Single-cell analysis of megakaryopoiesis in peripheral CD34+ cells: insights into ETV6-related thrombocytopenia

Timothée Bigot, Elisa Gabinaud, Laurent Hannouche, Véronique Sbarra, Elisa Andersen, Delphine Bastelica, Céline Falaise, Denis Bernot, Manal Ibrahim-Kosta, Pierre-Emmanuel Morange, et al.

Single-cell analysis of megakaryopoiesis in peripheral CD34$^+$ cells: insights into ETV6-related thrombocytopenia

Timothée Bigot[1,†], Elisa Gabinaud[1,†], Laurent Hannouche[1], Véronique Sbarra[1], Elisa Andersen[1], Delphine Bastelica[1], Céline Falaise[2], Denis Bernot[1], Manal Ibrahim-Kosta[1], Morange Pierre-Emmanuel[1], Marie Loosveld[2], Paul Saultier[1], Dominique Payet-Bornet[3,*], Marie-Christine Alessi[1,2,*], Delphine Potier[2,*] and Marjorie Poggi[1,*,#]

[1]Aix Marseille Univ, INSERM, INRAe, C2VN, Marseille, France

[2]APHM, C.H.U. Timone, French Reference Center on Inherited Platelet Disorders, Marseille, France

[3]Aix-Marseille Univ, CNRS, INSERM, CIML, Marseille, France

[†, *] These authors contributed equally to this work.

Corresponding author.

Running Head: Single cell analysis of ETV6-megakaryopiesis

Correspondence: Marjorie Poggi, C2VN, Faculté de Médecine, 27 Boulevard Jean Moulin, 13385 Marseille, France; E-mail: marjorie.poggi@univ-amu.fr; Phone: +33 491 324 452.

Abstract: 225; Main text: 4169; Figures: 7; Tables: 0; Supplementary data: 10 Figures and 8 Tables.

Essentials:

- ETV6-related thrombocytopenia mechanisms remain poorly understood
- Megakaryopoiesis was studied by single-cell RNA sequencing and defects were confirmed ex-vivo
- ETV6 deficiency led to the development of aberrant haemopoietic cells as early as the MEP stage
- Upregulation of ribosome biogenesis were documented in patient megakaryocytes and platelets

Abstract

Background. Germline mutations in the ETV6 transcription factor gene are responsible for familial thrombocytopenia and leukemia predisposition syndrome. Although previous studies have shown that ETV6 plays an important role in megakaryocyte (MK) maturation and platelet formation, the mechanisms by which ETV6 dysfunction promotes thrombocytopenia remain unclear.

Objective and methods. Presuming that ETV6 mutations result in selective effects at a particular cell stage, we applied single-cell RNA sequencing to understand gene expression changes during megakaryopoiesis in peripheral CD34+ cells from healthy controls and patients with ETV6-related thrombocytopenia.

Results. Analysis of gene expression and regulon activity revealed distinct clusters partitioned into seven major cell stages: hematopoietic stem/progenitor cells (HSPCs), common-myeloid progenitors (CMPs), MK-primed CMPs, granulocyte-monocyte progenitors, megakaryocyte-erythroid progenitors (MEPs), progenitor MKs/mature MKs (MKPs/MKs) and platelet-like particles (PLPs). We observed a differentiation trajectory in which MEPs developed directly from HSPCs and bypassed the CMP stage. ETV6 deficiency led to the development of aberrant cells as early as the MEP stage, which intensified at the MKP/MK stage, with a highly deregulated core "ribosome biogenesis" pathway. Indeed, increased translation levels have been documented in patient CD34+-derived MKs with overexpression of ribosomal protein S6 (RPS6) and phosphorylated-RPS6 in both CD34+-derived MKs and platelets. Treatment of patient MKs with the ribosomal biogenesis inhibitor CX-5461 resulted in an increase in PLPs.

Conclusions. These findings provide novel insight into both megakaryopoiesis and the link between ETV6, translation and platelet production.

Keywords

ETV6, Megakaryopoiesis, Ribosomal Protein S6, Single-cell RNA sequencing, Thrombocytopenia

Introduction

ETV6-related thrombocytopenia (ETV6-RT) is a highly penetrant form of inherited thrombocytopenia with autosomal dominant inheritance [1]. The common phenotype observed in ETV6-RT includes moderate thrombocytopenia sometimes associated with bleeding and predisposition to acute T- or B-cell lymphoblastic leukemia. The precise role that *ETV6* plays in megakaryocyte (MK) differentiation and pro-platelet formation currently remains poorly understood. Studying *Etv6* function in murine models is challenging because complete loss of the gene is lethal [2] and heterozygous *Etv6* mice have unperturbed hematopoiesis [3]. Human induced pluripotent stem cells (iPSCs) harboring a pathogenic heterozygous *ETV6* mutation do not give rise to an increase in hematopoietic progenitor cells and MKs. However, iPSCs carrying the homozygous *ETV6* mutation give rise to an increase in hematopoietic progenitor cells and immature MKs, as observed in heterozygous patients using an *in vitro* model of $CD34^+$-derived MKs [4,5]. Therefore, $CD34^+$-derived MK differentiation appears to be a valuable model to understand defective megakaryopoiesis associated with aberrant ETV6. This model has been successfully used to characterize acquired and inherited thrombocytopenia [6], investigate novel mechanisms in MK differentiation and platelet function [7–10], and evaluate the effect of infectious diseases or therapeutic agents on megakaryoblast differentiation [11,12].

Using $CD34^+$-derived cells from controls and patients harboring the *ETV6* variants, we examined the transcriptome of every cell type in the differentiation pathway from hematopoietic stem/progenitor cells (HSPCs) to MKs. We provide detailed insight into peripheral $CD34^+$ megakaryopoiesis. Furthermore, to fully characterize the megakaryocytic differentiation stages, regulon signatures and the corresponding regulatory networks were assessed. We found that ETV6 deficiency leads to the development of aberrant megakaryocyte-erythroid progenitor (MEP) and MK populations with some hyperactive

regulons and a highly deregulated core “ribosome biogenesis” pathway involving ribosomal protein S6 (RPS6). Our findings provide insight into the pathology of *ETV6* variants and potential targets for both diagnostic purposes and therapeutic approaches to restore pro-platelet formation.

Methods

Patients and DNA sequencing

Patients were recruited at the Center for the Investigation of Hemorrhagic and Thrombotic Pathologies (University Hospital of Marseille) after informing written consent was obtained, in accordance with protocols approved by the national review process (authorization number 20200T2-02). Two high-throughput sequencing methods were used to identify *ETV6* variants as previously described [5,13]. Two *ETV6* variants were evaluated in this study, the new variant pF417LTer4 (NM_001987, c.1251del) and a previously described variant p.P214L (NM_001987, c.C641T). Blood samples were collected from five different patients, on several occasions (at least twice for each, with an interval of six months minimum).

In vitro Megakaryocyte Differentiation

After density gradient separation (Eurobio), circulating $CD34^+$ cells were purified using positive selection with magnetic beads (Miltenyi-Biotec) and then cultured in StemSpan Serum-Free Expansion Medium II supplemented with Megakaryocyte Expansion Supplement (Stemcell Technologies) as previously described [13]. In some experiments, at day 11 or day 13, cells were pre-incubated for 48 hours with CX-5461 (Sigma Aldrich) or cycloheximide (Merck) as previously described [14,15]

Cell Preparation and Single-Cell RNA Sequencing

$CD34^+$-differentiated cells from two healthy controls (Control 1: male, born in 1996 and control 2: female, born in 1980) and two patients harboring *ETV6* variants (a male; born in 1985 and female, born in 1986) were harvested from culture at days 6 and 11 (Figure 1A). Cell concentrations and viability were assessed after trypan blue staining with an automated cell counter (EveTM NanoEntek). The cell samples from each volunteer were labeled with a

distinct hashtag oligo (TotalSeq, Biolegend) and pooled. Single-cell isolation was then carried out with the 10x Genomics Technology using the Chromium Next GEM Single Cell 5' Kit v2 (ref 1000263) according to the manufacturer's protocol. Single-cell cDNA synthesis and sequencing libraries (days 6 and 11) were prepared with a single cell 5' Library and Gel Bead kit. Libraries were sequenced applying a 75-bp paired end reads format with a Next-seq500 (GBiM platform) (parameters, read 1: 26 cycles, i7: 8 cycles, read 2: 57 cycles).

Statistical Analyses

Quantitative variables are expressed as the mean±SEM. Experimental data analyses were performed using GraphPad Prism software. Statistical differences were determined using one-way ANOVA with Kruskal-Wallis post-hoc test, Mann Whitney test or one-sample t-test when appropriate. P-value <0.05 was considered statistically significant.

Results

Characterization of differentiation in control CD34^{+}-cells

We performed UMAP (Uniform Manifold Approximation and Projection) dimensional reduction to analyze cell transcriptome heterogeneity. The variations between the transcriptomic profiles of cells after 6 and 11 days in culture indicated that distinct gene sets were involved in each stage of differentiation (Figure 1B).

Using unsupervised clustering, we identified a total of 15 clusters (Figure 1C), which were present in both controls (Supplementary figure 1A-C). We performed a detailed characterization of each cell cluster based on known gene sets (signatures) [14–18] (Figure 1D and supplementary Figure 1F) and the top differentially expressed genes (DEGs) (Figure 1E, Supplementary Figure 2, Supplementary table 1).

Cluster 10 was designated HSPC based on the expression profiles of *PROM1*, *CRHBP*, *FLT3*, *HOPX* and *AVP*. Similarly, clusters 11 and 7 were designated CMP based on the expression profiles of *CPA3*, *PRG2*, *CLC* and *TPSAB1*. Cluster 11 clearly differed from cluster 7 as it shared marker genes with the MKP/MK clusters (*VWA5A*, *LMNA*, *CAVIN2*, *LAT*, *CD9* and *ITGA2B*) and expressed the genes *KIT*, *KRT1*, *HPGDS* and *TPSB2*. We thus defined this cluster 11 as MK-primed CMP. Cluster 13 was designated GMP because it expressed specific markers such as *MPO*, *ELANE* and *PRTN3*. Clusters 9, 5, 4, 0 and 2 were designated MEP. Although these clusters shared a strong common MEP signature (*CD38*, *TFRC*, *DEPTOR*, *KLF1*, *TFR2*), we deepened the characterization of these five subpopulations. Cluster 9 was the most immature population and was designated “early MEP” based on the remaining expression levels of HSPC and CMP genes. Subsetting of cluster 9 confirms the presence of transitional steps with different levels of engagement towards MEP (HSPC-MEP and CMP-MEP; data not shown). Cluster 5 was designated a MEP-ERP subpopulation based on the

genes associated with erythroid progenitors (ERPs) such as *HBB*, *TFR2* and *ANK1*. Clusters 4, 0 and 2 were designated a MEP-MK subpopulation based on the expression levels of MK genes, such as *GP9*, *GP6*, *GP1BA* and *MPIG6B*, and the gradual increase in *MYH9* expression levels between clusters 4, 0 and 2 (Supplementary figure 1D). Clusters 1, 6, 3, 8 and 12 were primarily observed at day 11 and were designated MK based on the expression of genes associated with classical MK. Specific MK markers (*GP6*, *GP9*, *PF4*, *P2RY1*) gradually increased from clusters 1 to 12 (Figure 1D, Supplementary figures 1D). Of these five clusters, cluster 1 was the most immature stage (MKP) and cluster 12 was the most mature. Cluster 14 was mainly observed at day 11 and was designated platelet-like particles (PLPs) based on the low number of genes expressed (mean ± SD: 1,418±347 vs. 4,208 ±1,399), the low overall RNA count (mean ± SD: 2,243±751 vs. 23,855±14,826) as compared with the MKP/MK populations (Figures 1F-G, Supplementary figure 1D). Also, their proportion increases from day 6 to day 11 (0.04 % to 1.55 % of total cells). Furthermore, cluster 14 expressed the same genes as cluster 12, with higher expression levels of *STXBP5* and *RAB27b* involved in platelet granule metabolism and *RGS18*, *GP6* and *ITGA2B* (Figures 1D-E, Supplementary figures 1D-F). For each cell type, we proposed new enriched gene signatures that included known cell-type specific genes and top-ranked DEGs (Supplementary table 2). Lineage signature scores were then computed in each cell and cell types were defined accordingly (Figure 1H-I).

Inference of Megakaryopoiesis Regulon Activity in Control Cells

The cell state transitions in megakaryopoiesis are tightly controlled by transcription factors (TFs). Regulons are inferred groups of genes controlled as a unit by the same repressor or activator TF [19]. Cell type-specific regulons provide a great opportunity to identify key regulators of cell fate decisions and to improve signatures. Analysis of our control data set

using the single-cell regulatory network inference and clustering (SCENIC) workflow provided new insight into cell type-specific regulons that drive cellular heterogeneity and megakaryocytic differentiation. For each regulon, the activity score was calculated in each cell based on the number of transcripts for all genes included in the regulon (TFs and targets). Comparing each cell type, we analyzed the five most active regulons to isolate key regulons at each stage of differentiation and identify a regulon signature (Figure 2 A-B, Supplementary table 2). Using Cytoscape (v.3.9.1), we then constructed the TF gene regulatory network (TF-GRN) based on the top five regulons of each cell type and their direct TF targets (Figure 2C). Next, we established a list of specific and common active TFs for each cell type (Supplementary table 3). Among the hyperactive TFs that characterize the MK-primed CMP-biased population, we unexpectedly observed a TF that has never been associated with MK differentiation (ZMAT4; zinc finger matrin, type 4) and two highly specific TFs, AhR (aryl hydrocarbon receptor) and BATF (basic leucine zipper ATF-like transcription factor).

Characterization of the ETV6 Variants

We performed a functional study of the novel *ETV6* variant pF417LTer4 compared with the previously described p.P214L [5] (Figure 3A). The nonsense mutation p.F417Lter4 is located in the ETS domain, while the missense mutation p.P214L is localized in the linker domain (Figure 3B, lower panel). The clinical and laboratory characteristics of the patients are shown in Supplementary table 4. We analyzed the repressive activity of the two variants using a dual-luciferase reporter assay. Co-transfection of the reporter plasmid containing the ETS-binding site along with the expression of a plasmid encoding wild type (WT) *ETV6* resulted in almost 90% inhibition of luciferase activity. Substitution of WT *ETV6* with any of the *ETV6* variants led to a significant reduction in repressive activity (85% to 100%) (Figure 3B, left panel). Western-blot analysis showed that mutant ETV6 protein was expressed in the

GripTite293 macrophage scavenger receptor (MSR) cell line (figure 3B). Subcellular fractionation of GripTite293 MSR cells showed increased ETV6 protein levels in the cytoplasmic fraction and decreased levels or absence of ETV6 in the nuclear fraction in cells expressing the p.P1214L or p.F417Lter4 variants compared with cells expressing the WT protein (Supplementary figure 3A). Microscopy confirmed that WT ETV6 concentrated primarily in cell nuclei, whereas both ETV6 variants were predominantly located in the cytoplasm (Supplementary figure 3B).

Single-cell transcriptional profiling of *ETV6*-variant CD34^{+} cells during megakaryopoiesis

We performed UMAP dimensional reduction to analyze cell transcriptome heterogeneity and clusters. The day 6 and day 11 transcriptome profiles of patient cells overlapped considerably more than those of control cells, thereby suggesting that differentiation was delayed with accumulation of early-stage cells in *ETV6*-variant carriers (Figure 3C, left panel). Using unsupervised clustering, a total of 11 clusters were identified (Figure 3C, right panel). As previously, we performed a detailed characterization of each cell cluster based on known gene sets and the top DEGs (Figure 3D-E, Supplementary Figure 4, Supplementary table 5).

The cell types found in controls were also observed in patients. Clusters 3 and 7 corresponded to HSPC; cluster 4 corresponded to CMP; cluster 10 corresponded to MK-primed CMP; cluster 9 corresponded to GMP; clusters 5, 2, 6, 8 and 0 corresponded to MEP; and cluster 1 corresponded to MKP/MK. No PLP cluster was observed (figure 3D-E, Supplementary figure 4). The proposed lineage genetic signatures (Supplementary table 2, left panel) were computed in each cell and cell types were assigned (Figures 3F-G).

Developmental Trajectory of Megakaryocyte Differentiation

We then assessed the single-cell transcriptome for pseudotemporal ordering of differentiation stages during megakaryopoiesis in controls and patients. Using Slingshot, we inferred differentiation trajectories in cells from controls and *ETV6*-variant carriers independently. For each condition, we determined the overall trajectory structure of each lineage (rooted tree) by generating a transcriptomic distance matrix using the manually set root cluster (HSPCs; cluster 10 for controls and cluster 3 for patients). We observed an unexpected lineage of MK differentiation that bypassed the CMP cell type and differentiated directly from HSPCs into MEPs (Figures 4A).

Analyzing the differentiation trajectory in control cells, stem cell markers such as *CD34, CD38* and *HLA-DRA* displayed decreased expression, while the MK markers *GP9*, *PF4* and *ITGA2B* were upregulated (Figure 4B). The *GATA1* and *FLI1* TFs displayed increased expression, while *GATA2* displayed high expression levels in immature cells and reduced expression in later stages. During differentiation, *ETV6* and *RUNX1* expression levels were low and stable (Figure 4B).

By contrast, *ETV6*-deficient cells displayed stable expression of *CD34*, *HLA-DRA* and *GATA2* (Figure 4B) during MK differentiation; these genes remained highly expressed at the MKP/MK stage (MKP/MK: fold change patients vs. controls = 2.5 (*CD34*), 1.5 (*HLA-DRA*), 1.6 (*GATA2*), adjusted p-value <0.05). The expression levels of *GATA1* and *FLI1* were lower than that observed in controls. Analyzing *GP9* and *PF4*, we observed lower expression levels and a delayed response time (Figure 4B). These results indicate that *ETV6*-variant carriers exhibit a delay in differentiation.

Aberrant populations in *ETV6*-variant cells identified via single-cell RNA sequencing

The full data set (including both control and patients) was analyzed to compare controls and patients. The cell types identified in control and patient data sets were transferred to the full data set using the cell identity barcode (Figure 5A). In the early cell stages (HSPCs, CMPs, MK-primed CMPs and GMPs), the transcriptome profiles of controls and patients were markedly similar. By contrast, distinct gene expression profiles were observed as early as the MEP stage, which intensified at the MKP/MK stage.

Cell type distribution differed between controls and patients (Figures 5B). Patients harboring *ETV6* variants displayed an increased proportion of HSPC (D6: 17.6±6.1 vs 5.2±1.1%; D11: 16±8 vs 0.6±0.04%) and a decline in MKP/MK (D11: 37±1 vs 78±0.6%) compared with controls (Figure 5B).

We applied the SCENIC workflow to further characterize the aberrant MEP and MKP/MK populations. The UMAP reduction using the 312 combined regulon activity scores in each cell is shown in Figure 5C. Compared to the early stages, we observed marked differences between patients and controls at the MEP and MKP/MK stages. The top 20 differentially active regulons are shown in Figure 5D. Some of the top 20 regulons were common for the two cell stages and demonstrated hyperactivity in patient cells (Supplementary Table 3). After analyzing the regulons using tree hierarchical algorithms with Cytoscape software, we found that SPI1 and GATA2 hyperactivity represented initial events in the modification of regulon activity observed in patients (Supplementary figure 5). Overall, these results indicate that *ETV6* mutations alter megakaryopoiesis, resulting in the development of aberrant MEP and MK populations as well as the disruption of key TF activity, with GATA2 and SPI1 hyperactivity representing initial events.

Deregulated "DNA repair" and "translation" pathways in *ETV6*-variant cells

When comparing controls and patients, the number of DEGs (fold change > 1.3, adjusted p-value < 0.05) increased over the course of cell differentiation, especially at the MEP and MK stages (HSPC, DEGs=38, 23 up-regulated genes and 15 down-regulated genes; CMP, DEGs=56, 29 up-regulated genes and 27 down-regulated genes; MK-primed CMP, DEGs=30, 15 up-regulated genes and 15 down-regulated genes; GMP, DEGs=32, 21 up-regulated genes and 11 down-regulated genes; MEP, DEGs=239, 100 up-regulated genes and 139 down-regulated genes; and MKP/MK, DEGs=942, 339 up-regulated genes and 603 down-regulated genes) (Figure 6A). Focusing on MEP and MKP/MK populations, DEGs were enriched in various biological pathways in multiple gene set databases (GO Biological process (Figures 6B), KEGG (Supplementary figure 6) and Reactome (Supplementary figure 7). Top 10 GO deregulated pathways are available in the supplementary tables 6 and 7.

Several pathways associated with DNA repair and cellular response to DNA damage were downregulated in patient MEPs and MKs (Supplementary table 8). Some genes involved in major DNA repair pathways displayed reduced expression levels: *FEN1* (base excision repair), *MGMT* (direct reversal of DNA damage), *RAD23A* and *RAD23B* (nucleotide excision repair), and *XRCC6* and *PRKDC* (non-homologous end joining) [20]. Caretaker genes indirectly involved in maintaining genomic stability (e.g., *TTK, NUDT1, DUT* and *UBE2V2*) were also downregulated in MKs harboring *ETV6* variants (Figure 6C).

The top 10 upregulated pathways in *ETV6*-variant MKs were associated with "translation genes" (Supplementary table 7). The most upregulated genes, *RPS* and *RPL*, code for ribosomal protein small subunits and ribosomal protein large subunits, respectively (Supplementary figure 8).

Increased Translation and RPS6 in Patients Harboring an *ETV6* Variant

Translation was the strongly dominant upregulated pathway. Translation levels were evaluated in HEL cells transduced with P214L or WT *ETV6* before and after PMA-induced differentiation. As expected, PMA stimulation led to reduced translation in all samples (Figure 7A). Translation levels were significantly higher in *ETV6*-deficient HEL cells at the non-stimulated basal state and after PMA stimulation (Figure 7A). Compared with control cells, higher translation levels were also observed in $CD34^{+}$-derived MK from *ETV6*-variant carriers (Figure 7B).

Among all upregulated genes in ETV6-variant MKPs/MKs, 1,488 genes contained 2,189 sequences that correspond to the canonical ETV6-binding site (i-cisTarget: normalized enrichment score 4.85, rank 40e/512, Supplementary table 9), including RPS6, which is a cornerstone of translation [21] (Figure 7C). Furthermore, the ETV6-binding site in RPS6 is highly conserved across seven species (2^{nd}/2,189 detected ETV6-binding regions). Analyzing the differentiation trajectory in control cells, we found that RPS6 mRNA levels gradually decreased, with the lowest levels observed in the most mature MKs (cluster 12). By contrast, ETV6-deficient cells displayed stable RPS6 expression levels throughout the differentiation pathway (Figure 7D).

Our data suggest that ETV6 regulates *RPS6* expression. Therefore, we obtained ETV6 chromatin immunoprecipitation (ChIP)-sequence data on the GM12876 cell line from the Gene Expression Omnibus database (GSM2574795, GSM2534228, GSM2574796). The peak enrichment on chromosome 9, which includes the RPS6 locus, was extracted and displayed in the UCSC genome browser (Figure 7E). This analysis showed one enriched binding region for ETV6. The binding peak is located in the promoter.

The expression of total RPS6 and phosphorylated RPS6 (p-RPS6, Ser235/236) were analyzed by western blot and flow cytometry in washed platelets, $CD34^+$-derived MKs from *ETV6*-variant carriers, and the respective age- and gender-matched controls. RPS6 and p-RPS6 were overexpressed in native platelets issued from *ETV6*-variant carriers compared with controls (Figures 7F, left panel and Supplementary figure 9). We confirmed RPS6 and p-RPS6 overexpression in $CD34^+$-derived MKs issued from three *ETV6*-variant carriers using flow cytometry (Figures 7F, right panel).

To investigate the consequences of increased translation, we treated control and patient $CD34^+$-derived cell cultures with CX-5461, a specific RNA Pol I inhibitor. Healthy control $CD34^+$-derived MKs (day 11) treated with CX-5461 exhibited reduced translation and p-RPS6 levels, whereas the expression of total RPS6 was unaffected (Figure 7G). In $CD34^+$-derived cells issued from three *ETV6*-variant carriers, CX-5461 treatment increased the production of PLPs by $CD42a^+$ MKs in all cases (Figure 7G). In addition, cells exposed to cycloheximide (CHX) to inhibit protein synthesis, exhibited reduced translation, RPS6 and phopspho-RPS6 levels. CHX treatment was able to increase the production of PLPs (Supplementary Figure 10).

Discussion

Using single-cell transcriptome profiling of MK cell cultures derived from peripheral $CD34^+$ cells, we established a specific signature for each cell stage based on both distinct gene expression profiles and regulon activities. Cells harboring *ETV6* variants displayed distinct transcriptomic/regulon profiles starting at the MEP stage and intensifying at the MKP/MK stage, with SPI1 and GATA2 regulon hyperactivity representing early events. Translation pathways were upregulated in patient MEPs and MKs. Puromycin incorporation results as well as RPS6 and p-RPS6 immunoblots were consistent with these findings. Overall, ETV6 deficiency affects key cellular processes that may represent promising mechanistic targets.

In $CD34^+$-derived MK differentiation, we observed a biased megakaryopoiesis pathway that may be due to stress-driven hematopoiesis as a result of the culture conditions [22]. HSPCs directly gave rise to MEPs without passing through the CMP stage, as previously suggested [23]. Furthermore, it has been hypothesized that MK-biased hematopoietic stem cells (HSCs) represent the hierarchical apex with MEPs developing directly from HSCs [24], although this theory has been challenged. We did not observe a MK-biased HSPC population [25] because either the analysis time points were too late to detect these cell types or mRNA did not enable detection of MK-biased HSPCs, as previously reported [26]. However, we identified a MK-primed CMP population that co-expressed megakaryocytic genes and displayed high KIT expression levels (27% of the CMP population). High KIT expression levels represented a good marker of MK-biased CMPs. Previous murine *in vitro* and *in vivo* functional studies have demonstrated that HSCs with higher levels of c-Kit signaling preferentially differentiate into MKs [27]. In human bone marrow, MK-primed CMPs likely represented the major megakaryopoiesis pathway independent of the canonical MEP lineage [28]. This *in vitro* model is a valuable tool to investigate these biased pathways.

Hematopoiesis is a progressive process controlled by an elaborate network of TFs. We characterized each cell type by analyzing the activity of specific regulons, to better define hematopoietic populations and establish a TF-GRN of megakaryopoiesis. HSPCs were characterized by the expression of the regulons HLF [29], HOXA9 [18] and TCF4 [30], which are involved in stem cell maintenance and myeloid commitment. C/EBP family members are critical for myelopoiesis [31,32]. Accordingly, we observed noticeable C/EBP activity at the CMP and GMP stages. Interestingly, ZMAT4 was newly identified as an active TF in MK-primed CMP populations. ZMAT4 is a TF of the zinc finger protein family and is associated with "programed cell death" [33], innate lymphoid cells and natural killer cells [34]. Although copy number variants of ZMAT4 are associated with leukemia [35], the role that ZMAT4 plays in myeloid differentiation remains unclear. The MK-primed CMP population was also characterized by AhR expression, which is involved in MK maturation [36–38] and BATF hyperactivation. BATF hyperactivation was evidenced as a marker of MK commitment. We also observed KLF1 activity in MEPs, which is involved in MEP lineage decisions and commitment [39]. Finally, NFE2 and GATA1 were expected markers of MKs [40,41]. Overall, these findings demonstrate that characterizing TF activity at the single-cell level is useful to phenotype cell types and decipher unexpected key functions of new TFs in each cell type.

Applying the lineage signature used in control cells, we found that ETV6-deficient cells displayed the same cell types with a higher proportion of HSPCs, a lower proportion of MKs and a lack of PLPs, which correlates with the thrombocytopenia phenotype observed in patients harboring an *ETV6* variant. We also detected marked differences in regulon activities. Several regulons displayed higher activity levels in the MEP and MK stages. This finding may be due to a loss of ETV6 repressor activity as documented in all *ETV6*-mutant carriers

[5]. We observed GATA2 hyperactivity in MKs from *ETV6*-mutant carriers, directly regulating SPI1 activity. *SPI1* has previously been described as a common downstream target gene of GATA2 in human and murine embryonic stem cells[52,53]. High GATA2 expression levels have been described in REH cells, a B-cell acute lymphoblastic leukemia cell line harboring the ETV6/RUNX fusion transcript, which is the most frequent structural cytogenetic abnormality in children with acute lymphoblastic leukemia [42]. Beyond this observation, the connection between ETV6 deficiency and GATA2 hyperactivity remains poorly understood. This relationship merits further investigation as deficient or excess GATA-2 expression are both pathogenetic [43].

Analysis of several gene enrichment databases displayed significant differences. Assessing the ten highest scores and after grouping the pathways with common genes, we observed deregulation of pathways linked with translation and DNA repair in MEP and MKP/MK patient cells. Translation was the most upregulated pathway. Altered expression levels of numerous ribosomal protein genes were observed, including upregulation of more than 30 ribosomal proteins. Recent data has emphasized the unexpected contribution of ribosomal biogenesis in hematopoiesis and megakaryopoiesis. Additionally, ribosome biogenesis is involved in mediating the transition between proliferation and differentiation of erythroid progenitors [44]. GTPase-Dynamin-2 deletion in platelets and MKs induces severe thrombocytopenia and bleeding diathesis in mice and results in the upregulation of genes involved in ribosome biogenesis in erythroblasts [45].

Furthermore, increased translation was confirmed at the functional level. Protein synthesis was increased in $CD34^+$-derived MKs in patients and hematopoietic cell lines transduced with an *ETV6* variant. Analysis of the *RPS6* gene revealed a putative ETV6-binding site, which was confirmed via ChIP-sequencing [46]. As *Rps6*-deficient mice display features of

megakaryocytic dysplasia with thrombocytosis, this gene likely plays a role in thrombopoiesis [47,48]. Moreover, overexpression of RPS6 or p-RPS6 in murine Sertoli cells cultured *in vitro* has been shown to disrupt cell function through changes in the organization of actin and microtubule-based cytoskeletons via AKT inhibition [49,50]. In *ETV6*-variant carriers, overexpression of RPS6 was observed in platelets and may act on pro-platelet formation, which is dependent on massive reorganization of the actin cytoskeleton. Translation inhibition led to reduced RPS6 phosphorylation and increased numbers of PLPs in ETV6-deficient cultures, thereby suggesting that overexpression of p-RPS6 in patient cells causes a decrease in actin polymerization rates and pro-platelet formation.

Overall, our study demonstrates the heterogeneity of human $CD34^+$ cell-induced MK differentiation *in vitro* and a major differentiation trajectory in which HSPCs develop into MEPs, bypassing the CMP stage. We found that ETV6 variations result in aberrant MEP and MK populations, which are associated with GATA2 hyperactivity and deregulated translation pathways, with RPS6 and p-RPS6 overexpression in $CD34^+$-derived MKs and platelets. These findings provide novel insight into megakaryopoiesis and ETV6 function that may be applied to develop targeted therapeutic strategies to alleviate platelet defects.

Competing interests

The authors declare that they have no competing interests.

Authorship contributions

TB, LH and DP performed bioinformatic analyses. EG, EA, VS and DB performed the culture and functional experiments. CL, MIK, ML, PEM and PS performed the clinical and biological characterization of patients. DPB and MCA conceived and supervised the project. DP and MP directed the project, designed the study, analyzed the data and wrote the manuscript.

Acknowledgments

The authors acknowledge the members of the French Reference Center for Inherited Hereditary Platelet Disorders for their contribution regarding clinical analyses, the high-performance computing center of Aix-Marseille for granting access to its high-performance computing resources, the GBiM platform for the sequencing and Dr Argüello for providing reagents to study translation. This work was supported by Aix-Marseille University (AMIDEX “Emergence et innovation” ngSUMMIT), the Agence Nationale de la Recherche (JCJC MOST) and the Institut National de la Santé et de la Recherche Médicale (PIA Biofit).

References

1 Di Paola J, Porter CC. ETV6-related thrombocytopenia and leukemia predisposition. *Blood* 2019; **134**: 663–7.

2 Wang LC, Kuo F, Fujiwara Y, Gilliland DG, Golub TR, Orkin SH. Yolk sac angiogenic defect and intra-embryonic apoptosis in mice lacking the Ets-related factor TEL. *EMBO J* 1997; **16**: 4374–83.

3 Zhou C, Uluisik R, Rowley JW, David C, Jones CL, Scharer CD, Noetzli L, Fisher MH, Kirkpatrick GD, Bark K, Boss JM, Henry CJ, Pietras EM, Di Paola J, Porter CC. Germline ETV6 mutation promotes inflammation and disrupts lymphoid development of early hematopoietic progenitors. *Experimental Hematology* 2022; **112–113**: 24–34.

4 Borst S, Nations CC, Klein JG, Pavani G, Maguire JA, Camire RM, Drazer MW, Godley LA, French DL, Poncz M, Gadue P. Study of inherited thrombocytopenia resulting from mutations in ETV6 or RUNX1 using a human pluripotent stem cell model. *Stem Cell Reports* 2021; **16**: 1458–67.

5 Poggi M, Canault M, Favier M, Turro E, Saultier P, Ghalloussi D, Baccini V, Vidal L, Mezzapesa A, Chelghoum N, Mohand-Oumoussa B, Falaise C, Favier R, Ouwehand WH, Fiore M, Peiretti F, Morange PE, Saut N, Bernot D, Greinacher A, et al. Germline variants in ETV6 underlie reduced platelet formation, platelet dysfunction and increased levels of circulating CD34+ progenitors. *Haematologica* 2017; **102**: 282–94.

6 Grodzielski M, Goette NP, Glembotsky AC, Constanza Baroni Pietto M, Méndez-Huergo SP, Pierdominici MS, Montero VS, Rabinovich GA, Molinas FC, Heller PG, Lev PR, Marta RF. Multiple concomitant mechanisms contribute to low platelet count in patients with immune thrombocytopenia. *Sci Rep* 2019; **9**: 2208.

7 Kollmann K, Warsch W, Gonzalez-Arias C, Nice FL, Avezov E, Milburn J, Li J, Dimitropoulou D, Biddie S, Wang M, Poynton E, Colzani M, Tijssen MR, Anand S, McDermott U, Huntly B, Green T. A novel signalling screen demonstrates that CALR mutations activate essential MAPK signalling and facilitate megakaryocyte differentiation. *Leukemia* 2017; **31**: 934–44.

8 Zeddies S, Jansen SBG, Summa F di, Geerts D, Zwaginga JJ, Schoot CE van der, Lindern M von, Thijssen-Timmer DC. MEIS1 regulates early erythroid and megakaryocytic cell fate. *Haematologica* 2014; **99**: 1555–64.

9 Basak I, Bhatlekar S, Manne BK, Stoller M, Hugo S, Kong X, Ma L, Rondina MT, Weyrich AS, Edelstein LC, Bray PF. miR-15a-5p regulates expression of multiple proteins in the megakaryocyte GPVI signaling pathway. *J Thromb Haemost* 2019; **17**: 511–24.

10 Song B, Miao W, Cui Q, Shi B, Zhang J, Qiu H, Zhang L, Han Y. Inhibition of ferroptosis promotes megakaryocyte differentiation and platelet production. *Journal of Cellular and Molecular Medicine* 2022; **26**: 3582–5.

11 Lin G-L, Chang H-H, Lien T-S, Chen P-K, Chan H, Su M-T, Liao C-Y, Sun D-S. Suppressive effect of dengue virus envelope protein domain III on megakaryopoiesis. *Virulence* 2017; **8**: 1719–31.

12 Hu L, Yin X, Zhang Y, Pang A, Xie X, Yang S, Zhu C, Li Y, Zhang B, Huang Y, Tian Y, Wang M, Cao W, Chen S, Zheng Y, Ma S, Dong F, Hao S, Feng S, Ru Y, et al. Radiation-induced bystander effects impair transplanted human hematopoietic stem cells via oxidative DNA damage. *Blood* 2021; **137**: 3339–50.

13 Saultier P, Vidal L, Canault M, Bernot D, Falaise C, Pouymayou C, Bordet J-C, Saut N, Rostan A, Baccini V, Peiretti F, Favier M, Lucca P, Deleuze J-F, Olaso R, Boland A, Morange PE, Gachet C, Malergue F, Fauré S, et al. Macrothrombocytopenia and dense granule deficiency associated with FLI1 variants: ultrastructural and pathogenic features. *Haematologica* 2017; **102**: 1006–16.

14 Schmidt EK, Clavarino G, Ceppi M, Pierre P. SUnSET, a nonradioactive method to monitor protein synthesis. *Nat Methods* 2009; **6**: 275–7.

15 Adamik J, Munson PV, Hartmann FJ, Combes AJ, Pierre P, Krummel MF, Bendall SC, Argüello RJ, Butterfield LH. Distinct metabolic states guide maturation of inflammatory and tolerogenic dendritic cells. *Nat Commun* 2022; **13**: 5184.

16 Psaila B, Wang G, Rodriguez-Meira A, Li R, Heuston EF, Murphy L, Yee D, Hitchcock IS, Sousos N, O'Sullivan J, Anderson S, Senis YA, Weinberg OK, Calicchio ML, NIH Intramural Sequencing Center, Iskander D, Royston D, Milojkovic D, Roberts I, Bodine DM, et al. Single-Cell Analyses Reveal Megakaryocyte-Biased Hematopoiesis in Myelofibrosis and Identify Mutant Clone-Specific Targets. *Mol Cell* 2020; **78**: 477-492.e8.

17 Chen L, Kostadima M, Martens JHA, Canu G, Garcia SP, Turro E, Downes K, Macaulay IC, Bielczyk-Maczynska E, Coe S, Farrow S, Poudel P, Burden F, Jansen SBG, Astle WJ, Attwood A, Bariana T, de Bono B, Breschi A, Chambers JC, et al. Transcriptional diversity during lineage commitment of human blood progenitors. *Science* 2014; **345**: 1251033.

18 Estevez B, Borst S, Jarocha DJ, Sudunagunta VS, Gonzalez M, Garifallou J, Hakonarson H, Gao P, Tan K, Liu PP, Bagga S, Holdreith N, Tong W, Speck NA, French DL, Gadue P, Poncz M. RUNX1 haploinsufficiency causes a marked deficiency of megakaryocyte-biased hematopoietic progenitor cells. *Blood* 2021; .

19 Pellin D, Loperfido M, Baricordi C, Wolock SL, Montepeloso A, Weinberg OK, Biffi A, Klein AM, Biasco L. A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nat Commun* 2019; **10**.

20 Roy A, Wang G, Iskander D, O'Byrne S, Elliott N, O'Sullivan J, Buck G, Heuston EF, Wen WX, Meira AR, Hua P, Karadimitris A, Mead AJ, Bodine DM, Roberts I, Psaila B, Thongjuea S. Transitions in lineage specification and gene regulatory networks in hematopoietic stem/progenitor cells over human development. *Cell Reports* 2021; **36**: 109698.

21 Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine J-C, Geurts P, Aerts J, van den Oord J, Atak ZK, Wouters J, Aerts S. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* 2017; **14**: 1083–6.

22 Chatterjee N, Walker GC. Mechanisms of DNA damage, repair and mutagenesis. *Environ Mol Mutagen* 2017; **58**: 235–63.

23 Chauvin C, Koka V, Nouschi A, Mieulet V, Hoareau-Aveilla C, Dreazen A, Cagnard N, Carpentier W, Kiss T, Meyuhas O, Pende M. Ribosomal protein S6 kinase activity controls the ribosome biogenesis transcriptional program. *Oncogene* Nature Publishing Group; 2014; **33**: 474–83.

24 Psaila B, Mead AJ. Single-cell approaches reveal novel cellular pathways for megakaryocyte and erythroid differentiation. *Blood* 2019; **133**: 1427–35.

25 Adolfsson J, Månsson R, Buza-Vidas N, Hultquist A, Liuba K, Jensen CT, Bryder D, Yang L, Borge O-J, Thoren LAM, Anderson K, Sitnicka E, Sasaki Y, Sigvardsson M, Jacobsen SEW. Identification of Flt3+ Lympho-Myeloid Stem Cells Lacking Erythro-Megakaryocytic Potential: A Revised Road Map for Adult Blood Lineage Commitment. *Cell* 2005; **121**: 295–306.

26 Sanjuan-Pla A, Macaulay IC, Jensen CT, Woll PS, Luis TC, Mead A, Moore S, Carella C, Matsuoka S, Bouriez Jones T, Chowdhury O, Stenson L, Lutteropp M, Green JCA, Facchini R, Boukarabila H, Grover A, Gambardella A, Thongjuea S, Carrelha J, et al. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature* 2013; **502**: 232–6.

27 Noetzli LJ, French SL, Machlus KR. New Insights into the Differentiation of Megakaryocytes from Hematopoietic Progenitors. *Arterioscler Thromb Vasc Biol* 2019; **39**: 1288–300.

28 Haas S, Hansson J, Klimmeck D, Loeffler D, Velten L, Uckelmann H, Wurzer S, Prendergast ÁM, Schnell A, Hexel K, Santarella-Mellwig R, Blaszkiewicz S, Kuck A, Geiger H, Milsom MD, Steinmetz LM, Schroeder T, Trumpp A, Krijgsveld J, Essers MAG. Inflammation-Induced Emergency Megakaryopoiesis Driven by Hematopoietic Stem Cell-like Megakaryocyte Progenitors. *Cell Stem Cell* 2015; **17**: 422–34.

29 Shin JY, Hu W, Naramura M, Park CY. High c-Kit expression identifies hematopoietic stem cells with impaired self-renewal and megakaryocytic bias. *J Exp Med* 2014; **211**: 217–31.

30 Miyawaki K, Iwasaki H, Jiromaru T, Kusumoto H, Yurino A, Sugio T, Uehara Y, Odawara J, Daitoku S, Kunisaki Y, Mori Y, Arinobu Y, Tsuzuki H, Kikushige Y, Iino T, Kato K, Takenaka K, Miyamoto T, Maeda T, Akashi K. Identification of unipotent megakaryocyte progenitors in human hematopoiesis. *Blood* 2017; **129**: 3332–43.

31 Lehnertz B, Chagraoui J, MacRae T, Tomellini E, Corneau S, Mayotte N, Boivin I, Durand A, Gracias D, Sauvageau G. HLF expression defines the human hematopoietic stem cell state. *Blood* 2021; **138**: 2642–54.

32 In 't Hout FEM, van Duren J, Monteferrario D, Brinkhuis E, Mariani N, Westers TM, Chitu D, Nikoloski G, van de Loosdrecht AA, van der Reijden BA, Jansen JH, Huls G. TCF4 promotes erythroid development. *Exp Hematol* 2019; **69**: 17-21.e1.

33 Radomska HS, Huettner CS, Zhang P, Cheng T, Scadden DT, Tenen DG. CCAAT/enhancer binding protein alpha is a regulatory switch sufficient for induction of granulocytic development from bipotential myeloid progenitors. *Mol Cell Biol* 1998; **18**: 4301–14.

34 Shyamsunder P, Shanmugasundaram M, Mayakonda A, Dakle P, Teoh WW, Han L, Kanojia D, Lim MC, Fullwood M, An O, Yang H, Shi J, Hossain MZ, Madan V, Koeffler HP. Identification of a novel enhancer of CEBPE essential for granulocytic differentiation. *Blood* 2019; **133**: 2507–17.

35 Apostolidis PA, Lindsey S, Miller WM, Papoutsakis ET. Proposed megakaryocytic regulon of p53: the genes engaged to control cell cycle and apoptosis during megakaryocytic differentiation. *Physiological Genomics* American Physiological Society; 2012; **44**: 638–50.

36 Allan DSJ, Cerdeira AS, Ranjan A, Kirkham CL, Aguilar OA, Tanaka M, Childs RW, Dunbar CE, Strominger JL, Kopcow HD, Carlyle JR. Transcriptome analysis reveals similarities between human blood CD3− CD56bright cells and mouse CD127+ innate lymphoid cells. *Sci Rep* Nature Publishing Group; 2017; **7**: 3501.

37 Wan J, Gao Y, Zhao X, Wu Q, Fu X, Shao Y, Yang H, Guan M, Yu B, Zhang W. The association between the copy-number variations of ZMAT4 and hematological malignancy. *Hematology* 2011; **16**: 20–3.

38 Lindsey S, Papoutsakis ET. The aryl hydrocarbon receptor (AHR) transcription factor regulates megakaryocytic polyploidization. *Br J Haematol* 2011; **152**: 469–84.

39 Strassel C, Brouard N, Mallo L, Receveur N, Mangin P, Eckly A, Bieche I, Tarte K, Gachet C, Lanza F. Aryl hydrocarbon receptor-dependent enrichment of a megakaryocytic precursor with a high potential to produce proplatelets. *Blood* 2016; **127**: 2231–40.

40 Smith BW, Rozelle SS, Leung A, Ubellacker J, Parks A, Nah SK, French D, Gadue P, Monti S, Chui DHK, Steinberg MH, Frelinger AL, Michelson AD, Theberge R, McComb ME, Costello CE, Kotton DN, Mostoslavsky G, Sherr DH, Murphy GJ. The aryl hydrocarbon receptor directs hematopoietic progenitor cell expansion and differentiation. *Blood* 2013; **122**: 376–85.

41 Tallack MR, Perkins AC. Megakaryocyte-erythroid lineage promiscuity in EKLF null mouse blood. *Haematologica* 2010; **95**: 144–7.

42 Tsang AP, Fujiwara Y, Hom DB, Orkin SH. Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. *Genes Dev* 1998; **12**: 1176–88.

43 Levin J, Peng J-P, Baker GR, Villeval J-L, Lecine P, Burstein SA, Shivdasani RA. Pathophysiology of Thrombocytopenia and Anemia in Mice Lacking Transcription Factor NF-E2. *Blood* 1999; **94**: 3037–47.

44 Gaine ME, Sharpe DJ, Smith JS, Colyer HAA, Hodges VM, Lappin TR, Mills KI. GATA2 regulates the erythropoietin receptor in t(12;21) ALL. *Oncotarget* Impact Journals; 2017; **8**: 66061–74.

45 Katsumura KR, Bresnick EH, the GATA Factor Mechanisms Group. The GATA factor revolution in hematology. *Blood* 2017; **129**: 2092–102.

46 Le Goff S, Boussaid I, Floquet C, Raimbault A, Hatin I, Andrieu-Soler C, Salma M, Leduc M, Gautier E-F, Guyot B, d'Allard D, Montel-Lehry N, Ducamp S, Houvert A, Guillonneau F, Giraudier S, Cramer-Bordé E, Morlé F, Diaz J-J, Hermine O, et al. p53 activation during ribosome biogenesis regulates normal erythroid differentiation. *Blood* 2021; **137**: 89–102.

47 Eaton N, Boyd EK, Biswas R, Lee-Sundlov MM, Dlugi TA, Ramsey HE, Zheng S, Burns RT, Sola-Visner MC, Hoffmeister KM, Falet H. Endocytosis of the thrombopoietin receptor Mpl regulates megakaryocyte and erythroid maturation in mice. *Frontiers in Oncology* 2022; **12**.

48 ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489**: 57–74.

49 McGowan KA, Pang WW, Bhardwaj R, Perez MG, Pluvinage JV, Glader BE, Malek R, Mendrysa SM, Weissman IL, Park CY, Barsh GS. Reduced ribosomal protein gene dosage and p53 activation in low-risk myelodysplastic syndrome. *Blood* 2011; **118**: 3622–33.

50 Keel SB, Phelps S, Sabo KM, O'Leary MN, Kirn-Safran CB, Abkowitz JL. Establishing Rps6 hemizygous mice as a model for studying how ribosomal protein haploinsufficiency impairs erythropoiesis. *Exp Hematol* 2012; **40**: 290–4.

581 51 Mok K-W, Chen H, Lee WM, Cheng CY. rpS6 regulates blood-testis barrier dynamics through
582 Arp3-mediated actin microfilament organization in rat sertoli cells. An in vitro study.
583 *Endocrinology* 2015; **156**: 1900–13.

584 52 Li L-X, Wu S-W, Yan M, Lian Q-Q, Ge R-S, Cheng CY. Regulation of blood-testis barrier dynamics
585 by the mTORC1/rpS6 signaling complex: An in vitro study. *Asian J Androl* 2019; **21**: 365–75.

586

Figure Legends

Figure 1. Single-cell RNA sequencing analysis in megakaryopoiesis cell stages in CD34$^+$-cells isolated from healthy volunteers.

A- Schematic diagram of the experimental design. Fresh CD34$^+$ cells from human peripheral blood were isolated via density gradient and magnetic cell sorting (n=2 healthy volunteers and n=2 patients). CD34$^+$ cells were differentiated into megakaryocytes using serum-free StemSpan SFEM II medium with megakaryocyte expansion supplement (SCF, IL6, TPO and IL9) (STEMCELL Technologies). Cells from different samples are incubated with DNA-barcoded antibodies recognizing ubiquitous cell surface proteins. Distinct barcodes (referred to as hashtag-oligos, HTO) on the antibodies allow pooling of multiple samples into one single-cell RNA (scRNA) sequencing experiment. The cells were analyzed at day 6 and day 11 of culture using 10X genomics technology.

B- UMAP of the control cells at days 6 and 11 after merging the two data sets. Day 6 cells are shown in orange and day 11 cells are shown in blue.

C- UMAP of the control cells after clustering performed using a resolution of 1.2. Each cluster is indicated with a distinct color.

D- Dot plot showing the expression levels of known specific hematopoietic cell marker genes in each cluster of controls. The color is representative of relative average expression of the cluster compared to all others. Circle size is associated with the percentage of cells expressing the gene in the cluster.

E- Dot plot of the top five differentially expressed genes (DEGs) in each cluster compared to all others in controls. The color is representative of relative average expression of the cluster compared to all others. Circle size is associated with the percentage of cells expressing the gene in the cluster.

F- UMAP showing the number of expressed genes detected in hematopoietic cells derived from normal $CD34^+$ cells. The gradient of color is associated to the number of expressed genes.

G- UMAP showing the number of overall mRNA counts detected in hematopoietic cells derived from normal $CD34^+$ cells. The gradient of color is associated to the number of overall mRNA count.

H- UMAP of control cells after assignation to hematopoietic cell types. Each cell type is distinguished from another by color. Clusters from figure 1C are superimposed over cell types.

I- UMAP of control cells after computation and scoring of proposed signatures (Supplementary table 2). Each plot represents the computation for one cell type signature. Color gradient is associated to the relative score of the signature.

Figure 2. Inference of genes regulatory network during megakaryopoiesis: control $CD34^+$-cell culture.

A- Heatmap of top 5 differentially active regulons for each cell type in control $CD34^+$-cell culture. Color gradient is representative of the relative activity of regulons. Each cell type is distinguished from another by color. Black rectangles stake out the top 5 regulons of the corresponding cell type.

B- UMAP of control cells after computation and scoring of proposed regulon signatures (Supplementary table 2, right panel). Each plot represents the computation for one cell type regulon signature. Color gradient is associated to the relative activity score of the signature.

C- Transcription factors gene regulatory network (TF-GRN) of control $CD34^+$ cell culture. Each hematopoietic cell type is represented by a colored rectangle. Each arrow represents known (solid) or supposed (dashed) differentiation steps. Following a cross-

reading of active regulons in populations, we were able to propose lists of specific (in rectangles) or shared (outside rectangles) TFs.

Figure 3. Single-cell RNA sequencing of CD34^{+}-cells from patients with an *ETV6* variant.

A- Upper panel: schematic pedigrees of *ETV6*-variant inheritance. Squares represent males, circles represent females, and slashes represent deceased family members. Solid black symbols represent family members with thrombocytopenia. The F1 family carried the *ETV6* p.P214L variant, while the F2 family carried the p.F417LTer4. Lower panel: schematic diagram of the ETV6 protein structure. The functional N-terminal pointed domain (PNT) and C-terminal ETS DNA-binding domain (ETS) are depicted. The positions of the *ETV6* mutations are indicated.

B- GripTite 293 MSR cells were co-transfected with an empty vector, wild type (WT) or *ETV6* variant construct with the luciferase reporter plasmid containing three tandem copies of the ETS binding site upstream of the HSV TK promoter (E743tk80Luc) and the pGL4.73 Renilla luciferase control vector. Firefly:Renilla luminescence ratios (Fluc/Rluc) were calculated to assess transfection efficiency and are expressed as fold change relative to the empty vector. The data represent the mean ± SEM of four independent experiments; **P-value <0.01 vs. empty vector (one-way ANOVA with Kruskal-Wallis post-hoc test). Representative western blot analysis of ETV6 expression in GripTite 293 MSR cells transfected with an empty vector, WT or *ETV6* variant constructs using a goat anti-ETV6 antibody. GAPDH was used as a protein loading control. Quantification of band intensity is shown on the right.

C- Left panel: UMAP of the patient cells at days 6 and 11 after merging the two data sets (patients D6 and patients D11, respectively). Day 6 cells are shown in orange and day

11 cells are shown in blue. Right panel: UMAP of the patient cells after clustering performed using a resolution of 0.8. Each cluster is indicated with a distinct color.

D- Dot plot showing the expression levels of known specific hematopoietic cell marker genes in each cluster of patients. The color is representative of relative average expression of the cluster compared to all others. Circle size is associated with the percentage of cells expressing the gene in the cluster.

E- Dot plot of the top five DEGs in each cluster compared to all others in patients. The color is representative of relative average expression of the cluster compared to all others. Circle size is associated with the percentage of cells expressing the gene in the cluster.

F- UMAP of the patient's cells after computation and scoring of proposed signatures (Supplementary table 2). Each plot represents the computation for one cell type signature. Color gradient is associated to the relative score of the signature.

G- UMAP of the patient's cells after assignation to hematopoietic cell types. Each cell type is distinguished from another by color. Clusters from figure 3C are superimposed over cell types.

Figure 4. Megakaryocyte differentiation in healthy controls and *ETV6* patients: trajectory inference.

A- Slingshot trajectory inference in control and patient cells. Cell types are distinguished by color. Black lines represent the differentiation trajectory from an early stage (HSPC) to late stages (PLP or MKP/MK) in both control and patient data sets.

B- Violin plots of expression of specific genes (*CD34, CD38, HLA-DRA, GP9, PF4, ITGA2B, FLI1, GATA1, GATA2, ETV6, RUNX1*) in control and patient cells. Cells are regrouped by cluster and colored according to their assigned cell type. Box plots are

superimposed over violin plots. Upper limits of box plots correspond to quantile 75 % and lower limits are associated to quantile 25 %. Solid black lines correspond to the median. Black dots are outlier values.

Figure 5. Single-cell RNA sequencing of cells derived from *ETV6*-variant carriers and healthy controls.

A- UMAP of hematopoietic cells derived from control and *ETV6*-variant $CD34^+$ cells on day 6 and day 11. Cells are colored by cell type (HSPC, CMP, MK-primed CMP, GMP, MEP, MKP/MK and PLP).

B- Pie charts of hematopoietic populations distribution in controls and patients at day 6 and day 11. Cell types are associated to distinct colors. Percentages were calculated for each condition independently. Control cells on day 6: HSPC = 5.16 %, CMP = 11.71 %, MK-primed CMP = 3.80 %, GMP = 2.83 %, MEP = 72.57 %, MKP/MK = 3.91 %, PLP = 0.04 %. Control cells on day 11: HSPC = 0.58 %, CMP = 2.14 %, MK-primed CMP = 2.06 %, GMP = 0.42 %, MEP = 15.21 %, MKP/MK = 78.05 %, PLP = 1.55 %. Patient cells on day 6: HSPC = 17.60 %, CMP = 10.29 %, MK-primed CMP = 1.67 %, GMP = 3.42 %, MEP = 60.03 %, MKP/MK = 7.00 %, PLP = 0.00 %. Patient cells on day 11: HSPC = 15.98 %, CMP = 8.18 %, MK-primed CMP = 6.37 %, GMP = 3.89 %, MEP = 28.67 %, MKP/MK = 36.92 %, PLP = 0.00 %.

C- UMAP of hematopoietic cells derived from control and *ETV6*-variant $CD34^+$ cells on day 6 and 11. Cells are colored by cell type (HSPC, CMP, MK-primed CMP, GMP, MEP, MKP/MK and PLP). Cells are located on the UMAP using regulon matrix.

D- Heatmap of the top 20 differentially active regulons (patients against controls) in MEP and MKP/MK populations. The color gradient is associated to the scaled activity of the regulon.

Figure 6. Differentially expressed genes and deregulated pathways in *ETV6*-variant carriers.

A- Volcano plots of up regulated (red) and down regulated (blue) genes in *ETV6* patients vs. controls for each cell type (HSPC, CMP, MK-primed CMP, GMP, MEP and MKP/MK), FC> 1.3 [i.e., log2 FC>0.37], adjusted p-value < 0.05).

B- Bubble plot of the top enriched GO biological processes based on differential gene expression by cell type (classified by p values). The upper panel shows the upregulated pathways, while the lower panel shows the down-regulated pathways. Cells are color-coded by cell types (HSPC, CMP, MK-primed CMP, GMP, MEP, MKP/MK)

C- Dot plot showing the expression levels of known specific "DNA repair" marker genes in MEP and MKP/MK of patients and controls. The color is representative of relative average expression in the population compared to all others. Circle size is associated with the percentage of cells expressing the gene in the population.

Figure 7. Deregulated pathways in *ETV6*-variant MKP/MK.

A- Translation levels in HEL in basal conditions and after PMA (phorbol 12 myristate 13 acetate) stimulation (NT=non-transduced, ETV6-WT=transduced with wild-type *ETV6*, ETV6-P214L=transduced with *ETV6*-P214L variant). Translation levels were evaluated via flow cytometry after the incorporation of puromycin and staining with anti-puromycin antibody in three independent experiments involving 4 patients. *P<0.05, **P<0.01, ***P<0.001, one-sample t-test.

B- Translation levels in $CD34^{+}$-derived MK on day 11 for P214L variant (F1-III3, IV1, III8) and day 14 for F417Lter4 variant cells (F2 II-2). Translation levels were evaluated via flow cytometry after the incorporation of puromycin and staining with

anti-puromycin antibody in three independent experiments involving 4 patients and their respective control.

C- Canonical ETV6 binding site contained in RPS6 gene. The list of up-regulated genes in ETV6-MKP/MK was obtained with cistarget (NES 4.85, rank 40e/512).

D- Violin plot showing the RPS6 expression levels for each cluster on control and ETV6-variant cells.

E- Visualization of RPS6 regulatory region with ETV6 binding peak identified via chromatin immunoprecipitation-sequencing. Part of the *DENND4C* gene is shown after the RPS6 gene using the UCSC genome browser (chr9: 19,372,807 – 19,387,833 using GRCh38/hg38).

F- The left panel: Western-blot analysis of RPS6 and phosphorylated RPS6 (p-RPS6) expression in washed platelets from controls and ETV6-variant carriers. GAPDH was used as a protein loading control. Quantification of band intensity is shown on the right. Right panel: Flow-cytometry assessment of RPS6 and p-RPS6 in $CD34^+$-derived MK from controls and ETV6-variant carriers at day 13 (F1-III3) or 14 (F1-III8, F2-II2) of differentiation. Each patient was compared to matched control subject on age and sex. For each patient and matched control, blood samples were collected at the same time and manipulated in parallel to avoid any bias.

G- Treatment of $CD34^+$-derived MK with CX-5461 at day 11. Effects of treatment were analyzed after 48h. The left panel: translation levels, RPS6 and p-RPS6 were analyzed by flow cytometry in healthy cells. The right panel: PLP and p-RPS6 were quantified by flow cytometry in *ETV6*-variant carriers. PLP were analyzed in supernatant in "platelet gate" (FSC low SSC low) and by expression of CD42a and MitoGreen, and were normalized on MK number (FS high SSC high, and CD42a expression).

758

759

Figure 1

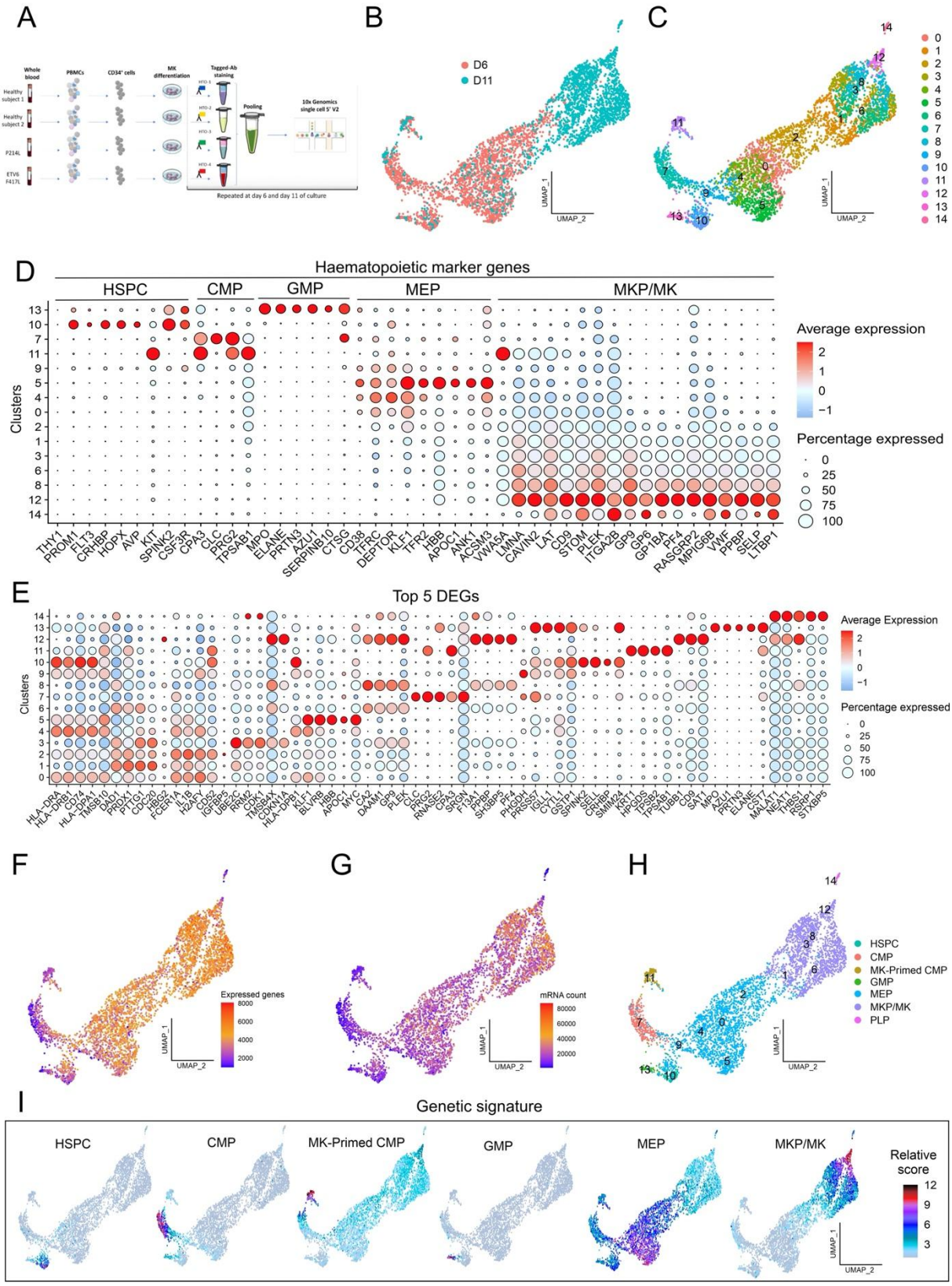
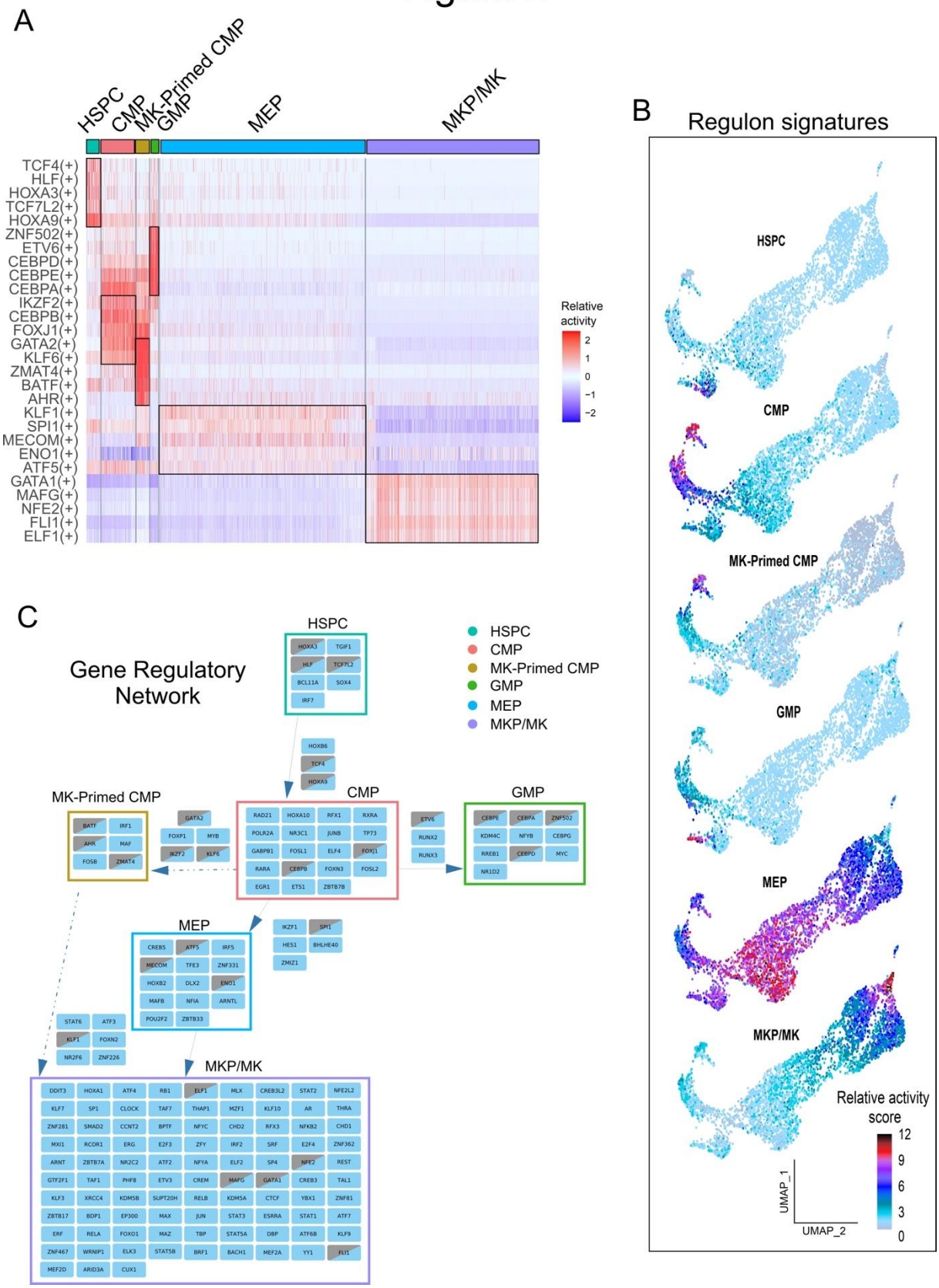

A
Whole blood
PBMCs
CD34+ cells
MK differentiation
Tagged-Ab staining
Pooling
10x Genomics single cell 5' V2
Repeated at day 6 and day 11 of culture
B
D6
D11
C
D
Haematopoietic marker genes
HSPC
CMP
GMP
MEP
MKP/MK
Clusters
Average expression
Percentage expressed
E
Top 5 DEGs
Average Expression
Percentage expressed
F
Expressed genes
G
mRNA count
H
HSPC
CMP
MK-Primed CMP
GMP
MEP
MKP/MK
PLP
I
Genetic signature
HSPC
CMP
MK-Primed CMP
GMP
MEP
MKP/MK
Relative score

Figure 2



761

762

Figure 3

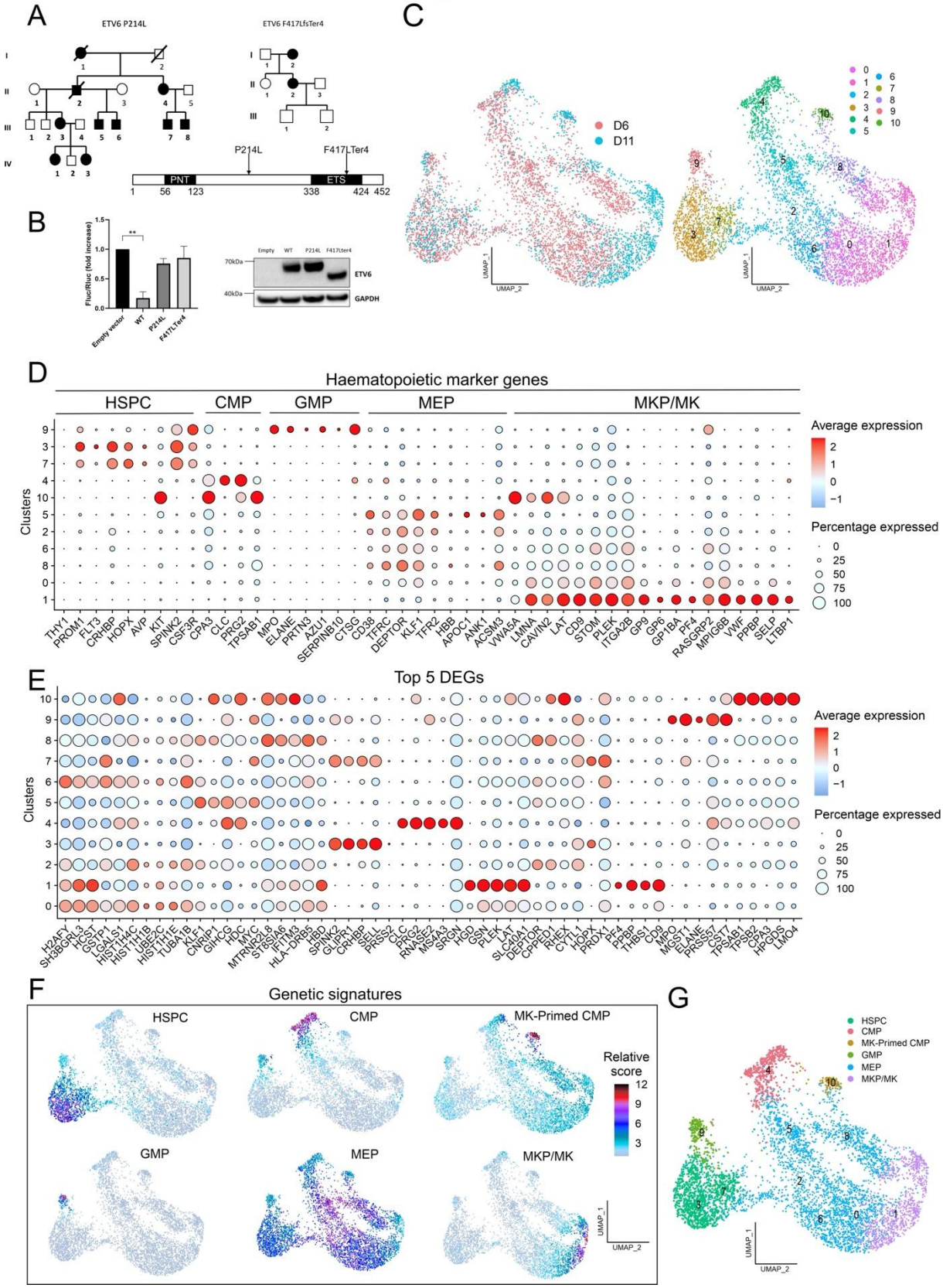
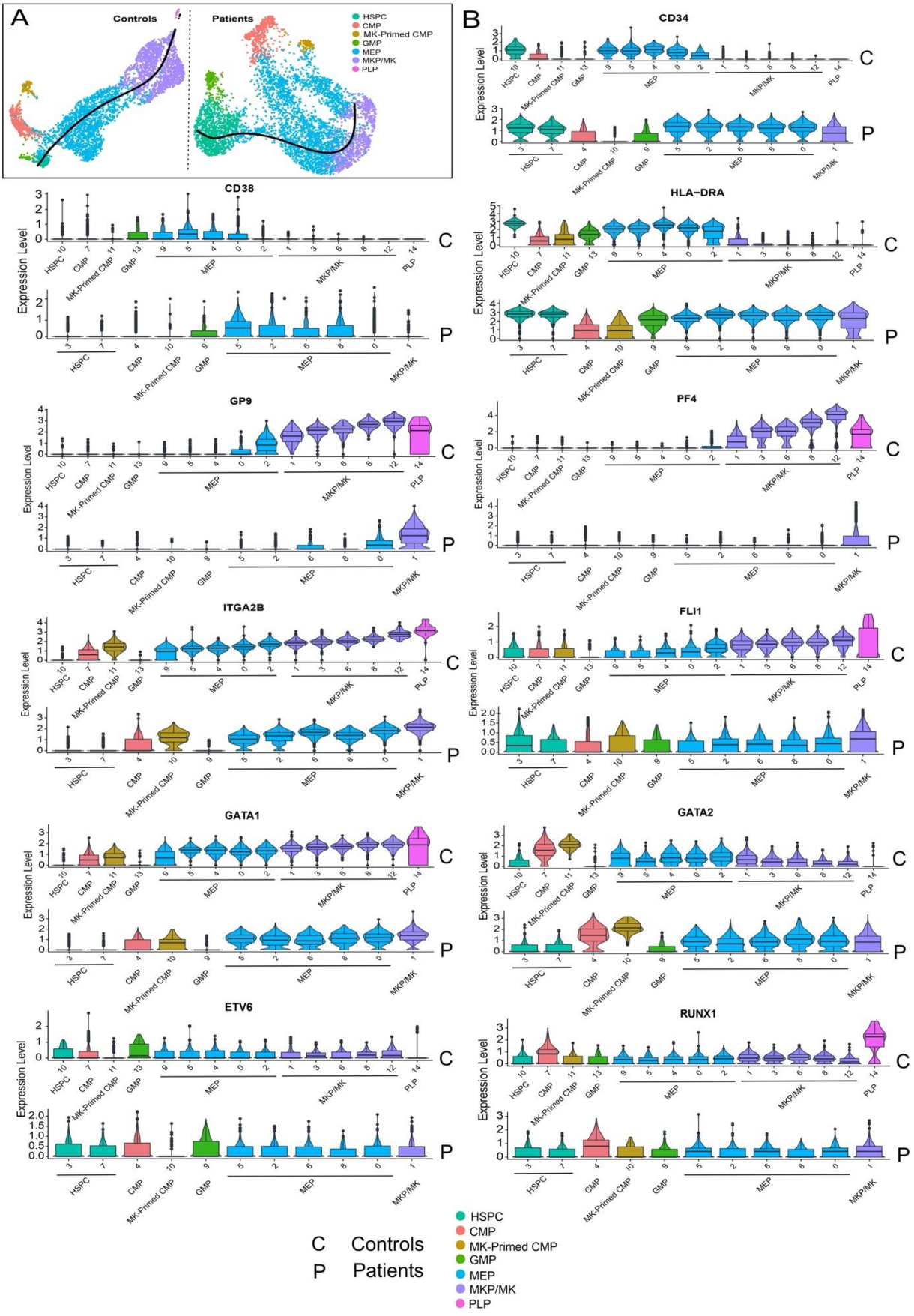

A
ETV6 P214L
ETV6 F417LfsTer4
P214L
F417LTer4
PNT
ETS
B
Fluc/Rluc (fold increase)
Empty vector
WT
P214L
F417LTer4
ETV6
GAPDH
C
D6
D11
D
Haematopoietic marker genes
HSPC
CMP
GMP
MEP
MKP/MK
Clusters
Average expression
Percentage expressed
E
Top 5 DEGs
F
Genetic signatures
HSPC
CMP
MK-Primed CMP
GMP
MEP
MKP/MK
Relative score
G
HSPC
CMP
MK-Primed CMP
GMP
MEP
MKP/MK

Figure 4


A
Controls
Patients
HSPC
CMP
MK-Primed CMP
GMP
MEP
MKP/MK
PLP
B
CD34
CD38
HLA-DRA
GP9
PF4
ITGA2B
FLI1
GATA1
GATA2
ETV6
RUNX1
Expression Level
C
P
C Controls
P Patients

765

Figure 5

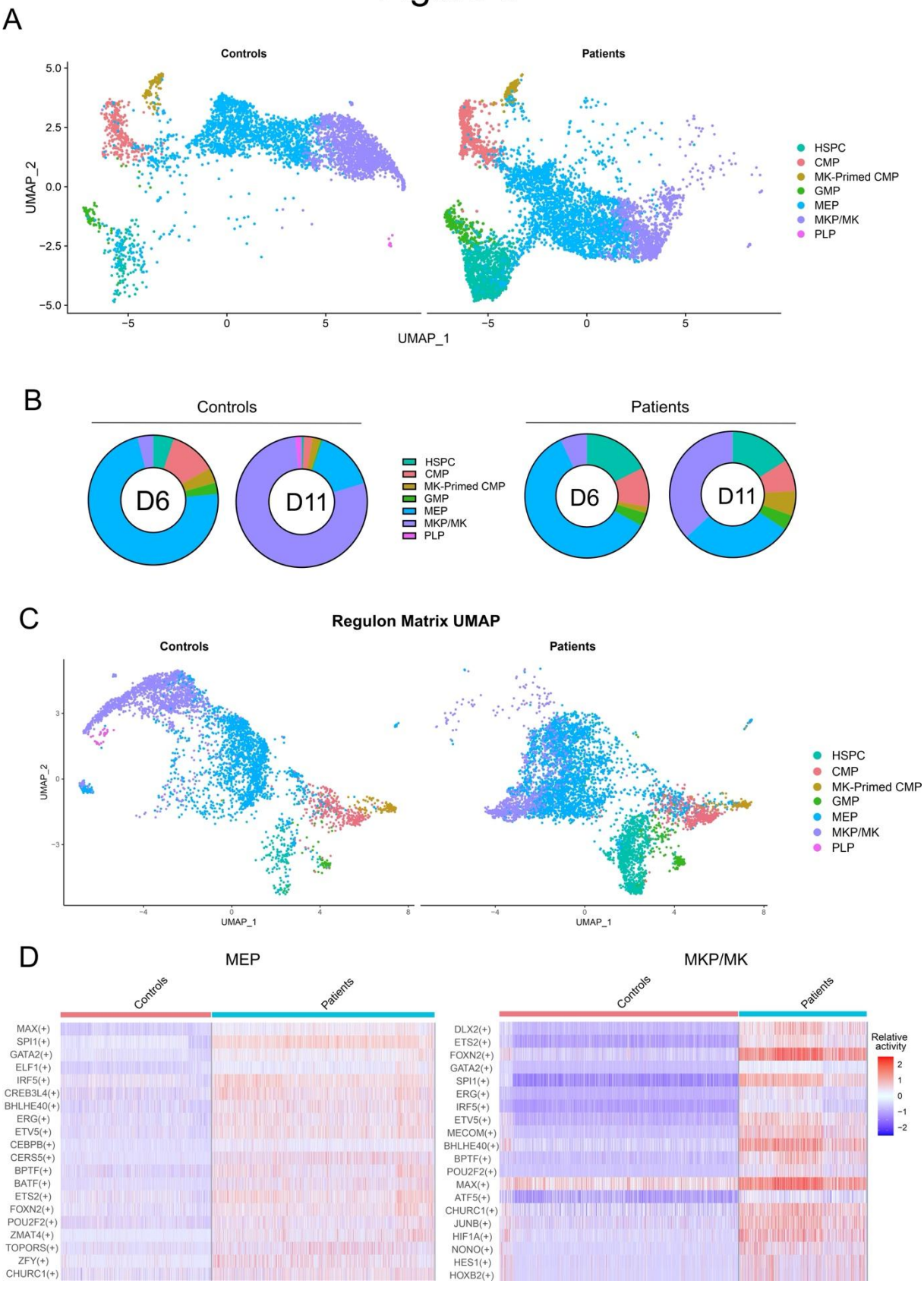

A
Controls
Patients
UMAP_2
UMAP_1
HSPC
CMP
MK-Primed CMP
GMP
MEP
MKP/MK
PLP
B
Controls
Patients
D6
D11
C
Regulon Matrix UMAP
D
MEP
MKP/MK
MAX(+)
SPI1(+)
GATA2(+)
ELF1(+)
IRF5(+)
CREB3L4(+)
BHLHE40(+)
ERG(+)
ETV5(+)
CEBPB(+)
CERS5(+)
BPTF(+)
BATF(+)
ETS2(+)
FOXN2(+)
POU2F2(+)
ZMAT4(+)
TOPORS(+)
ZFY(+)
CHURC1(+)
DLX2(+)
ETS2(+)
FOXN2(+)
GATA2(+)
SPI1(+)
ERG(+)
IRF5(+)
ETV5(+)
MECOM(+)
BHLHE40(+)
BPTF(+)
POU2F2(+)
MAX(+)
ATF5(+)
CHURC1(+)
JUNB(+)
HIF1A(+)
NONO(+)
HES1(+)
HOXB2(+)
Relative activity

Figure 6

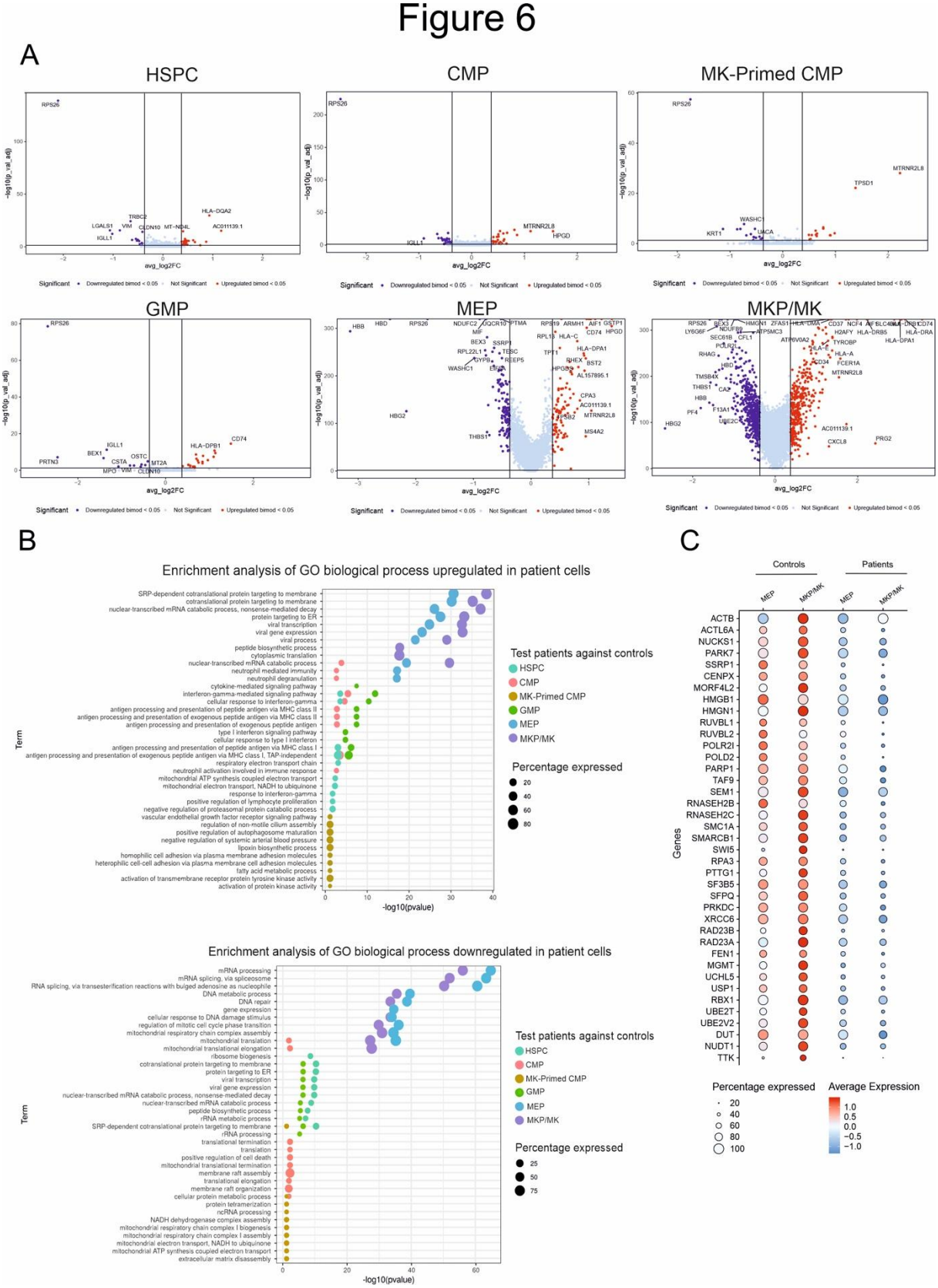
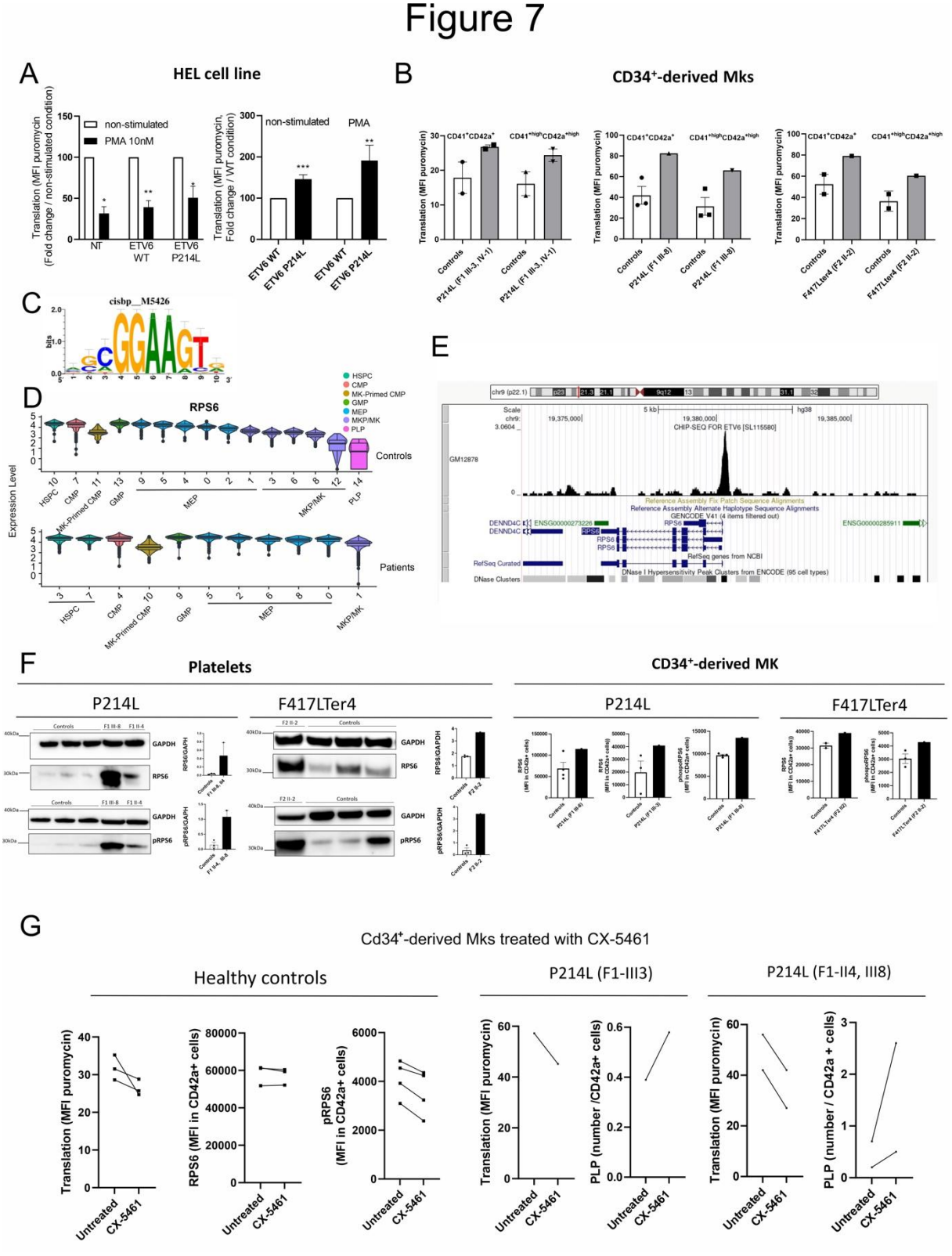
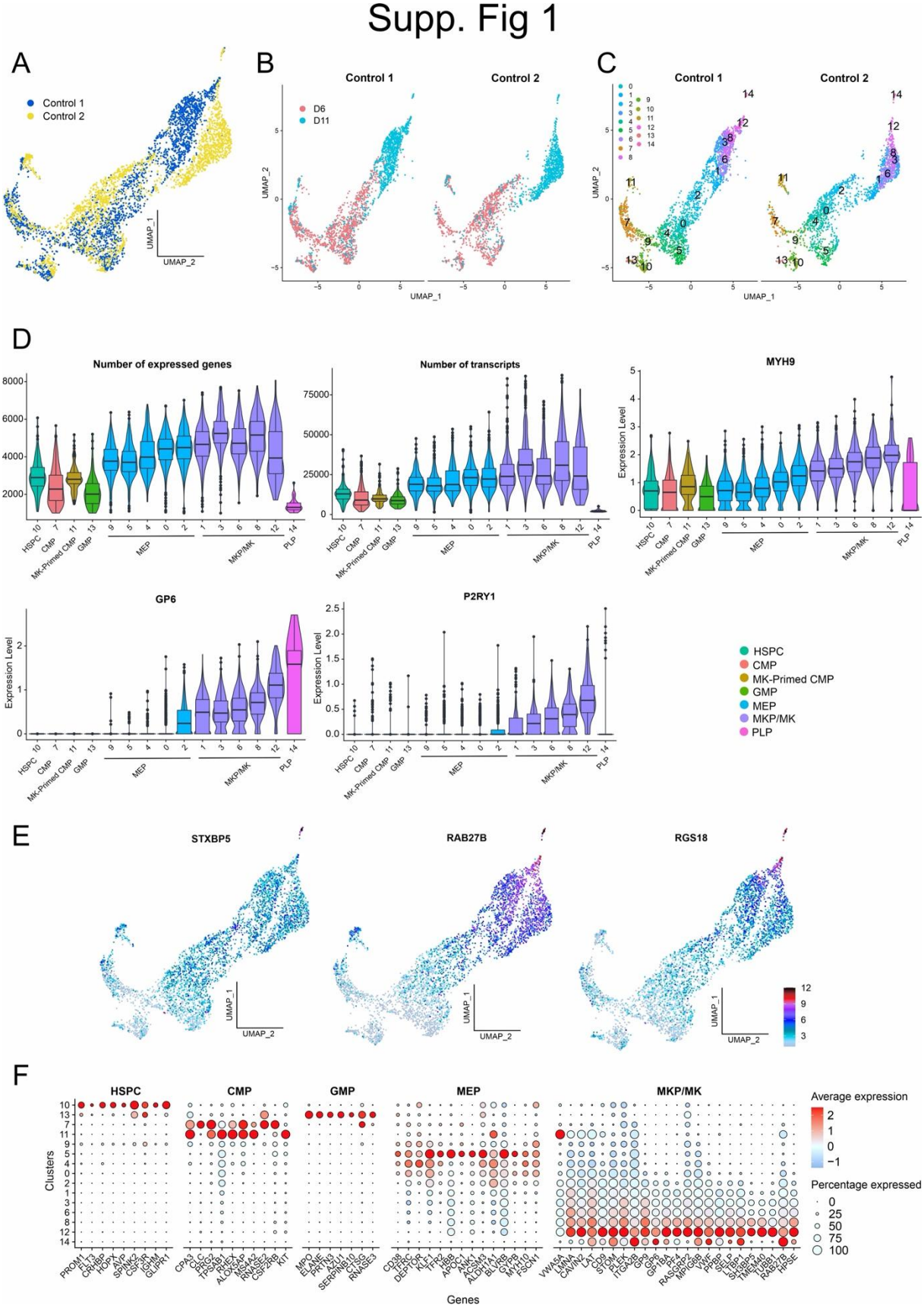

A
HSPC
CMP
MK-Primed CMP
GMP
MEP
MKP/MK
avg_log2FC
-log10(p_val_adj)
Significant
Downregulated bimod < 0.05
Not Significant
Upregulated bimod < 0.05
B
Enrichment analysis of GO biological process upregulated in patient cells
Term
-log10(pvalue)
Test patients against controls
HSPC
CMP
MK-Primed CMP
GMP
MEP
MKP/MK
Percentage expressed
Enrichment analysis of GO biological process downregulated in patient cells
C
Controls
Patients
Genes
Percentage expressed
Average Expression

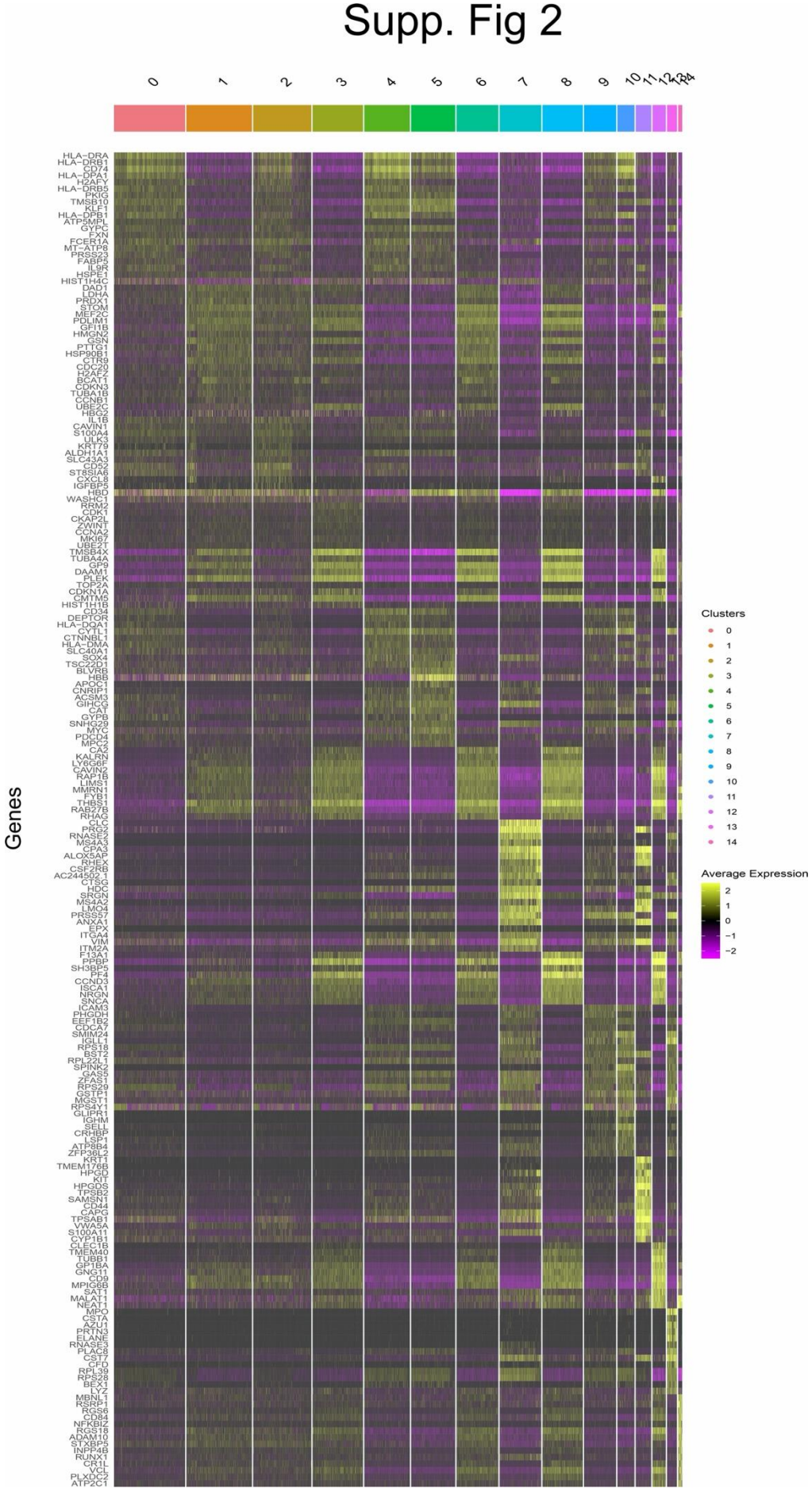
Figure 7


A
HEL cell line
non-stimulated
PMA 10nM
Translation (MFI puromycin (Fold change / non-stimulated condition)
NT
ETV6 WT
ETV6 P214L
Translation (MFI puromycin, Fold change / WT condition)
non-stimulated
PMA
ETV6 WT
ETV6 P214L
B
CD34+-derived Mks
Translation (MFI puromycin)
Controls
P214L (F1 III-3, IV-1)
P214L (F1 III-8)
F417Ter4 (F2 II-2)
C
cisbp__M5426
D
RPS6
HSPC
CMP
MK-Primed CMP
GMP
MEP
MKP/MK
PLP
Controls
Patients
Expression Level
E
GM12878
CHIP-SEQ FOR ETV6 [SL115580]
RefSeq Curated
DNase Clusters
F
Platelets
CD34+-derived MK
P214L
F417LTer4
GAPDH
RPS6
pRPS6
G
Cd34+-derived Mks treated with CX-5461
Healthy controls
P214L (F1-III3)
P214L (F1-II4, III8)
Translation (MFI puromycin)
RPS6 (MFI in CD42a+ cells)
pRPS6 (MFI in CD42a+ cells)
PLP (number /CD42a+ cells)
PLP (number / CD42a + cells)
Untreated
CX-5461

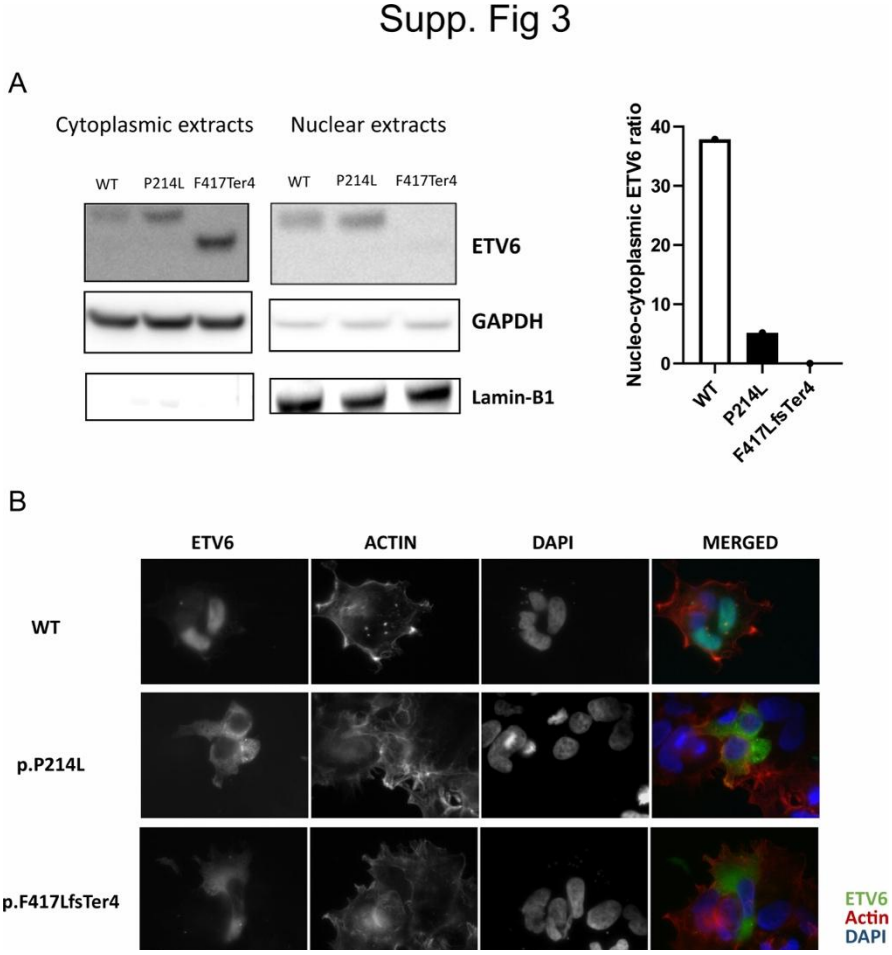
Supp. Fig 1


A
Control 1
Control 2
UMAP_1
UMAP_2
B
Control 1
Control 2
D6
D11
C
D
Number of expressed genes
Number of transcripts
MYH9
Expression Level
GP6
P2RY1
HSPC
CMP
MK-Primed CMP
GMP
MEP
MKP/MK
PLP
E
STXBP5
RAB27B
RGS18
F
HSPC
CMP
GMP
MEP
MKP/MK
Clusters
Genes
Average expression
Percentage expressed

Supp. Fig 2

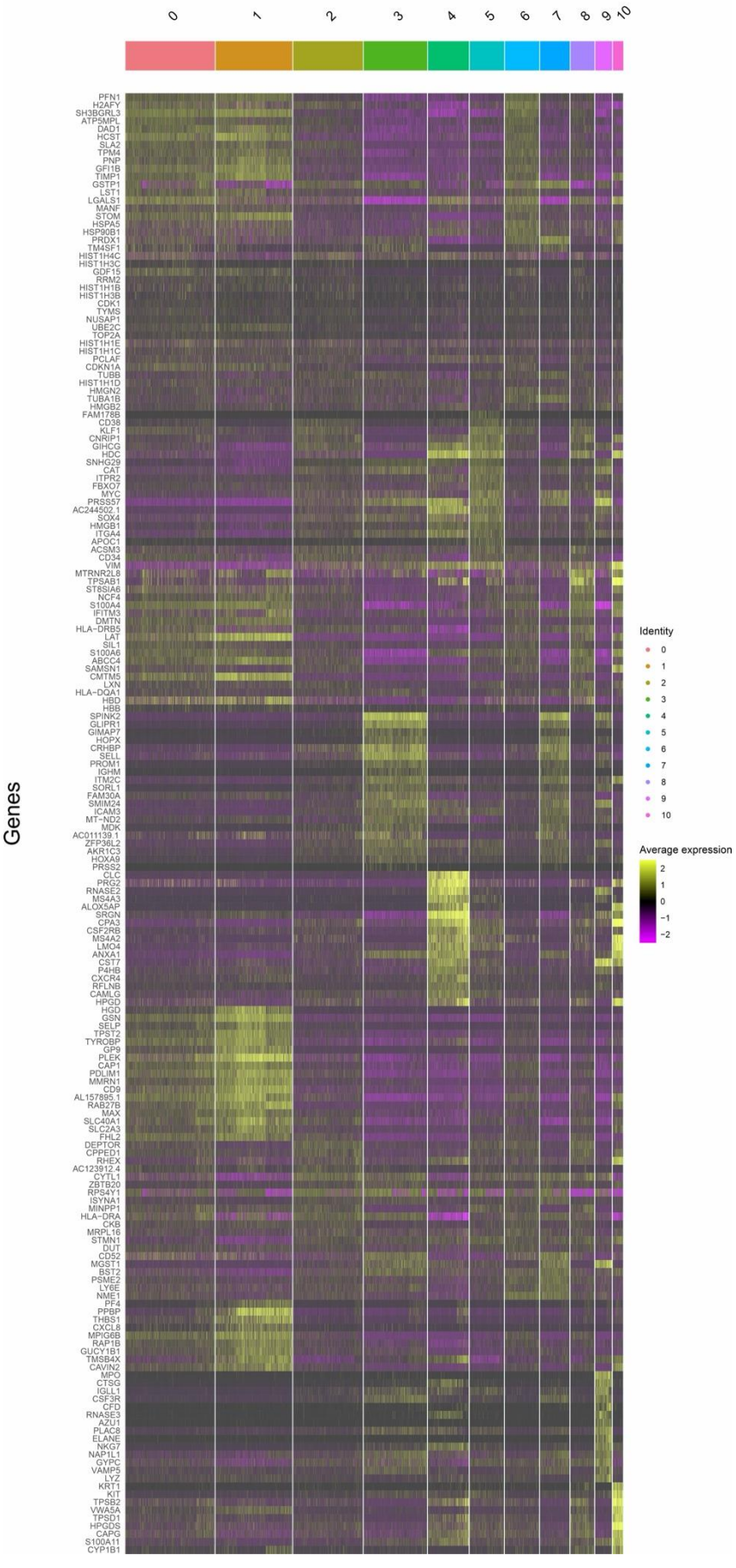

0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Genes
Clusters
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Average Expression
2
1
0
−1
−2

Supp. Fig 3

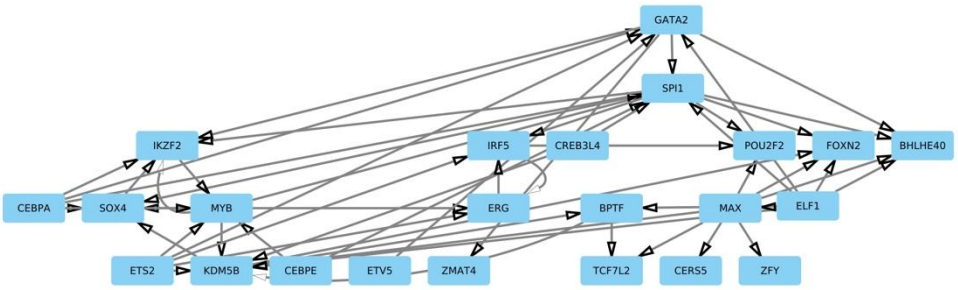


777

Supp. Fig 4


0
1
2
3
4
5
6
7
8
9
10
Genes
PFN1
H2AFY
SH3BGRL3
ATP5MPL
DAD1
HCST
SLA2
TPM4
PNP
GFI1B
TIMP1
GSTP1
LST1
LGALS1
MANF
STOM
HSPA5
HSP90B1
PRDX1
TM4SF1
HIST1H4C
HIST1H3C
GDF15
RRM2
HIST1H1B
HIST1H3B
CDK1
TYMS
NUSAP1
UBE2C
TOP2A
HIST1H1E
HIST1H1C
PCLAF
CDKN1A
TUBB
HIST1H1D
HMGN2
TUBA1B
HMGB2
FAM178B
CD38
KLF1
CNRIP1
GIHCG
HDC
SNHG29
CAT
ITPR2
FBXO7
MYC
PRSS57
AC244502.1
SOX4
HMGB1
ITGA4
APOC1
ACSM3
CD34
VIM
MTRNR2L8
TPSAB1
ST8SIA6
NCF4
S100A4
IFITM3
DMTN
HLA-DRB5
LAT
SIL1
S100A6
ABCC4
SAMSN1
CMTM5
LXN
HLA-DQA1
HBD
HBB
SPINK2
GLIPR1
GIMAP7
HOPX
CRHBP
SELL
PROM1
IGHM
ITM2C
SORL1
FAM30A
SMIM24
ICAM3
MT-ND2
MDK
AC011139.1
ZFP36L2
AKR1C3
HOXA9
PRSS2
CLC
PRG2
RNASE2
MS4A3
ALOX5AP
SRGN
CPA3
CSF2RB
MS4A2
LMO4
ANXA1
CST7
P4HB
CXCR4
RFLNB
CAMLG
HPGD
HGD
GSN
SELP
TPST2
TYROBP
GP9
PLEK
CAP1
PDLIM1
MMRN1
CD9
AL157895.1
RAB27B
MAX
SLC40A1
SLC2A3
FHL2
DEPTOR
CPPED1
RHEX
AC123912.4
CYTL1
ZBTB20
RPS4Y1
ISYNA1
MINPP1
HLA-DRA
CKB
MRPL16
STMN1
DUT
CD52
MGST1
BST2
PSME2
LY6E
NME1
PF4
PPBP
THBS1
CXCL8
MPIG6B
RAP1B
GUCY1B1
TMSB4X
CAVIN2
MPO
CTSG
IGLL1
CSF3R
CFD
RNASE3
AZU1
PLAC8
ELANE
NKG7
NAP1L1
GYPC
VAMP5
LYZ
KRT1
KIT
TPSB2
VWA5A
TPSD1
HPGDS
CAPG
S100A11
CYP1B1
Identity
0
1
2
3
4
5
6
7
8
9
10
Average expression
2
1
0
−1
−2

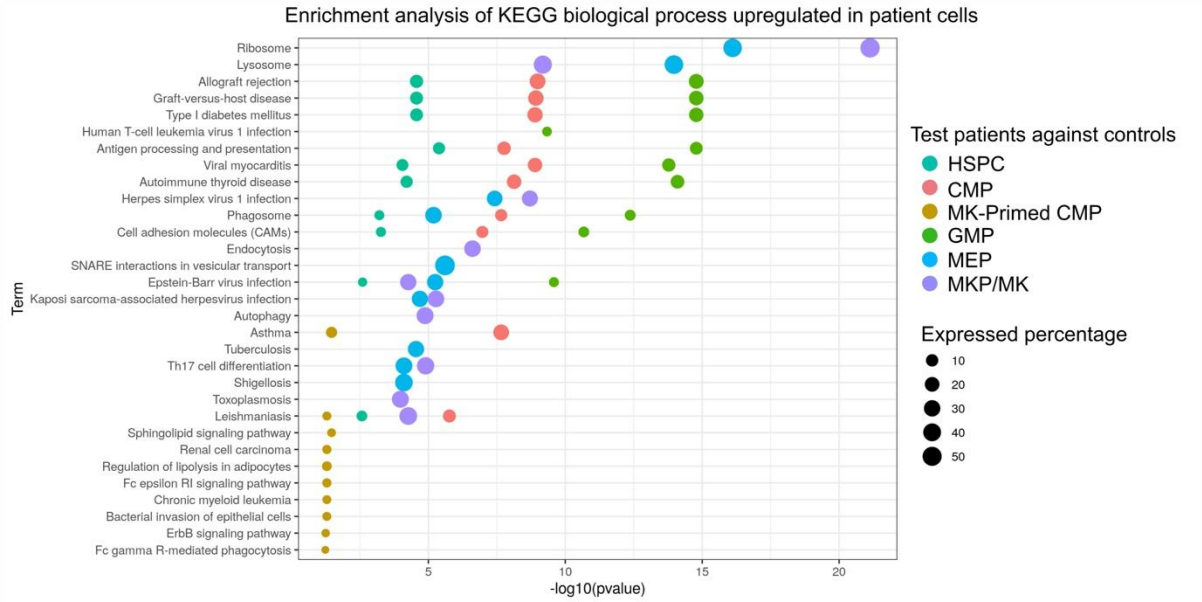
Supp. Fig 5



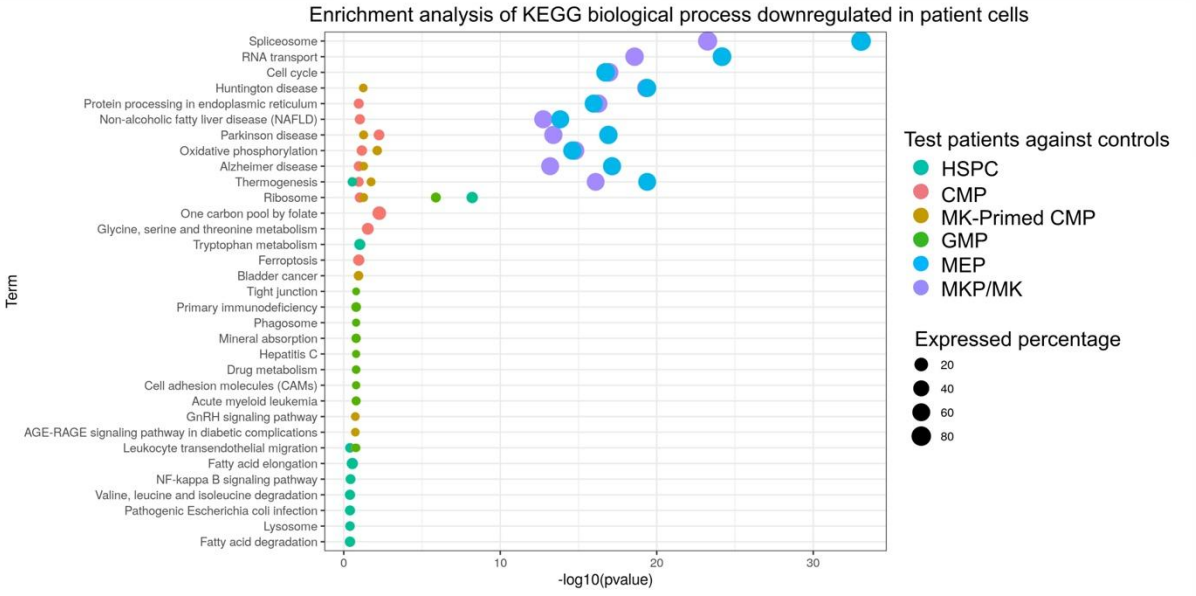
779

Supp. Fig 6

A



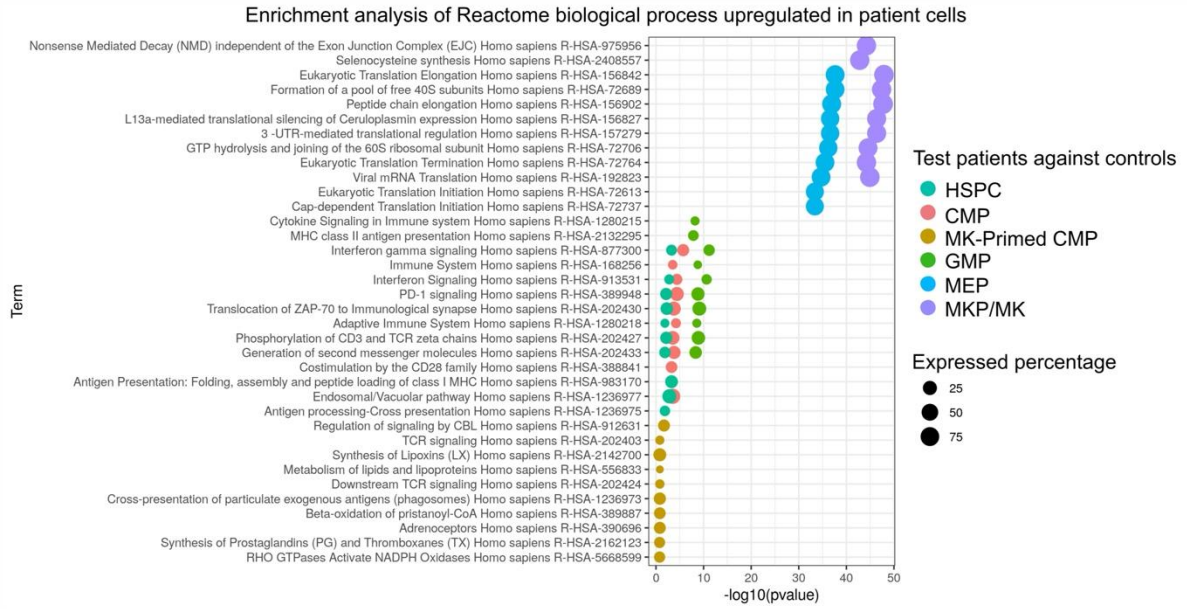
B



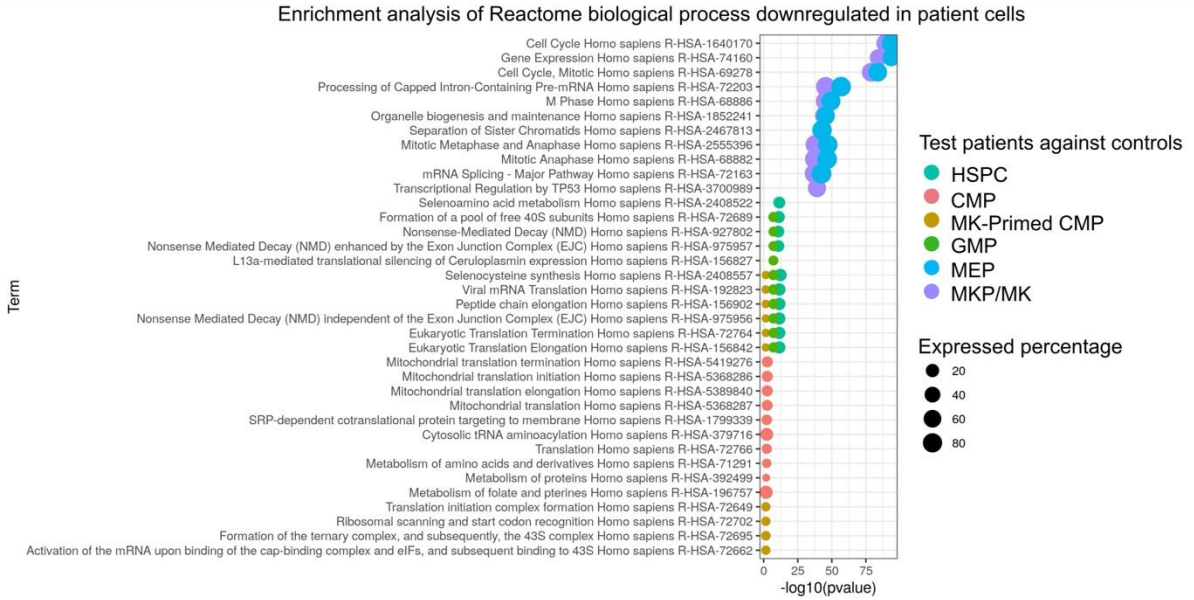
780

Supp. Fig 7

A

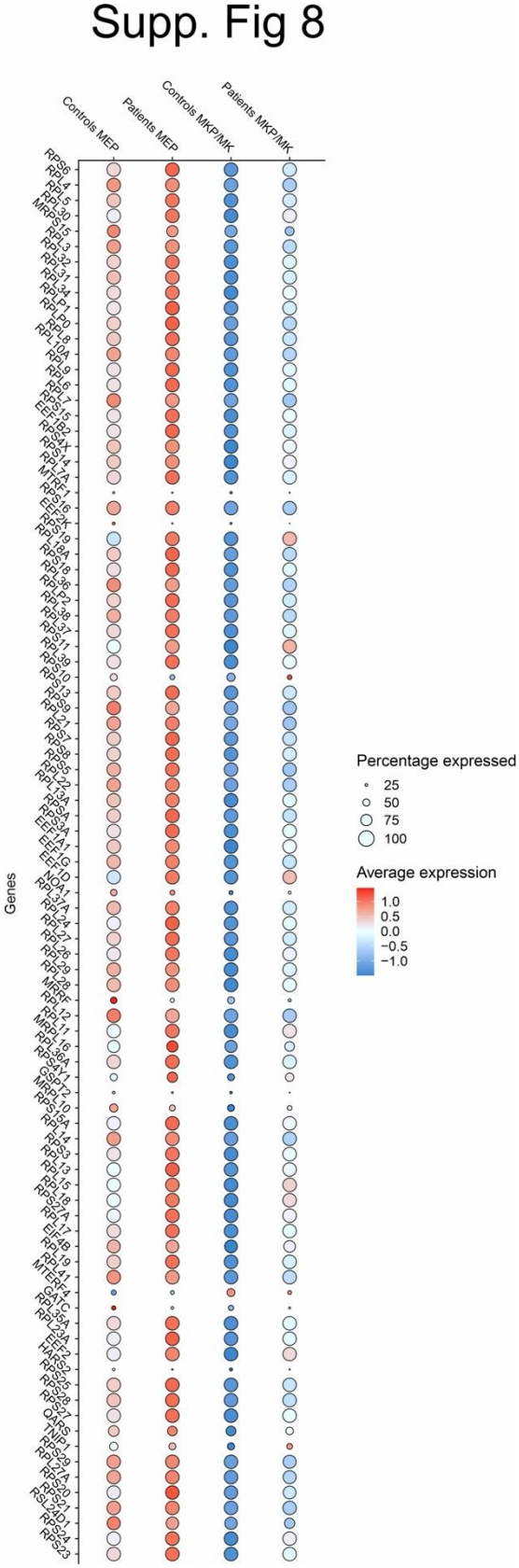


B



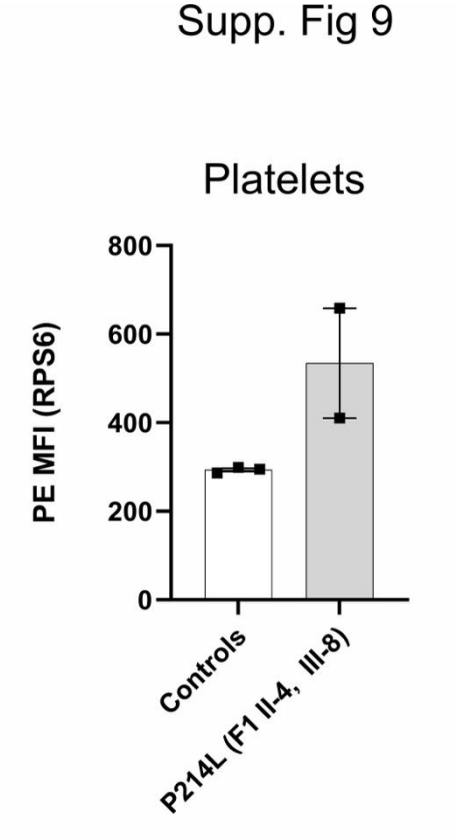
781

Supp. Fig 8



782

Supp. Fig 9



783

784

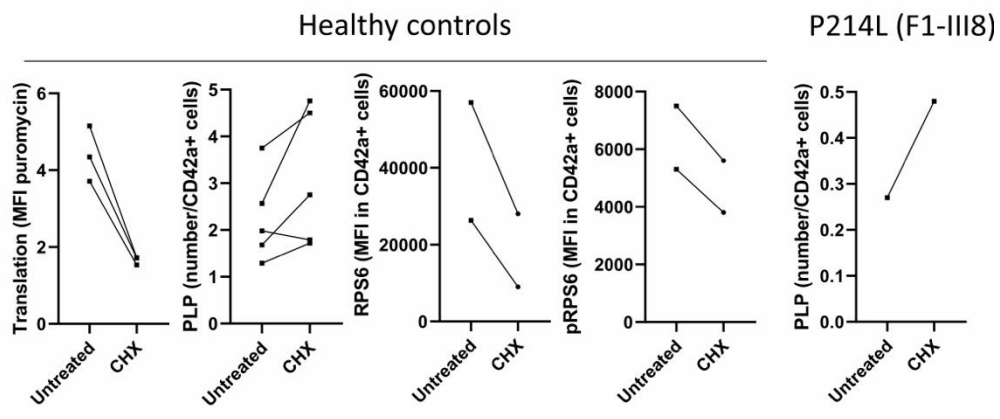
785

786

787

788

Supp. Fig 10



789