Unique single-nucleotide variations, insertions/deletions and copy number variations in myelodysplastic syndrome during treatment resistance and progression revealed by a single-cell DNA sequencing platform

Paul Lee
The University of Hong Kong
Rita Lok Hay Yim
The University of Hong Kong
Sin-Hang Fung
The Chinese University of Hong Kong
Ka-Kai Miu
The Chinese University of Hong Kong
Zhangting Wang
The Chinese University of Hong Kong
Ka Chun Wu
The University of Hong Kong
Lester Au
The University of Hong Kong
Garret Leung
University of Hong Kong
Victor Lee
The University of Hong Kong
Harinder Gill (gillhsh@hku.hk)
University of Hong Kong https://orcid.org/0000-0002-9551-4893

Article

Keywords:

Posted Date: February 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1366306/v1

TITLE

Unique single-nucleotide variations, insertions/deletions and copy number variations in myelodysplastic syndrome during treatment resistance and progression revealed by a single-cell DNA sequencing platform

AUTHORS AND AFFILIATIONS

Paul Lee [1#], Rita Yim [1#], Sin-Hang Fung [2#], Kai-Kei Miu [2#], Zhangting Wang[2], Ka-Chun Wu[3 4], Lester Au[1], Garret M.K. Leung[1], Victor Ho-Fun Lee[4], Harinder Gill [1*]

1: Department of Medicine, School of Clinical Medicine, LKS Faculty of Medicine, the University of Hong Kong, Hong Kong SAR, China.
2: School of Biomedical Sciences, Faculty of Medicine, the Chinese University of Hong Kong, Hong Kong SAR, China.
3: Laboratory for Synthetic Chemistry and Chemical Biology Limited, Hong Kong SAR, China.
4: Department of Clinical Oncology, School of Clinical Medicine, LKS Faculty of Medicine, the University of Hong Kong, Hong Kong SAR, China

*Correspondence to: Harinder Gill, MD, FRCP, FRCPath, 4/F Professorial Block, Department of Medicine, Queen Mary Hospital, 102 Pokfulam Road, Pokfulam, Hong Kong, Hong Kong SAR, China; E-mail: gillhsh@hku.hk ; Tel.: +852 2255 4542; Fax.: +852 2816 2863
# equal contribution

MANUSCRIPT INFORMATION

Abstract word count: 195

Word Count: 3197

Figures: 6

Supplementary files: 6

Reference count: 31

ABSTRACT

Myelodysplastic syndrome (MDS) is a clonal myeloid neoplasm characterized by ineffective hematopoiesis, cytopenia, dysplasia and clonal instability leading to leukemic transformation. Hypomethylating agents are the mainstay of treatment in higher-risk MDS. However, treatment resistance and disease transformation into acute myeloid leukemia (AML) is observed in the majority patients and portend a dismal outcome. The residual cell clones resistant to therapy or cell clones acquiring new genetic aberrations are two of the key events responsible for drug resistance. Bulk tumor sequencing often fail to detect these rare subclones that confer resistance to therapy. In this study, we employed a single-cell DNA (sc-DNA) sequencing approach to study the clonal heterogeneity and clonal evolution in two MDS patients refractory to HMA. In both patients, different single nucleotide variations (SNVs) or insertions and deletions (INDELs) were detected concordant with bulk tumor sequencing. Rare cell clones with mutations undetectable on bulk tumor sequencing, were detected on single-cell DNA sequencing. In addition to SNVs and short INDELs, this study also revealed the presence clonal copy number loss of *DNMT3A*, *TET2*, and *GAT2* as standalone events or in associated with the small SNVs or INDELs detected during HMA resistance and disease progression.

INTRODUCTION

Myelodysplastic syndrome (MDS) is a myeloid neoplasm associated with complex clonal architecture and is characterized by ineffective hematopoiesis, cytopenia, dysplasia and clonal instability [1, 2]. Common chromosomal aberrations include loss of chromosome 5q and 7 and gains of chromosomes 8, 19 and 21. Mutations of *SF3B1*, *TET2*, *ASXL1*, *SRSF2* and *DNMT3A* are reported in more than 10% of all MDS [3, 4]. These genetic lesions contribute the clinical presentations, risk of progression to acute myeloid leukemia (AML), treatment responses and survivals [4-7]. Prognosis is conventionally determined by International Prognostic Scoring System-Revised (IPSS-R) and more recently personalized prognostic models incorporating somatic mutations [4, 6-8]. Resistance to the hypomethylating agents (HMAs) azacitidine (AZA) and decitabine (DEC) in higher-risked MDS (HR-MDS) represent a major unmet need [3]. Resistance to HMA portends a poor prognostic with the median overall survival (OS) of only 4.3 months in HR-MDS [9].

Somatic mutations and epigenetic alterations contribute to HMA resistance and clonal evolution in response to treatment remains a challenge in the management of MDS [10-12]. The use of next-generation sequencing (NGS) may help predict response to HMA and guide the clinical use of novel targeted therapy such as *FLT3*, BCL2, *IDH1 and IDH2* inhibitors [13, 14]. While bulk tumor sequencing may help detect somatic mutations that predict treatment outcomes, the clonal heterogeneity characteristic of MDS remains a challenge for the detection of somatic mutations in small subclones

that may confer treatment resistance. In small studies, single-cell RNA (sc-RNA) sequencing of CD34+ cell have identified sub-populations with distinct gene expression profiles and demonstrated myeloid-biased hematopoiesis in patients with MDS [15, 16]. This single cell-based approach increases the sensitivity and allows sub-clonal analysis. Nevertheless, CNV detection at RNA level is still challenging as bioinformatic inference of CNV lacks accuracy.

To study clonal evolution and changes in clonal architecture during resistance to HMA at single-cell level, we performed sc-DNA sequencing in two patients with MDS who acquired resistance to HMA. In addition to studying SNV, INDEL and CNV serially at single-cell levels, we also evaluated the sensitivity of single-cell sequencing technology in detecting rare sub-population of cells harboring pathogenic mutations.

METHODS

Patients

Bone marrow samples of two patients treated failing treatment with HMA were collected serially. Mononuclear cells were isolated from by density separation prior to sequencing. This study was approved by the Institutional Review Board (IRB) of the University of Hong Kong and Hong Kong West Cluster and written informed consents were obtained.

Targeted next-generation sequencing (NGS) of the bulk tumor sample

DNA was extracted from bone marrow samples serially using DNA blood mini extraction kit. NGS was performed serially using 69-gene customized myeloid panel as previously described [17, 18]. Enrichment of 69 genes was performed and they comprised *ABL1, ANKRD26, ACD, ATRX, BRAF, BCOR, BCORL1, ASXL1, CALR, CBL, CBLB, CDKN2A, CEBPA, CBLC, CREBBP, CSF3R, CUX1, DNMT3A, DDX41, ETV6, EZH2, FLT3, FBXW7, GATA1, GATA2, GNAS, GNB1, HRAS, IDH1, IDH2, JAK2, JAK3, IKZF1, KIT, KRAS, KMT2A, KDM6A, KMT2D, KMT2B, MPL, MYD88, NF1, NOTCH1, NPM1, NRAS, PDGFRA, PTEN, PTPN11, RAD21, PPM1D, RUNX1, ROBO1, ROBO2, SMC1A, SMC3, SETDB1, SF3B1, SETBP1, SETD2, SRSF2, SRP72, TERT, STAG2, TET2, TP53, U2AF1, WT1, ZRSR2* and *PHF6*. The enriched libraries were sequenced pair-ended on the Illumina MiSeq System (Illumina, San Diego, California, USA) followed by in-house analyses involving Trimmomatic and BWA [19, 20]. GATK and VarScan2 were used for variant calling and detection of *FLT3-ITD* was performed using PINDEL [21-23]. The resulting variants were annotated by ANNOVAR and SnpEff followed by manual evaluation [24, 25].

Targeted single-cell DNA (sc-DNA) sequencing

Single-cell DNA sequencing was performed using a targeted AML panel detecting exonic mutations according to the manufacturer's protocol (Mission Bio, USA). Cryopreserved cells were thawed and

counted before loading 35μL of cell at a concentration of 3500 cells/μL onto the Tapestri microfluidic cartridge. Cells were emulsified with lysis buffer and incubated at 50°C for 1 hour followed by thermal inactivation of the protease. The emulsion containing the lysates from protease-treated single cells was then microfluidically combined with targeted gene-specific primers, PCR reagents, and cell-identifying molecular barcodes beads using the same cartridge. The PCR primers targets 20 genes including *ASXL1, GATA2, KIT, PTPN11, TET2, DNMT3A, IDH1, KRAS, RUNX1, TP53, EZH2, IDH2, NPM1, SF3B1, U2AF1, FLT3, JAK2, NRAS, SRSF2* and *WT1*. Upon cell barcoding, the emulsion was amplified to incorporate the barcode identifiers into amplified DNA from the targeted genomic loci. The emulsions were then broken and the aqueous fraction was purified. Further downstream purification was performed using magnetic beads. Sample indexing PCR was performed and Illumina adaptor sequences underwent 10 additional cycles of PCR. The final libraries were purified and sequenced on Illumina NovaSeq with V4 150bp paired-end chemistry.

Sequencing data were processed using Mission Bio's Tapestri Pipeline V2.0. This pipeline comprised the following key steps: (1) adapter-trimming using Cutadapt, (2) reference genome alignment to hg19, (3) cellular barcode demultiplexing, (4) cell-based genotype calling using GATK/Haplotypecaller. Detection of *FLT3*-ITD was done using additional scripts provided by Mission Bio.

Analyzed data in h5 format were further analyzed using Mission Bio's Tapestri Insights using the local computer. SNV and short INDELs were filtered and analyzed using Mission Bio's Tapestri Insights software while CNV were analyzed using Mission Bio's Mosaic python package. In Tapestri Insights, high-quality cells and variants were filtered using the following criteria: (1) genotype quality score $\geqslant$ 30, (2) $\geqslant$10 reads per cell per amplicon, (3) variant allele frequency (VAF) for mutant genotype $\geqslant$ 15%, (4) variant genotyped in $\geqslant$ 50% of cells, and (6) $\geqslant$ 1% mutant cells detected. The variant significance was predicted using COSMIC, ClinVar, gnomad, and DANN prediction tools and the variant pathogenicity was predicted using Varsome [26-31]. Variants clustering analysis was performed in a pooled manner using all time points. SNVs or INDELs with clinical implications confirmed from databases (ClinVar and dbSNP) and/or verified from previous bulk-tumor NGS were selected to assist identification of pathogenic cell clones.

For CNV detection, Clonal CNV analysis was performed following Mission Bio's mosaic Version 1.5 tertiary pipeline in python. Two analytic approaches were employed to complete this analysis with both involving defining a diploid clone as baseline according to selected SNV and INDEL genotype. The first approach defined the diploid clone using pathogenic genotypes determined from the previous step detecting CNV associated with pathogenic SNV or INDEL. Aiming to detect CNV

independent from any pathogenic SNV or INDEL, the second analytic approach defined the diploid control clone by selecting SNV or INDEL predicted to be: (1) intronic, (2) functionally benign and/or (3) with stable VAF across most time points in $\geqslant$ 5% of cells carrying variants.

Data sharing

All high-throughput sequencing data supporting the findings of this study were deposited in the Sequence Read Archive (SRA) with the BioProject number PRJNA748569. The sc-DNA raw reads were deposited to SRA under the accession codes SRR15209068, SRR15209069, SRR15209070, SRR15209071, SRR15209072, SRR15209073, SRR15209076, SR15209077 and SRR15209078. The bulk tumor sequencing raw data were deposited under the accession code SRR15209074, SRR15209075, SRR15209079 and SRR15209080.

RESULTS

Mutation dynamics on bulk tumor sequencing

Two patients with MDS were first studied by bulk sequencing of the diagnostic bone marrow and the serial reassessment bone marrow samples. Patient 1 was a 53-years-old man diagnosed with MDS with excess blasts-2 (MDS-EB-2). The blast percentage at diagnosis was 12% and karyotype was normal. Bulk sequencing using a 69-gene custom panel detected 37 variants at diagnosis with the major pathogenic mutation being *FLT3-ITD* with a VAF of 50% (supplemental file 1). This patient was treated with azacitidine (100mg/m^2/day on days 1-7 per cycle) every 28 days. After 14 cycles of azacitidine, Patient 1 displayed a different mutation signature characterized by a drop in the *FLT3*-ITD VAF to 15.18% and a change in list of variants detected. Among the variants, the pathogenic *IDH2* R140H mutation was detected with a VAF of 30.64%. The patient subsequently progressed to secondary AML and was treated with induction with daunorubicin (90mg/m^2/d on days 1-3), cytarabine (100mg/m^2/day on days 1-7) and the *FLT3* inhibitor midostaurin (50mg twice per day on days 8-21). He achieved a complete remission with incomplete hematologic recovery (CRi). He subsequently received one cycle of high dose cytarabine (3g/m^2/dose every 12 hours for 4 doses) and midostaurin (50mg twice per day on days 8-21). He relapsed again and was refractory to treatment with venetoclax (400mg/day on days 1-28)-decitabine (20mg/m^2/day on days 1-5) and Gilteritinib (120mg/day) monotherapy. He succumbed secondary to refractory AML.

Patient 2 was a 51-years-old women with MDS-EB-2 with 13% bone marrow blasts at diagnosis. The karyotype at diagnosis showed a complex karyotype:
46,XX,add(1)(p11),add(5)(q11.2),add(6)(p23),-8,+mar[11]/46,XX,add(1)(p11),add(5)(q11.2),

add(6)(p23),-8,+r[2]/46,XX[6]. Bulk sequencing showed two frameshift mutations of *KMT2A* and *KMT2D* and were predicted to be functionally disruptive (supplemental file 1). This patient refused treatment at initial presentation and azacitidine (100mg/m^2/day on days 1-7) every 28 days was started 7 months after first diagnosis. A bone marrow examination at the start of treatment showed a bone marrow blast percentage of 17%. After 3 cycles of azacitidine, the bone marrow blast percentage reduced to 13% with the bulk sequencing showing the same *KMT2A* and *KMT2D* mutations. This patient refused induction chemotherapy or allogeneic hematopoietic stem cell transplantation (HSCT). After 6 further cycles of azacitidine, this patient progressed to secondary AML with bone marrow blast percentage of 21%. At progression to secondary AML, the karyotype showed: 46,XX, add(1)(p11), der(5)t(1;5)(p31;q11.2), ad1d(6)(p23), -8,+mar[13]/46,XX[2].

Clonal architecture of SNVs and INDELs on single-cell DNA sequencing

For Patient 1, a total number of 34,931 cells from 5 time points were sequenced using sc-DNA sequencing platform. There were 98 variants detected across 5 time points with good panel uniformity and overall coverage of 31X – 600X per cell per amplicon (supplemental file 2). Due to poor cell viability, the diagnostic sample of Patient 1 did not complete sc-DNA sequencing and sc-DNA analysis started after 4 cycles of azacitidine. Clustering analysis of pooled variants showed multiple cells clusters with gradual reduction of clonal heterogeneity and dominance by 1 to 2 clusters during the treatment course. Despite having fewer cells successfully sequenced at 25 months from diagnosis, the dominant cell clone showed unique features compared with previous time points (Figure 1A).

Among the list of variants, two pathogenic mutations were identified with high genotype quality (supplemental files 3 and 4) and a dynamic mutation burden was observed: (1) *chr15:90631934:C/T* (*IDH2*:p.R140H) and (2) chr13:28608262:./CTGAAATCAACGTAGAAG (*FLT3*-ITD) (Figure 2A). The average VAF of *IDH2* remained stable at approximately 40% and decreased significantly at 25 months with only minor fraction of cells retaining the mutation. On the other hand, the VAF of *FLT3*-ITD mutation fluctuated between 18-22 months when Patient 1 received midostaurin with induction/consolidation chemotherapy, and venetoclax-decitabine. The VAF of *FLT3*-ITD decreased significantly when Patient 1 received Gilteritinib. At clonal level, 94% of cells were *wild type* (WT) during initial treatment with azacitidine, with a minor clone carrying a heterozygous *IDH2* mutation (Figure 2B and 2C). During transformation into AML at 18 months from diagnosis (after 14 cycles of azacitidine), this clone expanded together with another clone carrying the double heterozygous for *IDH2* and *FLT3*-ITD mutations. Other minor subclones with homozygous mutation either or both of the two genes were also detected, and the homozygous *FLT3* mutant clones expanded despite treatment with midostaurin and induction/consolidation chemotherapy. When gilteritinib was started

at 22 months, all mutant clones were suppressed and the WT clone became the major clone again. Nevertheless, none of the mutant clones were completely eradicated.

Rare cell clones with SNVs and INDELs undetectable with bulk tumor sequencing

For Patient 2, there were 130 variants detected in 37,710 cells sequenced across four time points with good panel uniformity and an overall coverage of 30X – 76X per cell per amplicon (supplemental file 5). Total variant clustering showed a distinct profile across the 4 time-points (at diagnosis, and at 7, 10 and 16 months from diagnosis). Clustering homologies were observed between the diagnostic time point and at 10 months (after 3 cycles of azacitdine) where the bone marrow blast percentage was the same at 13% (Figure 1B). On the other hand, clustering profile at 7 months (at the start of azacitidine) reassembled that at 16 months post-diagnosis when she completed 9 cycles of azacitidine.

Among the variants evaluated for pathogenicity, 4 missense and 2 frameshift variants were shortlisted: (1) chr12-25398284-C-T (*KRAS*:p.G12D), (2) chr4-106196792-T-C (*TET2*:p.C1709R), (3) chr3-128200690-G-A (*GATA2*:p.A372V), (4) chr4-55569900-A-T (*KIT*:p.Q256L), (5) chr7-148515102-C- (*EZH2*:p.R369Sfs*55) and (6) chr12-25378716-GA-G (*KRAS*: c.291-10del). The five variants *KRAS*:p.G12D, *TET2*:p.C1709R, *GATA2*:p.A372V, *EZH2*:p.R369Sfs*55 and *KRAS*:c.291-10del were known pathogenic mutations, while *KIT*:p.Q256L was predicted to be tentatively pathogenic based on Varsome. A generally low VAF was observed for most variants but a low fraction of cells (<1% of total) showed high VAFs in *KRAS*, *GATA* and *KIT* with high genotype quality (Figure 3A). For the *GATA2*, *TET2* and *EZH2* variants, a more diverse VAF heterogeneity was observed (Figure 3A). Longitudinal analysis of *GATA2* and *EZH2* mutations showed that the VAF of *GATA2* mutation was highest at diagnosis and decreased after 3 cycles of azacitidine while the VAF of *EZH2* mutation remained constant initially and increased after 9 cycles of azacitidine (Figure 3B).

Downstream clonal analysis of *GATA2* and *EZH2* showed that these mutations were derived from different heterozygous cell clones harboring either one or both of the mutations (Figure 3C). The two *GATA2* mutant clones ($GATA2^{Het}$ single mutant and $GATA2^{Het}/EZH2^{Het}$ double mutant clones) diminished at disease progression but remained as minor clones throughout the remaining time points while the smallest *EZH2* homozygous clone was no longer detectable at disease progression (Figure 3C). When reviewing the rare mutations, there were minor fractions of *GATA2* and *EZH2* mutant cells that harbored *KIT* or *TET2* mutation. Most of these triple mutant cells persisted across all time point with no specific time-points being enriched for these minor subclones (Figure 3E).

Clonal CNVs associated with pathogenic SNVs or INDELs during HMA resistance

In addition to studying the dynamics of clonal SNV and INDEL, CNVs of all 20 genes (127 amplicons) were studied and correlated with the *IDH2* and *FLT3*-ITD mutations detected in Patient 1. In Patient

1, cluster analysis of all amplicons pooled from all time-points showed no significant differences during treatment with azacitidine and at progression to secondary AML (Figure 4A). Unbiased clustering analysis showed several CNVs in *ASXL1*, *GATA2*, *DNMT3A*, *IDH1/2*, *KIT* and *TET2.* These CNVs were observed serial but *ASXL1*, *DNMT3A*, *IDH1/2* and *KIT* deletion were enriched during at 21 and 25 months when the patient failed treatment with midostaurin-based induction/consolidation, venetoclax-decitabine and gilteritinib (Figure 5A). Despite observing a population of WT clone, loss of *DNMT3A*, *GATA2* and *TET2* were enriched in clones harboring *FLT3*-ITD and/or *IDH2* mutations (supplemental file 6). The magnitude of *DNMT3A* and *GATA2* loss in *FLT3* and/or *IDH2* mutant clones also showed sequential changes as treatment resistance developed while *TET2* CNV remained stable (Figure 5B).

For Patient 2 where rare mutant clones were detected, CNV analysis was performed to detect copy number abnormalities independent of the rare SNVs and INDELs to maximize sensitivity. Overall clustering in patient 2 showed distinct profiles at diagnosis compared with that at subsequent disease stages and homology was observed between 7 months (pre-treatment) and 10 months (after 3 cycles of azacitidine) from diagnosis. The profile at 16 months (at progression to secondary AML after 9 cycles of azacitdine) was somewhat intermediate between 7 and 10 months (Figure 4B). Analysis of all time points in association with *GATA2* and *EZH2* mutations did not detect any large clonal specific CNVs compared with WT cells, except for the amplicon loss of *TET2* exon 3 *(106158314)* in the $GATA2^{Het}$ single mutant clone at 7 months diagnosis (Figure 6A and 6B). Consistent with longitudinal analysis of *GATA2* mutation, this clone was undetectable during treatment with azacitidine and at subsequent progression to secondary AML.

Clonal CNVs as independent events during HMA resistance

To further detect clonal CNVs that were independent from *GATA2* and *EZH2* mutations in patient 2, unbiased clustering with pooled time-points was performed. Putative deletions of *TET2*, *ASXL1*, *DNMT3A* and potential artifacts of *FLT3* and *SRSF2* deletions were detected across multiple time-points (Figure 6C). *FLT3* and *SRSF2* were later confirmed to be allele-dropouts due to poor PCR efficiency of amplicons and they were discarded from subsequent analysis (Figure 6C). To confirm the putative CNVs, repeated analyses normalized to cells with neutral polymorphisms were performed (Figure 6D). Among the three candidate CNVs, only *DNMT3A* and *TET2* deletions were confirmed. A consistent drop of all *DNMT3A* amplicons were observed at 7 months (at the start of azacitidine) and at 16 months (at progression to secondary AML) together with a homozygous loss of *TET2*.

DISCUSSION

In this study, we used a single-cell-based approach to detect mutations that were associated with disease progression in MDS and secondary AML. Using this high-resolution and sensitive method the complex clonal heterogeneity of MDS cells during treatment resistance and progression was deciphered demonstrating a distinct clonal architecture involving both SNVs, INDELs and CNVs. In addition to the good correlation between the VAF on bulk tumor sequencing and the VAF of the sub-clones harboring *FLT3*-ITD and *IDH2* mutations, rare mutations that were below detection limit of bulk tumor sequencing was detectable with sc-DNA sequencing. Although under-sequencing was observed at the last time-point in Patient 1 and at diagnosis in Patient 2 compared with other time points, the high-sensitivity to detect rare mutations, *KIT* and was achieved when integrated analyses of all serial time-points were performed. *KIT* and *TET2* mutations have been implicated in cellular differentiation and DNA methylation in MDS [1]. We also demonstrated that this sc-DNA platform allows definitive determination of co-occurring mutations within the same cell.

Excluding the low fraction of allele-dropout events that led to false positive CNVs of *SRSF2* and *FLT3*, pathogenic loss of *DNMT3A* and *TET2* were detected in both patients. With *DNMT3A* showing a progressive loss in both patients, this suggested continuous treatment with azacitidine or other DNMT inhibitors created a selection pressure for cell clones with *DNMT3A* loss or promote clonal acquisition of *de novo DNMT3A* mutations. While both *TET2* and *DNMT3A* were tumor suppressive in MDS, this study showed that *DNMT3A* loss being the primary or causative genomic abnormality to disease progression and treatment resistance [32, 33]. In patient 1, while *FLT3*-ITD clones were suppressed during potent FLT3 inhibition with Gilteritinib, disease remission was not achieved. In addition to the persistence of *IDH2* R140H clones and other rare cell clones, our study showed that copy number loss of *DNMT3A*, *TET2* and *GATA2* may play a role in resistance to FLT3 inhibitors.

In conclusion, this study has demonstrated clonal evolution of MDS at single cell level treatment resistance to HMA. The underlying evolutionary process involves not only SNVs and short INDELs but also acquired pathogenic CNVs of *GATA2*, *DNMT3A* and *TET2* that are associated with HMA resistance. These CNVs could be coupled with SNVs or small INDELs of *FLT3* and *IDH2*. Nevertheless, the effect of these mutations on cellular differentiation and proliferation, and the expression of cell surface markers will require a multiomic approach to simultaneously detect mutations and phenotypic changes at protein level. The resolution of clonal heterogeneity in MDS may allow better disease monitoring and early detection of resistance clones. Resistant clones pripr to therapy may also be detected. This will facilitate design of combinatorial treatment approaches in patients with MDS.

FUNDING SUPPORT AND ACKNOWLEDGEMENTS

The work was supported by the General Research Fund (GRF), Research Grant Council (RGC) of Hong Kong, China (project code: 17118914) and the Hong Kong Anti-Cancer Society General Award (project code: AR190027). We would like to acknowledge the Genomic Core at the Center for PanorOmic Sciences of Hong Kong University for performing NovaSeq sequencing of the single-cell DNA libraries.

AUTHORS' CONTRIBUTION

Lee P, Yim R, Miu KK and Fung SH performed the experiments and bioinformatic analyses, wrote and approved the manuscript. Wang Z, Wu KC performed the experiments and bioinformatic analyses, and approved the manuscript. Au L collected the clinical data, wrote and approved the manuscript, Leung GMK treatment the patients, wrote and approved the manuscript, Li VHF wrote and approved the manuscript. Gill H conceived the study, treated the patients, designed the experiments, wrote and approved the manuscript.

CONFLICT-OF-INTEREST DISCLOSURES

The authors declare no relevant conflicts of interest to disclose.

REFERENCES

1. Gill H, Leung AY, Kwong YL. Molecular and Cellular Mechanisms of Myelodysplastic Syndrome: Implications on Targeted Therapy. *Int J Mol Sci* 2016; **17**(4)**:** 440.

2. Walter MJ, Shen D, Ding L, Shao J, Koboldt DC, Chen K*, et al.* Clonal architecture of secondary acute myeloid leukemia. *The New England journal of medicine* 2012 Mar 22; **366**(12)**:** 1090-1098.

3. Garcia-Manero G, Chien KS, Montalban-Bravo G. Myelodysplastic syndromes: 2021 update on diagnosis, risk stratification and management. *Am J Hematol* 2020 Nov; **95**(11)**:** 1399-1420.

4. Traina F, Visconte V, Elson P, Tabarroki A, Jankowska AM, Hasrouni E*, et al.* Impact of molecular mutations on treatment response to DNMT inhibitors in myelodysplasia and related neoplasms. *Leukemia* 2014 Jan; **28**(1)**:** 78-87.

5. Woll PS, Kjallquist U, Chowdhury O, Doolittle H, Wedge DC, Thongjuea S*, et al.* Myelodysplastic syndromes are propagated by rare and distinct human cancer stem cells in vivo. *Cancer Cell* 2014 Jun 16; **25**(6)**:** 794-808.

6. Bersanelli M, Travaglino E, Meggendorfer M, Matteuzzi T, Sala C, Mosca E*, et al.* Classification and Personalized Prognostic Assessment on the Basis of Clinical and Genomic Features in Myelodysplastic Syndromes. *J Clin Oncol* 2021 Apr 10; **39**(11)**:** 1223-1233.

7. Nazha A, Komrokji R, Meggendorfer M, Jia X, Radakovich N, Shreve J*, et al.* Personalized Prediction Model to Risk Stratify Patients With Myelodysplastic Syndromes. *J Clin Oncol* 2021 Nov 20; **39**(33)**:** 3737-3746.

8. Platzbecker U. Treatment of MDS. *Blood* 2019 Mar 7; **133**(10)**:** 1096-1107.

9. Jabbour E, Garcia-Manero G, Batty N, Shan J, O'Brien S, Cortes J*, et al.* Outcome of patients with myelodysplastic syndrome after failure of decitabine therapy. *Cancer* 2010 Aug 15; **116**(16)**:** 3830-3834.

10. Santini V. How I treat MDS after hypomethylating agent failure. *Blood* 2019 Feb 7; **133**(6)**:** 521-529.

11. Meldi K, Qin T, Buchi F, Droin N, Sotzen J, Micol JB*, et al.* Specific molecular signatures predict decitabine response in chronic myelomonocytic leukemia. *J Clin Invest* 2015 May; **125**(5)**:** 1857-1872.

12. Tobiasson M, Abdulkadir H, Lennartsson A, Katayama S, Marabita F, De Paepe A*, et al.* Comprehensive mapping of the effects of azacitidine on DNA methylation, repressive/permissive histone marks and gene expression in primary cells from patients with MDS and MDS-related disease. *Oncotarget* 2017 Apr 25; **8**(17)**:** 28812-28825.

13. Nazha A, Sekeres MA, Bejar R, Rauh MJ, Othus M, Komrokji RS*, et al.* Genomic Biomarkers to Predict Resistance to Hypomethylating Agents in Patients With Myelodysplastic Syndromes Using Artificial Intelligence. *JCO Precis Oncol* 2019; **3**.

14. Bewersdorf JP, Zeidan AM. Management of patients with higher-risk myelodysplastic syndromes after failure of hypomethylating agents: What is on the horizon? *Best Pract Res Clin Haematol* 2021 Mar; **34**(1)**:** 101245.

15. Zhao X, Gao S, Wu Z, Kajigaya S, Feng X, Liu Q*, et al.* Single-cell RNA-seq reveals a distinct transcriptome signature of aneuploid hematopoietic cells. *Blood* 2017 Dec 21; **130**(25)**:** 2762-2773.

16. Ganan-Gomez I, Yang H, Ma F, Pellegrini M, Clise-Dwyer K, Garcia-Manero G*, et al.* Single-Cell RNA Sequencing Reveals Distinct Hematopoietic Stem Cell Hierarchies in MDS. *Blood* 2019; **134**(Supplement_1)**:** 771-771.

17. Gill H, Leung GMK, Yim R, Lee P, Pang HH, Ip HW*, et al.* Myeloproliferative neoplasms treated with hydroxyurea, pegylated interferon alpha-2A or ruxolitinib: clinicohematologic responses, quality-of-life changes and safety in the real-world setting. *Hematology* 2020 Dec; **25**(1)**:** 247-257.

18. Gill H, Ip HW, Yim R, Tang WF, Pang HH, Lee P*, et al.* Next-generation sequencing with a 54-gene panel identified unique mutational profile and prognostic markers in Chinese patients with myelofibrosis. *Ann Hematol* 2019 Apr; **98**(4)**:** 869-879.

19. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009 Jul 15; **25**(14)**:** 1754-1760.

20. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014 Aug 1; **30**(15)**:** 2114-2120.

21. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A*, et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013; **43:** 11 10 11-11 10 33.

22. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L*, et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012 Mar; **22**(3)**:** 568-576.

23. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 2009 Nov 1; **25**(21)**:** 2865-2871.

24. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L*, et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 2012 Apr-Jun; **6**(2)**:** 80-92.

25. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010 Sep; **38**(16)**:** e164.

26. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R*, et al.* VarSome: the human genomic variant search engine. *Bioinformatics* 2019 Jun 1; **35**(11)**:** 1978-1980.

27. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S*, et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018 Jan 4; **46**(D1)**:** D1062-D1067.

28. Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res* 1999 Aug; **9**(8)**:** 677-679.

29. Smigielski EM, Sirotkin K, Ward M, Sherry ST. dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res* 2000 Jan 1; **28**(1)**:** 352-355.

30. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q*, et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020 May; **581**(7809)**:** 434-443.

31. Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* 2015 Mar 1; **31**(5)**:** 761-763.

32. Feng YM, Li XP, Cassady K, Zou ZM, Zhang X. TET2 Function in Hematopoietic Malignancies, Immune Regulation, and DNA Repair. *Front Oncol* 2019 Apr 2; **9**.

33. Saliba AN, John AJ, Kaufmann SH. Resistance to venetoclax and hypomethylating agents in acute myeloid leukemia. *Cancer Drug Resist* 2021; **4:** 125-142.

FIGURE LEGENDS

Figure 1. Clonal heterogeneity during treatment by density-based spatial clustering of cells with noise (DBSCAN) based on the allelic burden of single nucleotide variants (SNV) and insertions-delections (INDELs) allele burden in 2 patients with myelodysplastic syndrome (MDS). A: Patient 1 at 4 months (after 4 cycles of azacitidine), at 18 months (at progression to secondary acute myeloid leukemia after 18 cycles of azacitidine), at 21 months (at relapse after the first cycle of midostaurin and high-dose cytarabine consolidation), at 22 month (non-remission to one cycle of venetoclax-decitabine) and at 25 months (non-remission after 3 months of Gilteritinib); **B:** Patient 2 at diagnosis, at 7 months (before the first cycle of azacitidine), at 10 months (after 3 cycles of

azacitidine), and at 16 months (at progression to secondary acute myeloid leukemia after 9 cycles of azacitidine).

Figure 2. Clonal evolution of *IDH2* R140H and *FLT3*-ITD mutations in Patient 1. A: Overall variant allele frequencies (VAFs) for *IDH2* R140H and *FLT3*-ITD across different time points evaluated; **B:** Fish plot showing evolution of *wild type* (WT), *FLT3*-ITD and *IDH2* R140H clones serially with treatment; **C:** VAFs of *IDH2* R140H and *FLT3*-ITD mutations at sub-clonal level. Red and green lines on the violin plots represent median and mean VAFs respectively.

Figure 3. Detection and clonal evolutions of mutations in Patient 2. A: Variant allele frequencies (VAFs) and genotype quality (GQ) of rare pathogenic single nucleotide variants (SNVs) and insertions-deletions (INDELs) which were present in <1% of total cells sequenced; **B:** VAFs of major *GATA2* and *EZH2* mutations across different time points evaluated; **C:** Fish plot evolution of *wild type* (WT), *GATA2* and *EZH2* clones serially with treatment; **D:** VAFs of *EZH2* and *GATA2* mutations at sub-clonal level; **E:** VAFs of rare *KIT* and *TET2* mutations in association with *GATA2* and *EZH2* mutations across different time points. Red and green lines on violin plots represented median and mean VAF respectively while color of each cell represented corresponding time points.

Figure 4. Heterogeneity of overall copy numbers during treatment by Principal Component Analysis (PCA) and Uniform Manifold Approximation Projection (UMAP) analysis. A: Patient 1 at 4 months (after 4 cycles of azacitidine), at 18 months (at progression to secondary acute myeloid leukemia after 18 cycles of azacitidine), at 21 months (at relapse after the first cycle of midostaurin and high-dose cytarabine consolidation), at 22 months (non-remission to one cycle of venetoclax-decitabine) and at 25 months (non-remission after 3 months of Gilteritinib); **B:** Patient 2 at diagnosis, at 7 months (before the first cycle of azacitidine), at 10 months (after 3 cycles of azacitidine), and at 16 months (at progression to secondary acute myeloid leukemia after 9 cycles of azacitidine).

Figure 5. Co-occurrence of single nucleotide variants (SNVs) and short insertions-deletions (INDELs) with copy number variations (CNVs) in Patient 1. A: merged CNV analysis of all time points; **B:** Gradual copy number loss of *DNMT3A*, *GATA2* and *TET2* at 18 months (at progression to secondary acute myeloid leukemia after 18 cycles of azacitidine), at 22 months (non-remission to one cycle of venetoclax-decitabine) and at 25 months (non-remission after 3 months of Gilteritinib).

Figure 6. Co-occurrence of single nucleotide variants (SNVs) and short insertions-deletions (INDELs) with copy number variations (CNVs) in Patient 2 as independent cell clones. A: merged CNV analysis of all time points; **B:** CNV analysis at individual time points in association with *GATA2* p.A372V and *EZH2* p.R369Sfs*55 mutations; **C:** Unbiased clustering of CNV analysis of all time points; **D:** Gradual copy number loss of *DNMT3A* and *TET2* at 7 months (before the first

cycle of azacitidine), at 10 months (after 3 cycles of azacitidine), and at 16 months (at progression to secondary acute myeloid leukemia after 9 cycles of azacitidine).

Figures

Figure 1

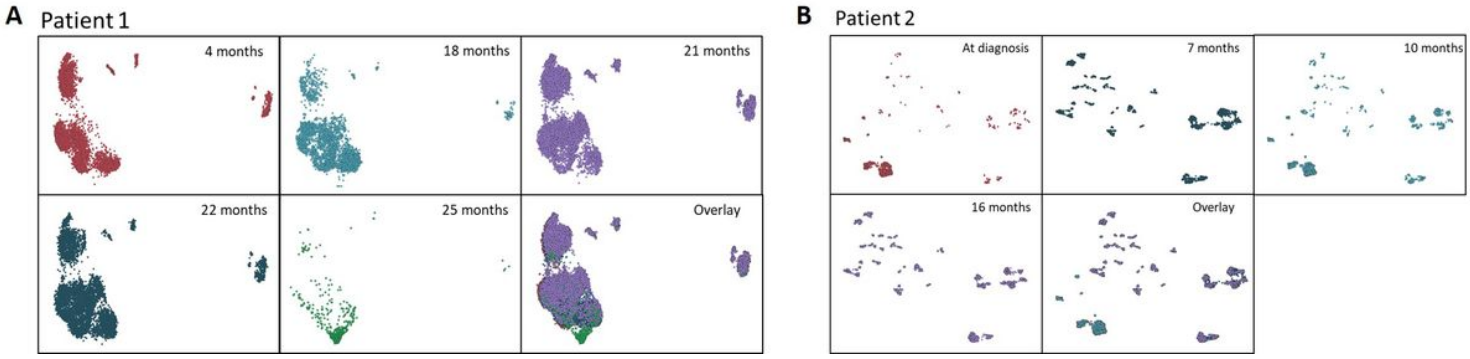

Figure 1

Clonal heterogeneity during treatment by density-based spatial clustering of cells with noise (DBSCAN) based on the allelic burden of single nucleotide variants (SNV) and insertionsdelections (INDELs) allele burden in 2 patients with myelodysplastic syndrome (MDS). A: Patient 1 at 4 months (after 4 cycles of azacitidine), at 18 months (at progression to secondary acute myeloid leukemia after 18 cycles of azacitidine), at 21 months (at relapse after the first cycle of midostaurin and high-dose cytarabine consolidation), at 22 month (non-remission to one cycle of venetoclax-decitabine) and at 25 months (non-remission after 3 months of Gilteritinib); B: Patient 2 at diagnosis, at 7 months (before the first cycle of azacitidine), at 10 months (after 3 cycles of 14 azacitidine), and at 16 months (at progression to secondary acute myeloid leukemia after 9 cycles of azacitidine).

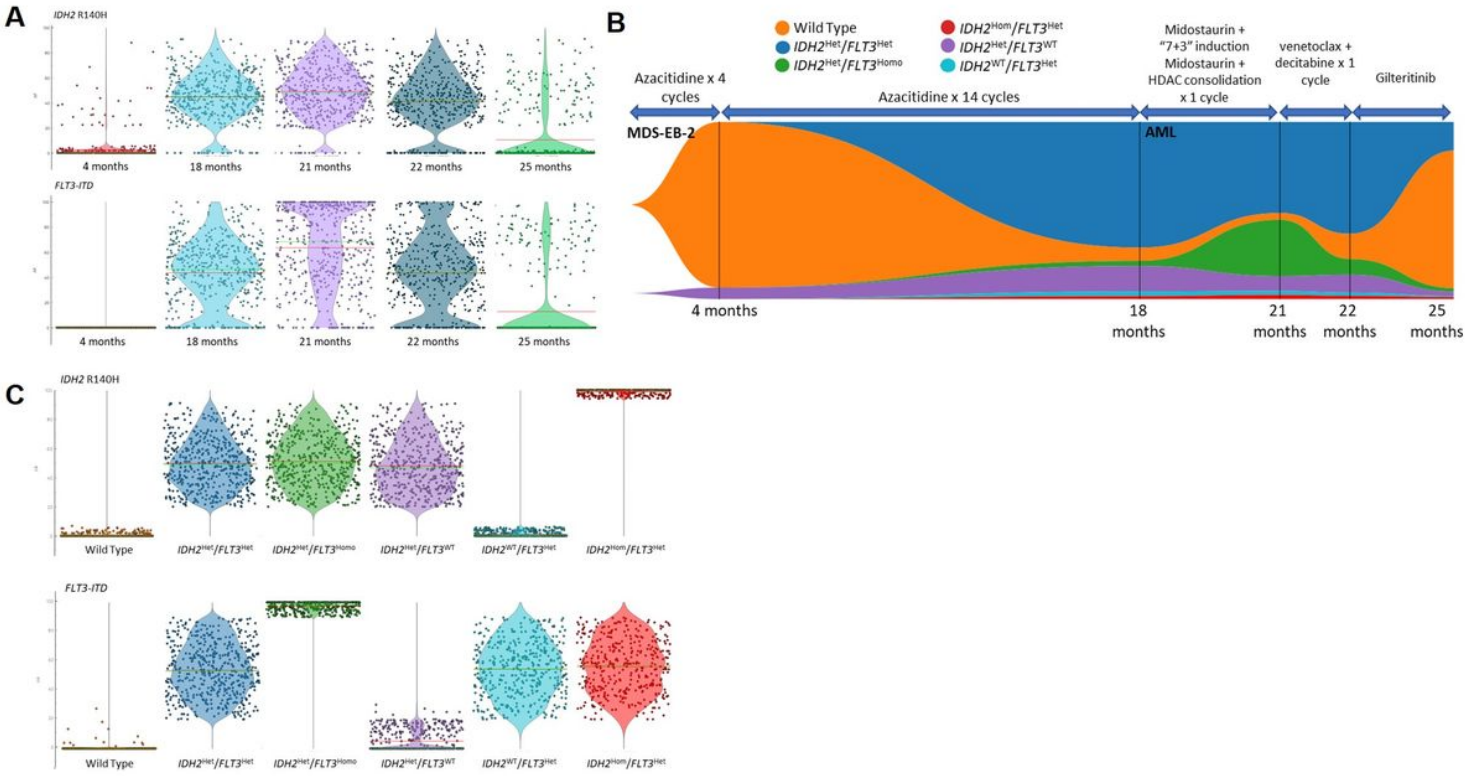

Figure 2

Clonal evolution of IDH2 R140H and FLT3-ITD mutations in Patient 1. A: Overall variant allele frequencies (VAFs) for IDH2 R140H and FLT3-ITD across different time points evaluated; B: Fish plot showing evolution of wild type (WT), FLT3-ITD and IDH2 R140H clones serially with treatment; C: VAFs of IDH2 R140H and FLT3-ITD mutations at sub-clonal level. Red and green lines on the violin plots represent median and mean VAFs respectively.

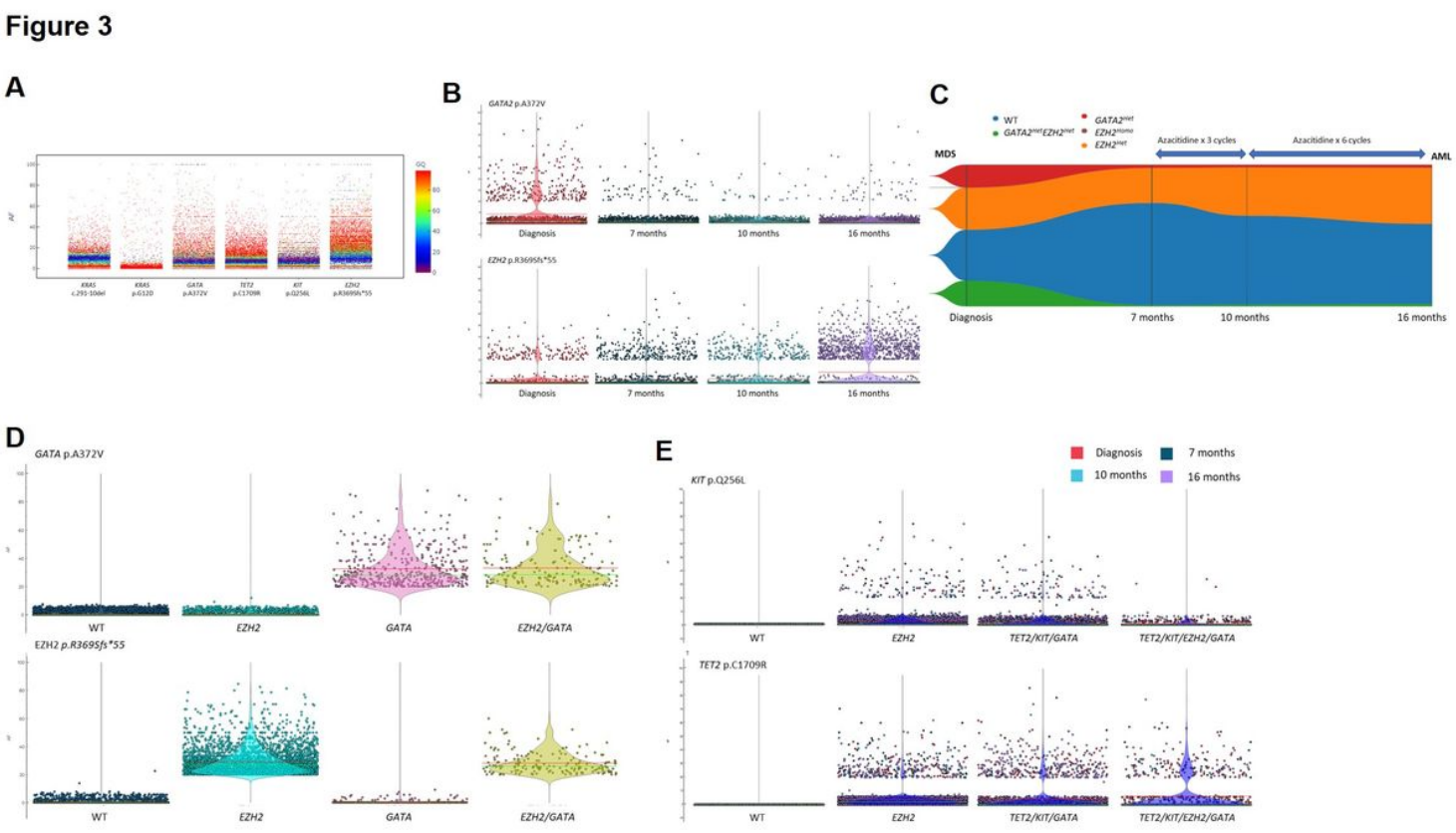

Figure 3

Detection and clonal evolutions of mutations in Patient 2. A: Variant allele frequencies (VAFs) and genotype quality (GQ) of rare pathogenic single nucleotide variants (SNVs) and insertions-deletions (INDELs) which were present in <1% of total cells sequenced; B: VAFs of major GATA2 and EZH2 mutations across different time points evaluated; C: Fish plot evolution of wild type (WT), GATA2 and EZH2 clones serially with treatment; D: VAFs of EZH2 and GATA2 mutations at sub-clonal level; E: VAFs of rare KIT and TET2 mutations in association with GATA2 and EZH2 mutations across different time points. Red and green lines on violin plots represented median and mean VAF respectively while color of each cell represented corresponding time points.

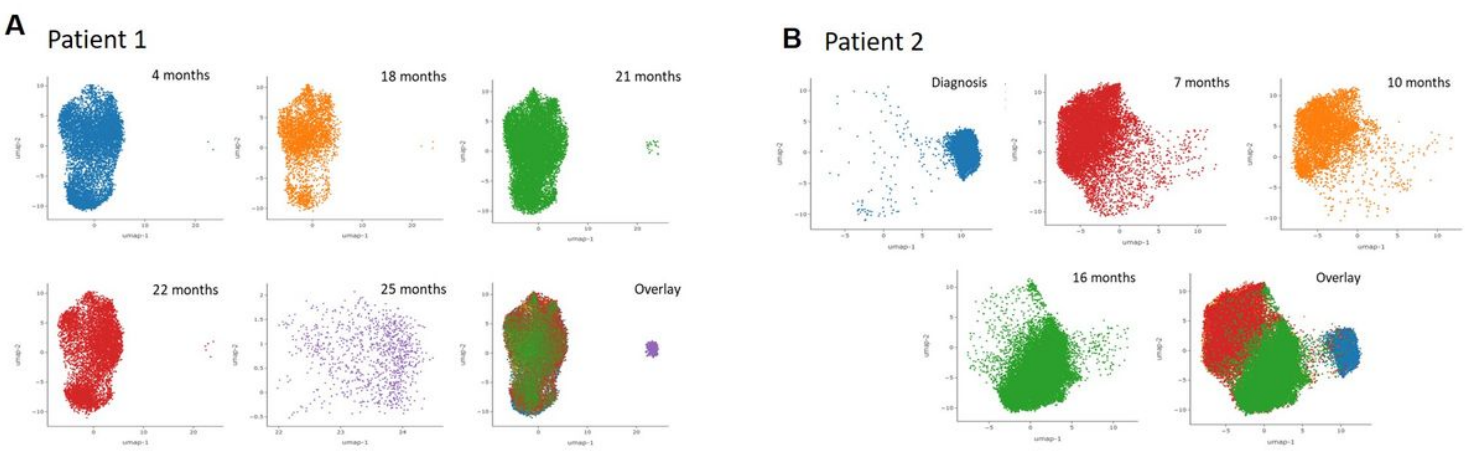

Figure 4

Heterogeneity of overall copy numbers during treatment by Principal Component Analysis (PCA) and Uniform Manifold Approximation Projection (UMAP) analysis. A: Patient 1 at 4 months (after 4 cycles of azacitidine), at 18 months (at progression to secondary acute myeloid leukemia after 18 cycles of azacitidine), at 21 months (at relapse after the first cycle of midostaurin and high-dose cytarabine consolidation), at 22 months (non-remission to one cycle of venetoclaxdecitabine) and at 25 months (non-remission after 3 months of Gilteritinib); B: Patient 2 at diagnosis, at 7 months (before the first cycle of azacitidine), at 10 months (after 3 cycles of azacitidine), and at 16 months (at progression to secondary acute myeloid leukemia after 9 cycles of azacitidine).

Figure 5

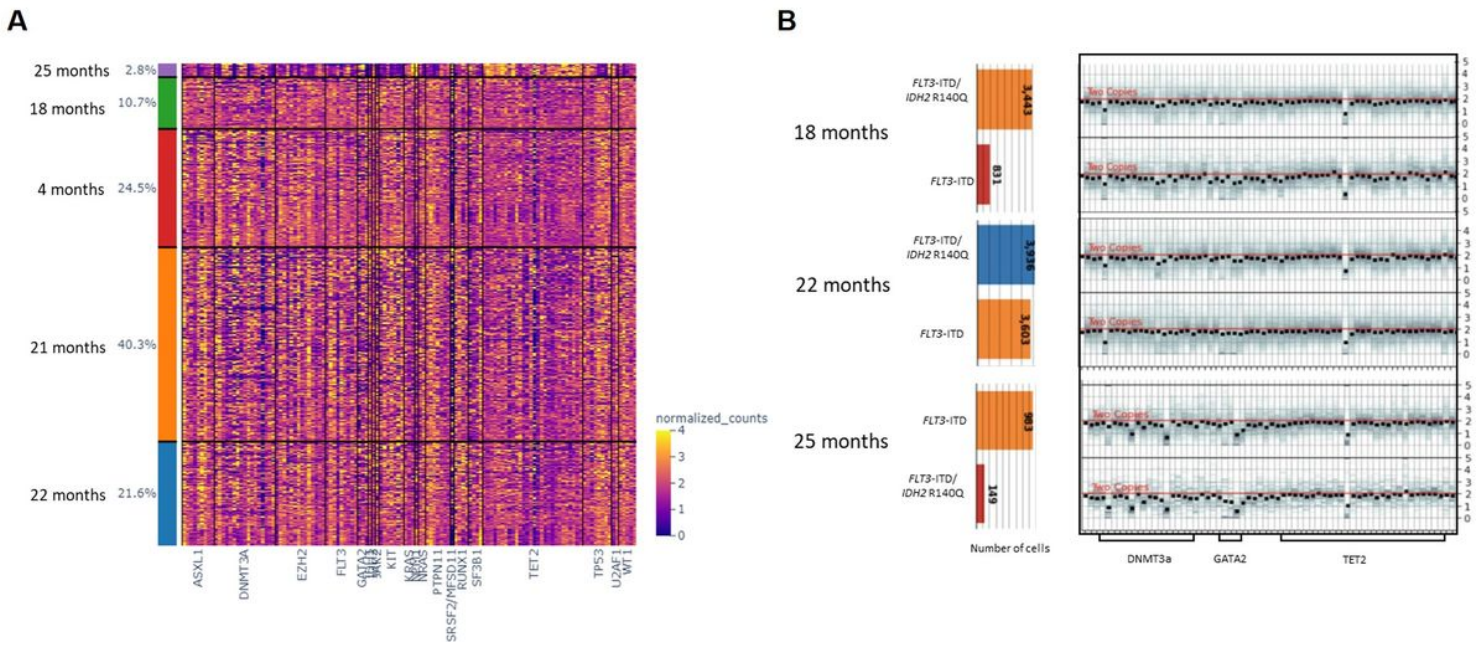

Figure 5

Co-occurrence of single nucleotide variants (SNVs) and short insertions-deletions (INDELs) with copy number variations (CNVs) in Patient 1. A: merged CNV analysis of all time points; B: Gradual copy number loss of DNMT3A, GATA2 and TET2 at 18 months (at progression to secondary acute myeloid leukemia after 18 cycles of azacitidine), at 22 months (non-remission to one cycle of venetoclax-decitabine) and at 25 months (non-remission after 3 months of Gilteritinib).

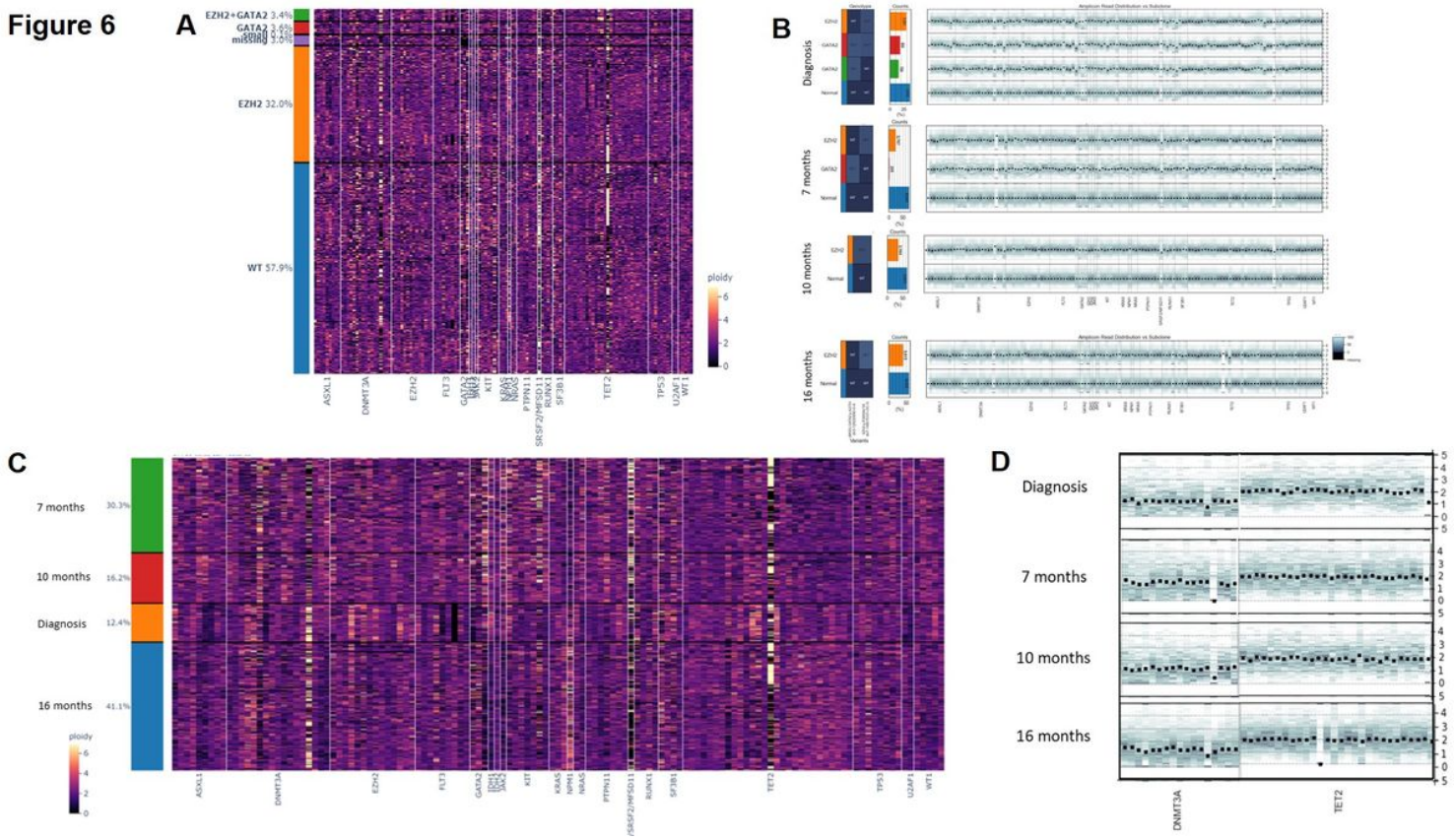

Figure 6

Co-occurrence of single nucleotide variants (SNVs) and short insertions-deletions (INDELs) with copy number variations (CNVs) in Patient 2 as independent cell clones. A: merged CNV analysis of all time points; B: CNV analysis at individual time points in association with GATA2 p.A372V and EZH2 p.R369Sfs*55 mutations; C: Unbiased clustering of CNV analysis of all time points; D: Gradual copy number loss of DNMT3A and TET2 at 7 months (before the first 15 cycle of azacitidine), at 10 months (after 3 cycles of azacitidine), and at 16 months (at progression to secondary acute myeloid leukemia after 9 cycles of azacitidine).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Legendsforsupplementalfiles.docx
- Supplementalfile1.xlsx
- Supplementalfile3.jpg
- Supplementalfile4.jpg
- Supplementalfile6.jpg
- Supplmentalfile2.xlsx
- Supplmentalfile5.xlsx