MANUSCRIPT

Molecular characterization of *EZH2*-mutant patients with myelodysplastic/myeloproliferative neoplasms

Jenny Rinke[1], Jörg P. Müller[2], Markus F. Blaess[3], Andrew Chase[4,5], Manja Meggendorfer[6], Vivien Schäfer[1], Nils Winkelmann[1], Claudia Haferlach[6], Nicholas C.P. Cross[4,5], Andreas Hochhaus[1], Thomas Ernst[1]

[1] Abteilung Hämatologie und Internistische Onkologie, Klinik für Innere Medizin II, Universitätsklinikum Jena, Jena, Germany

[2] Institut für Molekulare Zellbiologie, Center for Molecular Biomedicine, Universitätsklinikum Jena, Jena, Germany

[3] Klinik für Anästhesiologie und Intensivmedizin, Universitätsklinikum Jena, Jena, Germany

[4] National Genetics Reference Laboratory (Wessex), Salisbury District Hospital, Salisbury, UK

[5] Faculty of Medicine, University of Southampton, Southampton, UK

[6] MLL Münchner Leukämielabor, Munich, Germany

Correspondence: PD Dr. Thomas Ernst, Abteilung Hämatologie und Internistische Onkologie, Klinik für Innere Medizin II, Universitätsklinikum Jena, Am Klinikum 1, 07740 Jena, Germany; Phone: +49 3641 9324563; Fax: +49 3641 9324202; E-Mail: thomas.ernst@med.uni-jena.de

Keywords: *EZH2*, mutation, CMML, MDS/MPN, *ETS1*, target genes

Running Title: *EZH2* dysregulation in MDS/MPN

Word count:

Abstract

Mutations in the epigenetic regulator gene *EZH2* are frequently observed in patients with myelodysplastic/myeloproliferative neoplasms (MDS/MPN; 10-13%) and are associated with a poor outcome. To gain more insight into EZH2 pathology we sought to genetically characterize a cohort of 41 *EZH2*-mutated MDS/MPN patients using targeted deep next generation sequencing (NGS), colony forming progenitor assays and transcriptome analysis. Stable shRNA-mediated downregulation of *EZH2* was performed in MDS derived F-36P, MOLM-13 and OCI-M2 cells to study EZH2 specific changes. Targeted NGS revealed a complex pattern of mutations with a total of 190 individual mutations. *EZH2* mutations frequently co-occur with *TET2* (58%), *RUNX1* (40%) and *ASXL1* (34%) mutations. Colony assays indicated *EZH2* mutations to be mostly early events in leukemogenesis and showed a complex mutational hierarchy. Gene expression data revealed a number of differently expressed genes between *EZH2*-wild type and mutant patients including known *EZH2* targets. Comparison of patient transcriptome to *EZH2* downregulated cell line data revealed several genes as novel *EZH2* targets, showing opposite as well as unidirectional regulation between cell lines and patients. Some genes, such as *CXXC5*, *ETS1* and *VAV3* have previously been implied to play a role in leukemogenesis. Their precise role in MDS/MPN needs to be further investigated.

Introduction

Discoveries like *BCR-ABL1* in chronic myeloid leukemia (CML)[1] and *JAK2-V617F* in polycythemia vera (PV)[2] have shown that single molecular aberrations can bring about a better understanding of a disease entity and aid classification and treatment.[3,4]

However, in recent years, through advances in sequencing technology, particularly next generation sequencing (NGS), the ongoing effort to better define the molecular background of myeloid malignancies has revealed a perhaps unexpected complexity. A large number of somatic mutations affecting multiple pathways have been identified with varying frequencies and combinations that overlap the different disease entities.[5] Originally aberrations were discovered in genes that confer a growth advantage by altering signaling pathways and expression of key transcriptional targets.[6] Over time additional pathways such as epigenetic modification, RNA splicing, and as recently discovered the cohesin complex, in particular *STAG2*, were found to be involved in myeloid leukemogenesis.[5,7,8] The task at hand is to unravel these heterogeneous molecular patterns and hierarchies, to study their effects on disease phenotype and ultimately to identify new therapeutic targets. Epigenetic targeted therapy is of particular interest because epigenetic changes – unlike somatic gene mutations – are potentially reversible and offer promising therapeutic alternatives even in elderly patients.[9]

This study focuses on myeloproliferative/myelodysplastic neoplasms (MDS/MPN) in adults, as this overlap category encompasses both the dysplastic and the hyperproliferative features seen in other myeloid malignancies. Chronic myelomonocytic leukemia (CMML) is the most common entity in this category with a heterogeneous genetic background that is reflected in the diverse clinical and pathological features of this disease. Other diseases in the MDS/MPN category show

a similar complexity and include *BCR-ABL1*-negative atypical chronic myeloid leukemia (aCML), MDS/MPN, unclassifiable (MDS/MPN-U) and juvenile myelomonocytic leukemia (JMML).[3]

Amongst the most frequently mutated signaling genes in MDS/MPN are *CBL* (8-20%), *KRAS* (7-11%), *NRAS* (4-30%) and *SETBP1* (6-25%). Aberrations of components of the RNA splicing machinery are common, most frequently involving *SRSF2* with mutations detectable in 36-46% of CMML patients and less frequently involving *U2AF35* (5-15%) and *ZRSR2* (8-10%). *RUNX1* which encodes a transcription factor involved in cell growth, survival and differentiation is mutated in 6-20% of MDS/MPN cases. Thirty to 60% of patients show *TET2* mutations and less than 10% *DNMT3A* and *IDH1/2* mutations, all of which impair DNA methylation. Other epigenetic changes affecting histone modifications are frequently detected in *ASXL1* (20-44%) and *EZH2* (6-13%).[5]

EZH2 is an epigenetic regulator gene encoding the catalytic subunit of the polycomb repressive complex 2 (PRC2), a histone H3 lysine 27 (H3K27) methyltransferase. Gain-of-function mutations in this gene have previously been observed in prostate, breast and bladder cancer amongst others.[10] Remarkably, loss-of-function mutations have been detected in myeloid malignancies, highlighting the ambiguous role of *EZH2* in cancer. Such mutations are associated with poor overall and progression-free survival and occur most frequently in the MDS/MPN category (10-13%).[11]

In light of these findings we aimed to genetically characterize a cohort of 41 *EZH2*-mutated MDS/MPN patients to investigate mutational patterns associated with *EZH2*. To better understand the complex clonal hierarchies of detected mutations, colony assays were carried out. Additionally, to investigate the effect of loss-of-function *EZH2* mutations on gene expression we carried out whole transcriptome analysis on a subset of our patient cohort compared to *EZH2*-wild type MDS/MPN patients. To

identify specific *EZH2* targets, we compared results to cell lines F-36P, MOLM-13 and OCI-M2, where suppression of *EZH2* translation was achieved by stable transduction of *EZH2*-targeting shRNAs.

Methods and Patients

Patients

Bone marrow (BM) samples were collected from 41 *EZH2*-mutation positive MDS/MPN patients (CMML, n = 25; aCML, n = 11; MDS/MPN-U, n = 5; male, n = 25). Patients were classified according to the 2008 WHO classification of myeloid neoplasms and acute leukemia.[3] Bone marrow samples of a cohort of 12 *EZH2*-wildtype MDS/MPN patients (A - L; male, n = 9) were used as a control group. Clinical information is given in Supplemental Table 1. The study has been approved by the institutional ethics committee and written informed consent was provided according to the Declaration of Helsinki.

Sample Preparation

Total BM leukocytes were isolated after erythrocyte lysis according to standard protocols. Leukocyte RNA extraction was performed using TRIzol® reagent, as described previously.[12] Genomic DNA from leukocytes was extracted using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's guidelines. Mononuclear cells were isolated according to standard procedures using Ficoll density gradient centrifugation, and CD34+ selected by MACS® separation with the CD34 MicroBead Kit (Miltenyi Biotec, Bergisch Gladbach, Germany) according to manufacturer's instructions.

Next-generation sequencing (NGS)

NGS, including data analysis was performed for 30 leukemia-associated target genes on the 454 GS Junior platform with 454 GS Junior Titanium chemistry for amplicon sequencing (Roche Diagnostics, Basel, Switzerland) as previously described.[13,14] In

total, 231 amplicons were prepared for each sample and processed in a single NGS run. Mutations with a frequency >15% were confirmed using Sanger Sequencing with standard techniques on an ABI 3500 Genetic Analyzer (Thermo Fisher Scientific, Waltham, MA, USA) as described previously.[14] DNA-sequencing raw data is available on reasonable request.

Colony Assays and Sanger Sequencing

Cryoconserved CD34+ cells of patients #24, 37, 40 and 41 were plated and processed for analysis by colony assay as previously described.[15] To check the specific mutational status of individual colonies, Sanger sequencing of PCR products covering the relevant genetic regions was performed according to standard techniques.[14] Mutational analysis was performed using the Mutation Surveyor Software (SoftGenetics, State College, PA, USA). A dropping factor (drop in height of the normal peak at the position of the mutation relative to the neighboring peaks) of <30% was considered indicative of wild type, a dropping factor between 30% and 70% was considered to be a heterozygous mutation and a factor >70% was considered to be homozygous.

Stable shRNA-mediated downregulation of EZH2

Two shRNA encoding oligonucleotides targeting *EZH2* (shEZH2_1: 5' -GAGAGATTATTTCTCAAGATG- 3'; shEZH2_2: 5' -TGGAAAGAACGGAAATCTTAA -3'), or a non-targeting control shRNA (Sigma-Aldrich, St. Louis, MO, USA) were inserted between AgeI and EcoRI restriction sites into plasmid pLKO.1 (Sigma). Expression constructs were transduced and selected in cell lines F-36P (ACC-543), MOLM-13 (ACC-554) and OCI-M2 (ACC-619), purchased from The Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, as described previously[16] and tested for mycoplasma contamination. Growth curves of cell lines

were generated using the CellTitre-Glo cell viability assay (PROMEGA, Fitchburg, WI, USA), according to manufacturer instructions.

Whole transcriptome analysis and qRT-PCR

Total RNA (200 ng) of 24 MDS/MPN BM leukocyte samples (n = 12 *EZH2*-mutated and wild type cases, respectively) as well as shEZH2_1 and shEZH2_2 downregulated F-36P, MOLM-13 and OCI-M2 cell lines and respective control cells transduced with non-targeting shRNA (4 each, collected at 4 time points) were processed and analyzed as previously described.[17] Control probe quality check, background subtraction, signal averaging and data analysis were performed using Illumina GenomeStudio-Software v.2011.1 (Illumina, San Diego, CA, USA) and GraphPad Prism v6.01 (GraphPad Software, Inc., San Diego, CA, USA). Expression data obtained from *EZH2*-wild type vs. mutant MDS/MPN patients for genes showing FC values ≥2 and signal intensities ≥100, as well as detection p-values <0.05 in at least one group, were subjected to statistical analysis (Mann-Whitney test) and compared to F-36P, MOLM-13 and OCI-M2 expression data (Student's t-test). RNA-seq data have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-5766. Results were verified using qRT-PCR according to standard protocols with SYBR-green dye (Roche) and *beta glucuronidase* (*GUSB*) as a reference gene. Primer sequences are available in Supplemental Table 2.

Western blot analysis

Protein blot analysis was performed for genes showing average expression signals >100 in cell lines and patients, using primary antibodies against EZH2 (#3147), histone H3 (#9717), H3K27me3 (#9733), CXXC5 (#84546), ETS1 (#14069), VAV3 (#2398), GAPDH (#2118) (Cell Signaling Technology, Boston, MA, USA), FAM133B (HPA043901) (Atlas Antibodies, Bromma, Sweden) and STS-1 (ab197027) (Abcam,

Cambridge, UK). Protein isolation was carried out using 5 × 10^6 cells, according to standard procedures using radioimmunoprecipitation assay buffer (RIPA) with standard protease inhibitors and 100 mM sodium-ortho-vanadate. For each sample protein lysates were separated, blotted and blocked as previously described.[18] Blots were incubated with the indicated primary antibodies for 16 h at 4°C. After incubation with respective secondary antibodies, chemiluminescent HRP substrate (Millipore, Billerica, MA, USA) was added. Proteins were visualized with ImageQuant LAS-4000 (GE Healthcare, Chicago, IL, USA) and quantified using ImageJ (National Institute of Health, Bethesda, MD, USA).

Results

Cooperating mutations in leukemia-associated genes in 41 EZH2-mutated MDS/MPN patients

In total 190 individual mutations affecting 19 genes were detected, including 53 mutations of *EZH2*. In addition to *EZH2* mutations, 23/41 patients (56%) showed cooperating *TET2* mutations, followed by *RUNX1* (39%), *ASXL1* (34%), *CBL* (24%), *ZRSR2* (20%), *SRSF2* (15%), *NRAS* (15%), *STAG2* (12%) and *SETBP1* (12%), affecting all five aforementioned pathways. The majority of patients with mutations in *TET2*, *RUNX1* and/or *ASXL1* carried frameshift, nonsense or splice site mutations in these genes (83%, 81%, 93%, respectively), which often have severe functional consequences. Multiple mutations in the same gene were most commonly observed in *TET2* with 16/23 (70%) affected patients showing up to three mutations. Approximately 30% of affected patients carried multiple mutations in *EZH2*, *RUNX1* and *CBL* (Fig 1). Detailed DNA-sequencing data is given in Supplemental Table 3. As some mutations were located on the same amplicon biallelic mutations could be distinguished from monoallelic mutations. Biallelic mutations were detected in patients #4 (*CBL*), 13 (*CBL*) and 26 (*TET2*) whilst patients #30 (*RUNX1*) and 40

(*EZH2*) carried monoallelic mutations and patient #32 (*RUNX1*) showed overlapping mutations with indistinct allelic origin (Supplemental Fig 1).

Clonal hierarchies of patients #24, 37, 40 and 41

Analysis of 216 individual colonies for mutations detected by NGS revealed various mutational patterns. The detection of clones carrying only *EZH2* mutations in patients #37 and 40 indicates that *EZH2* precedes the acquisition of *SETBP1* mutations, followed by mutations of *NRAS* and *CBL* respectively (Figs 2A, 2B). In patient #37 the low-level *KRAS*, *RUNX1*, *STAG2* and *ZRSR2* mutations previously detected by NGS were not found in the analyzed colonies. The putative evolutionary tree generated for patient #24 also indicates *EZH2* and *SETBP1* mutations to be early events in leukemogenesis (Fig 2C). Overall, a more complex pattern of mutations was detected in this patient, with various subclones, changing from heterozygous to homozygous or wild type states for certain mutations. Interestingly, colony analysis of patient #41 indicates the early acquisition of an *ASXL1* mutation, followed by *EZH2* mutations p.T467Hfs*16 and p.Y579X, respectively (Fig 2D).

Transcriptome analysis of EZH2-wild type and mutant MDS/MPN patients

Initial comparison of transcriptome-wide expression data of *EZH2*-wild type and mutant MDS/MPN patients showed clear differences in overall gene expression, particularly for a number of highly expressed genes, albeit not above 2 FC (Fig 3A). In total 1,375 genes with significant ≥2 FC differences between the *EZH2*-wild type and mutant group with average signal intensities above 100 in at least one group were detected. Detailed expression data is given in Supplemental Table 4. Correlation cluster analysis of patient data showed a distinct separation of expression patterns according to *EZH2*-status. Heat map data revealed that most *EZH2*-wild type MDS/MPN patients displayed higher expression levels in the majority of differently expressed genes (Fig 3B). Some known *EZH2*-targets, such as *ADRB2*

and the NOTCH- and JAK-STAT pathway genes, *MYC*, *JAK1* and *JAK2*[10] were found to have significantly different expression levels between *EZH2*-mutant and wild type patients (Fig 3C).

Transcriptome analysis of the EZH2 downregulated F-36P, MOLM-13 and OCI-M2 cell lines

For each cell line we compared transcriptome data of genes that showed significant differences between EZH2-wild type and mutant patients as shown above, to gene expression data of shEZH2-downregulated cell lines with the respective control cell line, expressing non targeting control shRNA. All cell lines showed significant differences in expression levels for *CXXC5* and *FAM113B*. For F-36P we found 237 genes which showed p-values <0.05 for one or both shEZH2 manipulated cell lines. Four of these genes (*PM20D2*, *PRAGMIN*, *RAB3IL1* and *STS-1*) showed ≥2 FC, significant differences in both downregulated F-36P lines compared to the control. In MOLM-13 we found 91 genes which showed p-values <0.05 for one or both shEZH2 manipulated cell lines. Two genes (*ETS-1* and *VAV3*) showed ≥2 FC, significant differences in both downregulated MOLM-13 lines compared to the control. OCI-M2 cell lines showed 318 genes with p-values <0.05 for one or both manipulated cell lines. Significant differences in expression for both downregulated OCI-M2 lines were seen in *VAV3*, albeit not ≥2 FC. As seen with patient data, a number of previously identified EZH2 targets[10,19] showed significant differences in shEZH2 downregulated OCI-M2 lines compared to controls. Detailed expression data is given in Supplemental Table 4. Significant differences in *EZH2* expression were only observed in cell lines (Fig 4), as EZH2 may be expressed but functionally compromised in patients carrying missense mutations. Interestingly, some genes such as *CXXC5*, *ETS-1*, *STS-1* and *VAV3* showed opposite directions of regulation in patients compared to cell models, whilst others such as *FAM113B* and *PRAGMIN*

showed regulation in the same direction. Results were confirmed by quantitative real time PCR (qRT-PCR). For cell lines and patients, log2 array and qRT-PCR data comparing gene expression according to *EZH2* status are shown in Figure 4. Cell viability of cell lines was not affected by EZH2-manipulation (Supplemental Fig 2).

Histone H3 trimethylation and protein expression

Protein blots of all control and shEZH2 cell lines confirmed downregulation of EZH2 expression, resulting in reduced H3K27 trimethylation, as a result of impaired EZH2 function. Additionally, changes in protein levels of ETS-1, STS-1 and VAV3 were detected for respective cell lines. CXXC5 and FAM113B were on the resolution limit of protein blot analysis, but showed differences <2FC in line with RNA expression data (Fig 5A). Semi-quantified cell line data is shown in optical density charts in Figure 5B. Exemplary patient data showed no EZH2 in patients #27 and 32 as a result of homozygous frameshift mutations (p.K685Rfs*12 and p.K17Sfs*3, respectively). Patients # 24 and 34 carry heterozygous frameshift (p.N263Qfs*8) and nonsense mutations (p.Y292X), respectively. Downregulation of H3K27 trimethylation was observed in EZH2-mutant patients #24, 27, 32 and 34 compared to EZH2-wild type patients D and F though overall EZH2 protein levels were lower (Fig 5C).

Discussion

NGS of *EZH2*-mutant MDS/MPN patients revealed various mutational patterns. *TET2*, *RUNX1* and *ASXL1* were most frequently mutated together with *EZH2*, which is in agreement with previous findings.[19] Interestingly, 39% of patients showed *RUNX1* mutations (Fig 1). This is in line with recent MDS studies showing a significant association of *EZH2* and *RUNX1*.[20,21] These studies also reported that *EZH2* and *SRSF2* are rarely mutated together. We found *SRSF2* mutations in only 15% of our patients. Splicing mutations in *SF3B1* and *U2AF1* did not co-occur with *EZH2* mutations or were present at very low levels (2.4%), respectively (Fig 1). Khan

et al. have demonstrated that splicing abnormalities are often responsible for loss of *EZH2* function.[19] This may explain why splicing mutations occur rarely in *EZH2* positive patients, as epistasis or synthetic lethality can cause mutual exclusion of certain gene mutations.[22] However, heterozygous mutations for multiple genes that regulate PRC2 function can cooperate in leukemic transformations through additive effects.[23]

Previous findings of *EZH2* mutations in patients with refractory anemia, suggest that *EZH2* mutations are acquired early in leukemogenesis.[11] Our analysis of clonal hierarchies confirmed that *EZH2* mutations can indeed be early events, at least in MDS/MPN cases harboring additional *SETBP1* mutations as seen in patients #24, 37 and 40. In the latter two acquisition of *NRAS* and *CBL* mutations occurred last (Figs 2A, 2B), supporting previous findings that signaling genes are often affected later on in disease progression.[24] Results showed a clonal architecture with moderate to frequent branching. Particularly patient #24 displayed a complex pattern of disease evolution (Fig 2C). The progression to homozygous and reversion to wild type states for certain loci can be explained through loss of heterozygosity and reflect the current view of vast genetic diversity in leukemic cells as opposed to the merely linear acquisition of mutations.[24,25] In support of this mono- and biallelic, as well as overlapping mutations were detected in some patients (Supplemental Fig 1). The clonal hierarchy of patient #41 showed *EZH2* mutations as secondary events, preceded by the acquisition of an *ASXL1* mutation (Fig 2D). A recent study proposed that CMML arises through accumulation of mostly age-related somatic mutations that ultimately convert a skewed myelomonocytic hematopoiesis into leukemia.[26] In line with these findings, it has been shown that early clonal dominance, particularly of *TET2* or *ASXL1*-mutant myeloid progenitors is followed by selection of clones carrying secondary lesions during myeloid differentiation.[24] Both *TET2* and *ASXL1*

have been implicated in age related clonal hematopoiesis (ARCH), thus their role as early key players in the development of CMML is not surprising. In contrast *EZH2* (and *SETBP1*) mutations have not been shown to be involved in ARCH.[26] Thus EZH2 mutations may reflect an alternative mode of MDS/MPN initiation or, in cases where ARCH genes are involved, could provide the growth advantage that is essential for leukemic transformation. In any case the total number of mutated genes seems to be a strong prognostic factor in CMML, indicating that clonal complexity might affect therapy.[27] Thus, the delineation of clonal architecture remains of prime importance to ensure therapy is directed at primary lesions.

Whole transcriptome analysis of *EZH2*-wild type and mutant MDS/MPN patients showed distinct expression patterns (Fig 3A), suggesting that the *EZH2* mutational status has a global effect on gene expression. Interestingly, *EZH2*-wild type patients show the most upregulation in differently expressed genes, when looking at cluster correlation analysis (Fig 3B). These findings underline the complex role of EZH2 as an epigenetic regulator that may silence or activate targets by pathway dependent mechanisms. Since multiple mechanisms can affect H3K27me3 (ref. 19) it is not surprising that two patients showed expression patterns distinct from their respective group (Fig 3B). Looking at known EZH2 targets[10] we found significant differences in *ADRB2*, *JAK1*, *JAK2* and *MYC* between *EZH2*-wild type and mutant patients (Fig 3C). Expression levels of genes related to known *EZH2*-targets, such as *HOXA5*, *HOXB5* and *cyclins A1, A2*, *B1*, *D2* and *F,* were also affected (Supplemental Table 4). However, it is important to note that MDS/MPN patients show vastly heterogeneous genetic lesions, which is reflected in the large number of differently expressed genes and the clinical heterogeneity of these diseases.

Therefore *EZH2*-knock-down models were established in sAML cell lines F-36P, MOLM-13 and OCI-M2. We found two genes to be regulated in patients and all cell

lines (*CXXC5, FAM113B*), and a number of genes that were regulated in patients and at least one cell line, such as *ETS-1*, *STS-1* and *VAV3* (Fig 4). The Wnt-signaling inhibitor *CXXC5* plays a role in myelopoiesis and is a known tumor suppressor candidate, located on 5q31.2, a region recurrently deleted in AML and MDS. Its downregulation has been associated with a better outcome in AML, pointing towards the potential benefits of Wnt-inhibitor treatment.[28] ETS proteins are transcription factors that can alter the expression of genes that are involved in various biological processes, such as hematopoiesis, cellular proliferation, differentiation, development and apoptosis.[29] ETS1 has long been implied to play a role in hematologic malignancies.[30,31] Molecular allelokaryotyping of T-cell prolymphocytic leukemia cells revealed recurrent microdeletions targeting *ETS1*.[32] Additionally, analyses of low-risk MDS cases found *ETS1* to have tumor suppressor function as the gene was hypermethylated and showed low expression levels, with various ETS1 targets being downregulation as a result.[33] The protein tyrosine phosphatase STS-1 has recently been shown to interact with CBL, regulating myeloid proliferation in human AML1-ETO.[34] An interaction between VAV3 and EZH2 has previously been revealed for putative homologues in *Saccharomyces cerevisiae*.[35] Studies have shown that VAV3 plays a role in lymphocyte activation and development.[36] Interestingly genetic deficiency of *VAV3* delays the progression of *BCR-ABL*-positive lymphoblastic leukemia and increases leukemic cell death making the gene a potential therapeutic target.[37]

The apparent paradox of opposite regulation of some genes between cell lines and patients may occur for a number of reasons. (a) Artificial introduction of *EZH2* dysfunction in sAML cell lines might bring about different effects, compared to *EZH2* mutations in less proliferative MDS cells, whilst still indicating potential targets. (b) *EZH2* expression was overall higher in cell lines compared to patients (Fig 5,

Supplemental Table 4). The different levels of EZH2 protein may alter the spectrum of function and modify PRC2 activity, as highlighted by the arbitrary role of *EZH2* as a tumor suppressor and oncogene. (c) The results indicate that *CXXC5*, *ETS-1*, *STS-1* and *VAV3* may not be direct targets of EZH2 but rather fall under the regulatory control of secondary pathways. For example, Viré *et al.*[38] have shown that the two epigenetic systems of histone and DNA methylation are connected. Thus DNA methylation could serve as a switch for gene expression, depending on the level of EZH2 activity.

In conclusion, we found a spectrum of concurrent and mutually exclusive mutations of *EZH2*-positive MDS/MPN patients similar to molecular patterns observed in MDS. Further studies with larger cohorts will reveal how the mutational landscape affects disease phenotype and response to therapy. *EZH2* mutations seem to be early events in leukemogenesis, particularly when ARCH genes are not involved. Comparison of patient data with *EZH2*-downregulated sAML cell lines revealed *CXXC5, FAM113B, ETS1, STS-1 and VAV3* expression to be associated with *EZH2* mutational status, amongst other genes with lower expression levels (*PM20D2*, *PRAGMIN*, *RAB3IL1*). Whether these targets are up- or downregulated may depend on initial EZH2 expression and cellular context. At this point the exact pathway, through which EZH2 may exert regulatory control, remains unknown. Nonetheless, to our knowledge this is the first study that has uncovered these genes as EZH2 targets in myeloid malignancies. Further study will reveal their precise role in myeloid leukemogenesis.

Acknowledgements

The excellent technical assistance of Mrs. Anja Waldau is gratefully acknowledged. The study was supported by the Deutsche José Carreras Leukämie-Stiftung e.V. (DJCLS R15/20) and the Interdisziplinäres Zentrum für Klinische Forschung (IZKF, Jena, Germany).

Conflict of Interest

The authors declare no conflict of interest.

References

1 Shtivelman E, Lifshitz B, Gale RP, Canaani E. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. *Nature* 1985; 315(6020):550-4.

2 James C, Ugo V, Le Couédic JP, Staerk J, Delhommeau F, Lacout C et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature* 2005; 434(7037):1144-8.

3 Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A *et al*. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* 2009; 114:937-51.

4 Druker BJ, Tamura S, Buchdunger E, Ohno S, Segal GM, Fanning S et al. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nature Medicine* 1996; 2(5):561-6.

5 Zoi K, Cross NC. Molecular pathogenesis of atypical CML, CMML and MDS/MPN-unclassifiable. *International Journal of Hematology* 2015; 101(3):229-42.

6 Shih AH, Abdel-Wahab O, Patel JP, Levine RL. The role of mutations in epigenetic regulators in myeloid malignancies. *Nature Reviews Cancer* 2012; 12(9):599-612.

418 7 Kon A, Shih LY, Minamino M, Sanada M, Shiraishi Y, Nagata Y et al.
419 Recurrent mutations in multiple components of the cohesin complex in myeloid
420 neoplasms. *Nature Genetics* 2013; 45(10):1232-7.
421 8 Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R et al.
422 Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*
423 2011; 478(7367):64-9.
424 9 Fathi AT, Abdel-Wahab O. Mutations in epigenetic modifiers in myeloid
425 malignancies and the prospect of novel epigenetic-targeted therapy. *Advances*
426 *in Hematology* 2012; 2012:469592.
427 10 Kim KH, Roberts CW. Targeting EZH2 in cancer. *Nature Medicine* 2016;
428 22(2):128-34.
429 11 Ernst T, Chase AJ, Score J, Hidalgo-Curtis CE, Bryant C, Jones AV et al.
430 Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid
431 disorders. *Nature Genetics* 2010; 42(8):722-6.
432 12 Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid
433 guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*
434 1987; 162(1):156-9.
435 13 Schäfer V, Ernst J, Rinke J, Ziermann J, Winkelmann N, Hochhaus A et al.
436 EZH2 mutations and promoter hypermethylation in childhood acute
437 lymphoblastic leukemia. *Journal of Cancer Research and Clinical Oncology*
438 2016; 142(7):1641-50.
439 14 Rinke J, Schäfer V, Schmidt M, Ziermann J, Kohlmann A, Hochhaus A et al.
440 Genotyping of 25 leukemia-associated genes in a single work flow by next-
441 generation sequencing technology with low amounts of input template DNA.
442 *Clinical Chemistry* 2013; 59(8):1238-50.

15 Schmidt M, Rinke J, Schäfer V, Schnittger S, Kohlmann A, Obstfelder E et al. Molecular-defined clonal evolution in patients with chronic myeloid leukemia independent of the *BCR-ABL* status. *Leukemia* 2014; 28(12): 2292-9.

16 Sauer H, Ruhe C, Müller JP, Schmelter M, D'Souza R, Wartenberg M. Reactive oxygen species and upregulation of NADPH oxidases in mechanotransduction of embryonic stem cells. *Methods in Molecular Biology* 2008; 477:397-418.

17 Rittirsch D, Schoenborn V, Lindig S, Wanner E, Sprengel K, Günkel S et al. An Integrated Clinico-transcriptomic Approach Identifies a Central Role of the Heme Degradation Pathway for Septic Complications after Trauma. *Annals of Surgery* 2016; 264(6):1125-1134.

18 Zirm E, Spies-Weisshart B, Heidel F, Schnetzke U, Böhmer FD, Hochhaus A et al. Ponatinib may overcome resistance of FLT3-ITD harbouring additional point mutations, notably the previously refractory F691I mutation. *British Journal of Haematology* 2012; 157(4):483-92.

19 Khan SN, Jankowska AM, Mahfouz R, Dunbar AJ, Sugimoto Y, Hosono N et al. Multiple mechanisms deregulate EZH2 and histone H3 lysine 27 epigenetic changes in myeloid malignancies. *Leukemia* 2013; 27(6):1301-9.

20 Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* 2013; 122(22):3616-27

21 Bejar R, Stevenson KE, Caughey BA, Abdel-Wahab O, Steensma DP, Galili N et al. Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes. *Journal of Clinical Oncology* 2012; 30(27):3376-82.

22 Murati A, Brecqueville M, Devillier R, Mozziconacci MJ, Gelsi-Boyer V, Birnbaum D. Myeloid malignancies: mutations, models and management. *BMC Cancer* 2012; 12:304.

23 Abdel-Wahab O, Adli M, LaFave LM, Gao J, Hricik T, Shih AH et al. ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. *Cancer Cell* 2012; 22(2):180-93.

24 Itzykson R, Kosmider O, Renneville A, Morabito M, Preudhomme C, Berthon C et al. Clonal architecture of chronic myelomonocytic leukemias. *Blood* 2013; 121(12):2186-98.

25 Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature* 2011; 469(7330):356-61.

26 Mason CC, Khorashad JS, Tantravahi SK, Kelley TW, Zabriskie MS, Yan D et al. Age-related mutations and chronic myelomonocytic leukemia. *Leukemia* 2016; 30(4):906-13.

27 Jankowska AM, Makishima H, Tiu RV, Szpurka H, Huang Y, Traina F et al. Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX, EZH2, and DNMT3A. *Blood* 2011; 118(14):3932-41.

28 Kühnl A, Valk PJ, Sanders MA, Ivey A, Hills RK, Mills KI et al. Downregulation of the Wnt inhibitor CXXC5 predicts a better prognosis in acute myeloid leukemia. *Blood* 2015; 125(19):2985-94.

29 Seth A, Watson DK. ETS transcription factors and their emerging roles in human cancer. *European Journal of Cancer* 2005; 41(16):2462-78.

30 Collyn d'Hooghe M, Galiègue-Zouitina S, Szymiczek D, Lantoine D, Quief S, Loucheux-Lefebvre MH et al. Quantitative and qualitative variation of ETS-1 transcripts in hematologic malignancies. *Leukemia* 1993; 7(11):1777-85.

31 Kerckaert JP, Duterque-Coquillaud M, Collyn-d'Hooghe M, Morel P, Majérus MA, Laï JL et al. Polymorphism of the proto-oncogene ETS-1 in hematological malignancies. *Leukemia* 1990; 4(1):16-9.

32 Nowak D, Le Toriellec E, Stern MH, Kawamata N, Akagi T, Dyer MJ et al. Molecular allelokaryotyping of T-cell prolymphocytic leukemia cells with high density single nucleotide polymorphism arrays identifies novel common genomic lesions and acquired uniparental disomy. *Haematologica* 2009; 94(4):518-27.

33 del Rey M, O'Hagan K, Dellett M, Aibar S, Colyer HA, Alonso ME et al. Genome-wide profiling of methylation identifies novel targets with aberrant hypermethylation and reduced expression in low-risk myelodysplastic syndromes. *Leukemia* 2013; 27(3):610-8.

34 Goyama S, Schibler J, Gasilina A, Shrestha M, Lin S, Link KA et al. UBASH3B/Sts-1-CBL axis regulates myeloid proliferation in human preleukemia induced by AML1-ETO. *Leukemia* 2016; 30(3):728-39.

35 Hobert O, Jallal B, Ullrich A. Interaction of Vav with ENX-1, a putative transcriptional regulator of homeobox gene expression. *Molecular Cell Biology 1996*; 16(6):3066-73.

36 Fujikawa K, Miletic AV, Alt FW, Faccio R, Brown T, Hoog J et al. Vav1/2/3-null mice define an essential role for Vav family proteins in lymphocyte development and activation but a differential requirement in MAPK signaling in T and B cells. *The Journal of Experimental Medicine* 2003; 198(10):1595-608.

517 37 Chang KH, Sanchez-Aguilera A, Shen S, Sengupta A, Madhu MN, Ficker AM
518 et al. Vav3 collaborates with p190-BCR-ABL in lymphoid progenitor
519 leukemogenesis, proliferation, and survival. *Blood* 2012; 120(4):800-11.
520 38 Viré E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C et al. The
521 Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 2006;
522 439(7078):871-4. Erratum in: *Nature* 2007; 446(7137):824.

523

524 **Figure legends**

525 **Figure 1** *Cooperation of distinct and individual mutations detected by NGS in a*
526 *cohort of 41 EZH2-positive MDS/MPN patients.* A total of 190 individual mutations
527 were detected. Patients carried between 1 and 9 mutations of up to 7 different genes.
528 * frameshift, nonsense and/or splice donor/acceptor mutation(s); ^{0} two detected
529 mutations; ˜ three detected mutations. Detailed information is given in Supplemental
530 Table 4. Patient information such as karyotype and diagnostic category can be
531 viewed in Supplemental Table 1.

532 **Figure 2** *Clonal branching and mutational frequencies detected in patients #24, 37,*
533 *40 and 41.* Putative evolutionary trees generated from targeted analysis of CFU-GM
534 derived colonies from CD34+ cells, accounting for the zygosity status at each locus
535 (italicized: heterozygous, bold & underlined: homozygous). Only mutated genes are
536 indicated in each subclone. (A) Patient #40: Analysis of 80 individual colonies for
537 mutations previously detected by NGS, indicate initial *EZH2* mutation followed by
538 *SETBP1* and *CBL* mutations. The dominant clone (78%) is homozygous for both
539 *EZH2* and *CBL* and heterozygous for *SETBP1*, branching of into clones that carry
540 either wild type (6%) or homozygous (4%) *SETBP1* alleles. (B) Patient #37: Analysis
541 of 49 colonies showing initial homozygous *EZH2* mutation followed by *SETBP1* and
542 *NRAS* mutations. The dominant clone (49%) is homozygous for *EZH2* and

heterozygous for *SETBP1* and *NRAS,* branching of into clones that are either homozygous for *SETBP1* (10%) or *NRAS* (20%) and finally, a clone homozygous for all said mutations. Low-level *KRAS*, *RUNX1*, *STAG2* and *ZRSR2* mutations previously identified by NGS were not detected in colonies. (C) Patient #24: Analysis of 41 colonies indicates an initial heterozygous *EZH2* and *SETBP1* positive clone, followed by the acquisition of heterozygous *CBL* and *SRSF2* mutations and finally *STAG2*, which is homozygous in the dominant clone (54%). The complex pattern of mutations shows allele losses at all loci, with reversion to homozygous and/or wild-type state. (D) Patient #41: Analysis of 46 colonies indicates the early acquisition of an *ASXL1* mutation, followed by *EZH2* mutations p.T467Hfs*16 and p.Y579X, respectively, with frequent allele loss. The dominant clone (35%) is heterozygous for all mutations.

Figure 3 *Whole transcriptome analysis of EZH2-wild type and mutant MDS/MPN patients (n = 12, each)*. (A) Cloud diagram of average signal intensities detected for 34.694 genes in *EZH2*-wild type and mutant patients shows distinct differences in gene expression for genes with signal intensities above ~5000. (B) Correlation Cluster analysis (hierarchical clustering; nesting with average linking method; Pearson correlation using 1-r distance measure) of patient data and heat map showing ≥2 FC differently expressed genes and background corrected average signal intensities >200. Clustered columns, sorted by average signal of wild type group. Patients #34 and A did not match the gene expression pattern of their respective group. The majority of genes show upregulated expression levels in the *EZH2*-wild type group. (C) Microarray average signal intensities and Log2 data of known EZH2 targets *ADRB2*, *JAK1*, *JAK2* and *MYC* for *EZH2*-wild type and mutant patients; (*) p≤0.05; (**) p≤0.005. Detailed information is given in Supplemental Table 4.

Figure 4 *Microarray and qRT-PCR expression data of genes regulated in both patients and cell lines in relation to EZH2-aberration.* Genes showing differences between patients depending on *EZH2* status in comparison to cell line data; (*) p≤0.05; (**) p≤0.005; (***) p≤0.0005; Mean and standard deviation of microarray average signal intensities and Log2 expression ratios of (i) Microarray and (ii) qRT-PCR data (A) Expression data of genes with significant ≥2FC differences between *EZH2*-wild type and mutant patients (n = 12, each); (B) Expression data of genes with significant differences between control and shRNA manipulated F-36P cell lines (n = 4, each); (C) Expression data of genes with significant differences between control and shRNA manipulated MOLM-13 cell lines (n = 4, each); (D) Expression data of genes with significant differences between control and shRNA manipulated OCI-M2 cell lines (n = 4, each). Differences between microarray and qRT-PCR data are due to inter-assay variations. Detailed information is given in Supplemental Table 4.

Figure 5 *Protein blot analysis of F-36P, MOLM-13, OCI-M2 and patient data.* Cell lines expressing *EZH2*-targeting shRNA genes 1 or 2 (sh 1, sh 2) or a nontargeting control shRNA and patient cell samples were lysed and processed via SDS-PAGE and immunoblotted. Histone H3 and GAPDH were used as loading controls. (A) Protein blots show a reduced EZH2 expression in manipulated cells as well as a reduced H3K27 trimethylation. ETS-1, STS-1 and VAV3 show upregulation in manipulated cell lines. Marginal downregulation was observed for CXXC5 and FAM113B for all cell lines. (B) Cell line protein blots, semi-quantified using H3 as a reference for H3K27me3 and GAPDH as a reference for all other proteins (C) Protein blots of *EZH2*-mutant patients #24 (p.N263Qfs*8), 27 (p.K685Rfs*12), 32 (p.K17Sfs*3) and 34 (p.Y292X) show a reduced EZH2 expression, particularly for

homozygous patients #27 and 32, and a reduced H3K27me3 expression, compared to *EZH2*-wild type patients D and F.

Figure 1

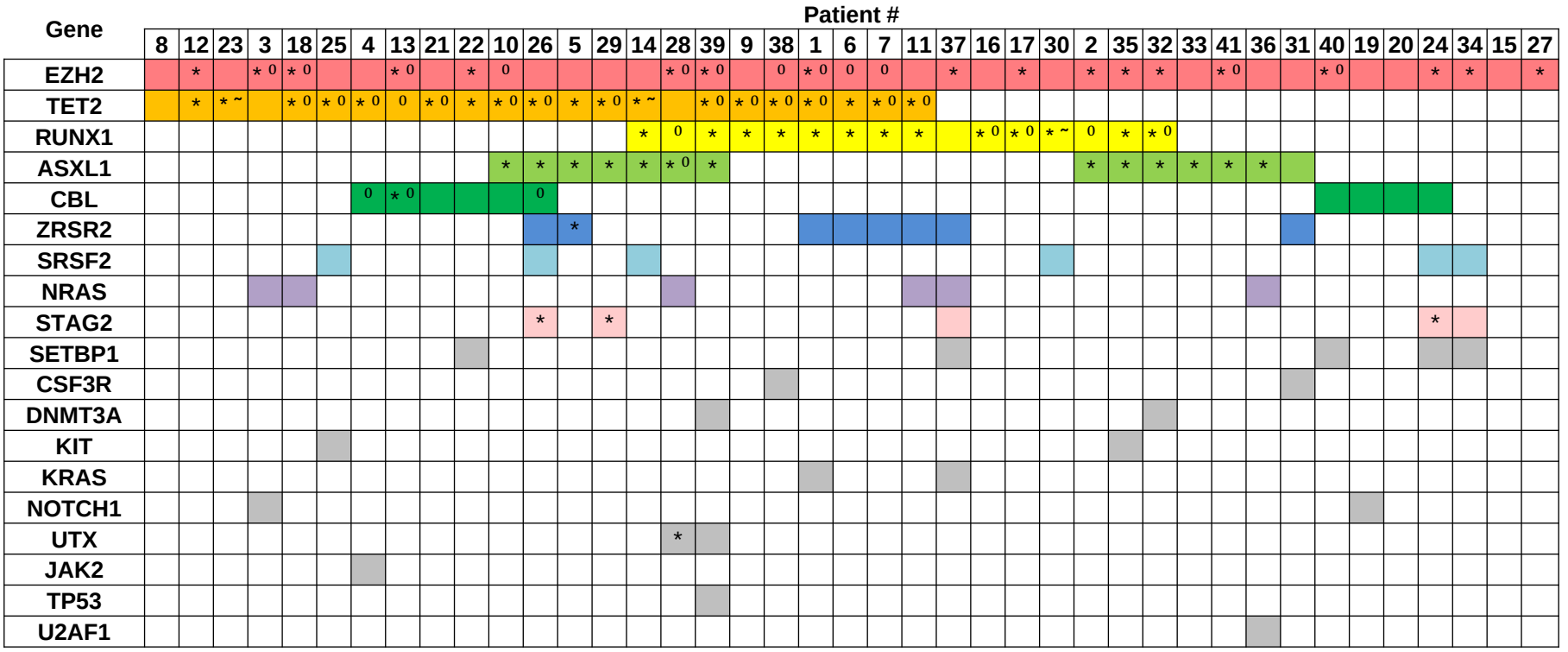
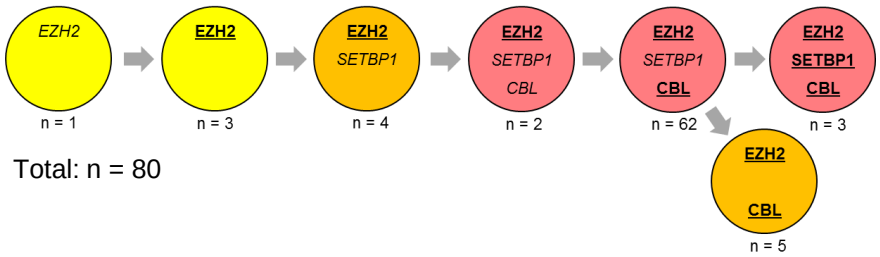
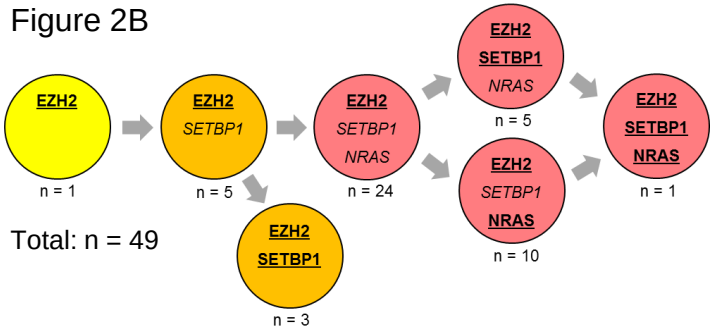

Gene
Patient #
8 12 23 3 18 25 4 13 21 22 10 26 5 29 14 28 39 9 38 1 6 7 11 37 16 17 30 2 35 32 33 41 36 31 40 19 20 24 34 15 27
EZH2
TET2
RUNX1
ASXL1
CBL
ZRSR2
SRSF2
NRAS
STAG2
SETBP1
CSF3R
DNMT3A
KIT
KRAS
NOTCH1
UTX
JAK2
TP53
U2AF1

Figure 2A



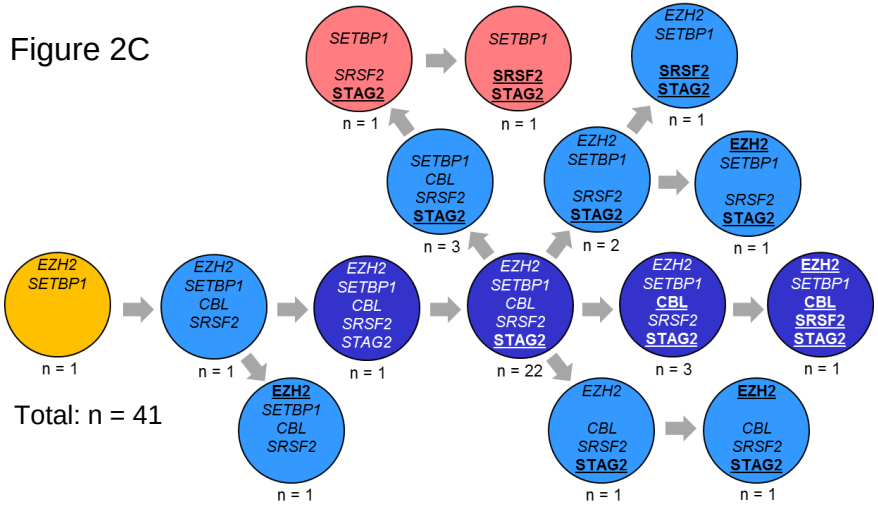
Total: n = 80

Figure 2B



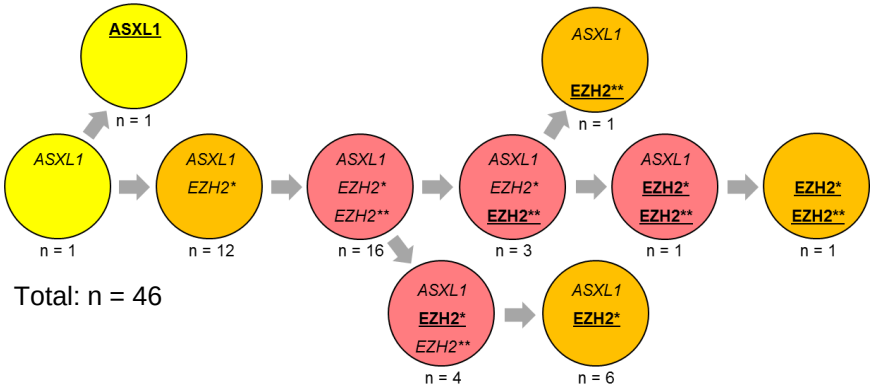
Total: n = 49

Figure 2C



Total: n = 41

Figure 2D



Total: n = 46

Legend

- 1 Mutation/colony
- 2 Mutations/colony
- 3 Mutations/colony
- 4 Mutations/colony
- 5 Mutations/colony
- *K* Heterozygous mutation
- **B** Homozygous mutation

Figure 3A

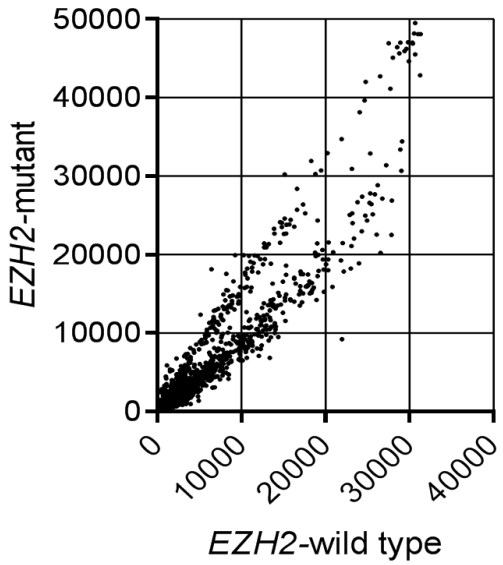


Figure 3B

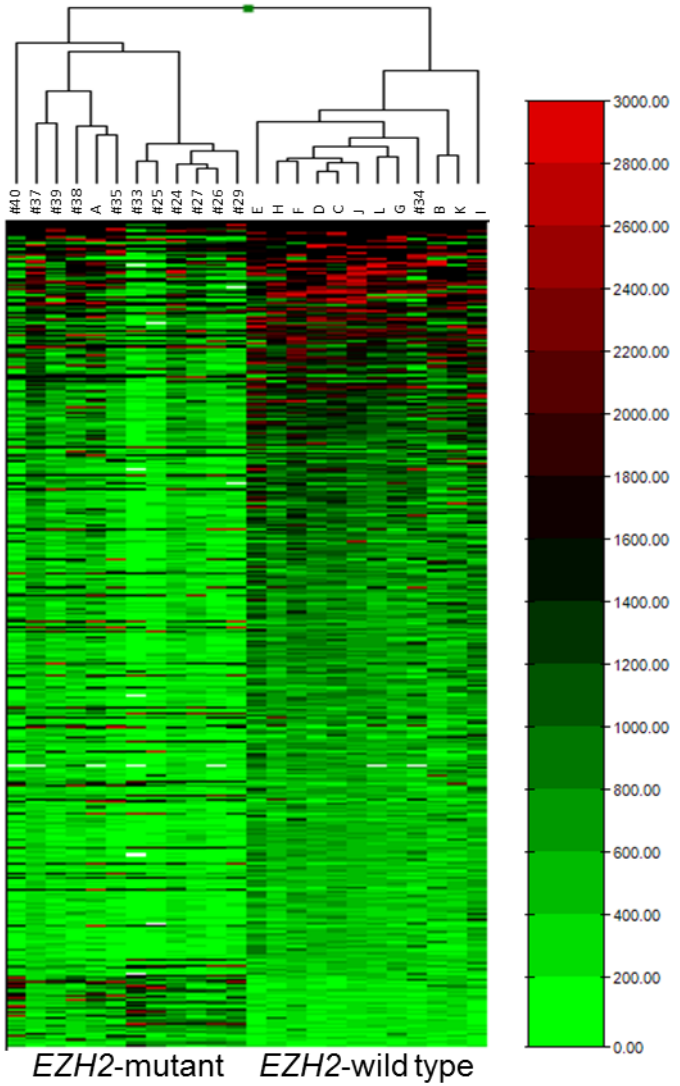


Figure 3C

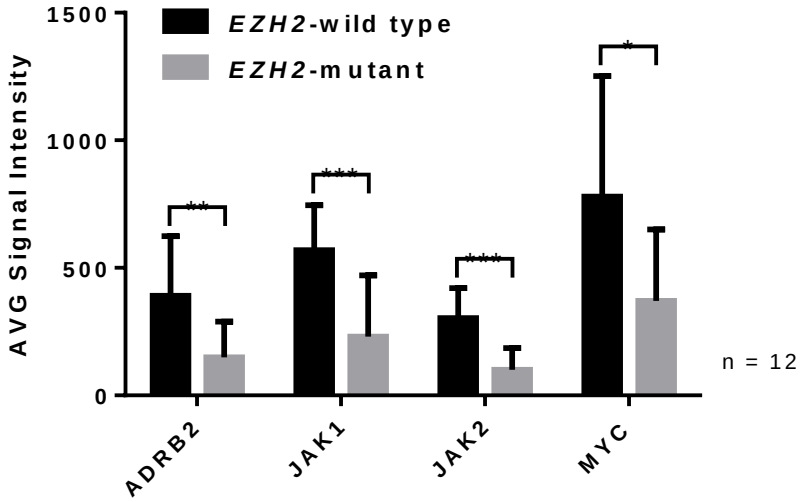
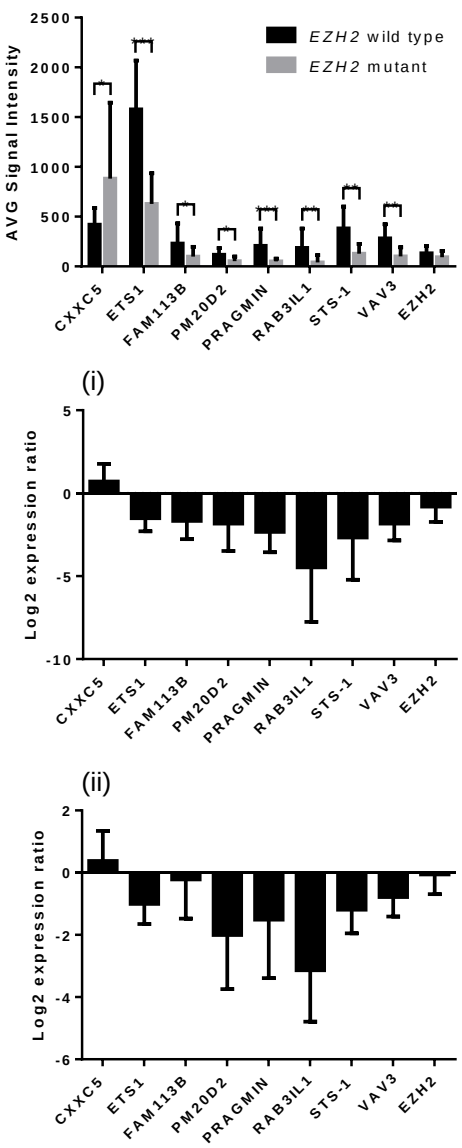
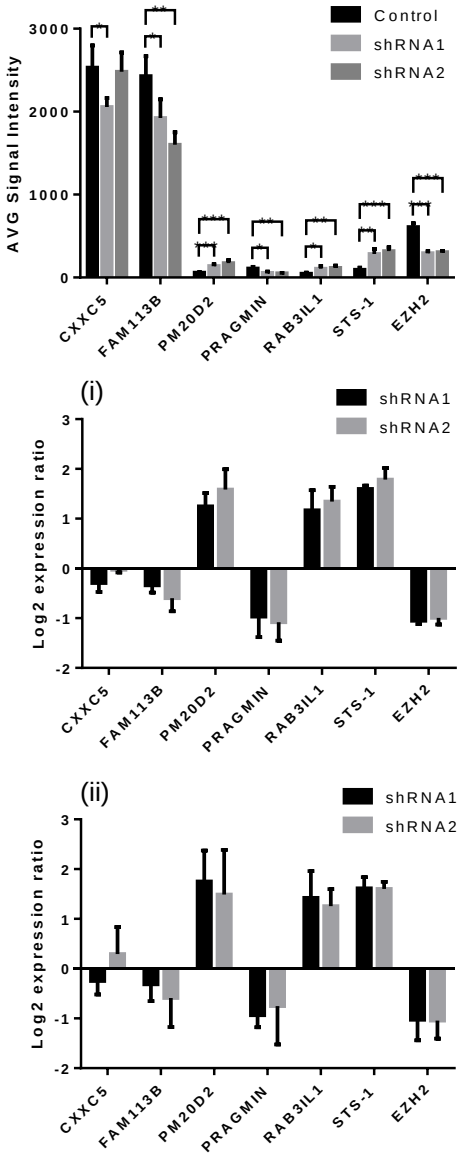


Figure 4A



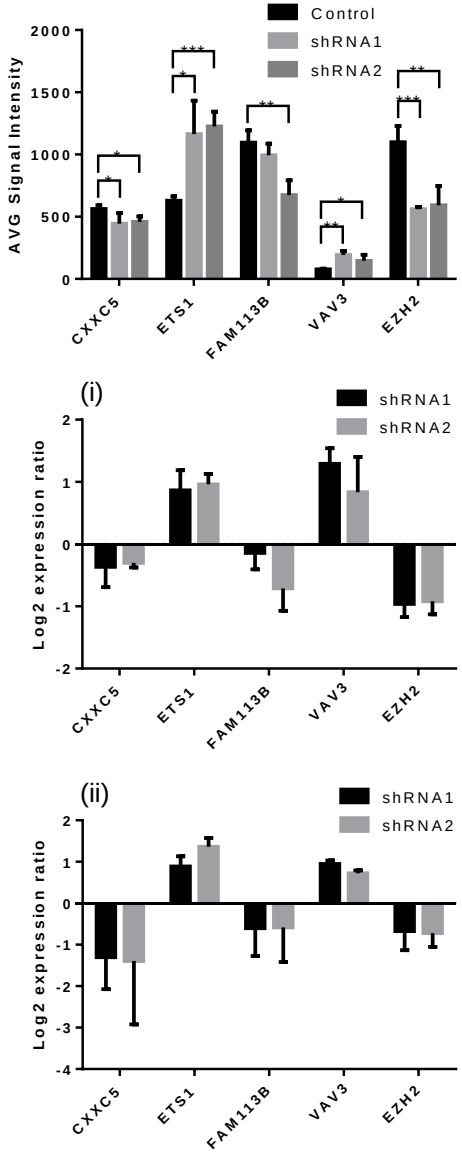
Patients

Figure 4B



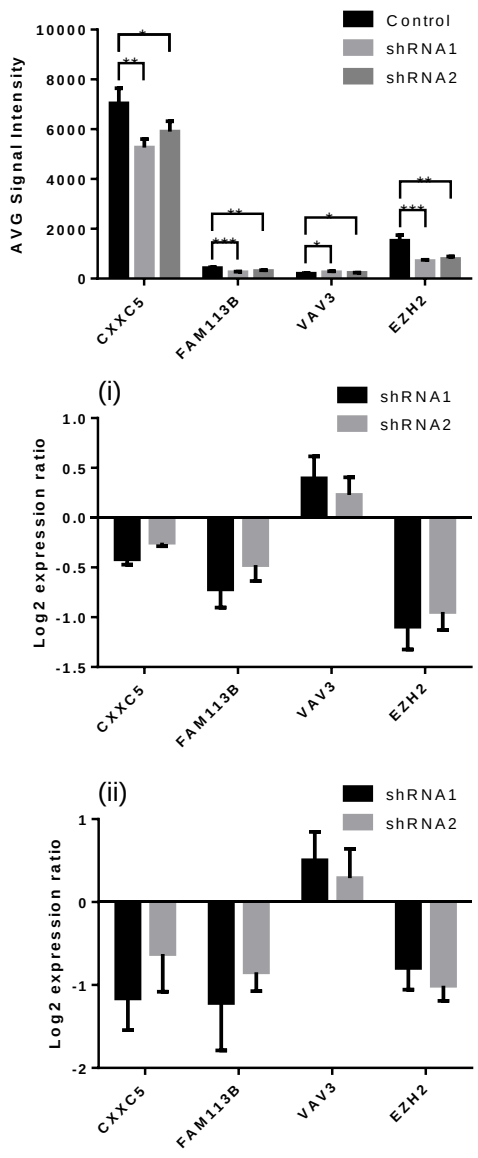
F-36P

Figure 4C



MOLM-13

Figure 4D



OCI-M2

Figure 5A

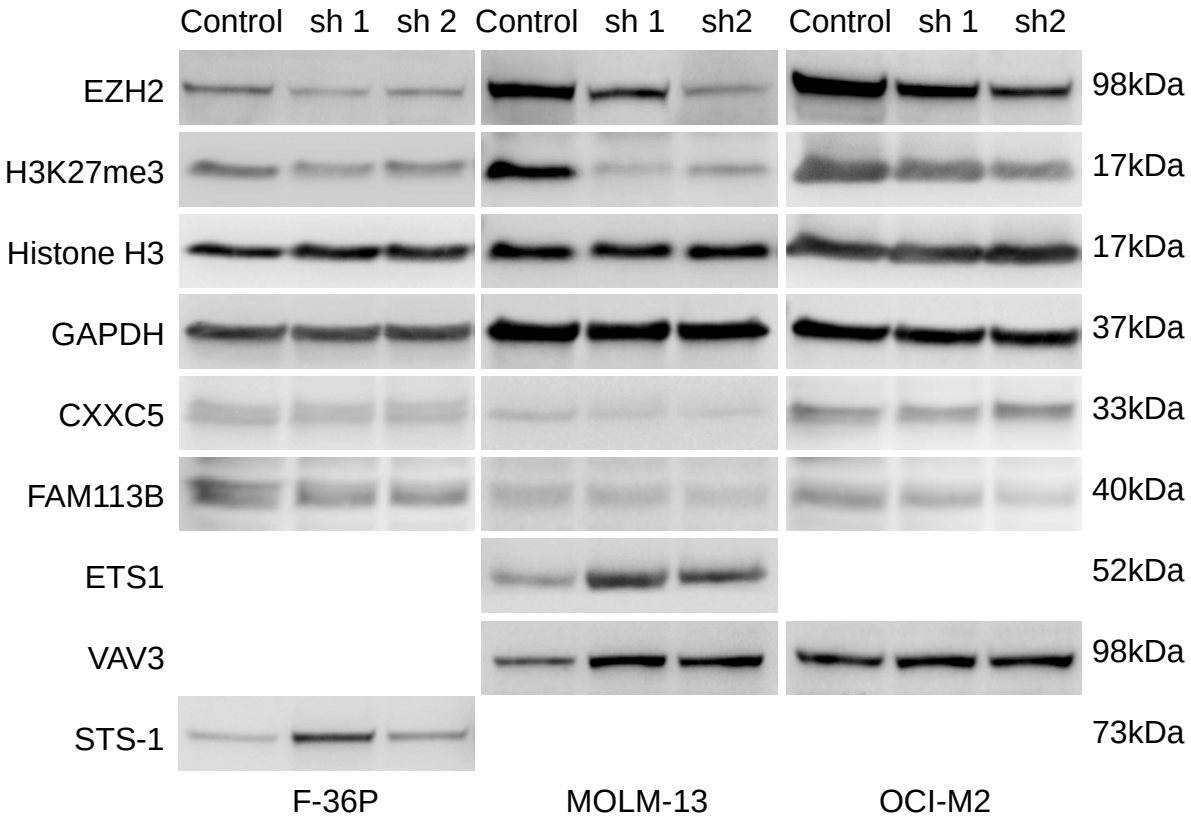


Figure 5B

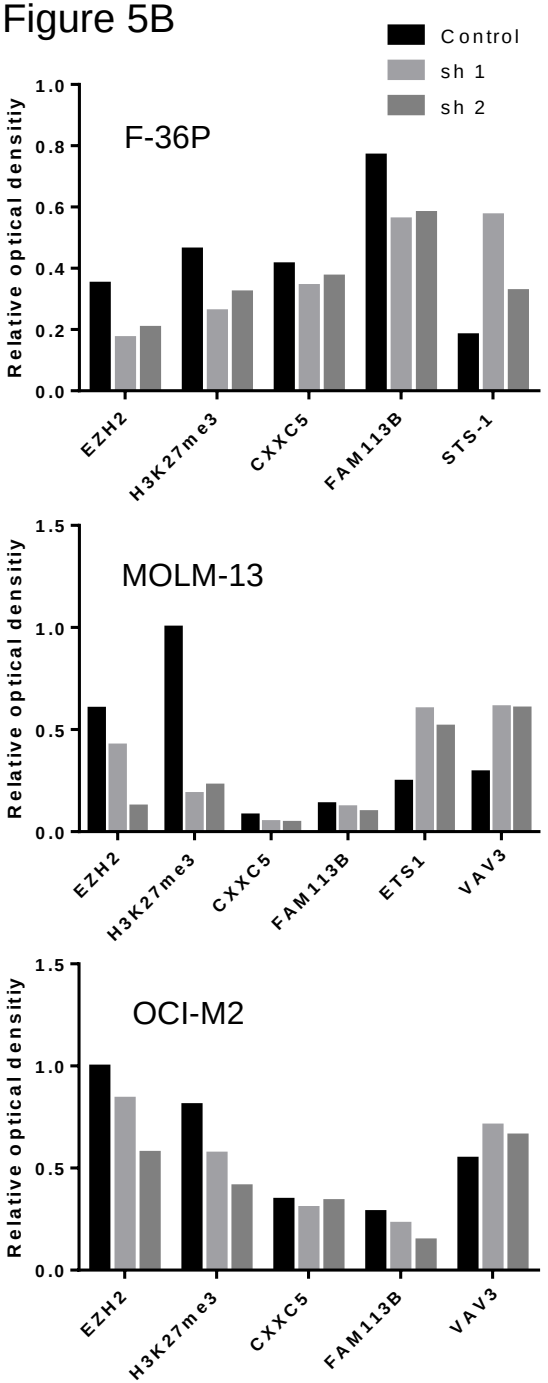


Figure 5C