We cannot rule out the possibility of occult disease influencing some of our findings, particularly for indolent malignancies such as CLL. Indeed, we observed strong sCD23 associations for CLL/SLL diagnosed closer in time to phlebotomy, which may be at least partly disease induced, as circulating sCD23 in CLL patients has been associated with tumor burden.[14] However, our sCD23 findings for the more aggressive DLBCL are consistent with etiologic effects, given the consistency in associations across follow-up periods and especially the strong risk present even among cases diagnosed 15–23 years after phlebotomy. The weaker NHL associations for sCD27 and sTNFR2 $\geqslant 8$ years postphlebotomy may reflect disease-induced effects, odds ratio attenuation due to a decline in representativeness of one-time analyte measurements as surrogates for usual levels, or both.

The generalizability of these findings is questionable, given that the study population consisted entirely of male Finnish heavy smokers. However, similar NHL and subtype associations with sCD23 and sCD30 have previously been reported, although with fewer years of follow-up, in population-based cohorts that include women (including one all-female cohort) and non-smokers.[3,4,6]

Our findings raise the possibility that serum sCD23 and sCD30 may hold promise as surrogate end points for cross-sectional studies of suspected B-cell-stimulative risk factors. It is also possible, although speculative, that these markers may have future clinical value in developing risk prediction models. In particular, given the strong association with the CLL/SLL risk observed for elevated sCD23 in serum collected $\leqslant 15$ years before diagnosis, there is a need to investigate whether the elevated levels of this marker are predictive of future progression to CLL among patients with monoclonal B-cell lymphocytosis, an asymptomatic precursor condition to this malignancy.[15] Studies addressing these questions, particularly using serially collected specimens, are warranted.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the Intramural Research Program of the National Institutes of Health and the National Cancer Institute.

MP Purdue[1,2], Q Lan[1], TJ Kemp[3], A Hildesheim[1], SJ Weinstein[1], JN Hofmann[1], J Virtamo[4], D Albanes[1], LA Pinto[3] and N Rothman[1]
[1]*Division of Cancer Epidemiology and Genetics, Department of Health and Human Services, National Cancer Institute, NIH, Bethesda, MD, USA;*
[2]*Ontario Institute for Cancer Research, Toronto, ON, Canada;*
[3]*HPV Immunology Laboratory, Leidos Biomedical Incorporated, Frederick National Laboratory for Cancer Research, Frederick, MD, USA and*
[4]*Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland*
E-mail: mark.purdue@ontariohealthstudy.ca

REFERENCES

1 Hartge P, Wang SS, Bracci PM, Devesa SS, Holly EA. Non-Hodgkin lymphoma. Schottenfeld D, Fraumeni JF (eds). *Cancer Epidemiology and Prevention*. 3rd edn. Oxford University Press: New York, NY, USA, 2006; 898–918.
2 Gu Y, Shore RE, Arslan AA, Koenig KL, Liu M, Ibrahim S *et al.* Circulating cytokines and risk of B-cell non-Hodgkin lymphoma: a prospective study. *Cancer Causes Control* 2010; **21**: 1323–1333.
3 Purdue MP, Lan Q, Martinez-Maza O, Oken MM, Hocking W, Huang WY *et al.* A prospective study of serum soluble CD30 concentration and risk of non-Hodgkin lymphoma. *Blood* 2009; **114**: 2730–2732.
4 Vermeulen R, Hosnijeh FS, Portengen L, Krogh V, Palli D, Panico S *et al.* Circulating soluble CD30 and future risk of lymphoma; evidence from two prospective studies in the general population. *Cancer Epidemiol Biomarkers Prev* 2011; **20**: 1925–1927.
5 Purdue MP, Lan Q, Bagni R, Hocking WG, Baris D, Reding DJ *et al.* Prediagnostic serum levels of cytokines and other immune markers and risk of non-hodgkin lymphoma. *Cancer Res* 2011; **71**: 4898–4907.
6 De Roos AJ, Mirick DK, Edlefsen KL, LaCroix AZ, Kopecky KJ, Madeleine MM *et al.* Markers of B-cell activation in relation to risk of non-Hodgkin lymphoma. *Cancer Res* 2012; **72**: 4733–4743.
7 Purdue MP, Hofmann JN, Kemp TJ, Chaturvedi AK, Lan Q, Park JH *et al.* A prospective study of 67 serum immune and inflammation markers and risk of non-Hodgkin lymphoma. *Blood* 2013; **122**: 951–957.
8 The ATBC Cancer Prevention Study. The Alpha-Tocopherol, Beta-Carotene Lung Cancer Prevention Study: design, methods, participant characteristics, and compliance. The ATBC Cancer Prevention Study Group. *Ann Epidemiol* 1994; **4**: 1–10.
9 Vendrame E, Martinez-Maza O. Assessment of pre-diagnosis biomarkers of immune activation and inflammation: insights on the etiology of lymphoma. *J Proteome Res* 2011; **10**: 113–119.
10 Gordon J. CD23 and B cell activation. *Clin Exp Allergy* 1992; **22**: 199–204.
11 Fournier S, Rubio M, Delespesse G, Sarfati M. Role for low-affinity receptor for IgE (CD23) in normal and leukemic B-cell proliferation. *Blood* 1994; **84**: 1881–1886.
12 Del PG, De CM, Almerigogna F, Daniel CK, D'Elios MM, Zancuoghi G *et al.* Preferential expression of CD30 by human CD4+ T cells producing Th2-type cytokines. *FASEB J* 1995; **9**: 81–86.
13 Croft M. Co-stimulatory members of the TNFR family: keys to effective T-cell immunity? *Nat Rev Immunol* 2003; **3**: 609–620.
14 Reinisch W, Willheim M, Hilgarth M, Gasché C, Mader R, Szepfalusi S *et al.* Soluble CD23 reliably reflects disease activity in B-cell chronic lymphocytic leukemia. *J Clin Oncol* 1994; **12**: 2146–2152.
15 Mowery YM, Lanasa MC. Clinical aspects of monoclonal B-cell lymphocytosis. *Cancer Control* 2012; **19**: 8–17.

Supplementary Information accompanies this paper on the Leukemia website (http://www.nature.com/leu)

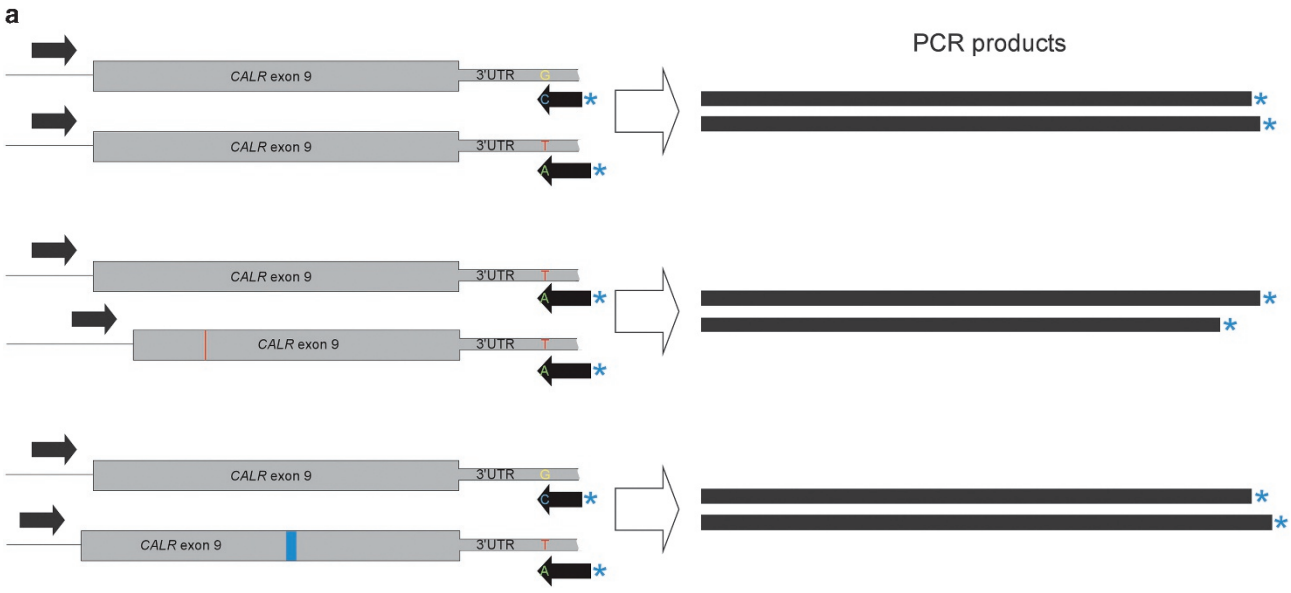
OPEN

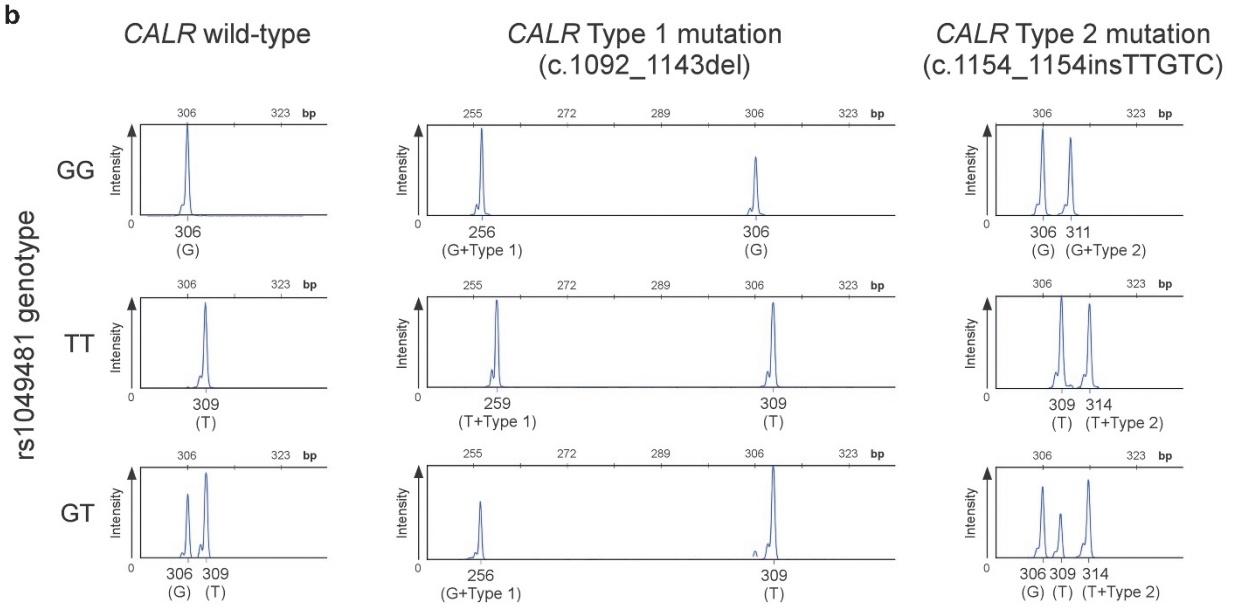
Allelic imbalance in *CALR* somatic mutagenesis

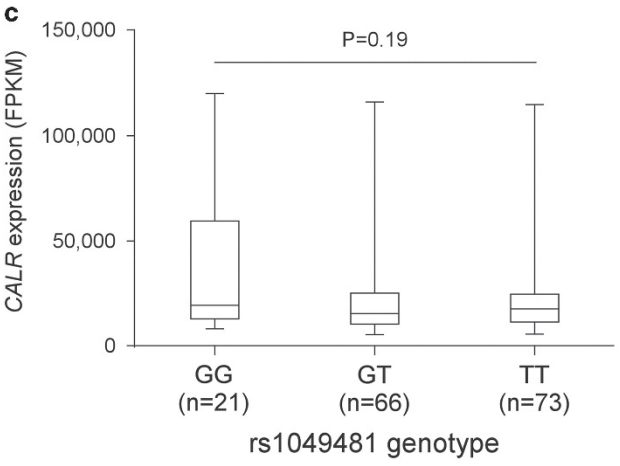
Leukemia (2015) **29,** 1431–1435; doi:10.1038/leu.2015.3

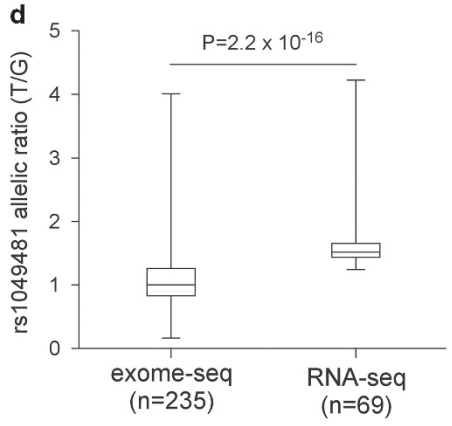
Myeloproliferative neoplasms (MPNs) are clonal hematological malignancies characterized by excessive production of terminally differentiated myeloid cells. MPNs comprise nine disease entities; three of which, namely, polycythemia vera (PV), essential thrombocythemia (ET) and primary myelofibrosis (PMF), share major molecular and pathological features.[1,2] The MPN phenotype is mainly defined by mutually exclusive oncogenic mutations in

Accepted article preview online 8 January 2015; advance online publication, 30 January 2015









the genes *JAK2*,[3–7] *MPL*[8,9] and, as recently discovered, *CALR*.[10,11] In most cases, the *JAK2* and *MPL* mutations are single-nucleotide substitutions at key amino-acid positions. *JAK2*-V617F is the most common mutation in MPN accounting for >60% of cases. *CALR* mutations consist exclusively of insertions and deletions in the last exon of the gene resulting in a frameshift to a specific alternative reading frame.[10]

A germline haplotype spanning the *JAK2* gene (GGCC) is a major predisposition factor for developing *JAK2* mutation-positive MPN.[12–14] Interestingly, in heterozygote patients, the *JAK2*-V617F mutation is acquired preferentially on the GGCC haplotype of *JAK2*.[14] This observation led to the two different hypotheses of hypermutability vs fertile ground to explain the germline predisposition.[12,14] However, these hypotheses have neither been proven nor been disproven to date.

After the recent discovery of *CALR* mutations in MPN,[10,11] we hypothesized that allelic preference in mutation acquisition similar to *JAK2*-V617F may also occur at *CALR* locus. Moreover, as all *CALR* somatic mutations in MPN are insertions and deletions, specific mutational mechanisms might be of relevance. As is the case of *JAK2*, sequence variants such as single-nucleotide polymorphisms (SNPs) proximal to the mutational hotspot in the *CALR* gene may also predispose to MPN.

A common SNP is located close to the *CALR* mutational hotspot, 54 bp downstream of the *CALR* stop codon (rs1049481). Allele frequencies in the general population are about 40% G and 60% T. To investigate the possibility of preferential acquisition of *CALR* mutations on one of the alleles, we designed a PCR-based assay that allows detecting which allele of rs1049481 acquired the somatic mutation. The unlabeled forward primer (5′-GGCAAGGCC CTGAGGTGT-3′) binds to a sequence in intron 8 (last intron, 5′ to *CALR* mutations), whereas a pair of 6-FAM-labeled reverse primers (5′-AGACATTATTTGGCGCGGCC-3′ and 5′-TTTAGACATTATTTGGCG CGGCA-3′) bind at the 3′ untranslated region region, with the most 3′ base binding either to G or T allelic variants of rs1049481 (Figure 1a). As the PCR product includes the entire exon 9, somatic insertions and deletions can be detected by measuring the size of the labeled PCR product. In addition to the base difference, the reverse primer specific to the G allele is also three nucleotides shorter than the T allele primer. Thus, the assay simultaneously genotypes rs1049481 and discriminates between mutant and wild-type *CALR* (Figure 1b). The results of the assay were concordant with *CALR* mutation data obtained through Sanger sequencing previously in all cases.

If the *CALR* gene mutates randomly, somatic mutations are expected to be equally distributed between *CALR* alleles. We initially included 386 MPN patients and 163 healthy controls from Austria in this study. Of the 386 MPN cases, 70 had *CALR* mutations, of which 37 (52.9%) harbored type 1, 19 (27.1%) type 2 and 14 (20.0%) other mutation types. Later in the study, we expanded the *CALR*-mutated cohort by including additional 129 Austrian and Italian cases heterozygous for rs1049481. We observed unequal distribution of *CALR* mutations between the G and T alleles of rs1049481 in heterozygous cases. *CALR* mutations were more frequently acquired on the T allele ($N=23$) compared with the G allele ($N=10$) in patients with GT genotype of rs1049481 (Table 1). As the number of *CALR*-positive cases heterozygous for rs1049481 was low in the initial cohort, we tested additional *CALR*-positive patients from different cohorts using the same assay.

Overall, we observed a significantly higher number of *CALR* mutations in the T allele (100 cases) compared with G allele (62 cases) of rs1049481. Thus, the T allele of *CALR* mutates 1.61 times more frequently than the G allele ($P=0.0028$; one-proportion *z*-test). We subsequently performed subtype analysis; however, low sample size implicated low statistical power for potential trends. Nevertheless, we observed a distinct pattern of allelic bias in *CALR* mutation types. The allelic bias was most pronounced for *CALR* type 2 mutations (T/G: 32/15 cases, ratio 2.13, $P=0.013$), followed by type 1 mutations (T/G: 50/30 cases, ratio 1.67, $P=0.025$) and was nearly absent for all the other mutations combined (T/G: 18/17 cases, ratio 1.06, $P=0.866$).

As a similar allelic bias in the *JAK2* mutation acquisition confers predisposition to *JAK2*-positive MPN, we next examined whether rs1049481_T confers susceptibility to *CALR*-positive MPN. As shown in Table 1, genotypic frequencies between controls and MPN cases did not differ significantly ($P=0.8407$, χ^2 test for independence). Thus, despite the presence of an allelic bias in somatic mutagenesis of *CALR* (Table 1), rs1049481 does not exhibit statistically significant association with MPN in our patient cohort. Notably, a larger cohort might be necessary for *CALR* (allelic bias: 1.6-fold) compared with *JAK2* (allelic bias: 7.2-fold)[14] to have sufficient statistical power for observing disease association as a consequence of the mutation acquisition bias. Genotype frequencies were similar in different MPN mutational subtypes (*CALR* positive, *JAK2* positive) and diagnostic classes (PV, ET or PMF; data not shown). Although rs1049481 serves as a suitable tagging SNP for the *CALR* locus and allows for screening of a large number of heterozygote individuals owing to its high minor allele frequency, another variant in linkage disequilibrium might be causative for differential mutation acquisition. However, with the present cohort size it was not possible to dissect the haplotype for identifying variants with stronger linkage to *CALR* mutations than rs1049481 (data not shown).

The allelic bias of *CALR* mutation acquisition may be a result of specific mutational mechanisms responsible for *CALR* mutagenesis. Particularly interesting is the fact that *CALR* somatic mutations are restricted to insertions and deletions. The most frequent *CALR* mutations are the 52-bp deletion (type 1) and the 5-bp insertion (type 2). A close examination of the sequence of *CALR* exon 9 reveals a complex repetitive region with both trinucleotide repeats and longer repeat elements. The nucleotide sequences around the breakpoints of the 52-bp deletion consist of two

Figure 1. Investigation of potential allelic imbalances in *CALR* mutational acquisition and gene expression. (**a**) Assay used for simultaneous genotyping of rs1049481 and allelic localization of *CALR* mutations. The two allele-specific labeled reverse primers bind and amplify the chromosomal region bearing the complementary allele of the SNP resulting in products of different sizes. In addition, the size difference of *CALR* insertions and deletions allows for identification of the allele on which the mutations are acquired. (**b**) Capillary electrophoresis tracks of the amplified labeled product for different cases of rs1049481 genotypes and *CALR* mutations. Below each peak, the size, the corresponding genotype (in brackets) and the *CALR* mutational status are indicated. Coexistence of mutated and wild-type *CALR* on the T allele suggests incomplete clonality (bottom right panel). (**c**) The relationship between rs1049481 genotypes and *CALR* mRNA expression levels. RNA-Seq data on peripheral blood from 173 AML patients were downloaded from The Cancer Genome Atlas (TCGA) project (LAML data set). Normalized expression values were grouped by rs1049481 genotype. There is no detectable statistically significant difference in genotype-specific gene expression ($P=0.19$; Kruskal–Wallis test). (**d**) Evidence for allelic expression imbalance at the *CALR* locus. RNA-Seq and exome sequencing alignment data were downloaded from the TCGA data platform and the 1000 Genomes Project, respectively. Genomic DNA for exome sequencing was derived from normal control tissues, whereas RNA-Seq data was generated from AML peripheral blood enriched for myeloid progenitor cells. Allelic read depth ratios at rs1049481 (T/G) of heterozygous samples that passed genotyping quality filtering are shown. While the T/G ratio of exome sequencing calls is around 1, allelic ratios at RNA level significantly deviate from allelic balance ($P=2.2\times10^{-16}$; Mann–Whitney *U*-test).

Table 1. The distribution of rs1049481 SNP genotypes in *CALR* mutation-positive and -negative MPN cases and controls

rs1049481 genotype	*Control* (n = *163)*	CALR-*negative MPN* (n = *316)*	CALR-*positive MPN* (n = *70)*	CALR *mutation on T/G allele*
GG	17.2% (28)	16.8% (53)	15.7% (11)	
GT	49.7% (81)	47.8% (151)	47.1% (33)	23/10
TT	33.1% (54)	35.4% (112)	37.1% (26)	
Additional *CALR*-positive MPN cases with GT genotype (*n* = 129)				77/52
Total (*n* = 162)				100/62
CALR type 1 mutation (*n* = 80)				50/30
CALR type 2 mutation (*n* = 47)				32/15
Other *CALR* mutation types (*n* = 35)				18/17

Abbreviations: MPN, myeloproliferative neoplasm; SNP, single-nucleotide polymorphism.

imperfect direct repeats separated by a spacer sequence. The start positions of the two repeats are 52 bp apart. In case of the 5-bp insertion, the sequence of the insertion creates a 10-bp palindromic sequence (it is the inverted complementary copy of the preceding five bases). These facts indicate that the mutagenesis of *CALR* might be recombination mediated. In such a scenario, the actual sequence context of the region may have an important role in facilitating or impeding the mutation generation. Thus, owing to the complex mutational mechanisms, the *CALR* mutations may occur more frequently in the T allele of the rs1049481 compared with the G allele. This might be more relevant for type 1 and type 2 mutations, as specific mutational mechanisms resulting in increased mutation rate may be responsible for their high frequency. The observed trend in allelic bias specificity toward type 1 and type 2 mutations provides some support to this hypothesis.

An alternative mechanism potentially explaining the observed mutational imbalance might be allelic bias in gene expression. Higher expression from one of the alleles can be associated with a more open chromatin state, thus being more exposed to mutagenic stimuli. Furthermore, mutations occurring on the allele producing higher transcript levels can be more likely to result in clonal outgrowth, a concept underlying the 'fertile ground hypothesis'.[12–14] To test for potential allelic expression bias at the *CALR* locus, we made use of the combined genotyping and RNA-Seq data on peripheral blood from 173 acute myeloid leukemia (AML) patients provided by The Cancer Genome Atlas project.[15] The AML peripheral blood is enriched for myeloid progenitors and might therefore represent an adequate tissue for this analysis. Grouping *CALR* expression values by genotypes did not yield any statistically significant evidence for differential expression (Figure 1c). However, when examining the RNA-Seq data on a per-individual level, heterozygotes showed a modest but significant expression bias toward the T allele, compatible with allelic expression imbalance (Figure 1d). Whether allelic expression bias is responsible for the allelic bias in *CALR* mutational acquisition needs to be the subject of further investigation.

Cancer-associated genes showing allelic bias in their somatic mutagenesis have not been commonly reported. Interestingly, after *JAK2*, our observation on allelic imbalance of *CALR* mutations describes the second such case involved in MPN pathogenesis. It remains to be seen whether other cancer-associated loci exhibit biases in the acquisition of somatic mutations similar to *JAK2* and *CALR* in MPN.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

Results on allele-specific expression are based on the data generated by the TCGA Research Network: http://cancergenome.nih.gov/. The studies were supported by funding from the Austrian Science Fund (FWF4702-B20) granted to RK, Italian Ministry of Health (GR-2010-2312855) granted to ER and AIRC Special Program Molecular Clinical Oncology 5 per Mille, project 1005, granted to MC. RJ holds an Erwin Schrödinger Fellowship from the Austrian Science Fund (FWF).

AS Harutyunyan[1], R Jäger[1], D Chen[1], T Berg[1], E Rumi[2,3], B Gisslinger[4], D Pietra[2], H Gisslinger[4], M Cazzola[2,3] and R Kralovics[1,4]

[1]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria; [2]Department of Hematology Oncology, Fondazione Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Policlinico San Matteo, Pavia, Italy; [3]Department of Molecular Medicine, University of Pavia, Pavia, Italy and [4]Division of Hematology and Blood Coagulation, Department of Internal Medicine I, Medical University of Vienna, Vienna, Austria
E-mail: rkralovics@cemm.oeaw.ac.at

REFERENCES

1 Campbell PJ, Green AR. The myeloproliferative disorders. *N Engl J Med* 2006; **355**: 2452–2466.
2 Tefferi A, Vainchenker W. Myeloproliferative neoplasms: molecular pathophysiology, essential clinical understanding, and treatment strategies. *J Clin Oncol* 2011; **29**: 573–582.
3 Baxter EJ, Scott LM, Campbell PJ, East C, Fourouclas N, Swanton S *et al.* Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet* 2005; **365**: 1054–1061.
4 James C, Ugo V, Le Couedic JP, Staerk J, Delhommeau F, Lacout C *et al.* A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature* 2005; **434**: 1144–1148.
5 Kralovics R, Passamonti F, Buser AS, Teo SS, Tiedt R, Passweg JR *et al.* A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med* 2005; **352**: 1779–1790.
6 Levine RL, Wadleigh M, Cools J, Ebert BL, Wernig G, Huntly BJ *et al.* Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* 2005; **7**: 387–397.
7 Scott LM, Tong W, Levine RL, Scott MA, Beer PA, Stratton MR *et al.* JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. *N Engl J Med* 2007; **356**: 459–468.
8 Pikman Y, Lee BH, Mercher T, McDowell E, Ebert BL, Gozo M *et al.* MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. *PLoS Med* 2006; **3**: e270.
9 Staerk J, Lacout C, Sato T, Smith SO, Vainchenker W, Constantinescu SN. An amphipathic motif at the transmembrane-cytoplasmic junction prevents autonomous activation of the thrombopoietin receptor. *Blood* 2006; **107**: 1864–1871.
10 Klampfl T, Gisslinger H, Harutyunyan AS, Nivarthi H, Rumi E, Milosevic JD *et al.* Somatic mutations of calreticulin in myeloproliferative neoplasms. *N Engl J Med* 2013; **369**: 2379–2390.
11 Nangalia J, Massie CE, Baxter EJ, Nice FL, Gundem G, Wedge DC *et al.* Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N Engl J Med* 2013; **369**: 2391–2405.
12 Jones AV, Chase A, Silver RT, Oscier D, Zoi K, Wang YL *et al.* JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nat Genet* 2009; **41**: 446–449.

13 Kilpivaara O, Mukherjee S, Schram AM, Wadleigh M, Mullally A, Ebert BL *et al.* A germline JAK2 SNP is associated with predisposition to the development of JAK2 (V617F)-positive myeloproliferative neoplasms. *Nat Genet* 2009; **41**: 455–459.
14 Olcaydu D, Harutyunyan A, Jager R, Berg T, Gisslinger B, Pabinger I *et al.* A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms. *Nat Genet* 2009; **41**: 450–454.
15 Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013; **368**: 2059–2074.

The predominant myeloma clone at diagnosis, CDR3 defined, is constantly detectable across all stages of disease evolution

Leukemia (2015) **29,** 1435–1437; doi:10.1038/leu.2015.7

Multiple myeloma (MM) pathogenesis has been explained for many years by the cancer biology dogma introduced in 1976 by Peter Nowell:[1] first, a single plasma cell would be immortalized by an error in the immunoglobulin genes rearrangement process; then, a progressive stepwise acquisition of somatic cell mutations would induce a sequential selection, leading to domination by the fittest clone.[2] In line with this idea of 'myeloma stability', single-nucleotide polymorphism arrays studies in diagnostic-relapse paired samples have revealed the presence of common clonal characteristics.[3] Biologically, the M-protein remains usually constant across MM evolution, to the point that it is conventionally used to monitor treatment response and disease progress. Further, the variable domain of the rearranged immunoglobulin heavy-chain genes (or CDR3 region) has been used as a patient-specific myeloma fingerprint in minimal residual disease (MRD) studies.[4]

However, massive genome studies with next-generation sequencing have challenged this concept in MM, showing a significant intraclonal heterogeneity at diagnosis with the possible presence of several clonal progenitors or tumor-initiating cells. Therapeutic or ecosystem-dependent selection pressures would drive the alternating dominance of these clones over time.[5]

In this study, we have characterized and compared the highly tumor-specific CDR3 region in 52 paired samples from 26 patients with MM obtained at diagnosis and progression or at two different timepoints of progression, aiming (1) to assess monoclonality in MM evolution by means of the analysis of the CDR3 sequence, and (2) to further support ASO RQ-PCR approaches for MRD in MM, based on the constancy and specificity of the CDR3 region.

Genomic DNA from 52 unselected samples of 26 patients diagnosed with MM was isolated using conventional methods. Pairs corresponded to diagnosis/progression samples in 19 cases or to two different time points of progressive disease in seven cases; 50 samples proceeded from bone marrow and 2 from extramedullary tissues (testes and spleen). PCR amplification was performed according to the BIOMED-2 methods.[6] The clonal products were identified by fragment analysis and directly sequenced in an ABI 3130 DNA Sequencer (Applied Biosystems, Foster City, CA, USA). Sequencing was performed using Big-Dye terminators (Applied Biosystems). Germline VH, DH and JH segments from complete rearrangements were identified by comparison with public databases. CDR3 region was identified and highlighted for analysis and comparison between paired samples.

We analyzed 52 samples from 26 patients with MM obtained at diagnosis and progression or at two timepoints of progressive disease. Time between sampling varied from <1 year to >6 years (median, 2 years). The M-protein subtype remained stable in all cases except in one pair (Table 1, case 16), associated with a λ light-chain escape phenomenon. All pairs were obtained from bone marrow except for two (Table 1, cases 25 and 26), corresponding to bone marrow and extramedullary disease (spleen and testes). Also, two major cytogenetic changes were identified (Table 1; in case 8, an increased percentage of 13q14 deletion from 7 to 54% was detected and in case 24, of 17p (p53) deletion from 5 to 87%). Treatments administered between sampling are detailed in Table 1.

VDJH gene sequence analysis has been performed in MM, aiming to determine the B-cell origin of the clonal plasma cell and/or to assess the presence of intraclonal heterogeneity.[7] On the basis of the presence of somatic hypermutations, results consistently confirm the post-germinal origin of the pathological plasma cell, excluding intraclonal variation due to the constancy of the nucleotide sequence. Interestingly however, consecutive samples have been anecdotally analyzed, mainly to assess VDJH patterns of transition from MGUS (monoclonal gammopathy of undetermined significance) to MM.[8,9]

We identified the sequence of the CDR3 region in all samples and then compared the obtained results between the two samples in the 26 pairs. The sequence of nucleotides was constantly identical between all the samples, including cases with significant differences within the pair, such as major cytogenetic changes, light-chain escape, extramedullar *vs* bone marrow infiltration, and relapsed (and therefore, treatment selected) *vs* refractory disease (data not shown). Similar results were provided by Ralph *et al.*[9] in a series of 16 patients using size-based methods and partly confirmed by sequencing. From our study, using PCR amplification and direct Sanger sequencing, we can conclude that the main tumor clone in MM retains a specific signature across disease evolution that allows the identification of samples as evolutionary related.

Next-generation sequencing studies have revealed a highly complex clonal architecture in MM, with a Darwinian-like evolution in a branching manner, leading to substantial clonal diversity and genetic heterogeneity. Although the M-protein subtype remains unchanged in most of the patients with MM, the loss of the immunoglobulin heavy chain (escape phenomenon) has been described.[10] Further, major cytogenetic changes between diagnosis and relapse have been documented, especially in cases of extramedullary disease. Our study shows that the dominant myeloma clone, even in the setting of a clinically and genetically unstable picture, retains a unique and specific *IGH* rearrangement, unaffected by the selection pressure induced by the treatment/s exposure and constant across disease progress.

These results also provide further support for molecular MRD techniques. The high rate of complete responses (up to 50–60%) currently achieved in MM has prompted the use of new

Accepted article preview online 8 January 2015; advance online publication, 27 January 2015