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 scCD23 associations for CLL/SLL diagnosed closer in time to phlebotomy, which may be at least partly disease induced, as circulating scCD23 in CLL patients has been associated with tumor burden. However, our scCD23 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 scCD27 and sTNFR2 ≥ 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 scCD23 and scCD30 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.

Our findings raise the possibility that serum scCD23 and scCD30 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 scCD23 in serum collected ≤ 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. 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 Purdue1,2, Q Lan1, TJ Kemp3, A Hildesheim1, SJ Weinstein1, JN Hofmann1, J Virtamo4, D Albanes1, LA Pinto5 and N Rothman1
1Division of Cancer Epidemiology and Genetics, Department of Health and Human Services, National Cancer Institute, NIH, Bethesda, MD, USA;
2Ontario Institute for Cancer Research, Toronto, ON, Canada;
3HPV Immunology Laboratory, Leidos Biomedical Incorporated, Frederick National Laboratory for Cancer Research, Frederick, MD, USA and
4Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland
E-mail: mark.purdue@ontariohealthstudy.ca

REFERENCES
1 Harget 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, Martínez-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 W, 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: 4698–4707.
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, Martínez-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’Ellos 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 Reintisch W, Wilhelm M, Hilgath M, Gaiché 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.

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. The MPN phenotype is mainly defined by mutually exclusive oncogenic mutations in
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.12–18 Interestingly, in heterozygote patients, the JAK2-V617F mutation is acquired preferentially on the GGCC haplotype of JAK2. 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′-GGCAAGGGCC) 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′-AGACATTATGGCCGGCCGG-3′ and 5′-TTAGACATTATGGCCGGCCGGCCGA-3′) bind at the 3′ untranslated region region, with the most 3′ base binding either to G or T allelic variants of rs1049481 (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 rs1049481-positive 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, χ² 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 the CALR allelic bias (allelic bias: 1.6-fold) compared with JAK2 (allelic bias: 7.2-fold) 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 is 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 × 10⁻¹⁰; Mann–Whitney U-test).
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’. 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. 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, R Jäger, D Chen, T Berg, E Rumi, B Gisslinger, D Pietra, H Gisslinger, M Cazzola and R Kralovics

1CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria;
2Department of Hematology Oncology, Fondazione Istituto di Ricovero e Cur a Carattere Scientifico (IRCCS) Policlinico San Matteo, Pavia, Italy;
3Department of Molecular Medicine, University of Pavia, Pavia, Italy and
4Division of Hematology and Blood Coagulation, Department of Internal Medicine I, Medical University of Vienna, Vienna, Austria

E-mail: rkralovic@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, Fournouclas N, Swanton S et al. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. Lancet 2005; 365: 1056–1061.
4. James C, Ugo V, Le Couedic JP, Staelj J, Delhommeau F, Lacout C et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythemia vera. Nature 2005; 434: 1144–1148.
5. Kralovics R, Passamonti F, Buser AS, Tejedor 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, Wemig 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. Pilkan Y, Lee BH, Mercer T, McDowell E, Ebert BL, Gaoz M et al. MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. PLoS Med 2006; 3: e270.
9. Staelj J, Lacout C, Sato T, Smith SO, Vainchenker W, Constantinescu SN. An amphiphatic motif at the transmembrane-cytoplasmic junction prevents autonomous activation of the thrombopoietin receptor. Blood 2006; 107: 1864–1871.
10. Klampfi T, Gisslinger H, Harutyunyan AS, Nivarthi H, Rumi E, Miloisev 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, Gurnend 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.
Multiple myeloma (MM) pathogenesis has been explained for many years by the cancer biology dogma introduced in 1976 by Peter Nowell: 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. 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. 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.

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.

In this study, we have characterized and compared the highly 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

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

© 2015 Macmillan Publishers Limited

Letters to the Editor

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/