There are no validated molecular biomarkers to identify newly-diagnosed individuals with chronic-phase chronic myeloid leukemia likely to respond poorly to imatinib and who might benefit from first-line treatment with a more potent second-generation tyrosine kinase inhibitor. Our inability to predict these ‘high-risk’ individuals reflects the poorly understood heterogeneity of the disease. To investigate the potential of genetic variants in epigenetic modifiers as biomarkers at diagnosis, we used Ion Torrent next-generation sequencing of 71 candidate genes for predicting response to tyrosine kinase inhibitors and probability of disease progression. A total of 124 subjects with newly-diagnosed chronic-phase chronic myeloid leukemia began with imatinib (n=62) or second-generation tyrosine kinase inhibitors (n=62) and were classified as responders or non-responders based on the BCRABL1 transcript levels within the first year and the European LeukemiaNet criteria for failure. Somatic variants affecting 21 genes (e.g. ASXL1, IKZF1, DNMT3A, CREBBP) were detected in 30% of subjects, most of whom were non-responders (41% non-responders, 18% responders to imatinib, 38% non-responders, 25% responders to second-generation tyrosine kinase inhibitors). The presence of variants predicted the rate of achieving a major molecular response, event-free survival, progression-free survival and chronic myeloid leukemia-related survival in the imatinib but not the second-generation tyrosine kinase inhibitors cohort. Rare germline variants had no prognostic significance irrespective of treatment while some pre-leukemia variants suggest a multi-step development of chronic myeloid leukemia. Our data suggest that identification of somatic variants at diagnosis facilitates stratification into imatinib responders/non-responders, thereby allowing earlier use of second-generation tyrosine kinase inhibitors, which, in turn, may overcome the negative impact of such variants on disease progression.

**Introduction**

Although tyrosine kinase inhibitors (TKI) have profoundly changed the prognosis of chronic-phase chronic myeloid leukemia (CML-CP), some 10-15% of affected individuals do not respond and need other therapies. Four TKI are approved for use in newly-diagnosed CML, including imatinib and the second-generation TKI...
(2G-TKI) dasatinib, nilotinib and bosutinib. In randomized studies, 2G-TKI induce faster, deeper molecular responses than imatinib with a lower risk of progression to blast phase but no convincing evidence of better survival. Early molecular responses analyzed by reverse transcription-quantitative PCR (RT-qPCR) at 3, 6 and 12 months are widely used to direct therapy. Individuals failing to achieve these landmarks can be switched to a different TKI but there are no convincing data that the change of therapy changes their outcome. Because most progressions occur within two years of starting TKI-therapy, early identification of those at high risk of progression would facilitate more rapid decision-making regarding more aggressive therapy.

In other hematologic neoplasms, the presence of somatically mutated genes involving signaling, RNA splicing, transcriptional control, DNA damage response and epigenetic regulation is correlated with survival and sometimes drives treatment. However, no ‘mutator’ phenotype associated with clinical outcome has been described in CML, particularly in chronic phase, as earlier studies focused on blastic phase. Four recent whole-exome sequencing (WES) and/or whole-transcriptome (RNA-Seq) studies identified somatic variants in 24, 19, 13 and 65 exomes (or transcriptomes in one study) from newly-diagnosed CML-CP and in chronic phase and blast transformation. Studies using deeper, targeted sequencing described genetic variants in CML-CP especially in genes associated with epigenetic regulation. These variants were also present in Philadelphia (Ph)-chromosome-negative cells, suggesting that they antedated the BCR/ABL translocation event as an early step in developing CML. Finally, variants in hematologically normal elderly individuals are described, although the risk of conversion from age-related clonal hematopoiesis (ARCH) to leukemia was modest.

We recently identified differences in genome-wide DNA methylation patterns in CD34+ cells of CML-CP and in chronic phase compared with normal subjects. These differences were not observed at the time of complete cytogenetic remission, suggesting a role for epigenetic regulation in CML-CP. In this study, we interrogated genetic variants in pre-treatment CML-CP using a targeted panel of genes enriched in epigenetic modifiers and Ion Torrent Personal-Genome-Machine (PGM) next-generation sequencing (NGS). We then assessed the predictive value of these variants for diversity therapy outcomes in the context of different TKI therapies.

Methods

Study participants

We studied 124 untreated subjects with CML-CP and 14 normal individuals as negative controls, selecting CD34+ cells. Subjects were non-consecutive and selected for optimal response (n=69) or non-response (n=55) to TKI-therapy. Evolution of the somatic variants after treatment, was investigated using CD34+ (n=11) and whole-blood cells (n=4) from subjects in major molecular remission (MR3; 3-log reduction in BCR/ABL transcripts from baseline) and after progression to blast phase, respectively. We also used CD34+ cells from three subjects with somatic variants at diagnosis, to establish liquid cultures with in vitro TKI-treatment, as a biological validation of our findings (Online Supplementary Methods and Results, and Online Supplementary Figure S4). All subjects gave written informed consent and the local research Ethics Committee approved the study.

Definitions

Response was defined as BCR/ABL transcript levels according to the International Scale (IS) ≤0.1%, ≤1% and ≤0.1% at 3, 6 and 12 months after initiating TKI-therapy, respectively. Because some subjects did not have real-time qualitative polymerase chain reaction (RT-qPCR) sampling at pre-specified times during the first year, responders were required to have at least two of these results available. Non-responders satisfied the European LeukemiaNet criteria for failure.

DNA preparation

In 103 subjects, paired leukemia/control DNA was analyzed; in 44 and 59 subjects control DNA was obtained from diagnostic T cells expanded in vitro and from samples in MR4-molecular remission (4-log reduction from baseline), respectively. We measured BCR/ABL/ABL in 13 T-cell samples and confirmed very low expression. In seven subjects with somatic variants at diagnosis we compared detection of somatic variants in whole-blood and CD34+ cell populations (Online Supplementary Methods and Results).
Table 1. Seventy-one epigenetic modifiers grouped according to gene function.

| 71 epigenetic modifying genes | DNA methylation | Histone modification at lysine residues | Histone modification at arginine residues | Histone ubiquitination | Histone phosphorylation |
|------------------------------|-----------------|---------------------------------------|------------------------------------------|------------------------|------------------------|
| DNA methylation              | DNMT1           | DNMT2                                 | DNMT3A                                   | DNMT3B                 | DNMT3L                 |
| TET1                         | TET2            | IDH1                                  | IDH2                                     | IDH3B                  |                        |
| Histone methylation at lysine residues | EZH2          | SUZ12                                 | EED                                      | ASL1                   | SETD2                  |
| KMT2A (MLL)                  | KMT2B (MLL4)    | KMT2D (MLL2)                          | KMT2E (MLL5)                             | SETD1A                 |                        |
| EHHMT2                       | KMT5A (SETD8)   | SUZ39H1                               | SMYD2                                    | SETD1B                 |                        |
| DOT1L                        | AFF1            | SETD8                                 | SETD2                                    | SETD3                  |                        |
| KMT5B (SUV420H1)             |                 |                                       |                                          |                        |                        |
| Histone modification at arginine residues | PRMT2          | PRMT3                                 | PRMT5                                    | PRMT9 (PRMT10)         |                        |
| KDM1B                        | KDM5B           | KDM1A (LSD1)                          | KDM6A                                    | KDM4C                  |                        |
| JMJD6                        | JMJD8           |                                       |                                          |                        |                        |
| Histone acetylation          | CREBBP          | EP300                                 | KAT2A                                    | KAT2A                  | HAT1                   |
| NCOA2                        |                 | ATF2                                  |                                          |                        |                        |
| Histone deacetylation        | HDAC1           | HDAC2                                 | HDAC8                                    | HDAC7                  | SIRT1                  |
| SIRT2                        |                 |                                       |                                          |                        |                        |
| Histone ubiquitination       | RNF2 (RING18)   | BAPI                                  | BMII                                     | AURKB                  |                        |
| Other genes (transcription factors, signaling molecules) | RUNX1           | WT1                                   | IKZF1                                    | SET (2PP2A)            | SETBP1                 |
| NPM1                         | PHF5            | BCOR                                  | BRD1                                     | CALR                   |                        |

see Online Supplementary Methods). Logistic regression and Fisher’s exact test were used to calculate associations of variables and probability of greater-than-random overlaps, respectively. P<0.05 were considered significant.

Results

Subjects

Demographic, clinical and molecular data of the subjects are shown in Table 2. Sixty-two, 55, 4 and 3 subjects started treatment with imatinib, dasatinib, nilotinib and bosutinib. Subjects treated initially with 2G-TKI did so because they were enrolled in clinical trials. Among subjects treated with initial imatinib, 33 were responders (R) and 29 non-responders (NR). Of those who received 2G-TKI, 36 and 26 were classified as responders and non-responders, respectively.

Sequencing data

A mean depth of coverage of 302x (range: 85x-1088x) was achieved yielding a limit of detection of 4% variant allele frequency (VAF). After filtering (Online Supplementary Figure S3), 142 non-synonymous variants remained (in 51 of 71 genes), of which 45 were somatically acquired variants. Of these, 40 were classified as somatic if they were present only in leukemia DNA, and three as pre-leukemia (before BCR/ABL1) if VAF in leukemia DNA was >20% greater than VAF in control DNA. The remaining 99 were present at similar VAF (about 50%) in leukemia and control/matched DNA and were most likely germline variants.

Incidence of somatic variants in chronic-phase chronic myeloid leukemia

Forty-three somatic variants were observed 49 times (5 variants >1) in 37 of 124 subjects [30% (95% CI: 23, 39%)], including 18 of 62 subjects [29%, (95% CI:20, 43%)] in the imatinib cohort and 19 of 62 subjects [31% (95% CI:21, 45%)] in the 2G-TKI cohort. The incidence of subjects with at least one somatic variant (1 or ≥2 grouped together) was higher in non-responders [22 of 55; 40% (95% CI:28, 53%)] compared with responders from both imatinib- and 2G-TKI-treated cohorts [15 of 69; 22% (95% CI: 14, 38%); P=0.031] (Figure 1A). More than one somatic variant in the same or different genes was seen in three subjects in the imatinib cohort and in six of the 2G-TKI cohort and occurred more often in non-responders (Figure 1A).

Most of the 49 variants (26 missense, 14 nonsense, 3 splice-site, 5 frameshift insertions and 1 non-frameshift deletion) identified in 21 of 71 genes were in non-responders (Figure 1B and Online Supplementary Table S4). The most frequently altered genes were ASXL1 (n=10 in 9 subjects), IKZF1 (n=6 in 4 subjects), DNMT3A and CREBBP (n=4), KMT2D (MLL2), KMT2E (MLL5), and EP300 (n=3) (Figure 1C). VAF were 4.6-64% and for 28 of 49 variants were <20% (Figure 1B). In three subjects, two variants occurred in the same gene. Fifteen of 45 variants (33%) are
currently listed in the COSMIC v86 database (Figure 1B).

None of these variants were found in the paired-control DNA, apart from three variants, classed as pre-leukemia and detected in the diagnostic CD34+ cells at a markedly higher VAF than in control DNA. Two variants occurred in DNMT3A: a missense (p.Arg899Gly) and a splice-site (c.1123-2A>G) variant with VAF of 50% and 48%, respectively, at diagnosis, but reduced to 22% and 6% in paired remission samples collected at 55 and 47 months from starting therapy, respectively (with BCR/ABL1 of 0.0012% and 0.0009%, respectively). An ASXL1 nonsense variant (p.Tyr591*) occurred at 52% and 16% in leukemia and paired T cells, respectively.

Evolution of somatic variants after imatinib treatment

We next examined somatic variants in follow-up samples from the imatinib-treated subjects. In four subjects with somatic variants detected in CD34+ cells at diagnosis who progressed to blast phase (BP) we compared paired samples from whole-blood cells (as opposed to CD34+ cells) at diagnosis and in BP (median follow up 25 months). The somatic variants identified in diagnostic CD34+ cells (ASXL1 p.Gln780*, ASXL1 p.Gln594fs) at high level (VAF 51% and 40%, respectively) were also found in diagnostic whole-blood cells at similar VAF. On the contrary, those identified in diagnostic CD34+ cells (IKZF1 p.Arg184Trp and IKZF1 p.Arg213*; IKZF1 p.Tyr348*) at low levels (VAF 5.9%, and 6.9%/4.9%) were undetectable in diagnostic whole-blood cells. In one case of low-level variants (IKZF1 p.Arg213*; IKZF1 p.Tyr348*) identified in diagnostic CD34+ cells, these were undetectable in whole-blood samples from CP and BP; however, the clone with the low-level variant IKZF1 p.Arg184Trp, expanded during progression (from undetectable to 17% VAF). As for the high-level variants, in one case the variant ASXL1 p.Gln594fs remained at similar levels (from 40% to 43%) in both CP and BP, whereas in the second case, the variant ASXL1 p.Gln780* dropped to lower, but still high, levels in BP (from 45% to 27%) (Online Supplementary Table S2).

Of the 11 patients who achieved MR3 and in whom we had paired samples, only three had somatic variants at diagnosis (KMT2D p.Gln3946Leu, PRMT9 p.Phe591fs, IKZF1 p.Arg184Trp) with VAF of 54%, 42% and 16%, respectively. These variants were undetectable in the follow-up samples collected at a median of 20 months from starting therapy. In two subjects achieving MR3 we identified variants (DNMT3A p.Cys497Tyr and EHMT2 p.Pro196Gln) in follow-up samples in MR3 with median follow-up 18 months with VAF of 5% and 10%. These variants were undetectable at diagnosis, so the possibility arises that they are due to clonal evolution in Philadelphia negative clone (Online Supplementary Table S2).

The same DNMT3A variant was identified at 141 months from diagnosis, when the patient was in durable MR4 and the sample was collected as a control. The presence of eight missense variants identified only in the paired deep remission samples (with median follow up 72 months) was associated with increased age (P=0.029) (Online Supplementary Results and Online Supplementary Table S3).

Somatic variants and outcomes in subjects treated with imatinib

Imatinib-treated subjects with somatic variants had lower rates of 5-year MR3 compared to subjects without variants at diagnosis (47% [95%CI: 11, 68%] vs. 94% [67, 99%]; P=0.048) (Figure 2A). Similarly they had lower rates of...
Table 2. Demographics and clinical/molecular characteristics of subjects with chronic phase-chronic myeloid leukemia at diagnosis.

| Characteristic                                                                 | IM-treated subjects | 2G-TKI-treated subjects |
|-------------------------------------------------------------------------------|---------------------|-------------------------|
| N. of subjects (%)                                                            | 62 (50)             | 62 (50)                 |
| Subtype of frontline TKI therapy                                              | 62/0/0/0            | 0/55/4/3                |
| (IM/DAS/NIL/BOS) (%)                                                           | (100/0/0/0)         | (0/89/6/5)              |
| Responders to TKI therapy (according to the 2013 ELN criteria of 2013)* (%)   | 33/0/0/0            | 0/1/4/1                 |
| (% of patients per subtype of frontline TKI therapy)                          | (52/0/0/0)          | (0/56/100/3)            |
| Non-responders to TKI therapy (according to the ELN criteria of 2013) (%)      | 29/0/0/0            | 0/2/0/2                 |
| (% of patients per subtype of frontline TKI therapy)                          | (47/0/0/0)          | (0/44/0/67)             |
| Age (years) median (range)                                                    | 50 (20.8-80.1)      | 49.9 (20.2-85)          |
| Gender (male/female)                                                          | 38/24               | 40/22                   |
| (%)                                                                           | (61/39)             | (65/35)                 |
| Sokal risk group at diagnosis (n=122)                                         | N=62                | N=60                    |
| Low/intermediate/high risk (%)                                                | 23/20/19            | 19/21/20                |
| (%)                                                                           | (37/32/31)          | (32/35/33)              |
| ELTS risk group at diagnosis (n=122)                                          | N=62                | N=60                    |
| Low/intermediate/high risk (%)                                                | 32/18/12            | 29/23/8                 |
| (%)                                                                           | (52/29/19)          | (48/38/14)              |
| BCRABL1/ABL1 transcript before TKI-therapy median (range)                     | 20.1 (11.9-43.4)    | 22.9 (7.1-51.8)         |
| BCRABL1 Transcript type                                                       | 23/32/7/1/0         | 23/31/7/0/1             |
| e13a2e14a2/e13a2e14a2/e13a2e14a2/e13a3 (%)                                     | (36/51/11/2/0)      | (37/50/11/0/2)          |
| ACA besides Ph chromosome (n=61)*                                              | n=41                | n=20                    |
| classical Ph/additional ACA in Ph/variant Ph/Ph* (%)                           | 34/2/4/1            | 17/2/0                  |
| (%)                                                                           | (83/5/10/2)         | (85/10/50)              |
| Subsequent BCRABL1 KD mutations (n=58)                                        | 6/1                 |                          |
| (%)                                                                           | (10)                | (2)                     |
| Follow-up duration after TKI therapy (months) median (range)                  | 95.6 (18-196.1)     | 69.9 (9.5-112.8)        |

*Real-time qualitative polymerase chain reaction ≥10%, ≥1%, ≥0.1% (International Score) at 3, 6 and 12 months, respectively. Additional cytogenetic abnormalities (ACA) data available only for subjects treated at Hammersmith Hospital, London, UK. BOS: bosutinib; DAS: dasatinib; ELN: European LeukemiaNet; ELTS: European Treatment and Outcome Study long-term survival; IM: imatinib; KD: kinase domain; NIL: nilotinib; Ph: Philadelphia; TKI: tyrosine kinase inhibitors; 2G: second-generation.

of 8-year EFS, PFS and CML-related survival compared with the non-variant subjects (28% (13, 59%) vs. 68% (55, 85%); P=0.003 for EFS; 61% (42, 88%) vs. 85% (74, 97%); P=0.025 for PFS; 58% (37, 92%) vs. 84% (73, 97%); P=0.039 for CML-related survival) (Figure 2 B-D). Somatic variants and Sokal score were independently predictive for PFS and CML-related survival whereas only somatic variants predicted EFS (Table 3). When compared with European Treatment and Outcome Study (EUTOS) long-term survival (ELTS) score, only ELTS score predicted MR3, PFS and CML-related survival while both ELTS and somatic variants predicted EFS (Table 3). Subjects with a low Sokal score or a low ELTS score and somatic variants had worse EFS (P=0.015 for Sokal; P=0.021 for ELTS) and PFS (P=0.040 for Sokal; P=0.081 for ELTS) than those without somatic variants (Online Supplementary Figure S4A and S8). The trends towards poorer outcomes in subjects with somatic variants was also evident for intermediate- and high-Sokal/ELTS subjects. Somatic variants were a more accurate predictor of outcomes compared with BCRABL1 transcript type (Table 3) and similarly compared with BCRABL1/ABL1 transcript levels before TKI-therapy, age and gender (data not shown).

Molecular response within the 1st year (calculated by BCRABL1/ABL1 transcripts at 3, 6, 12 months) (P<0.001) and somatic variants (P=0.044) were independently predictive for EFS. Non-responders without somatic variants had a better EFS (23% (9, 59%)) compared to non-responders with somatic variants at 0% (Online Supplementary Figure S5A). No association was detected between somatic variants in epigenetic modifiers at diagnosis and subsequent BCRABL1 kinase domain (KD) mutations (P=0.81).

In all the above analyses, we included the subject with pre-leukemia variant. Excluding this subject did not alter our conclusions (Online Supplementary Table S4A).

**Somatic variants and outcomes in the second-generation tyrosine kinase inhibitor cohort**

There was no association between the presence of somatic variants and the 5-year rates of MRS and 6-year
rates of EFS, PFS and CML-related survival in 2G-TKI-treated subjects: 90% (39, 98%) vs. 100%, P = 0.25 for MR3; 61% (41, 91%) vs. 75% (62, 92%), P = 0.32 for EFS; 82% (66, 99%) vs. 89% (80, 99%), P = 0.46 for PFS; 81% (63, 99%) vs. 93% (85, 99%), P = 0.29 for CML-related survival in variant versus non-variant subjects (Figure 3). We found no association of the somatic variants with outcomes in multivariate analysis with Sokal score, ELTS score or type of BCRABL1 transcript (Online Supplementary Table S4A). When combined with molecular response within the 1st year, only BCRABL1/ABL1 transcripts within the 1st year but not somatic variants were predictive for MR3 and EFS. Non-responders with and without variants had no significant difference in rates of EFS: 38% (16, 87%) vs. 48% (27, 85%), respectively; P = 0.69 (Online Supplementary Figure S5B).

Excluding the two subjects with somatic pre-leukemia variants did not alter our conclusions (Online Supplementary Table S4B).

To reduce potential heterogeneity conferred by different 2G-TKI, we next investigated the association of somatic variants with outcomes in the dasatinib-treated cohort only (n=55). No association with outcomes was detected (Online Supplementary Table S5B).

### Germline variants in chronic-phase chronic myeloid leukemia subjects

We identified 99 missense variants classified as rare germline (with frequency <1% in the general population) (Online Supplementary Results, Online Supplementary Table S6 and Online Supplementary Figure S6), which did not affect clinical outcome in either cohort (data not shown).
Functional associations of variants with chronic myeloid leukemia

To further investigate any association of the altered genes with CML a protein-protein-interaction (PPI) network of P210\(_{BCRABL1}\) with the 21 coded proteins affected by the somatic variants was constructed. Twenty-one of 23 proteins were parts of the network with six proteins including those encoded by ASXL1, IKZF1, EP300 and RUNXt interacting directly with P210\(_{BCRABL1}\) suggesting a functional association (Online Supplementary Table S7 and Online Supplementary Figure S7).

Finally, we studied whether the presence of a somatic variant in epigenetic modifiers influenced the DNA methylation signature in the imatinib cohort. Hierarchical clustering based on 1,028 differentially methylated positions (DMP) (see criteria in the Online Supplementary Results) clearly separated the 12 variant and 30 non-variant CMP-CP subjects (Online Supplementary Figure S8). Functional annotation of DMP showed the imatinib pharmacokinetics/pharmacodynamics pathway being among the top ten hits of over-represented pathways (F=0.0016) (Online Supplementary Table S8).

Discussion

The successful introduction of TKI in CML therapy has resulted in an excellent outcome for approximately 90% of individuals, who have a life expectancy approaching that of unaffected individuals. However, the remaining 10% should ideally be identified at diagnosis and offered more potent TKI immediately or early allogeneic-stem cell transplantation if they demonstrate TKI resistance. The most widely used biomarker for outcomes in CML-CP, namely \(BCRABL1\) transcript levels after three months on TKI (\(BCRABL1\) ≤10%), identifies a cohort with an excellent prognosis. However, patients with \(BCRABL1\) >10% at three months may or may not respond to 2G-TKI. Our aim was to investigate associations between somatic variants in epigenetic modifiers and response to imatinib and 2G-TKI given from diagnosis of CML-CP.

Others have explored the predictive value of a number of different biomarkers at the time of diagnosis including gene expression, protein expression, DNA methylation, miRNA expression, and SNP analysis, but none has proved sufficiently accurate and precise for clinical decision-making. We previously identified different genome-wide DNA methylation and gene expression patterns between CML-CP and normal individuals, but this did not correlate clearly with TKI response. However, this prompted us to investigate the role of genetic variants in epigenetic modifiers in greater detail.

We used targeted amplicon sequencing to detect genetic variants that might correlate with TKI response. To optimize the opportunity to assess differences in genetic variants between responders and non-responders to TKI-ther-
apy, we selected a cohort enriched for non-responders (approx. 45% in each of the imatinib and 2G-TKI cohorts), which explains why the outcomes of patients in this study are inferior to those seen in unselected subjects. We believed our strategy of deliberately choosing equal proportions of responders and non-responders would maximize our chances of detecting a biomarker, if such existed. Moreover, the use of CD34 + progenitor cells from patients at diagnosis, the population in which the clone capable of progression resides, would maximize our chance of yielding meaningful results and reduce heterogeneity. We are aware that, for clinical utilization, our data should be validated in whole-blood samples and in a larger cohort of newly-diagnosed individuals with CML.

Progression to blast crisis in CML is often attributed to underlying ‘genetic instability’, in part because this is increased in stem/progenitor cells from individuals with CML (especially in blast phase) compared to normal individuals. BCRABL1-induced genomic aberrations and/or BCRABL1-independent pre-existing genetic lesions may then function as “amplifiers” of a genetically unstable phenotype and thereby predispose to blast transformation. However, our results, and those of others, suggest that the ‘mutator’ phenotype of CML is moderate compared with other cancers, particularly in chronic phase. We included genes in which somatic variants have been identified by WES including ASXL1, RUNXI, IKZF1, KDM2D, BCOR, IDH1/2, PHF6, TET2, KDM1A, KAT6A, SETBP1, SETD2 and found somatic variants in approximately 30% of newly-diagnosed individuals with CML-CP, similar to previous reports. We identified overlap with other studies of 14-30% but feel this can be explained, at least in part, by our focus on epigenetic regulators that resulted in the omission of some genes that have frequently been found mutated in CML, such as TP53, and also by the readily available technology of targeted NGS at the start of this project. Concordance with other studies regarding specific variants was also limited, while 15 somatic variants in our study were COSMIC, mostly identified in other hematologic neoplasms. Because most of the variants we identified affect epigenetic modifiers and genome-wide DNA methylation changes are reported in CML, a better understanding of the role of such epigenetic alteration should be complemented by genome-wide landscape of histone marks.

In this study, we report for the first time a correlation between somatic variants and survival of individuals with CML-CP. Others have reported the presence of somatic variants but have so far been unable to directly associate these with clinical outcomes or limited the assessment to achievement of major molecular remission. One study found that a subset of variants (16 of 73) affecting epigenetic modifiers had an adverse impact on cytogenetic/molecular responses. Because of our use of extreme responders, and because of the availability of prolonged follow up, our cohort contains patients who had experienced disease progression, with 20 patients developing blast crisis over the period of observation. Eleven of these (8 of 13 and 3 of 7 on imatinib and 2G-TKI, respectively) had somatic variants. Absence of variants in the remaining subjects does not exclude their presence in genes absent from our panel, who could have structural variants/copy number variations such as IKZF1 deletions that would be identified by WES/whole-genome sequencing (WGS).

Although the two patient cohorts (treated with imatinib or 2G-TKI) were similar in their clinical characteristics at diagnosis and in the proportion of responders and non-

Table 3. Multivariate analysis of somatic variants with Sokal score, European Treatment and Outcome Study long-term survival (ELTS) score and BCRABL1 transcript type in the imatinib cohort for cumulative incidence of 3-log reduction in BCRABL1 transcripts from baseline (MR3) (by the Fine-Gray model) and probabilities of event-free survival (EFS), progression-free survival (PFS) and chronic myeloid leukemia (CML)-related survival (by the Cox proportional hazard regression model).

|                          | MR3 cumulative incidence | EFS probability | PFS probability | CML-related survival probability |
|--------------------------|--------------------------|-----------------|-----------------|----------------------------------|
| **MVA**                  |                          |                 |                 |                                  |
| Somatic variants at Dx   | HR (95% CI)              | 0.43 (0.19, 1.02) | 2.92 (1.32, 6.49) | 3.15 (1.06, 9.40) | 3.09 (1.01, 9.69) |
|                          | p                        | 0.054           | 0.008 **        | 0.040 *                        | 0.049 *               |
| Sokal score at Dx        | HR (95% CI)              | 0.71 (0.48, 1.04) | 1.45 (0.88, 2.35) | 2.31 (1.08, 4.95) | 2.90 (1.27, 6.42) |
|                          | p                        | 0.080           | 0.14            | 0.032 *                        | 0.011 *               |
| **MVA**                  |                          |                 |                 |                                  |
| Somatic variants at Dx   | HR (95% CI)              | 0.48 (0.20, 1.15) | 2.53 (1.12, 5.71) | 2.57 (0.85, 7.81) | 2.60 (0.82, 8.25) |
|                          | p                        | 0.098           | 0.026 *         | 0.095                          | 0.11                  |
| ELTS score at Dx         | HR (95% CI)              | 0.53 (0.30, 0.93) | 1.77 (1.08, 2.90) | 2.10 (1.05, 4.22) | 2.73 (1.31, 5.67) |
|                          | p                        | 0.028 *         | 0.025 *         | 0.037 *                        | 0.007 **              |
| **MVA**                  |                          |                 |                 |                                  |
| Somatic variants at Dx   | HR (95% CI)              | 0.43 (0.19, 0.99) | 3.21 (1.43, 7.19) | 3.23 (1.87, 9.67) | 3.30 (1.03, 10.5)  |
|                          | p                        | 0.049 *         | 0.005 **        | 0.036 *                        | 0.044                 |
| BCRABL1 transcript type at Dx | HR (95% CI) | 0.90 (0.58, 1.39) | 1.15 (0.65, 2.05) | 0.97 (0.45, 2.11) | 1.2 (0.53, 2.88)  |

CI: 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. P < 0.05; ** P < 0.01.
responders, somatic variants impacted clinical outcome only in those treated with imatinib. This mirrors clinical experience in which the 2G-TKI can result in deep and durable responses in patients who were resistant to imatinib, and induce these responses more rapidly and in a larger proportion of patients when used as first-line therapy. One possible explanation is that the increased potency of the 2G-TKI results in the rapid eradication of the mutated clones and thus overcome the adverse prognostic impact. This hypothesis is supported by the fact that 90% of 2G-TKI-treated subjects with somatic variants achieved MR3 compared with <50% of subjects treated with imatinib. Data from in vitro liquid cultures corroborated our original findings, since at least cells containing some variants were eradicated on treatment with dasatinib but persisted or were eradicated more slowly on treatment with imatinib. We also have additional indirect evidence that our findings may predict response to imatinib. First, we were able to show distinct methylation patterns between imatinib-treated subjects with and without variants, and second, the PPI network indicates close interactions between p210BCRABL1 and proteins affected by somatic variants.

Somatic variants had better predictive power for outcomes than other widely used predictive variables such as the Sokal score4,14,47 and BCRABL1 transcript type48,49 but less compared with the newly defined ELTS score.10 Previous clinical risk scores identified individuals at high risk of early progression but were less successful in predicting poor-risk subjects in the low/intermediate cohorts. Combining a clinical risk score with somatic variants is a potentially promising approach, and may be particularly valuable in those with low Sokal/ELTS scores, who are heavily influenced by age, such that a young patient with inherently poor prognosis may be inappropriately classified as non-high risk.

By using paired leukemia and control DNA, we found most somatic variants were part of a Ph+ clone. Therapy with imatinib eradicated the Ph+ clones with somatic variants in responders achieving MR3 but not in non-responders who progressed. These persistent somatic variants may be implicated in disease evolution or may be passenger mutations and require confirmation in larger cohorts.

The presence of pre-leukemia variants was implicated in three subjects. DNMT3A, ASXL1 variants were found in Ph+ and, albeit at lower levels, in Ph- cells. This suggests that these variants preceded the acquisition of BCRABL1. Variants in DNMT3A, ASXL1 and TET2 are thought to be latent initiating mutations33 and are described in CML.37,28 However, ASXL1 mutations have also been found in children and young adults with CML.28 Rare germline variants had no impact on clinical outcomes in either of the imatinib or 2G-TKI cohorts, contrary to other reports.45 Variants detected only in Ph+ cells from subjects aged >60 years in molecular remission may have developed during therapy, or have been present at diagnosis and unmasked in remission.

In summary, we showed potentially pathogenic somatic variants of epigenetic modifiers are common in CML-CP at diagnosis, and when combined with other risk factors may be promising predictive biomarkers determining which is the best TKI for each individual.

Our study has some limitations, the most important of which are small sample size (although it is the largest study to date assessing the effect of genetic variants on survival), the potential exaggeration of the effect size due to the selection of the extreme responders/non-responders, the limited number of target genes, and the retrospective nature of our observations. Furthermore, we were unable to assess the impact of additional chromosomal abnormalities (ACA)46 due to the absence of cytogenetic data at diagnosis. The intriguing question remains as to whether any of the variants identified has more or less impact on prognosis, but because of the small numbers of each variant we were unable to explore this in more detail.

We now wish to see our panel enriched by the addition of genes found to be altered in CML-CP by targeted, exome or WGS, and validated in larger, unselected, consecutive cohorts of individuals with CML. Our findings, if confirmed in a prospective study, could assist in distinguishing individuals who would benefit starting therapy with a more potent 2G-TKI rather than imatinib.

**Funding**

JFA is a NIHR Senior Investigator and we are grateful for the support from the Imperial College NIHR Biomedical Research Centre.

**Acknowledgments**

We would also like to thank the Molecular Pathology Laboratory and the John Goldman Centre for Cellular Therapy for providing us with guanidinium thiocyanate (GTC) lysates and cells from CML and normal subjects’ samples. We are grateful to individuals with CML for participating in this study.

**References**

1. Innes AJ, Milojkovic D, Apperley JF. Allogeneic transplantation for CML in the TKI era: striking the right balance. Nat Rev Clin Oncol. 2016;13(2):79-91.
2. Brummendorf TH, Cortes JE, de Souza CA, et al. Bosutinib versus imatinib in newly diagnosed chronic-phase chronic myeloid leukemia: results from the 24-month follow-up of the BELA trial. Br J Haematol. 2015;168(1):69-81.
3. Cortes JE, Saglio G, Kantarjian HM, et al. Final 5-Year Study Results of DASISION: The Dasatinib Versus Imatinib Study in Treatment-Naive Chronic Myeloid Leukemia Patients Trial. J Clin Oncol. 2016;34(20):2333-2340.
4. Hochhaus A, Saglio G, Hughes TP, et al. Long-term benefits and risks of frontline nilotinib vs imatinib for chronic myeloid leukemia in chronic phase: 5-year update of the randomized ENESTnd trial. Leukemia. 2016;30(5):1044-1054.
5. Aichberger KJ, Herrdhofer S, Schernthaner GH, et al. Progressive peripheral arterial occlusive disease and other vascular events during nilotinib therapy in CML. Am J Hematol. 2011;86(7):533-539.
6. Montani D, Bergot E, Gathe J, et al. Pulmonary arterial hypertension in patients treated by dasatinib. Circulation. 2012;125(17):2128-2137.
7. Hasford J, Baccarani M, Hoffmann V, et al. Predicting complete cytogenetic response and subsequent progression-free survival in 2060 patients with CML on imatinib treatment: the EUTOS score. Blood. 2011;118(3):696-692.
8. Hasford J, Pﬁrrmann M, Hehlmann R, et al. A new prognostic score for survival of patients with chronic myeloid leukemia.
