Somatic variants in epigenetic modifiers can predict failure of response to imatinib but not to second-generation tyrosine kinase inhibitors

Georgios Nteliopoulos,[1] Alexandra Bazeos,[1*] Simone Claudiani,[1,2*] Gareth Gerrard,[1,3*] Edward Curry,[4] Richard Szydlo,[1] Mary Alikian,[1,2] Hui En Foong,[2] Zacharoula Nikolakopoulou,[1,5] Sandra Loaiza,[1,2] Jamshid S. Khorashad,[1,2] Dragana Milojkovic,[1,2] Danilo Perrotti,[1,6] Robert Peter Gale,[1] Letizia Foroni[1] and Jane F. Apperley[1,2]

[1]Centre for Haematology, Department of Medicine, Imperial College, London, UK; [2]Imperial College Healthcare NHS Trust, London, UK; [3]Sarah Cannon Molecular Diagnostics, HCA Healthcare UK, London, UK; [4]Department of Surgery and Cancer, Ovarian Cancer Action Research Centre, Imperial College, London, UK; [5]Lungs for Living Research Centre, UCL Respiratory, University College London, London, UK and [6]Greenebaum Cancer Center, University of Maryland School of Medicine, Baltimore MD, USA

AB, SC and GG contributed equally to this work.

Supplementary Material

Supplementary Methods:

Study participants

Cryopreserved mononuclear cells (MNCs) from untreated subjects with newly diagnosed CML in the chronic phase (CML-CP) were used. Subjects included in this study started treatment at the Hammersmith Hospital between May 2000 and October 2012 (n=55) or enrolled on the STI571 Prospective International Randomised Trial (SPIRIT2) Study (ClinicalTrials.gov: NCT01460693, EudraCT Number: 2007-006185-15), and samples were collected between 2009 and 2011 (n=62), the Evaluating Nilotinib Efficacy and Safety in Clinical Trials-Newly Diagnosed Patients (ENESTnd) study (ClinicalTrials.gov: NCT00471497, EudraCT Number: 2007-000208-34), samples collected in 2008 (n=3), the ENEST1 study (NCT01535391, EudraCT Number: 2011-002787-25), samples collected in 2011 (n=1) and the Bosutinib trial in First line chrOnic myelogenous leukemia tREatment (BFORE) study (ClinicalTrials.gov: NCT02130557, EudraCT Number: 2013-005101-31), samples collected 2014-2015 (n=3). SPIRIT 2 is a Phase III, multicentre, open-label, prospective randomized trial comparing imatinib (Glivec) 400mg/day versus dasatinib (Sprycel) 100mg/day. ENESTnd is a Phase III Multi-center, Open-label, Randomized Study comparing nilotinib (Tasigna) 300 mg twice daily and 400 mg twice daily, with imatinib (Glivec) 400 mg once daily. ENEST1 is a phase IIIb single-arm study of nilotinib, 300 mg twice daily. BFORE is a Phase III, two-arm, randomized, open label trial comparing bosutinib (Bosufil) 400 mg orally once daily or or imatinib 400 mg orally once daily.

Of the 124 CML-CP subjects, 62 started treatment on imatinib (IM) 400 mg/day, 55, on dasatinib (DAS) 100 mg/day, 4 on nilotinib (NIL) 600 mg/day and 3 on bosutinib (BOS) 400 mg/day. For classification of subjects as responders (n=74) or non-responders (n=50) we used mainly their *BCRABL1/ABL1* transcript levels (at international scale/ IS) 3, 6, 12 months after initiating TKI therapy (1). Because the role of molecular monitoring has evolved over the period of use of the TKI and the aim of this work was to evaluate the predictive value of the somatic variants at diagnosis rather than comparing the early molecular responses, we used a combination of the

European LeukemiaNet (ELN) criteria to define response. In particular, as TKI responders were considered the subjects with the majority of the available 3 real-time quantitative polymerase chain reaction (RT-qPCR) measurement within the first year, i.e. ≤10% at 3 months, ≤1% at 6 months or ≤0.1% at 12 months. As further evidence of their appropriate classification as responders, all responders either subsequently achieved a depth of response equivalent to or deeper than MR4, or sustained MR3 for more than a year. All non-responders met the definition of ELN failure, including primary (failure to achieve a given response at a given time) or secondary (loss of response) (2). As a result, IM subjects were classified as 33 responders (R) and 29 non-responders (NR), DAS subjects as 31 R and 24 NR, NIL subjects as 4 R and 0 NR and BOS subjects 1 R and 2 NR. Additionally, we used material from normal individuals consisted of cryopreserved granulocyte-colony-stimulating factor (G-CSF) mobilised peripheral blood MNCs collected (in excess requirement) from 14 normal donors collected between December 2000 and May 2013 destined for allogeneic stem cell transplant procedures. Samples were collected at diagnosis and stored before use and selection of CD34$^+$ progenitor cells.

Some subjects that had achieved MR3 (3-log reduction in *BCRABL1* transcripts from baseline), under IM treatment underwent G-CSF mobilisation of Philadelphia (Ph) chromosome-negative peripheral blood stem cells (with median follow-up 25.6 months after treatment initiation). This was an additional strategy for reducing the number of contaminating Ph-positive stem cells with the intention of improving autologous stem cells transplantation outcome if required sometime in the future (3). We examined the presence of somatic variants in paired-DNA samples from CD34$^+$ cells derived from 11 subjects in MR3. For some subjects with somatic variants in CD34$^+$ cells at diagnosis that progressed to blast phase (BP; n=4) after IM treatment, we examined the presence of somatic variants in paired CP-BP whole-blood samples, to investigate the evolution of the somatic variants after treatment. Totally we used whole-blood samples from 7 subjects with somatic variants at diagnosis, in order to compare the presence of those variants between whole-blood vs CD34$^+$ cell fraction. We also used CD34$^+$ cells from 3 subjects with somatic variants at diagnosis to establish some liquid cultures.

DNA preparation

MNCs were kept frozen in liquid nitrogen for a median duration of 6 years (range 1-13 years). Immediately prior to CD34+ cell selection, MNCs were thawed at 37°C and washed in a solution containing 50% fetal bovine serum (FBS; Biosera, Nuaillé, France), 50% RPMI 1640 medium (Gibco, Thermo Fisher Scientific, Paisley, UK), 10 IU/ml preservative-free heparin sodium (CP Pharmaceuticals Ltd, Wrexham, UK) and 1000 IU/ml DNase I (Merck Millipore, Billerica, MA, United States). Following washing a further 200 µl (20,000 IU) DNAse were added directly to the cell pellet for 5 minutes. The cells were then washed twice and re-suspended in Hank's balanced salt solution (HBSS) (Gibco) for Lymphoprep density gradient centrifugation (Axis-Shield, Dundee, UK). The samples were enriched for viable cells using the Dead Cell Removal Kit (Miltenyi Biotec). Next, they were labelled using CliniMacs Anti-human CD34 Microbeads (Miltenyi Biotec, Bergisch Gladbach, Germany) according to the manufacturer's instructions. The labelled cells were passed through a pre-washed LS Column (Miltenyi Biotec). The purity of the CD34+ fraction was consistently above 96% as measured by flow cytometry. Genomic DNA was isolated from CD34+ cells using QIAamp® DNA mini blood kit (Qiagen, Hilden, Germany) according to the manufacturers' instructions. DNA concentration and purity was assessed with the Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) using the Qubit® ds DNA BR assay kit (Thermo Fisher Scientific).

For T-cells isolation, CD3+ cells were separated using CD3 Microbeads (Miltenyi Biotec) and DNA was extracted after *in vitro* culture in RPMI 1640 medium (Gibco) supplemented with 10% FBS (Biosera), GlutaMAX™ I-supplement and 25 mM HEPES (Gibco), 5 mL penicillin/streptomycin (10,000 U/mL penicillin and 10,000 _g/mL streptomycin; Gibco), 5 mL sodium pyruvate MEM (100 mM; Gibco), 0.5 mL 2-ME (1000×, Gibco) and 100 µg/ml of IL-2 (Miltenyi Biotec) for 7 days. We measured *BCRABL1/ABL1* by RT-qPCR in 13 T-cell samples (median 0.12 (IS), range 0.01-0.68), and confirmed very low expression such that we were confident that any variants detected were not emanating from leukemic cells (see 'Sequencing data' section). DNA from routine peripheral blood draws from subjects in deep molecular remission or paired samples at diagnosis-progression was extracted from guanidinium thiocyanate (GTC) lysates pellets, stored at the Imperial College Molecular Pathology Laboratory.

Liquid culture of primary CD34⁺ cells

We studied 3 subjects with somatic variants at diagnosis who did not respond to imatinib and from whom we had available cells. We selected CD34⁺ cells and put them in liquid culture in IMDM medium (Gibco) supplemented with 10% FBS and 5 mL penicillin/streptomycin. The 3 different culture conditions were in medium with StemSpan™ CC100 cytokine mix (Stem Cell Technologies, Cambridge, UK), medium with CC100 + 1 µM of imatinib and medium with CC100 + 50 nM of dasatinib. Fresh medium + cytokines + drugs were added every 3 days and the cells were harvested at days 7 and 14. Cell numbers were counted and RNA and DNA were isolated for *BCRABL1/ABL*1 transcript measurement and mutation analysis using our panel (see below), respectively.

Targeted gene panel generation

The custom panel consisted of 2 pools of 1008 and 994 individual primer pairs and was generated using AmpliSeq Designer (v3.0.1), https://ampliseq.com; Thermo Fisher Scientific) to amplify 2002 amplicons. These amplicons covered all exons, 3'- and 5'- untranslated regions (UTRs) and exon-intron boundaries of the 71 selected genes (covered regions= 96.95% of total exons).

Semiconductor-based targeted resequencing (Ion Torrent PGM)

Amplicon library preparation was performed with Ion AmpliSeq Library Kit v2.0 (Thermo Fisher Scientific), according to the manufacturer's instructions but by using higher input mass (30ng instead of 10ng) of template DNA per pool, to increase the multiplex polymerase chain reaction (PCR) target amplification rate. This was followed by partial digestion of primer sequences using FuPa reagent and the ligation of barcoded sequencing adapters (Ion Xpress Barcode Adapters 1-16, Thermo Fisher Scientific) according to the manufacturer's instructions. Following AMPure bead (Beckman Coulter, Brea, CA, USA) purification, bar-coded libraries were quantified using the KAPA Library Quantification Kit (Kapa Biosystems, Merck Millipore). Libraries were combined to a final concentration of 100 pM and 8 patient samples (or 7 patients for the validation runs) were pooled for template preparation.

Templates were prepared using the Ion PGM Template OT2 200 Kit (Thermo Fisher Scientific) according to the Ion OneTouch 2 System protocol. Briefly, diluted libraries

were mixed with ion sphere particles (ISPs) in the emulsion PCR master mix. Emulsion PCR was performed on the Ion OneTouch 2 System, and as a result amplicons were clonally amplified generating clusters on the surface of the ISPs. Non templated-beads were removed from the solution and the recovered template-positive ISPs were enriched using Dynabeads MyOne Streptavidin C1 Beads in the Ion OneTouch Enrichment System (ES). ISP enrichment was verified using the Ion Sphere quality control kit (Thermo Fisher Scientific) and a Qubit 2.0 fluorometer. Sequencing of the 8 multiplexed samples (or 7 samples for the validation runs) was performed using the Ion Torrent Personal Genome Machine (Thermo Fisher Scientific) on Ion 318 v2 chips using the Ion PGM Sequencing 200 Kit v2 (200-base pair reads) according to the manufacturer's instructions, using 500 flows for 125 cycles. A 318 chip can obtain 4-5.5 million reads and generate 600 Mb – 1 Gb sequencing data for 200-bp reads. To validate somatic variants we prepared new libraries from subject samples and re-ran them multiplexing 7 samples to increase the read depth.

Ion PGM sequencing informatics

Data obtained from the PGM runs were initially processed using Torrent Suite™ version 4.0.2 (Thermo Fisher Scientific) software, which was used to generate sequence reads, trim adapter and primer sequences, filter and remove poor signal reads (low quality and polyclonal sequences) and de-multiplex the reads according to the barcode sequences. Sequencing reads were aligned to the human genome build 19 (hg19) by Torrent Mapping Alignment Program TMAP (v4.0) algorithm using default parameters, and the initial variant calling was performed using the Variant Caller plugin (v4.0-r76860) using high-stringency somatic variant settings. Amplicon coverage summary files were generated using the Torrent Suite Coverage Analysis plugin (v4.0-r77897). The run summary included statistics and quality metrics such as chip loading efficiency, total usable sequences, polyclonality, percentage of low quality reads, median read length, percentage of aligned reads/ bases, on target reads, and coverage mean depth, uniformity and length coverage at 20x. The percentage of usable reads was the percentage or Library IPSs that passed the polyclonal, low quality and primer dimer filters.

Variant Call Format (VCF) files from the Torrent Browser were transferred to Ion Reporter Software v5 for further analysis and filtering of the results. The analysis

included variant annotations for coordinate (position) and chromosome, type of variant (single-nucleotide variants – SNVs or small insertion or deletion of bases - INDELs), genotype, technical criteria such as read coverage, P-value, allele frequency, variant annotations for gene, transcript, location in exons or introns, molecular classification (e.g. synonymous, missense, nonsense, frameshift), amino acid and coding nucleotide change, SIFT (J. Craig Venter Institute) (4) and PolyPhen-2 (Harvard University) (5) functional predictions scores, Grantham and Phylop conservation scores (6), minor allele frequency (MAF) in normal populations from 5000 Exomes, 1000 Genomes projects (7), dbSNP database ID (build 151; National Center for Biotechnology Information, Bethesda, MD, USA) (8), Catalogue of Somatic Mutations in Cancer (COSMIC, v86, Welcome Trust Sanger Institute, Hinxton, UK) (9) and Online Mendelian Inheritance in Man (OMIM) classifications and finally pfam and GO analysis functional annotations.

The filtering steps started by setting a threshold for depth coverage of >40 reads totally and >5 reads for the mutated allele, excluding variants of low quality (with Ion Reporter base-calling error probabilities p value < 0.05). The following step was excluding common single nucleotide polymorphisms (SNPs) from the UCSC database (http://genome.ucsc.edu/). In addition, we used the Exome Aggregation Consortium (ExAC) (10), dbSNP, 5000 Exomes, 1000 Genomes projects for excluding variants with MAF >1%. Next we excluded variants found at any variant allele frequency (VAF) in age matched G-CSF mobilised CD34[+] cells from normal individuals, after applying the same filtering criteria for the healthy controls. For INDEL variants mainly and for some SNV variants we visualised, and manually reviewed them using the Integrative Genomics Viewer (IGV, Broad Institute, Cambridge, MA, USA), excluding strand-biased variants and variants present in homopolymers (false positives mainly for INDELs). Next we excluded intronic (except for the intron-exon boundaries), 5'- and 3'-UTRs variants and synonymous variants that did not cause amino acid change and kept all the non-synonymous variants that are predicted to affect the function of the protein after *in silico* analysis by the pathogenicity programs Mutation Taster (11), Polyphen-2, SIFT and Combined Annotation Dependent Depletion (CADD) (12). As pathogenic variants were those with at least half of Mutation Taster score predictive for disease causing variants, Polyphen-2 score >0.15 or >0.85 predictive for possibly or probably damaging variants, SIFT score ≤0.05 predictive for deleterious variants and CADD "PHRED-scaled" scores >10. CADD scores express the rank in order of

magnitude terms for damaging single nucleotide variants with the 10th-% of CADD scores are assigned to CADD-10, top 1% to CADD-20, top 0.1% to CADD-30. Finally, after comparing with the control DNA used as non-leukaemia control, we were able to detect somatic, germline or pre-leukaemia variants.

The presence of variants in the control DNA (taken from diagnostic T-cells or white cells from the same individuals in MR4; 4-log reduction in *BCRABL1* transcripts from baseline) may indicate except for genetic variant in the germline, pre-leukaemia variants acquired before the Philadelphia chromosome (Ph) formation, in other words the BCRABL1 translocation. Somatic variants were considered the ones not found in the control DNA. For the 14 samples we did not have control DNA, we assumed that all the disruptive variants (nonsense, frameshift, and splice-site) were somatic, whereas the missense variants were considered germline for VAF≥35%, somatic for lower VAF. Variants were classed as pre-leukaemia if VAF in leukaemia DNA was at least 20% greater than VAF in the control DNA. The remaining variants with similar VAF in leukaemia-control DNA (differences <20%) were classified as germline, although the only ones definitely germline were the ones of germline origin in the dbSNP database.

Statistical analysis

The data generated by Ion PGM sequencing (and the gene expression/methylation data) for the 71 genes were analysed in R programming language version 3.2.2 (r-project.org/) using the survival, cmprsk, limma and gplots packages (https://www.bioconductor.org/). Clinical and demographic data were available for all 124 subjects for whom sequencing was performed, including median follow up, age, gender, additional cytogenetic abnormalities (ACA) other than *Ph* translocation, *BCRABL1* transcript (IS), BCRABL1 transcript type, BCRABL1 kinase domain mutations, Sokal clinical risk score, which have been used the last 30 years for patients' classification as low, intermediate or high risk for disease progression at diagnosis and the newer EUTOS long-term survival (ELTS) clinical risk score, which also classifies subjects as low, intermediate or high risk (Table 2). Socal and ELTS scores were calculated on different algorithms based on the same parameters, age, spleen size, platelet count and percentage of myeloblasts. Sokal score values < 0.8, 0.8 - 1.2, > 1.2 define the low-risk, intermediate-risk and high-risk groups, respectively.

ELTS score values ≤ 1.5680, > 1.5680 but ≤ 2.2185, > 2.2185 define the low-risk, intermediate-risk and high-risk groups, respectively.

Event-free survival (EFS) was measured from the start of treatment to the date of any of the following events while on TKI-therapy: loss of haematological or cytogenetic responses, progression to accelerated or blast phases, or death at any time Progression-free survival (PFS) and CML-related survival were measured from the start of therapy to the date of progression to accelerated/blast phases, and date of death from CML, respectively. Multivariate analysis using Cox proportional hazard regression model or Fine-Gray sub-distribution hazards model was performed comparing only two parameters (somatic variants and one extra parameter), due to small subject size. Cumulative incidences were calculated under calculation of competing risks, which were defined as progression to accelerated/blast phases, death from any cause and drug failure due to intolerance. Because we were interested in relatively early endpoints, we did not calculate rates of MR4 and MR4.5 (*BCRABL1/ABL1* ≤0.0032%).

The sample size derived from power calculations based on the manuscript comparing subjects with more or less than 10% *BCRABL1/ABL1* (IS) transcript after 3 months on imatinib, since there was no previous study associating somatic variants at diagnosis with survival (1). Our null hypothesis was that patients with somatic variants would not have different clinical outcome compared to those without variants, which was rejected. Using 1% level of significance (α) and 90% power (β-1) and based on the effect sizes between good and bad responders for OS, PFS, EFS and MR3 which were 36.4, 35.8, 58.2 and 62.4%, the sample sizes were calculated as 78, 82, 36, 35, respectively with average 58. For more robustness we selected 62 and based on the calculations for the IM cohort we selected also 62 for the 2G-TKI cohort. There was no adjustment for multiple testing as all analyses were explorative, except for gene expression/DNA methylation comparisons where FDR (false discovery rate) multiple testing correction was used. p value < 0.05 was considered significant. P values: . <0.1, * <0.05, ** <0.01, *** <0.001

Protein-protein-interaction (PPI) network

A protein-protein-interaction (PPI) network of p210^{BCRABL1} itself was constructed through STRING database (v10.5- http://www.string-db.org/) (13). Using the coded proteins affected by the somatic variants and BCR+ABL1 as query, a PPI network was

formed. Protein interactions whose integrated scores were bigger than 0.4 (the default threshold in the STRING database) were predicted based on different sources, such as experiments, co-expression, pathway databases and PubMed literature (textmining).

DNA methylation profiling and initial analysis
Genomic DNA (500 ng) was bisulfite treated using the Zymo EZ DNA Methylation Kit (Zymo Research) according to the manufacturer's instructions. Illumina Infinium HD Methylation Assay was performed on 4 µl bisulfite-converted gDNA at 50 ng/µl using the manufacturers' protocol. Samples were hybridized to the HumanMethylation450 (HM450K) BeadChips and scanned using the iScan System (Illumina). Data were extracted from the supplied IDAT files and imported into R for use with the package minfi that was used for the initial analysis. All samples passed QC based on the array control probes. Probes with a detection p-value < 0.01 were removed from one sample. Sex chromosomes and those autosomal probes mapping with 90% accuracy to a sex chromosome were taken out from the main analysis. Samples were normalized by quantile normalization of Type I Methylated (Red), Type I Methylated (Green), Type I Unmethylated (Red), Type I Unmethylated (Green), Type II Methylated and Type II Unmethylated separately and beta values calculated. This initial analysis was not part of this manuscript.

RNA preparation and expression array profiling and initial analysis
RNA extraction from the 22 normal and 38 CML-CP subjects was performed using the RNeasy® Plus Mini or Micro Kit (Qiagen) depending upon starting CD34+ cell numbers. Samples were quantified using the Qubit® 2.0 Fluorometer (Invitrogen) and the quality were assessed by the 2200 TapeStation system (Agilent) where samples with RNA integrity numbers greater than 8.0 were used in subsequent expression analyses. RNA amplification and labelling were carried out using the Ambion Illumina TotalPrep-96 RNA Amplification Kit (Life Technologies, Ambion) using 500 ng of RNA per sample. Biotinylated cRNA (750 ng) were loaded on to Human HT-12 v4 Expression (HT-12) BeadChips (Illumina) and imaged using the iScan System (Illumina). Arrays were initially analysed using the lumi package. Four arrays failed initial QC and were removed from further testing. Data were transformed using the

variance stabilizing transform and normalised using quantile normalization. This initial analysis was not part of this this manuscript.

Informatics analysis: Differential methylation/ expression calling

The methylation and gene expression data previously generated were further analysed in R programming language version 3.2.2 (r-project.org/). Normalised Log2 transformed expression values were used for gene expression, whereas normalised beta (β) values (Methylated allele intensity (M) / (Unmethylated allele intensity (U) + Methylated allele intensity (M) + 100), ranging between 0-1) and M values (log2 (β/1-β)) were used for methylation.

For calling differential methylation probes (DMPs)/ differential gene expression between IM responders and non-responders and/or healthy controls the limma package was used to fit a linear regression model. M values were used in the analysis, since the model was based on a Bayesian Gaussian model (14). Since, we had also clinical data from the same patients, we were able to correct for age, so we fitted a linear regression model, including age as a cofounder factor.

The gplots package was used for heatmap construction, a plot with colours in a 100-point scale from blue (lowest) to red (highest). As clustering method, we used hierarchical clustering, based on pairwise Euclidean distances (i.e. the sum of the squared differences between them (15). The scale in the heatmap was set as "row", so as the rows were scaled to have mean zero and standard deviation one.

DNA methylation and gene expression statistical analysis

Differentially methylated positions and differentially expressed genes were calculated using an F-test with a genome-wide corrected p-value < 0.01. Multiple testing correction was used to reduce the likelihood of identifying false-positives. Any place where multiple hypothesis tests was carried out, we used FDR multiple testing correction to estimate family-wise error rates for assessing statistical significance. To have more robust results and identify the DMPs with biological significance, we used the thresholds of 10% methylation difference (delta β > 0.1) and FDR adjusted p-values < 0.01). However, for our comparisons between variant non-variant patients we also used less stringent criteria (non-adjusted p-values < 0.01).

Pathway analysis/ functional annotation

For pathway analysis and functional annotation, an updated (2015) version of ConsensusPathDB (CPDB) was used (16) (http://consensuspathdb.org/). CPDB integrates interaction networks in *Homo sapiens* including binary and complex protein-protein, genetic, metabolic, signalling, gene regulatory and drug-target interactions, and biochemical pathways originated from currently 32 public resources from the literature. Since 450K array has great heterogeneity regarding the CpG representation by gene specific regions, and the genes were not equally covered throughout the array, potential pathway analysis for 450K data would be biased by CpG selection. Therefore, we decided to use stratification by gene region (e.g. promoter) to decrease the potential for bias.

Supplementary Results:

Biological validation of TKI influence on somatic variants

We established 3 different liquid culture conditions using CD34$^+$ cells from 3 different subjects with somatic variants at diagnosis to assess the effects of imatinib and dasatinib on those variants and also on cell numbers and *BCRABL1/ABL1* transcript levels. The results are shown in Supplementary Figure S1. Supplementary Figure S1A shows that treatment with IM or DAS reduces the number of live cells to approximately 50% compared to cytokines control by day 14. In Supplementary Figure S1B we demonstrate that by day 14, *BCRABL1/ABL1* (IS) levels drop by 40% after IM treatment and by 70% after DAS treatment compared to cytokine control. As for the presence of the variants after TKI treatment (Supplementary Figure S1C), *IKZF1* p.Arg184Trp from patient NR17, remained detectable after IM treatment and in the untreated control, but became undetectable after DAS treatment by day 7. *IKZF1* p.Arg213* from patient NR2 was undetectable by day 7 after DAS treatment while it also became undetectable after IM treatment but at the later time-point of day 14 and persisted in the untreated control. *IKZF1* p.Tyr348* from the same patient was persistent in all conditions. Our results imply that variants with low VAF are eradicated by *in vitro* treatment with dasatinib but persist on treatment with imatinib or are eradicated more slowly. In contrast, under these *in vitro* conditions, a variant (*ASXL1* p.Gln780* from patient NR2) with high VAF was still detectable after exposure to either imatinib or dasatinib, but at a slightly reduced level. Eradication may require longer exposure and/or higher doses of TKI. In summary our data suggest that at least for some variants there is a differential response to imatinib and dasatinib.

Gene expression for the 71 epigenetic modifiers

Gene expression data from 70 genes (no data for *KDM6A* [*UTX*]) were compared between 19 IM non-responders (CML-CP-NR), 19 IM responders (CML-CP-R) and 22 normal individuals (Supplementary Figure S2). There was a clear separation between CML-CP (R+NR) and healthy subjects, as demonstrated by the two clusters in the heatmap. When we calculated differential gene expression, we found that there was significant difference in 51/70 genes, with 19 epigenetic modifiers downregulated in CML-CP and 32 upregulated. However, we could not identify clear separation between responders- non-responders since when we excluded from the analysis the healthy

12

controls and examined differential expression between R vs NR, we found no genes with significant differences based on the adjusted false discovery rate (FDR) p value.

Sequencing data - Run results and variant reducing pipeline

An average of 625,735 (183,362-1,742,697) mapped reads per sample with an average depth of coverage of 273x (85-805), average coverage uniformity (percentage of bases covered at ≥20% of the mean coverage) of 92.1 (69.1-96.9) % and average on-target reads (percentage of reads that mapped to target regions, out of total mapped reads per run) of 97.3 (89.5-98.9) % was obtained when we multiplexed 8 samples, yielding sufficient coverage for detection of variants with high sensitivity with VAF as low as 4%. When we multiplexed 7 samples for the validation runs, the number of reads [764,187 (220,562-2,465,814)], coverage depth [331x (98-1088)], uniformity [95.6 (93.2-96.7) %] and on-target reads [97.8 (96.8-98.9) %] were increased. The average number of variants derived from the PGM sequencing results (in high stringency configuration) was initially 177 (106-244) per sample. Filtering out the common in the population SNPs with MAF>0.01 reduced the number of variants/ sample to 29 (14-76) and use of healthy controls to 12 variants/ sample. Of these variants 6 on average were intronic, 3' or 5' UTR and excluded, apart from the sites in splice sites (splice donors, or splice receivers) that were kept. From the 6 exonic variants, 2 were synonymous, leaving 4. Of these, 3 were predicted as disease causing, deleterious and damaging by Mutation Taster, PolyPhen-2/CADD and SIFT respectively. After excluding the low quality variants the number was reduced to 1.2/ sample.

Incidence of somatic variants in CML-CP subjects

The majority of the 49 somatic variants were found in NR (14/21, 67% in the IM- and 16/28, 57% in the 2G-TKI-treated group). All somatic variants were detected once, except the missense variants in *KMT2E* (p.Met574Lys) and *IKZF1* (p.Arg184Trp) and the nonsense variants in *EP300* (p.Arg1312*), *ASXL1* (p.Tyr591*) and *ASXL1* (p.Arg693*) (n=3, 2, 2, 2 and 2, respectively). Apart from the 15 variants found in the COSMIC database v86, 5 of the variants identified were alternations at the same position as COSMIC variants.

Somatic variants in CD34+ and whole-blood cells at diagnosis - Evolution of variants after treatment

We performed a comparison of somatic variants between CD34+ and whole leukocytes cell fraction in 7 subjects at diagnosis. The subjects who had somatic variants in diagnostic CD34+ cells (*ASXL1* p.Gln780*, *ASXL1* p.Gln594fs) at high level (VAF 51% and 40%, respectively) had also variants in diagnostic whole-blood cells at similar VAFs (51% and 39%, respectively). On the contrary, the ones that had somatic variants in diagnostic CD34+ cells (*IKZF1* p.Arg184Trp, *KMT2E* p.Met574Lys, *KDM5C* p.Pro10Leu and *IKZF1* p.Arg213*/ *IKZF1* p.Tyr348*) at low levels (VAF 5.9%, 6.8%, 5% and 6.9%/4.9%, respectively) had no detectable variants in diagnostic whole blood cells. The low-level variants found in CD34+ cells are probably diluted in the whole blood. However, it may be important to detect these small clones at diagnosis, since when comparing paired whole-blood samples in CP-BP, we found that in one case the undetectable clone at diagnosis (but detectable in CD34+ cells, VAF 5.9%) with a sub-clonal variant (*IKZF1* p.Arg184Trp) expanded during progression (VAF 17%).

In the 3 subjects who achieved MR3 with detectable somatic variants in CD34+ cells at diagnosis but not at MR3, the *BCRABL1/ABL1* ratio at time of MR3 was measured as 0.003%, 0.01% and 0.024%, respectively (Supplementary Table S2).

The age related clonal haematopoiesis (ARCH) related variants, identified only in paired deep remission samples were found in samples collected at a median interval from diagnosis of ~6 years (range 1-16). Eight missense variants were found in 8 subjects with a median age of 60 years (57-72) in deep remission (7 in MR4, 1 in MR3). The presence of these variants affected mostly at low-level (VAF<10%) 8 genes, including *DNMT3A*, *TET2*, *ASXL1* (Supplementary Table S3).

Germline variants in CML-CP subjects

The 99 germline variants, with similar VAF between leukaemia-control samples, were found in 80/124 subjects (40/62, 64% in both the IM and the 2G-TKI cohort) in similar frequencies in the R and NR groups (Supplementary Figure S5). Eleven and 16 subjects had ≥3 and 2 germline variants (missense), respectively in 43 genes (e.g. *KMT2A (MLL), KAT6A, KMT2D, CREEBP, TET1* and *TET2*) (Supplementary Figure S6). The most frequent specific variants were *KAT6A* p.Ala1442Thr (6/124), *TET2* p.Met1701Ile, *SETBP1* p.Pro497Leu, *CREEBP* p.Asn1978Ser and *NSD3 (WHSC1L1)* p.Ser754Leu (3/124). In the dbSNP database 19/99 (19%) variants were of known

germline origin, including 7 variants present in COSMIC (Supplementary Table S6). The remaining variants were not confirmed as germline in the dbSNP database. Germline variants were not associated with age (p=0.89, median age of subjects with variants 49 (range, 21-79) vs 50 (20-80) without variants. In addition, the median age of the normal individuals used to filter-out variants, was also 49 (24-61).

Functional associations of variants with CML-DNA methylation

The last step of functional association of the affected genes with CML pathogenesis was to examine whether they have any effect on DNA methylation. We first called methylation differences between CML-CP subjects with and without variants using an FDR adjusted p value < 0.01, but no differences could be detected. However, using less stringent criteria, just a p-value <0.01, 6,226 differentially methylated probes (DMPs) separated them (Supplementary Figure S8). In order to include only DMPs of biological significance, we only included methylation differences >10% reducing number to 1028 DMPs. Both groups (12 variant and 30 non-variant CMP-CP patients) included R and NR (8NR (67%) and 4R (33%) amongst variant, 7NR (23%) and 23R (77%) amongst non-variant). Using the same criteria imatinib responders and non-responders separated less effectively based on 424 DMPs (data not shown). For functional annotation of DMPs using ConsensusPathDB (CPDB) we used stratification by gene region (promoter, gene body, 5-3- UTR or intergenic) and found over-representation of promoter DMPs for pathways including in the top 10 hits the Imatinib Pathway, Pharmacokinetics/Pharmacodynamics (p=0.0016). However, once again we used non FDR adjusted p value < 0.01 (Supplementary Table S8).

References:

1.   Marin D, Ibrahim AR, Lucas C, et al. Assessment of BCR-ABL1 transcript levels at 3 months is the only requirement for predicting outcome for patients with chronic myeloid leukemia treated with tyrosine kinase inhibitors. J Clin Oncol. 2012;30(3):232-238.

2.   Baccarani M, Deininger MW, et al. European LeukemiaNet recommendations for the management of chronic myeloid leukemia: 2013. Blood. 2013;122(6):872-884.

3.   Drummond MW, Marin D, Clark RE, et al. Mobilization of Ph chromosome-negative peripheral blood stem cells in chronic myeloid leukaemia patients with imatinib mesylate-induced complete cytogenetic remission. Br J Haematol. 2003;123(3):479-483.

4.   Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073-1081.

5.   Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-249.

6.   Siepel A Pollard K, Haussler D. New methods for detecting lineage-specific selection. Proceedings of the 10th International Conference on Research in Computational Molecular Biology. RECOMB 2006: Research in Computational Molecular Biology: pp 190-205

7.   Genomes Project C, Abecasis GR, Altshuler D, et al. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-1073.

8.   Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29(1):308-311.

9.   Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2015;43(Database issue):D805-811.

10.   Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016 Aug;536(7616):285-291.

11.   Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods. 2010;7(8):575-576.

12.   Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310-315.

13.   Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017;45(D1):D362-D368.

14.   Wettenhall JM, Smyth GK. limmaGUI: a graphical user interface for linear modeling of microarray data. Bioinformatics. 2004;20(18):3705-3706.

15.   Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998;95(25):14863-14868.

16.   Herwig R, Hardt C, Lienhard M, Kamburov A. Analyzing and interpreting genome data at the network level with ConsensusPathDB. Nat Protoc. 2016;11(10):1889-1907.

Supplementary Tables:

Supplementary Tables S1, S2, S3, S6, S7, S8 are submitted as excel files.

Supplementary Table S4A: Influence of pre-leukaemia variants on cumulative incidence of MR3 and probabilities of EFS, PFS and CML-related survival in IM-treated subjects

	MR3 cumulative incidence		EFS probability		PFS probability		CML related survival probability	
	p value	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)
All somatic variants in 18 of 62 IM subjects	0.048 *	0.43 (0.19, 0.99)	0.003 **	3.1 (1.4, 6.9)	0.025 *	3.2 (1.1, 9.7)	0.039 *	3.1 (1, 9.7)
Exclusion of 1 subject with pre-leukaemia variants	0.027 *	0.37 (0.15, 0.9)	0.006 **	3 (1.3, 6.6)	0.025 *	3.4 (1.1, 10.2)	0.027 *	3.3 (1.1, 10.3)

Supplementary Table S4B: Influence of pre-leukaemia variants on cumulative incidence of MR3 and probabilities of EFS, PFS and CML-related survival in 2G-TKI-treated subjects

	MR3 cumulative incidence		EFS probability		PFS probability		CML related survival probability	
	p value	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)
All somatic variants in 18 of 62 2GTKI subjects	0.25	0.66 (0.32, 1.3)	0.32	1.7 (0.6, 4.9)	0.46	1.7 (0.4, 7.8)	0.3	2.3 (0.5, 11.3)
Exclusion of 2 subject with pre-leukaemia variants	0.18	0.66 (0.27, 1.3)	0.2	2 (0.7, 5.7)	0.38	1.9 (0.4, 8.6)	0.24	2.5 (0.5, 12.6)

(*) p-values: **.**<0.1, *<0.05, **<0.01, ***<0.001

(†) Excluding subjects with pre-leukaemia variants is compared to the baseline (inclusion of all variants) depicted in the top raw

(‡) CI, confidence intervals, EFS, event-free survival; HR, hazard ratio; MR3, 3-log reduction in *BCRABL1* transcripts from baseline; PFS, progression-free survival

Supplementary Table S5A: MVA of somatic variants with Sokal score, ELTS score, BCRABL1 transcript type in the 2G-TKI cohort for cumulative incidence of MR3 and probabilities of EFS, PFS and CML-related survival

			MR3 cumulative incidence	EFS probability	PFS probability	CML-related survival probability
M	Somatic variants at Dx	HR (95% CI)	0.67 (0.34, 1.29)	1.75 (0.58, 5.26)	1.79 (0.40, 8.01)	2.29 (0.46, 11.4)
V		p value	0.23	0.32	0.45	0.31
A	Sokal score at Dx	HR (95% CI)	0.94 (0.64, 1.38)	1.50 (0.74, 3.01)	1.57 (0.61, 4.08)	1.66 (0.59, 4.68)
		p value	0.74	0.26	0.35	0.34
M	Somatic variants at Dx	HR (95% CI)	0.67 (0.34, 1.30)	1.74 (0.56, 5.39)	1.66 (0.36, 7.66)	2.12 (0.41, 11.1)
V		p value	0.24	0.34	0.52	0.37
A	ELTS score at Dx	HR (95% CI)	0.92 (0.59, 1.43)	1.28 (0.61, 2.66)	1.42 (0.53, 3.81)	1.37 (0.47, 4.06)
		p value	0.72	0.52	0.49	0.57
M	Somatic variants at Dx	HR (95% CI)	0.66 (0.32, 1.33)	1.86 (0.63, 5.51)	1.76 (0.39, 7.91)	2.31 (0.47, 11.5)
V		p value	0.24	0.26	0.46	0.31
A	BCRABL1 transcript type at Dx	HR (95% CI)	0.94 (0.62, 1.42)	1.49 (0.70, 3.17)	0.86 (0.29, 2.52)	0.71 (0.21, 2.42)
		p value	0.77	0.31	0.78	0.58

(*) p-values: .<0.1, *<0.05, **<0.01, ***<0.001
(†) CI indicates confidence intervals, Dx, diagnosis; EFS, event-free survival; ELTS, EUTOS long-term survival; HR, hazard ratio; MR3, 3-log reduction in *BCRABL1* transcripts from baseline; MVA, multivariate analysis; PFS, progression-free survival

Supplementary Table S5B: Association of occurrence of somatic variants with clinical outcome of subjects starting on dasatinib (excluding nilotinib and bosutinib frontline treatment)

|	MR3 cumulative incidence		EFS probability		PFS probability		CML related survival probability	
	p value	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)
All somatic variants in 18 of 62 2GTKI subjects	0.25	0.66 (0.32, 1.3)	0.32	1.7 (0.6, 4.9)	0.46	1.7 (0.4, 7.8)	0.3	2.3 (0.5, 11.3)
Exclusion of 7 subjects treated with NIL or BOS	0.15	0.55 (0.25, 1.2)	0.15	2.2 (0.7, 6.5)	0.43	1.8 (0.4, 8.1)	0.24	2.5 (0.5, 12.6)

(*) p-values: .<0.1, *<0.05, **<0.01, ***<0.001

(†) Excluding subjects threated with frontline NIL or BOS is compared to the baseline (inclusion of all variants) depicted in the top raw
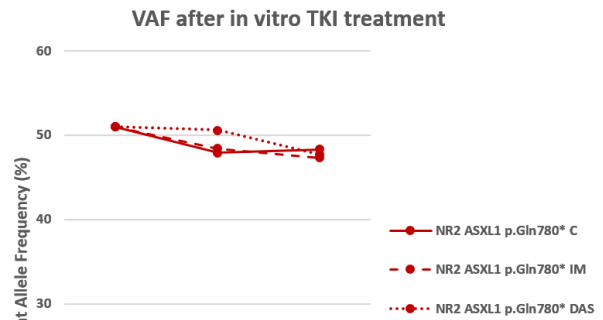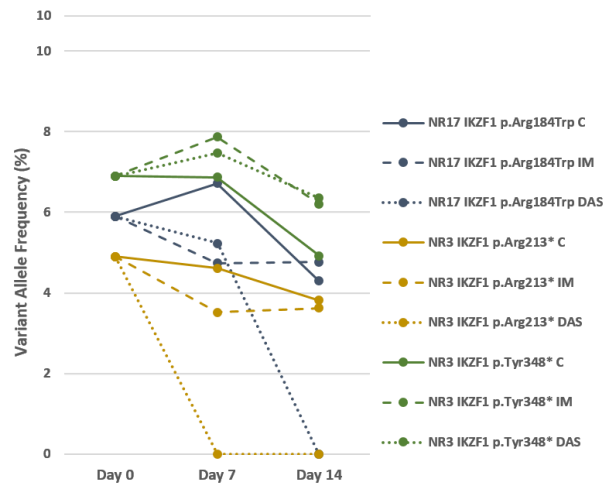
(‡) BOS indicates bosutinib; CI, confidence intervals, EFS, event-free survival; HR, hazard ratio; MR3, 3-log reduction in *BCRABL1* transcripts from baseline; NIL, nilotinib; PFS, progression-free survival

S1A

Cell numbers after in vitro TKI treatment
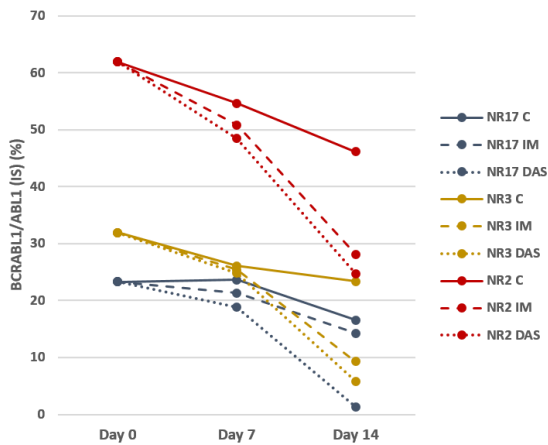
S1C

**VAF after in vitro TKI treatment**
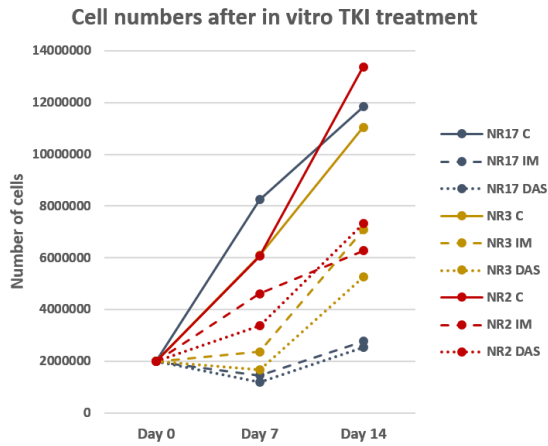
S1B BCRABL1/ABL1 after in vitro TKI treatment

Supplementary Figure S1: Liquid cultures of CD34+ cells from 3 subjects with somatic variants at diagnosis who did not respond to imatinib to assess the effects of *in vitro* TKI treatment. The 3 different culture conditions were in medium with StemSpan™ CC100 cytokine mix (C) medium with CC100 + 1 μM of imatinib (IM) and medium with CC100 + 50 nM of dasatinib (DAS). The cells were harvested at days 7 and 14 and cell numbers **(S1A),** *BCRABL1/ABL1* (IS) levels **(S1B)** and variant allele frequency (VAF) of the relevant somatic variants **(S1C)** were measured for each condition
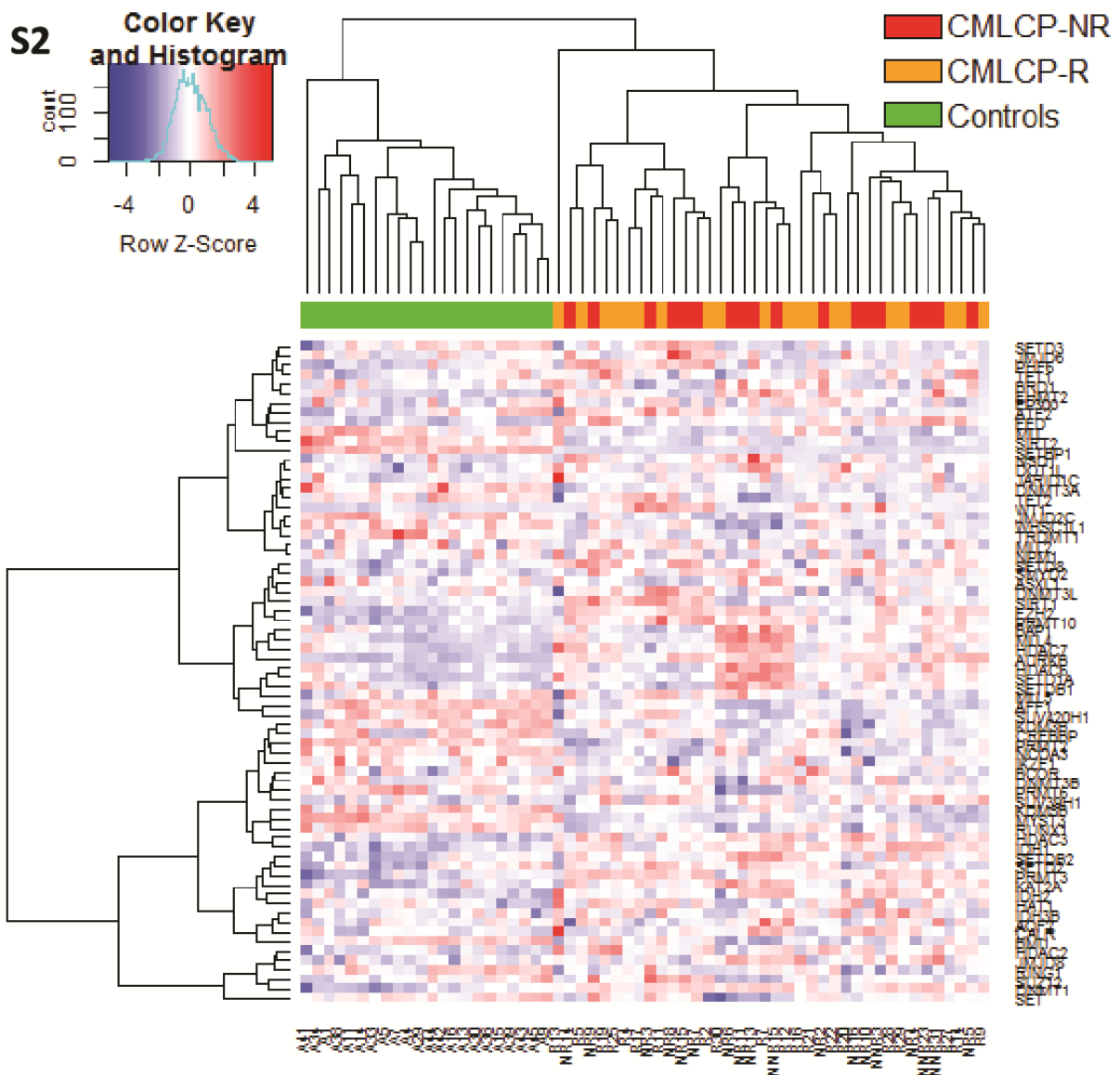
Supplementary Figure S2: Epigenetic modifiers gene expression. Heatmap of 71 epigenetic modifiers based on unsupervised hierarchical clustering of gene expression profiles of CD34-positive cells from CML-CP subjects (n=38) including 19 responders (CML-CPR; orange), 19 non-responders (CML-CPNR; red) and from 22 controls from normal individuals (green). The colours were set as red for relative increase and blue for relative decrease in gene expression
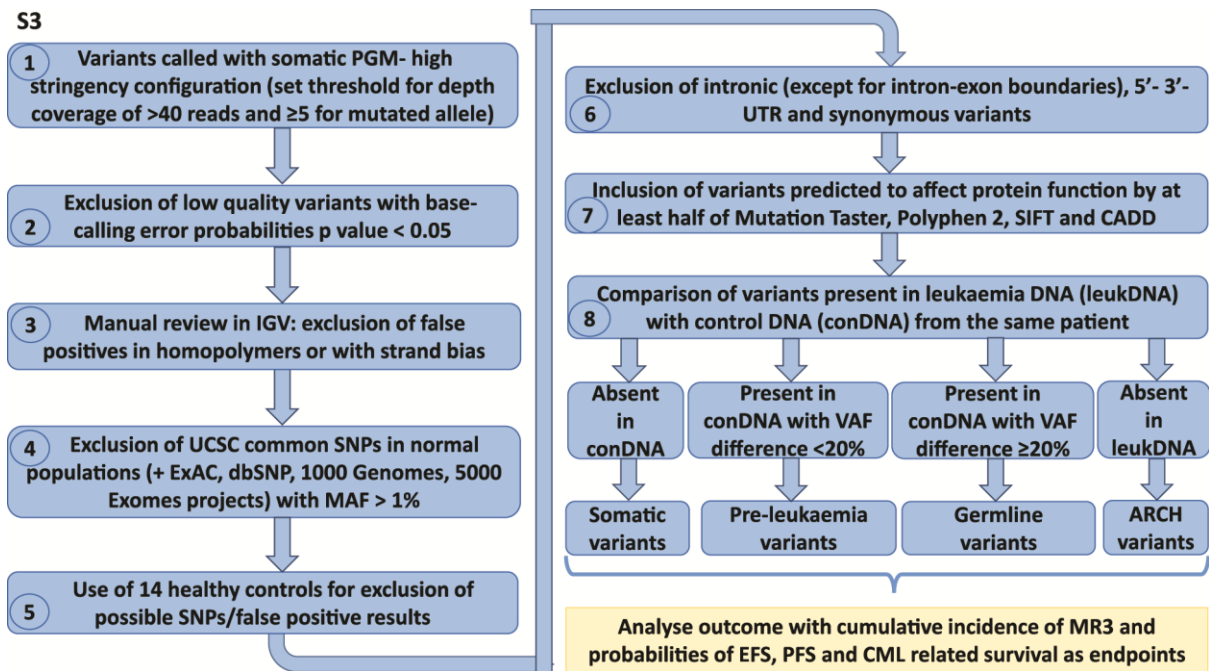
Supplementary Figure S3: Pipeline for variant analysis using Ion reporter v5.
Filtering steps for variant reduction (1-8) and selection criteria for identifying somatic, pre-leukaemia, germline and ARCH variants are depicted. Abbreviations: IGV: Integrative Genomics Viewer, SNV: Single Nucleotide Variants, INDELs: small Insertions/Deletions of bases, SNPs: Single Nucleotide Polymorphisms, MAF: Minor Allele Frequency, ExAC: Exome Aggregation Consortium, UTR: Untranslated Region, CADD: Combined Annotation Dependent Depletion, ARCH: age-related clonal haematopoiesis, MR3: 3-log reduction in *BCRABL1* transcripts from baseline, EFS: event-free survival, PFS: progression-free survival
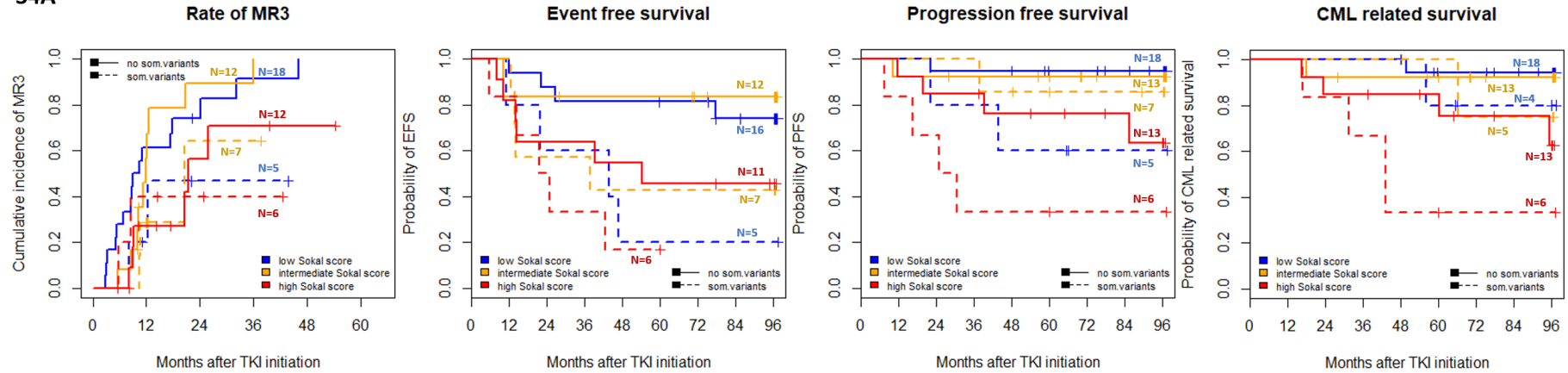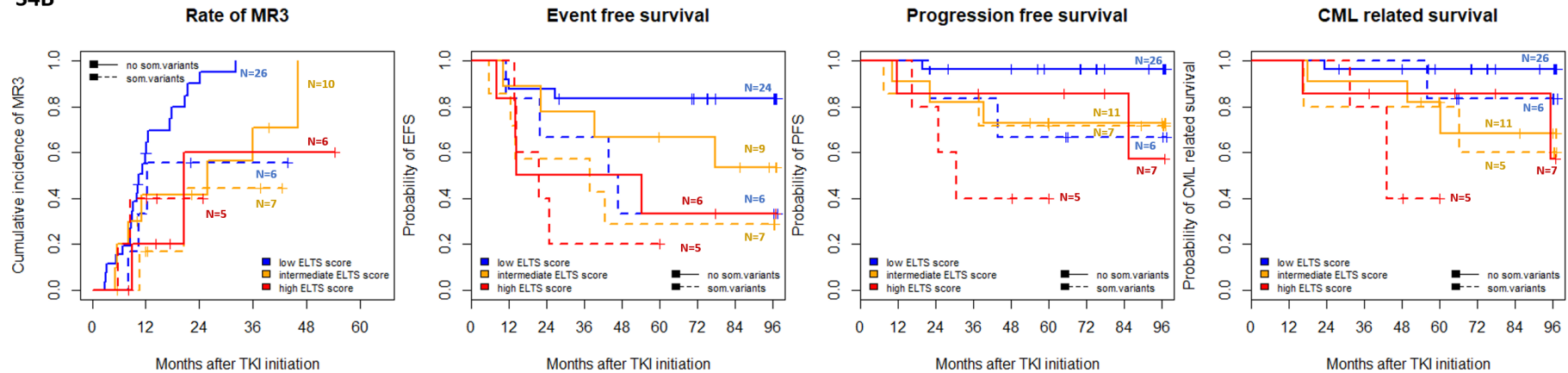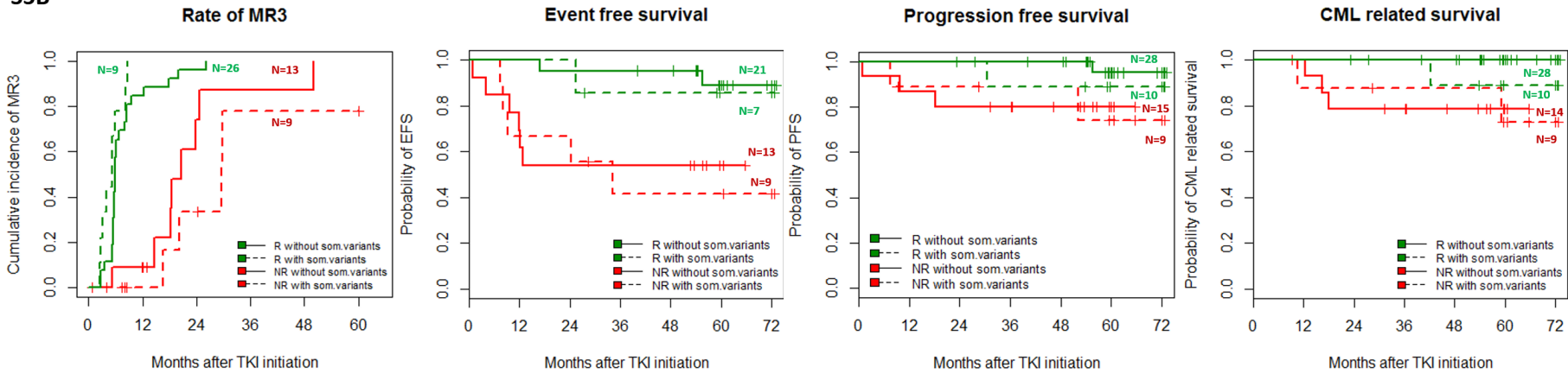
Supplementary Figure S4: Association of somatic variants and Sokal score (S4A) or EUTOS long-term survival (ELTS) score (S4B) with clinical outcome of subjects starting on IM treatment. Kaplan-Meier plots for visualising the survival curves of the variant (dashed lines) vs non-variant (solid lines) patients for low (blue), intermediate (orange) and high (red) Sokal/ ELTS score subjects. Number of subjects (N)/ group is also shown. The endpoints used were cumulative incidence of achieving MR3 and probability of EFS, PFS and CML related survival
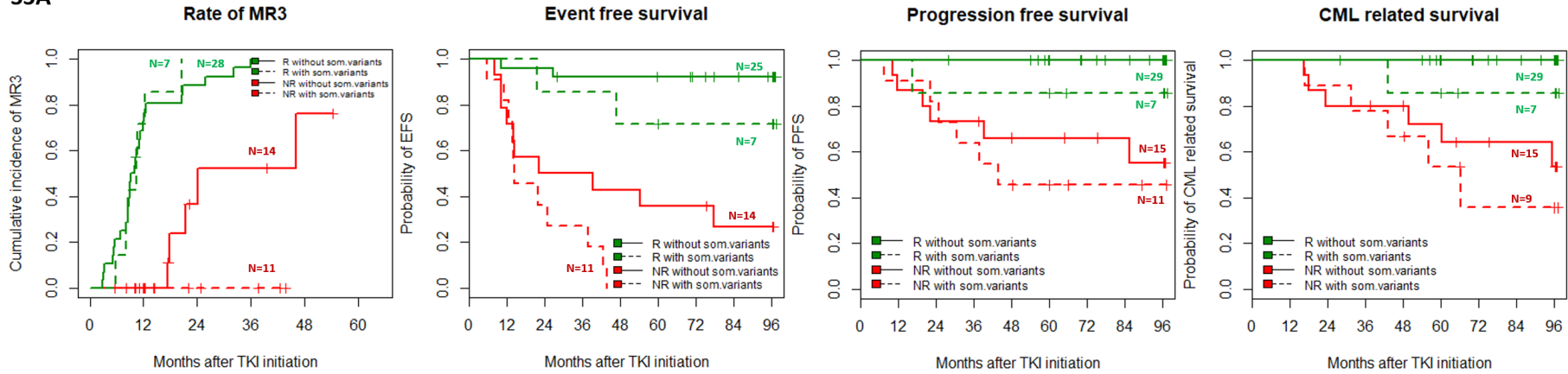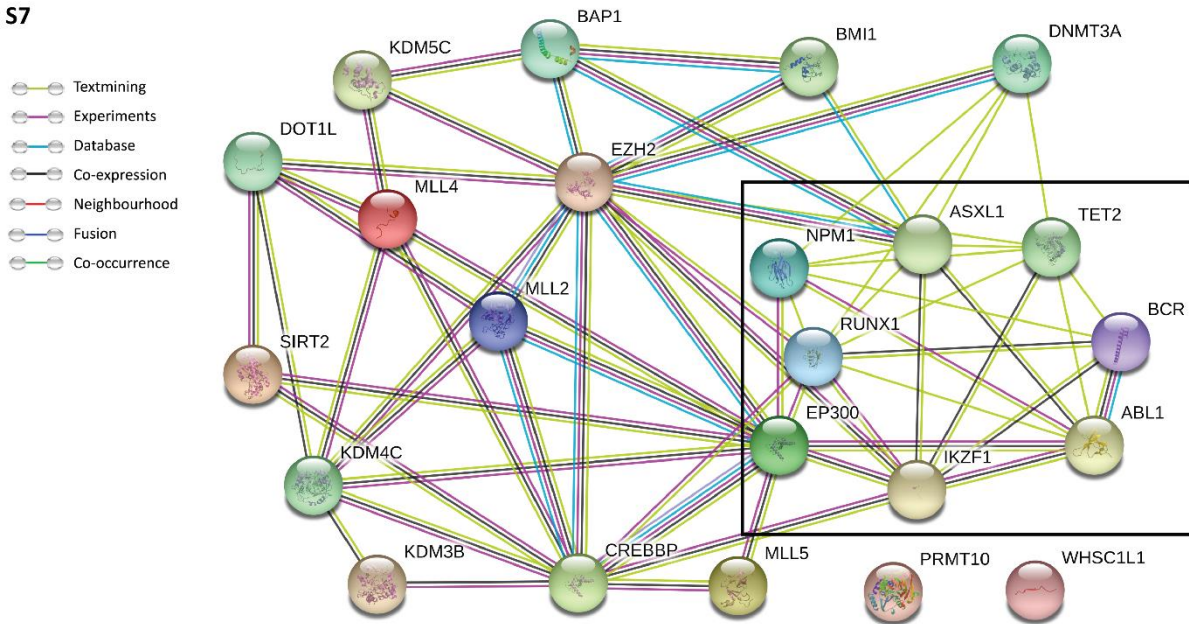
Supplementary Figure S5: Association of somatic variants and first year molecular responses with clinical outcome of subjects starting on IM (S5A) or 2G-TKI (S5B) treatment. Kaplan-Meier plots for visualising the survival curves of the variant (dashed lines) vs non-variant (solid lines) subjects for responders (green) and non-responders (red) to IM/2G-TKI, classified on 3/6/12 months *BCRABL1/ABL1* transcript. Number of subjects (N)/ group is also shown. The endpoints used were cumulative incidence of achieving MR3 and probability of EFS, PFS and CML related survival
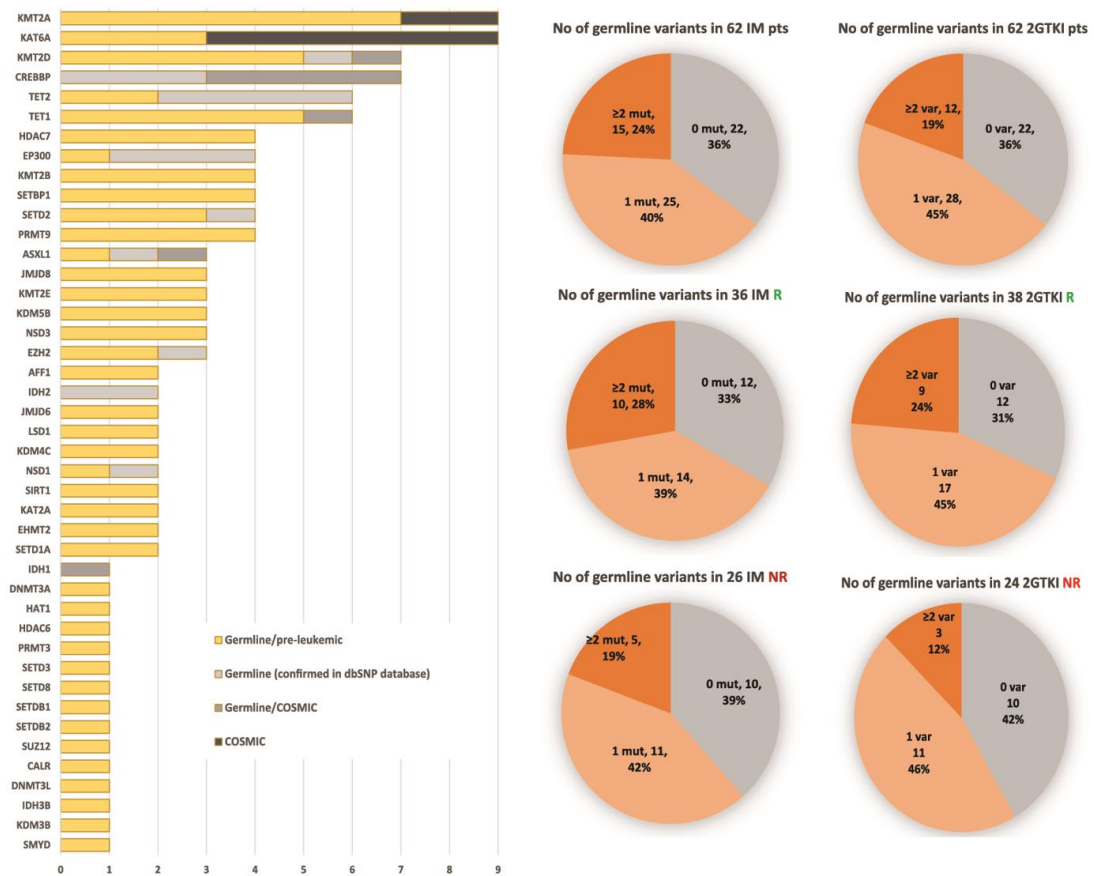
Supplementary Figure S6: Landscape of germline variants in 124 CML-CP subjects at the time of diagnosis. Bar plots indicate the number of germline variants affecting each gene. Genes are ordered by prevalence of variants/ gene in CML-CP. Colours used to distinguish variants based on their presence in the COSMIC and their origin in the dbSNP databases: Yellow indicates absence in both databases; light-grey indicates germline origin in dbSNP; grey COSMIC- germline variants; and dark-grey COSMIC variants. Pie charts showing the proportion/ different number of germline variants per subjects (R+NR) in the 124 CML-CP individuals. IM and 2G-TKI subjects are also demonstrated separately. The colours used were grey, light orange and dark orange for R and NR with no variants, one variant or with ≥2 variants respectively

Supplementary Figure S7: Predicted functional association of affected genes with p210^{BCRABL1}. PPI-network using the 21 coded proteins affected by somatic variants and p210^{BCRABL1} was constructed through STRING database_v10. Different types of predicted interactions among 21/23 proteins are demonstrated based on different sources, depicted with different colours. Black dashed box contains proteins described as direct interacting partners of p210^{BCRABL}. Gene symbols: *MLL2*: *KMT2D*, *MLL5*: *KMT2E*, *MLL4*: *KMT2B*, *WHSC1L1*: *NSD3* and *PRMT10*: *PRMT9*
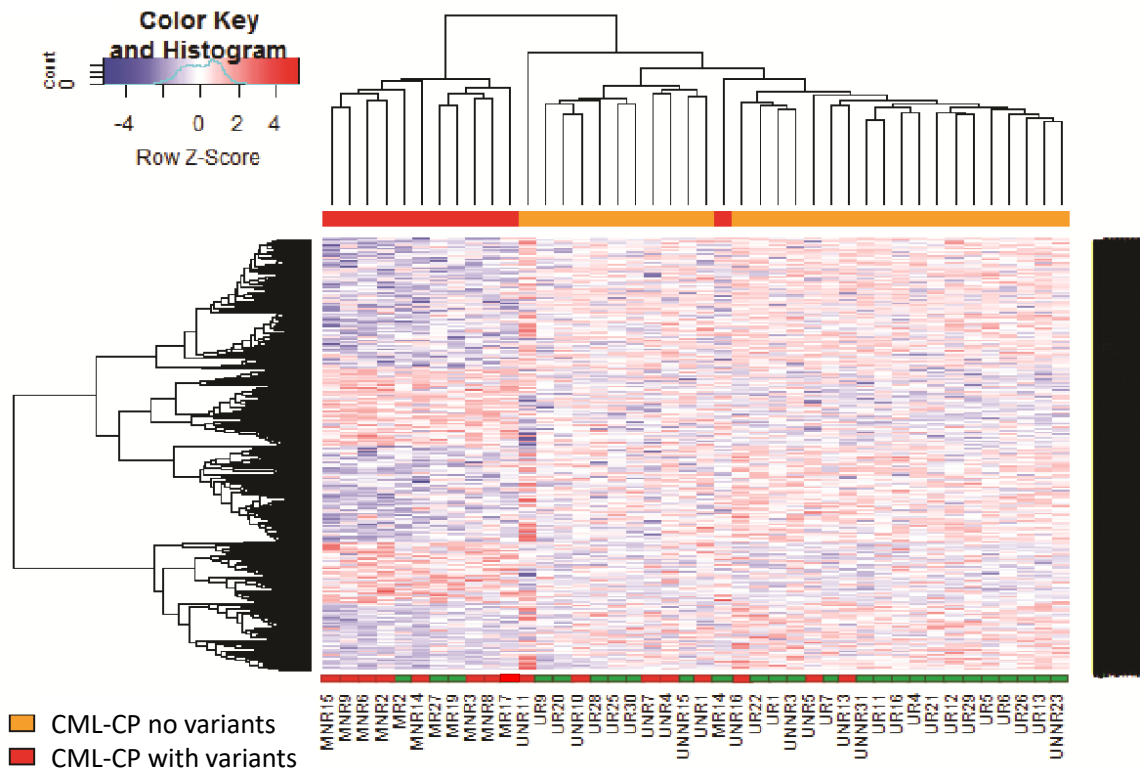
Supplementary Figure S8: Effects of somatic variants on DNA methylation patterns. Heatmap shows hierarchical clustering of DNA methylation profiles of CD34-positive cells from IM-treated individuals with CML-CP based on 1028 DMPs between variant (N=12, red) and non-variant subjects (N=30, orange). Non-adjusted p<0.01 and methylation differences>10% were used. Colours were set as red for relative increase and blue for relative decrease in methylation. Responders are illustrated with green and non-responders with red at the bottom of the heatmap