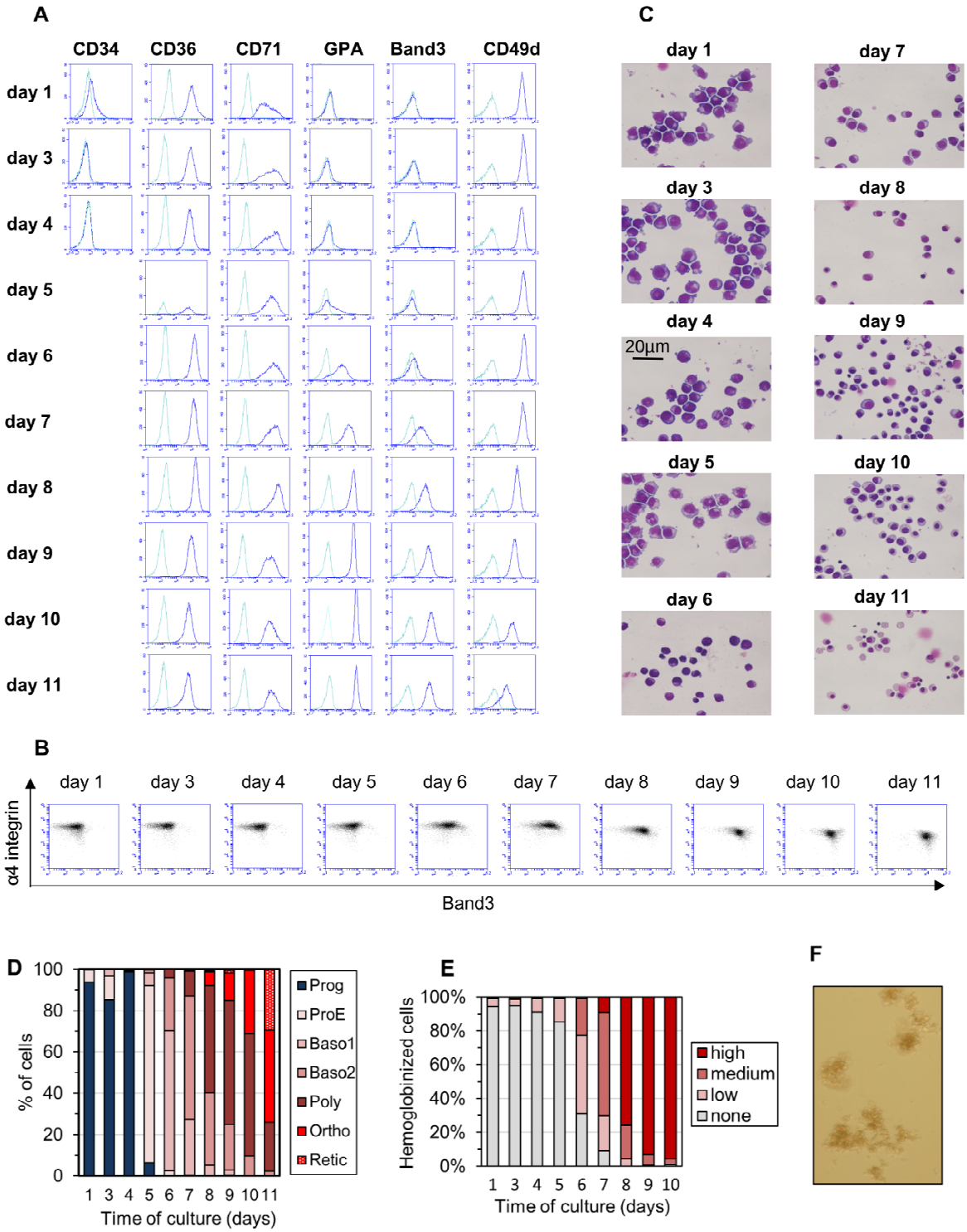


Figure S1 related to figure 2: Overall analysis of a typical culture

Complete analysis of culture 3 is presented. A: Complete phenotyping. B: Co-expression of Integrin α4 (CD49d) and Band3.C: MGG staining of cytocentrifuged cells. D: Repartition of cells among differentiation stages according to MGG staining. E: Cell hemoglobinization determined by benzidine staining.F: Clonogenic culture of D1 cells. The picture was recorded after 10 days of culture in methylcellulose in the presence of Epo alone.

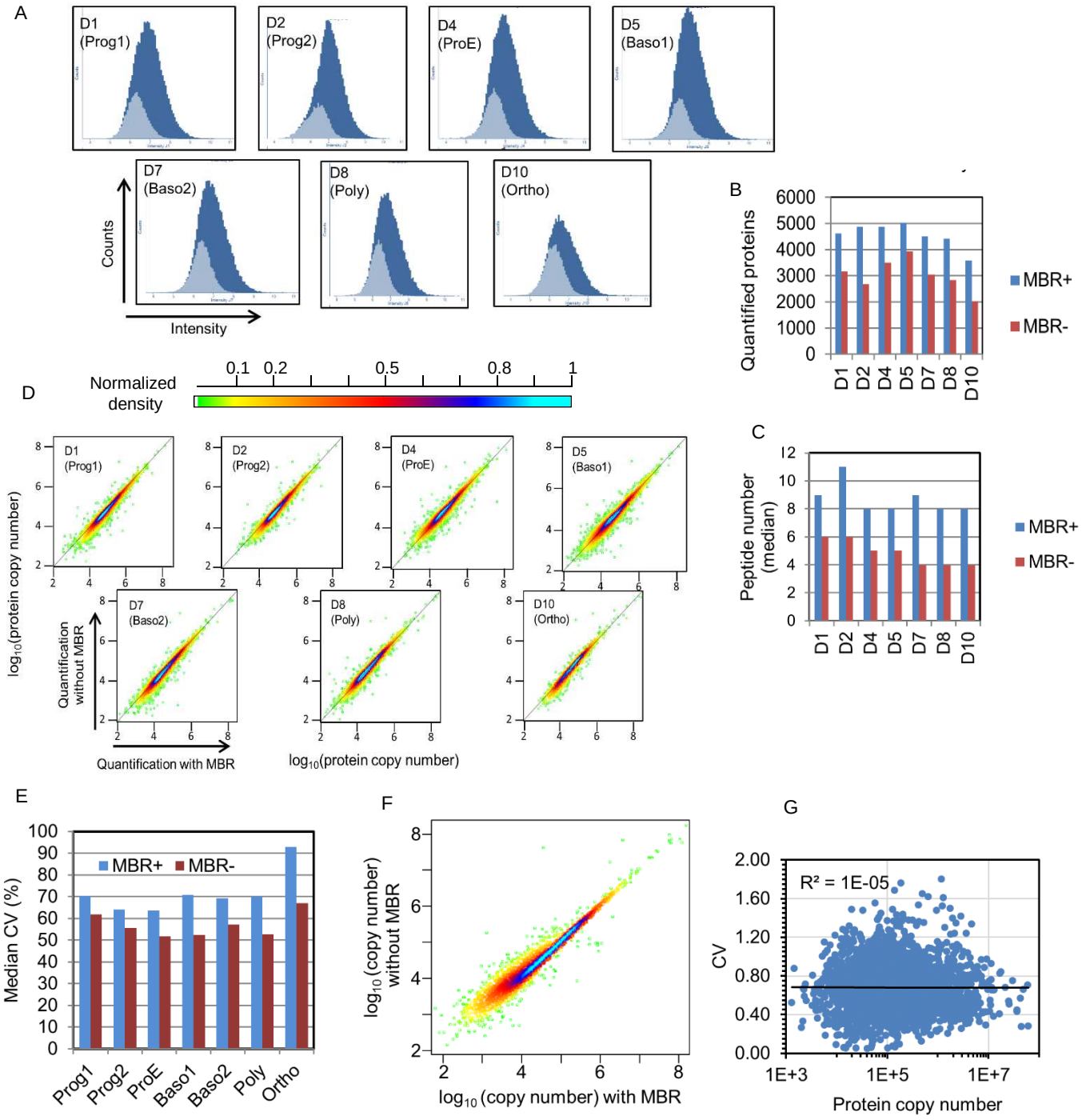


Figure S2 related to figure 3: Reliability of the MBR algorithm

A : Peptides used for LFQ quantification identified by MS/MS (dark blue) or by identity propagation using MBR (light blue). Data from experiment E1 are presented. B: Number of quantified proteins in the same samples after Maxquant analysis with (blue bars) or without (orange bars) MBR. C: Median number of peptides used for protein quantification in the same samples from experiment E1. Only proteins quantified both with and without MBR were taken into account. D: Comparison of absolute quantification values determined with and without MBR in the samples of experiment E1. E: The coefficient of variability (CV) of the protein copy number per cell was determined for each protein identified with or without MBR in the four experiments. The median CV value for each differentiation stage is presented. F: Comparison of absolute quantification values determined with and without MBR at the Ortho stage in the four experiments. G: Relationship between protein expression level and CV. Four independent quantifications of UT7 cell proteins were performed together with cellular extracts from K562 cells and from primary human erythroblasts. Data were analyzed using MaxQuant with the MBR algorithm. The CV of UT7 proteins quantified in a least 3 experiments was calculated and expressed as a function of the protein copy number per cell.

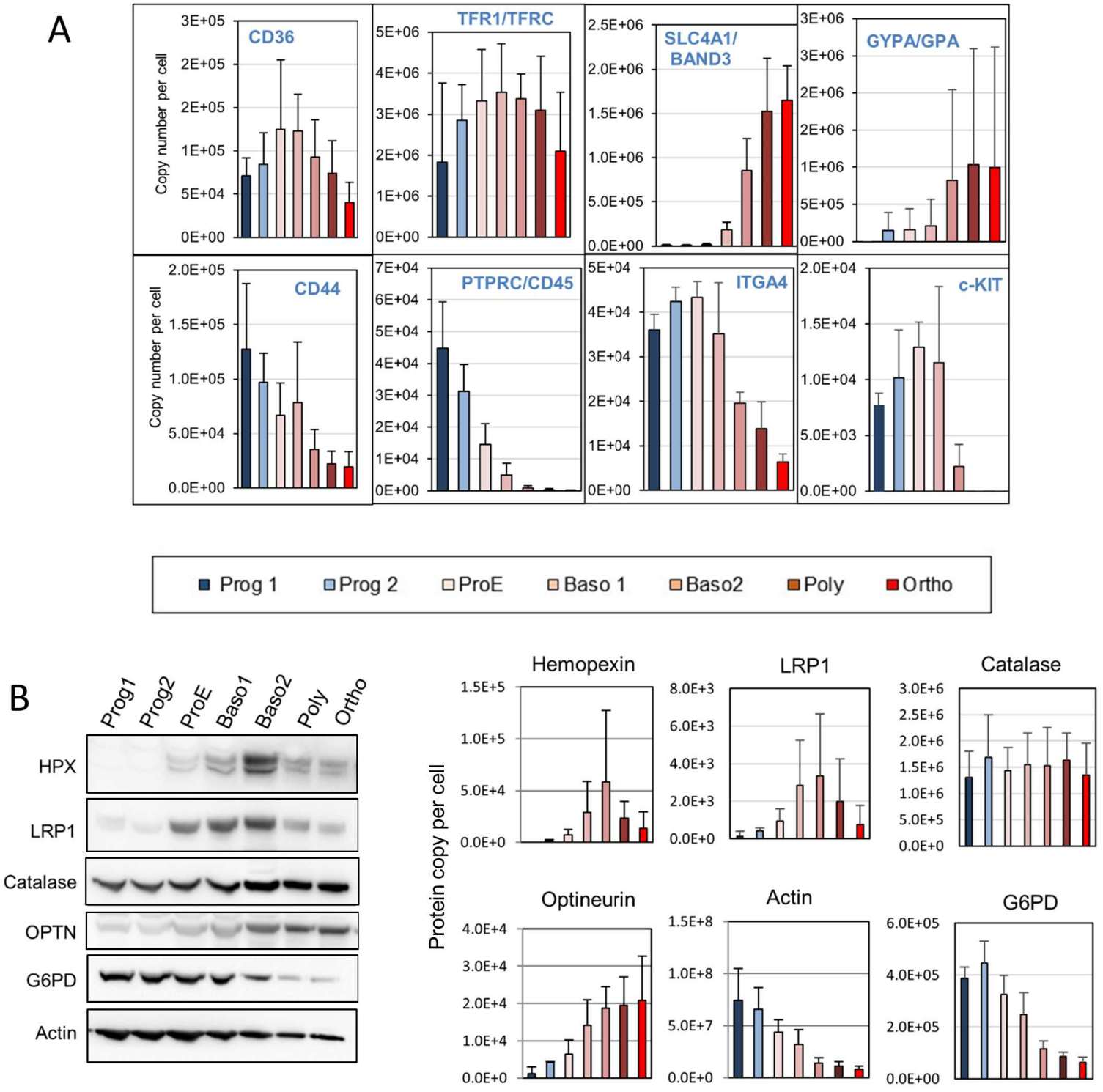
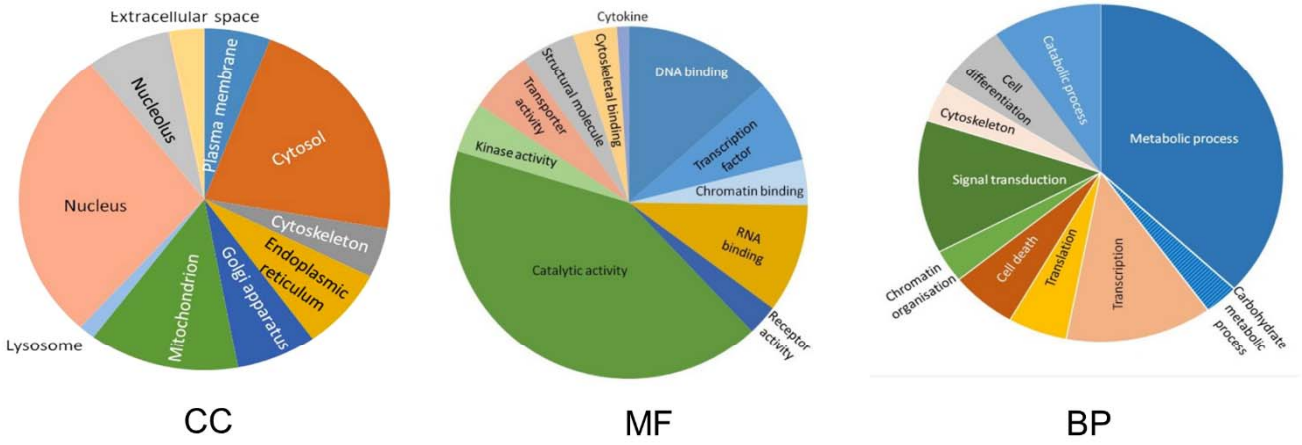


Figure S3 related to figure 4 : Global analysis of quantified proteins in erythroid cells

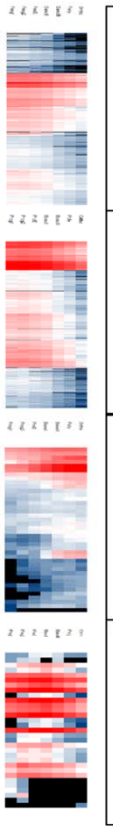
A: Patterns of quantitative expression of erythroid markers during differentiation

B: Western blot analysis of proteins expressed during erythropoiesis (left panel). Absolute protein quantifications (protein copy number per cell) derived from mass spectrometry data are indicated in the right panels.

A



B



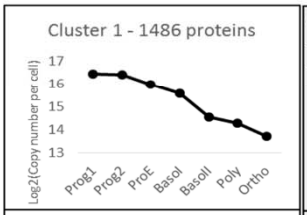
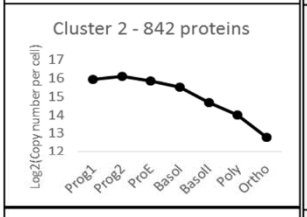
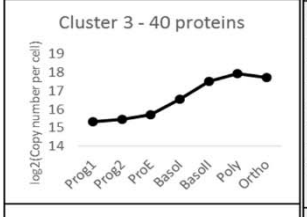
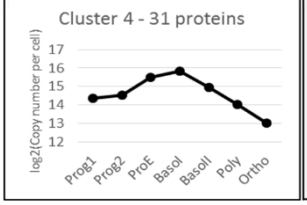
	**GOBP:** -RNA processing -RNA splicing -mRNA processing -mRNA metabolic process -RNA splicing, via transesterification reactions	**GOMF:** -RNA binding -Nucleotide binding -Translation factor activity, nucleic acid binding -ATPase activity -Purine nucleotide binding	**KEGG:** -Spliceosome -FC gamma R-mediated phagocytosis -Proteasome -Pathogenic Esherichia coli infection -Bcell receptor signaling pathway
	**GOBP:** -Ribonucleoprotein complex biogenesis -Translation -RNA processing -Ribosome biogenesis -ncRNA metabolic process	**GOMF:** -RNA binding -Structural constituant of ribosome -Nucleotide binding -Structural molecule activity -Adenyl nucleotide binding	**KEGG:** -Ribosome -Spliceosome -Mismatch repair -DNA replication -Nucleotide excision repair
	**GOBP:** -Oxygen transport -Gas transport -Negative regulation of organelle organization -Actine filament capping -Negative regulation of actin filament depolymerization	**GOMF:** -Oxygen transport activity -Oxygen binding -Iron ion binding -Heme binding -Tetrapyrrole binding	**KEGG:** -Porphyrin and chlorophyll metabolism
	**GOBP:** -Cofactor metabolic process	**GOMF:** -Oxidoreductase activity, acting on the CH-CH group donors, NAD or NADP as acceptor	

Figure S4 related to figure 4: Global analysis of proteins expressed during erythroid differentiation

A: Repartition of quantified proteins according to the Gene Ontology annotations. Annotations associated with Uniprot were recorded using the Perseus software. CC: Cellular Component, MF: Molecular Function, BP: Biological Process

B: Cluster analysis of differentially expressed proteins along erythroid differentiation. Differentially expressed proteins during erythroid differentiation were determined by a multi-sample ANOVA test and a k-mean clustering was performed using these differentially expressed proteins. The DAVID software was used to search for a significant enrichment of Gene Ontology (GO) and KEGG annotations.

Rank	BFU mRNA	CFU mRNA	Prog1 Protein	Prog2 Protein	ProE mRNA	ProE Protein	Baso1 mRNA	Baso1 Protein	Baso2 mRNA	Baso2 Protein	Poly mRNA	Poly Protein	Ortho mRNA	Ortho Protein
1	*RPS6*	*RPS18*	HIST1H4A	HIST1H4A	*HBG2*	HIST1H4A	*HBB*	HIST1H4A	*HBB*	HIST1H4A	*HBB*	HBB	*HBB*	HBB
2	*RPS24*	*RPS6*	ACTG1	ACTG1	*HBB*	ACTG1	*HBG2*	HBA1	*HBG2*	HBB	*HBG2*	HBA1	*HBG2*	HBA1
3	*RPL9*	*RPS3A*	PTMA	PTMA	*RPS6*	PTMA	*RPS6*	HBB	*HBA2*	HBA1	*HBA2*	HIST1H4A	*HBA2*	HBG2
4	*RPL37A*	*RPL3*	PFN1	HIST1H1C	*RPS18*	HIST1H1C	*RPS12*	ACTG1	*FTH1*	HBG2	*HBA1*	HBG2	*HBG1*	HIST1H4A
5	*RPLP2*	*RPS2*	HIST1H1C	PFN1	*RPL3*	S100A6	*RPS18*	HBG2	*HBG1*	HBG1	*HBG1*	HBG1	*HBA1*	HBG1
6	*RPLP1*	*TPT1*	RPLP2	TUBA1B	*RPS2*	HBA1	*RPL3*	HIST1H1C	*AHSP*	HIST1H1C	*FTH1*	HIST1H1C	*FTL*	HIST1H1C
7	*RPS3A*	*RPL37A*	GAPDH	RPLP2	*RPS12*	PFN1	*RPS2*	S100A6	*RPS12*	ACTG1	*FTL*	HBD	*FTH1*	HBD
8	*RPL7*	*RPL9*	PPIA	GAPDH	*RPLP1*	HBB	*RPLP1*	PTMA	*TPT1*	PRDX2	*SLC4A1*	PRDX2	*ALAS2*	HIST1H1B
9	*RPL35*	*RPLP1*	HSPA8	PPIA	*RPS3A*	RPLP2	*TPT1*	HBG1	*TFRC*	S100A6	*ALAS2*	ACTG1	*BSG*	PRDX2
10	*RPL12*	*RPL7*	ENO1	NPM1	*TPT1*	GAPDH	*RPS3A*	HIST1H1B	*PRDX2*	HIST1H1B	*AHSP*	HIST1H1B	*SLC4A1*	AHSP
11	*RPL38*	*HBG2*	TUBA1B	S100A6	*RPL7A*	PPIA	*FTH1*	RPLP2	*RPS6*	HBD	*TPT1*	AHSP	*TMCC2*	ACTG1
12	*RPL17*	*RPS29*	HSP90AB1	HSPA8	*RPL37A*	TUBA1B	*TFRC*	TUBA1B	*SLC4A1*	AHSP	*TFRC*	TUBA1B	*SLC25A37*	BLVRB
13	*RPS25*	*RPS8*	NPM1	ENO1	*RPS4X*	NPM1	*RPL7A*	GAPDH	*RPLP1*	TUBA1B	*HBM*	S100A6	*NCOA4*	TUBA1B
14	*RPS12*	*RPL7A*	CFL1	HSP90AB1	*RPL4*	ENO1	*RPL4*	PPIA	*RPS18*	PTMA	*BLVRB*	RPLP2	*SLC2A1*	S100A6
15	*RPL8*	*RPS12*	HIST1H1B	LDHB	*RPS8*	HIST1H1B	*RPL37A*	PFN1	*HBA1*	RPLP2	*PRDX2*	BLVRB	*TPT1*	PTMA
16	*TPT1*	*RPL8*	LDHB	HIST1H1B	*RPL7*	HSP90AB1	*RPS8*	NPM1	*RPS3A*	PPIA	*SLC25A37*	PPIA	*AHSP*	PPIA
17	*RPL11*	*RPL31*	RPS3	ALDOA	*FTH1*	ALDOA	*RPS4X*	PRDX2	*RPL3*	GAPDH	*BSG*	TUBB4B	*BLVRB*	RPLP2
18	*RPS8*	*RPS15A*	ALDOA	CFL1	*RPL8*	HSPA8	*RPL8*	HSP90AB1	*FTL*	HSPA8	*SLC2A1*	HSPA8	*BCL2L1*	HBM
19	*RPS4X*	*RPS24*	TPI1	TUBB4B	*RPS29*	S100A4	*PRDX2*	S100A4	*RPS2*	HSP90AB1	*HMBS*	PTMA	*FKBP8*	TUBB4B
20	*ACTB*	*RPL4*	S100A6	S100A4	*EEF1G*	LDHB	*AHSP*	HSPA8	*RPL7A*	TUBB4B	*RPS12*	PFN1	*BPGM*	TFRC

Table S1 related to figure 5: 20 most expressed genes during erythroid differentiation at the mRNA and protein levels

Supplemental Experimental Procedures

Erythroid cell culture and characterization

Cord bloods were obtained from human donors with informed consents in accordance with the Declaration of Helsinki Principles. CD34+ progenitors were purified using magnetic CD34 microbeads (Miltenyi Biotec Bergisch Glodbach, Germany) with a purity greater than 95%. A culture method adapted to Freyssinier et al. (Freyssinier et al., 1999) and to Leberbauer et al (Leberbauer et al., 2005) was used to obtain highly stage-enriched erythroblast populations. CD34+ cells were first cultured 7 days in Iscove Dulbecco's Modified Eagle Medium (IMDM) containing 15% BIT 9500 serum substitute (StemCell Technologies), 100 ng/ml Stem Cell Factor (SCF), 10 ng/ml interleukin (IL)-6 and 10 ng/ml IL-3. After 7 days, $CD36^+$ erythroid progenitors were purified using magnetic microbeads (CD36 FA6.152 from Beckman Coulter and anti-mouse IgG1 MicroBeads from Miltenyi Biotec) and cultured in the presence of 2.10^{-7} M dexamethasone (added two days before CD36 cell sorting), 100 ng/ml SCF, 10 ng/ml IL-3, 2U/ml EPO during 3 days for progenitors amplification and culture synchronization. To synchronously induce terminal differentiation, cells were washed with Phosphate Buffer Saline (PBS) and cultivated with 3µM of RU-486 (Mifeprestone, Sigma-Aldrich) to block dexamethasone effects, 40 ng/ml SCF, 4U/ml EPO and 3% human A/B sera (Sigma-Aldrich) in IMDM containing 15% BIT 9500. Then, cells were daily diluted to 0.5 million cells/ml with culture medium containing 3% human A/B sera and Epo only. Starting from day 6, 10% fetal calf serum was added to the culture medium. Erythroid differentiation was analyzed by May-Grünwald-Giemsa (MGG) staining of cytocentrifugated samples and by flow cytometry using the following antibodies: PE-conjugated anti-CD34, FITC-conjugated anti-CD36, FITC-conjugates anti-CD71, PC7-conjugated anti-GPA, APC-conjugated anti-CD49d (α4 integrin), or appropriated isotype control (all from Beckman Coulter) and PE-conjugated anti-BRIC6 (anti Band3) from International Blood Group Reference Laboratory NHS Blood and Transplant. FITC-conjugated annexinV was used to verify absence of apoptosis.

Clonogenic assays

At day 1 after CD36 cell sorting, between 1000 and 3300 cells were plated in duplicate in 1 ml growth factor-enriched methylcellulose medium (Methocult H4531, Stem Cell Technologies) in the presence of 2U/ml EPO. Colonies were checked at days 7, 10 and 14 of culture.

Reticulocyte and pyrenocyte sorting

To optimize the enucleation rate, a slightly modified culture protocol was used (Giarratana et al., 2011). Cord blood CD34+ cells were isolated as above using Mini-MACS columns (Miltenyi Biotech) and cultured in erythroid differentiation medium (EDM) based on IMDM (Iscove modified Dulbecco's medium, Biochrom, Germany), 330 µg/ml holo- human transferrin (BBI Solutions, Sittingbourne, UK), 10 µg/ml recombinant human insulin (FEF Chemicals, Koge, Denmark), 2 IU/ml heparin (Panpharma, Boulogne-Billancourt, France) and 5% solvent/detergent virus inactivated (S/D) plasma (Etablissement Français du Sang, France). The expansion procedure comprised three steps. In the first step (day 0 to day 8), CD34+ cells were cultured in EDM in the presence of 100 ng/ml SCF (PeproTech, Neuilly-sur-Seine, France), 5 ng/ml IL-3 (PeproTech) and 3 IU/ml Epo (Eprex, Janssen-Cilag, Issy-les-Moulineaux, France) at a density of 10^4 cells/ml and diluted with four volumes of fresh medium containing SCF, IL-3 and Epo on day 4. In the second step (day 8 to day 11), cells were resuspended at $5x10^4$ cells/ml in EDM supplemented with SCF and Epo. For the third step, cells were resuspended at day 11 at $7x10^5$ /ml in EDM supplemented with Epo alone. Cells were harvested at day 15/16, corresponding both to the peak of expelled nuclei present in the cell suspension and to a similar number of pyrenocytes and reticulocytes in the culture. At the time of cell collection, the suspension contained 5-6 $x10^6$ erythroid cells/ml, 23-27% of expelled nuclei and less than 6% of dead cells determined by trypan blue staining. Ten to 25ml of the cell suspension from day 15/16 were used for the cell sorting. Cells were filtered 70 µm, diluted at $4x10^6$/ml then incubated with Hoechst 33342 (10µg/ml) during 30 minutes at 37°C and then with GPA-PE antibodies (Beckman Coulter A07792 diluted 1/50). Cells were sorted with an ARIA III cytometer with the help of the "Cybio" facility of the Cochin Institute.

Imaging flow cytometry

The cells (1 to $2x10^6$) were labeled with anti-GPA antibodies (CD235a, BD Biosciences) and Hoechst 33342 (10µg/ml). Cells were resuspended in a total volume of 50µl and run on an Imagestream ISX MkII flow cytometer (Amnis Corp, EMD Millipore). Acquisitions were performed using INSPIRE software and 60x magnification. Between 30,000 and 50,000 events were collected in all experiments. Non stained and single stained controls were run for each fluorochrome used. Spectral compensation and sample analysis were performed in IDEASTM 6.2 software (Amnis Corp, EMD Millipore). Focus cells were selected with the Gradient RMS feature and single cells were gated with the area and aspect ratio features. For each cell, a mask

was defined using either GPA or Hoechst staining and the radius was determined using the diameter feature. Pyrenocytes, nuclei and reticulocytes were equated to spherical objects to calculate their surface and volume using the IDEASTM 6.2 software.

Peptide preparation for label-free proteomic analysis

Cells were counted and lysed by heating for 10 min at 95°C in 50mM Tris pH8.5, 2%SDS. Protein concentration was determined using bicinchoninic acid assay (BCA,Pierce). Peptides were prepared by the Filter-Aided Sample Preparation method (FASP) essentially as described (Wisniewski et al., 2009). Briefly, 50 µg of proteins from whole cell lysates were diluted to 100 µl in solubilization buffer (50 mM Tris/HCl, pH8.5, 2% SDS, 20mM TCEP, 50 mM chloroacetamide) and heated for 5 min at 95°C. After cooling to room temperature, extracts were diluted with 300 µl Tris Urea buffer (8M Urea, 50mM Tris/HCl pH 8.5), transferred onto 30 kDa centrifuged filters and prepared for digestion as described (Wisniewski et al., 2009). Proteins were digested for 14h at 37°C with 1µg trypsin (Promega). Peptides were desalted on C18 StageTips (Rappsilber et al., 2007) and fractionated by strong cationic exchange (SCX) StageTips, mainly as described (Kulak et al., 2014) except than fractions 1 and 2 were pooled in most experiments.

Label Free Quantification (LFQ) mass spectrometry analysis

Mass spectrometry analyses were performed on a Dionex U3000 RSLC nano-LC-system coupled to either a Q-Exactive or a LTQ Orbitrap-Velos mass spectrometer, all from Thermo Fisher Scientific. After drying, peptides from SCX StageTip fractions were solubilized in 10 µl of 0.1% TFA containing 2% acetonitrile (ACN). One µl was loaded, concentrated and washed for 3 min on a C18 reverse phase precolumn (3 µm particle size, 100 Å pore size, 75 µm inner diameter, 2 cm length, Dionex, Thermo Fischer Scientific). Peptides were separated on a C18 reverse phase resin (2 µm particle size, 100 Å pore size, 75 µm inner diameter, 15 cm length from Dionex) with a 3-hour gradient starting from 99% of solvent A containing 0.1% formic acid in H_2O and ending in 40% of solvent B containing 80% ACN and 0.085% formic acid in H_2O. The mass spectrometers acquired data throughout the elution process and operated in a data-dependent scheme with full MS scans acquired with the Orbitrap, followed by up to 20 MS/MS CID spectra in the Velos linear trap (Orbitrap Velos) or 10 MS/MS HCD fragmentations (Q-Exactive) on the most abundant ions detected. Settings for Orbitrap Velos were: full MS automatic gain control (AGC): 1.10^6, maximum ion injection time (MIIT): 500 ms, resolution: 6.10^4, m/z range 400-2000 and for MS/MS, AGC: 5.10^3, MIIT: 200 ms, minimum signal threshold: 500, isolation width: 2 Th, peptides with undefined charge state or charge state of 1 were excluded from fragmentation. Settings for Q-Exactive were: full MS AGC target 1.10^6 with 60ms MIIT and resolution of 70 000. The MS scans spanned from 350 to 1500 Th. Precursor selection window was set at 2Th. HCD Normalized Collision Energy (NCE) was set at 27% and MS/MS scan resolution was set at 17,500 with AGC target 1.10^5 within 60ms MIIT. For both Orbitrap Velos and Q-Exactive, dynamic exclusion time was set to 30 s and spectra were recorded in profile mode.

The mass spectrometry data were analyzed using Maxquant version 1.5.2.8 (Cox et al., 2014; Cox et al., 2009). The database used was a concatenation of human sequences from the Uniprot-Swissprot database (Uniprot, release 2015-02) and the list of contaminant sequences from Maxquant. The enzyme specificity was trypsin. The precursor mass tolerance was set to 4.5 ppm and the fragment mass tolerance to 0.5 Da for Orbitrap Velos data or 20 ppm for Q-Exactive data. Carbamidomethylation of cysteins was set as constant modification and acetylation of protein N-terminus and oxidation of methionines were set as variable modifications. Second peptide search was allowed and minimal length of peptides was set at 7 amino acids. False discovery rate (FDR) was kept below 1% on both peptides and proteins. Label-free protein quantification (LFQ) was done using both unique and razor peptides. At least 2 such peptides were required for LFQ. The “match between runs” (MBR) option was allowed with a match time window of 1 min and an alignment time window of 30 min.

Validation of the Match Between Runs (MBR) utilization

In a whole proteome analysis, the number of produced peptides exceeds by far the fragmentation capacity of the most recent mass spectrometers (Michalski et al., 2011). Due to the stochastic selection of the peptides to be fragmented and the efficiency of the fragmentation process that could be different from sample to sample, peptides are frequently identified in a fraction of the samples that are analyzed only, leading to the problem of the missing values in the data sheet. Nevertheless, in contrast to the lack of identifications linked to the fragmentation (MS/MS) process, the quantification information that is linked to the MS signal is generally present. MBR was developed by Cox et al (Cox et al., 2014) to propagate peptide identification from a run with valid MS/MS identification to the other runs of the same analytical set where identifications are lacking and thus to reduce the number of missing values. MBR uses both the highly accurate m/z value of a peptide that has been identified by fragmentation in a sample and its chromatographic retention time to search for the presence of the same peptide in the other samples of the same experiment and to retrieve the quantification values. We have

tested the accuracy of this method (figure S2). As expected, peptides quantified by MBR corresponded to the peptides with the lowest MS intensities and accordingly, MBR was especially used for the peptides from Ortho erythroblasts that present both the lowest expression level of many proteins and huge amounts of few proteins like globins (figure S2A). We quantified the proteins from the samples of the first experiment (E1) both with and without MBR. Using MBR allowed to increase the number of quantified proteins from 20% to 50% depending on the sample (figure S2B). We focused on proteins quantified both with and without MBR. The median number of peptides used to quantify these proteins was significantly lower in the analysis without MBR (Figure S2C). Nevertheless, in each sample, both analyses gave identical quantification values (figure S2D). Then, we calculated the coefficient of variability (CV) of protein quantifications determined with or without MBR in the different erythroid cell populations using the data from the four experiments. Comparison of the determinations made with and without MBR shows that MBR contributes only slightly to increase the overall CV (figure S2E). While this increase is more important at the Ortho stage, using MBR did not significantly modified the absolute quantification values (figure S2F). To determine the contribution of a possible biological variability to the CV, we performed four determinations of the UT7 cell line proteome. We determined a median CV of 67.21 %, a value close to the values determined for the primary erythroid cell populations (figure S2E), showing that little variability in our experiment can be attributed to the biological variability of the cord blood samples. We tested the relationship between the expression level of each quantified UT7 protein and the CV of the measurement and we did not notice an obvious relationship between the variability of the quantification and the expression level of a protein (figure S2G). Overall, our data show that using the MBR algorithm contributes only slightly to the technical variability of the quantification while it allows to significantly increase the number of quantified proteins. We have indicated in the main data sheet for each differentiation stage, the proteins that have been quantified using MBR only (see table S2).

LFQ data analysis

For analysis, LFQ results from MaxQuant were imported into the Perseus software (version 1.5.1.6). Reverse and contaminant proteins were excluded from analysis. Contaminating proteins from culture medium, essentially coming from added serum, such as albumin, immunoglobulins or transferrin were also removed from the protein list after manually checking. The full list of removed contaminants is presented in the supplementary table S2 ("Removed contaminants" worksheet). Protein copy numbers per cell were then calculated using the "Protein ruler" plugin of Perseus by standardization on total histone MS signal as described (Wisniewski et al., 2014) using a histone concentration of 6,5452 pg/human cell as implemented in the Perseus software. Statistical analysis and data comparison were done using Perseus, DAVID or Excel software. Enrichment of functional categories was performed using GeneTrail (http://genetrail.bioinf.uni-sb.de/index.php). Hits with <2-fold enrichment (observed / expected) were excluded.

iTRAQ mass spectrometry quantification

Equal numbers of purified pyrenocytes and reticulocytes were lysed in 50mM Tris SDS 2% pH8.5 and incubated for 10 minutes at 95°C. Proteins from whole cell lysates were digested using the FASP method as described above. Peptides were concentrated in a centrifugal vacuum concentrator, desalted on UptiTip Packed C18 10µl-200µl (Interchim Uptima BI5280) and dried. Peptides were solubilized in 34µl 500mM tri-ethyl ammonium bicarbonate and labeled according to the protocol of the iTRAQ Reagents 8plex Application Kit Protein (ABSciex). EN5 reticulocyte and pyrenocyte were labeled with isobaric 8plex tag 113 and 117, respectively. EN6 pyrenocyte, EN6 reticulocyte, EN7 pyrenocyte, EN7 reticulocyte were labeled with isobaric 4plex tag 114, 115, 116 and 117, respectively. Sample labeling was controlled by mass spectrometry before pooling and drying. Excess iTRAQ reagent was removed by SCX chromatography. Briefly, the SCX column (ABSciex) was prewashed with 2ml of cleaning buffer (25% ACN, 10mM H_3PO_4, 1M KCl, pH3.00) and equilibrated with 2ml of loading buffer (25% ACN, 10mM H_3PO_4, pH3.00). Peptides were solubilized in 1ml of loading buffer acidified to pH3.00 with H_3PO_4, injected on column and washed with 2ml loading buffer. Peptides were eluted with 500µl of 25% ACN, 10mM H_3PO_4, 350mM KCl, pH3.00 and desalted using a Sep-Pak C18 column (Waters). Peptides were then separated by isoelectrofocalisation on 13 cm pH 3-10 strips using an Agilent 3100 Off-Gel fractionator following manufacturer's instructions. After focusing, each fraction was collected. To extract peptides trapped in the strip gel, 200µl of 50% methanol in 1% formic acid were added to each tank of the frame and incubated for 30min. Methanol-extracted peptides were pooled with their respective fraction then dried in a vacuum concentrator.

After solubilization in 10µl of 10% ACN, 0.1% TFA, 1µl of each fraction was analyzed in LC-MS-MS using an Ultimate 3000 Rapid Separation liquid chromatographic system coupled to a hybrid LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific). Briefly, peptides were loaded and washed on a C18 reverse phase precolumn (3 µm particle size, 100 Å pore size, 75 µm i.d., 2 cm length). The loading buffer contained 98% H_2O, 2% ACN and 0.1% TFA. Peptides were then separated on a C18 reverse phase resin (2 µm particle

size,100 Å pore size, 75 µm i.d.,15 cm length) with a 97 min gradient from 99% A (0.1% formic acid in H_2O) to 7% B (80% ACN, 0.085% formic acid in H_2O) in 33 min then to 40% B in 64 min. The mass spectrometer acquired data throughout the elution process and operated in a data dependent scheme with full MS scans acquired with the Orbitrap, followed by up to 10 LTQ MS/MS CID spectra and 10 MS/MS HCD on the most abundant ions detected in the MS scan. Mass spectrometer settings were: full MS (AGC: $1x10^6$, resolution: $6x10^4$, m/z range 400-2000, maximum ion injection time: 500 ms); MS/MS (minimum signal threshold: 500, isolation width: 2Th, dynamic exclusion time setting: 30 s, Ion Trap MSn AGC Target: $5x10^3$ and maximum injection time: 200 ms, FTMS MSn AGC Target: $4x10^4$ and maximum injection time: 250 ms). The fragmentation was permitted for precursor with a charge state >1.

Proteome discoverer 1.3 was used to generate .mgf files for peptides with a signal to noise ratio >3 applying a spectrum grouper node to merge HCD and CID scans The data was analyzed by Protein Pilot version 4.5 (ABSciex) using the human database from Uniprot and a 5% local FDR. FDR calculations were performed using a reverse database.

Supplemental references

Cox, J., Matic, I., Hilger, M., Nagaraj, N., Selbach, M., Olsen, J.V., and Mann, M. (2009). A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. Nat Protoc *4*, 698-705.

Leberbauer, C., Boulme, F., Unfried, G., Huber, J., Beug, H., and Mullner, E.W. (2005). Different steroids co-regulate long-term expansion versus terminal differentiation in primary human erythroid progenitors. Blood *105*, 85-94.

Michalski, A., Cox, J., and Mann, M. (2011). More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. J Proteome Res *10*, 1785-1793.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc *2*, 1896-1906.

Supplemental data items

Table S2 Excel file related to figures 3 and 4: Overall data of the evolution of the erythroid cell proteome during differentiation

Table S3 Excel file related to figures 3 and 4: Cluster analysis of the evolution of the erythroid cell proteome during differentiation

Table S4 Excel file related to figure 5: Relationship between mRNA and protein expression during erythroid differentiation

Table S5 Excel file related to figure 7: Protein repartition between the reticulocytes and the pyrenocytes after enucleation

Table S6 Excel file related to figure 7: Subcellular localization of proteins quantified in reticulocytes and pyrenocytes according to Gene Ontology