Europe PMC Funders Group
Author Manuscript
Leukemia. Author manuscript; available in PMC 2015 July 21.

Published in final edited form as:
Leukemia. 2015 June ; 29(6): 1431–1435. doi:10.1038/leu.2015.3.

Allelic imbalance in *CALR* somatic mutagenesis

Ashot S. Harutyunyan[1], **Roland Jäger**[1], **Doris Chen**[1], **Tiina Berg**[1], **Elisa Rumi**[2,3], **Bettina Gisslinger**[4], **Daniela Pietra**[2], **Heinz Gisslinger**[4], **Mario Cazzola**[2,3], and **Robert 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

[4]Department of Internal Medicine I, Division of Hematology and Blood Coagulation, Medical University of Vienna, Vienna, Austria

Letter to the editor

Myeloproliferative neoplasms (MPN) are clonal hematological malignancies characterized by excessive production of terminally differentiated myeloid cells. MPN 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 the genes *JAK2,*[3-7] *MPL*[8,9] and, as recently discovered, *CALR.*[10,11] In most cases, *JAK2* and *MPL* mutations are single nucleotide substitutions at key amino acid positions. *JAK2*-V617F is the most common mutation in MPN accounting for more than 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 *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 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, since 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

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

[*]Correspondence should be addressed to Robert Kralovics (rkralovics@cemm.oeaw.ac.at).

Conflict of Interest
The authors declare no relevant conflicts of interest.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

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′-GGCAAGGCCCTGAGGTGT-3′) binds to a sequence in intron 8 (last intron, 5′ to *CALR* mutations), while a pair of 6-FAM-labeled reverse primers (5′-AGACATTATTTGGCGCGGCC-3′ and 5′-TTTAGACATTATTTGGCGCGGCA-3′) bind at the 3′UTR region, with the most 3′ base binding either to G or T allelic variants of rs1049481 (Figure 1A). Since 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 into the 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, respectively. Later in the study, we expanded the *CALR* mutated cohort by 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 to the G allele (N=10) in patients with GT genotype of rs1049481 (Table 1). Since the numbers of *CALR* positive cases heterozygous for rs1049481 were 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 on the T allele (100 cases) compared to 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, ratio 1.67, P=0.025) and was nearly absent for all the other mutations combined (T/G: 18/17, ratio 1.06, P=0.866).

Since 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, chi-square test for independence). Thus, despite the

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

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 to *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). While rs1049481 serves as a suitable tagging SNP for the *CALR* locus and allows for screening of a large number of heterozygote individuals due 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). 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 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 play an important role in facilitating or impeding the mutation generation. Thus, due to the complex mutational mechanisms, the *CALR* mutations may occur more frequently on the T allele of the rs1049481 than on the G allele. This might be more relevant for Type 1 and Type 2 mutations, since specific mutational mechanisms resulting in increased mutation rate may be responsible for their high frequency. The observed trend in allelic bias specificity towards 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 combined genotyping and RNA-Seq data on peripheral blood from 173 acute myeloid leukemia (AML) patients provided by the TCGA project.[15] AML peripheral blood is enriched for myeloid progenitors and might therefore represent an adequate tissue for this analysis. Grouping *CALR* expression values by genotypes, we did not find statistically significant evidence for differential expression (Figure 1C). However, when examining RNA-Seq data on a per-individual level, heterozygotes showed a modest but significant expression bias towards the T allele, compatible with allelic expression imbalance (Figure

Europe PMC Funders Author Manuscripts

1D). Whether allelic expression bias is responsible for the allelic bias in *CALR* mutational acquisition needs to be 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 if other cancer-associated loci exhibit biases in acquisition of somatic mutations similar to *JAK2* and *CALR* in MPN.

Acknowledgements

Results on allele specific expression are based upon 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).

References

1. Campbell PJ, Green AR. The myeloproliferative disorders. N Engl J Med. 2006; 355(23):2452–66. [PubMed: 17151367]
2. Tefferi A, Vainchenker W. Myeloproliferative neoplasms: molecular pathophysiology, essential clinical understanding, and treatment strategies. J Clin Oncol. 2011; 29(5):573–82. [PubMed: 21220604]
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(9464):1054–61. [PubMed: 15781101]
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(7037): 1144–8. [PubMed: 15793561]
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(17):1779–90. [PubMed: 15858187]
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(4):387–97. [PubMed: 15837627]
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(5):459–68. [PubMed: 17267906]
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(7):e270. [PubMed: 16834459]
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(5):1864–71. [PubMed: 16249382]
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(25):2379–90. [PubMed: 24325356]
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(25): 2391–405. [PubMed: 24325359]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

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(4):446–9. [PubMed: 19287382]
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(4):455–9. [PubMed: 19287384]
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(4):450–4. [PubMed: 19287385]
15. Cancer Genome Atlas Research N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013; 368(22):2059–74. [PubMed: 23634996]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

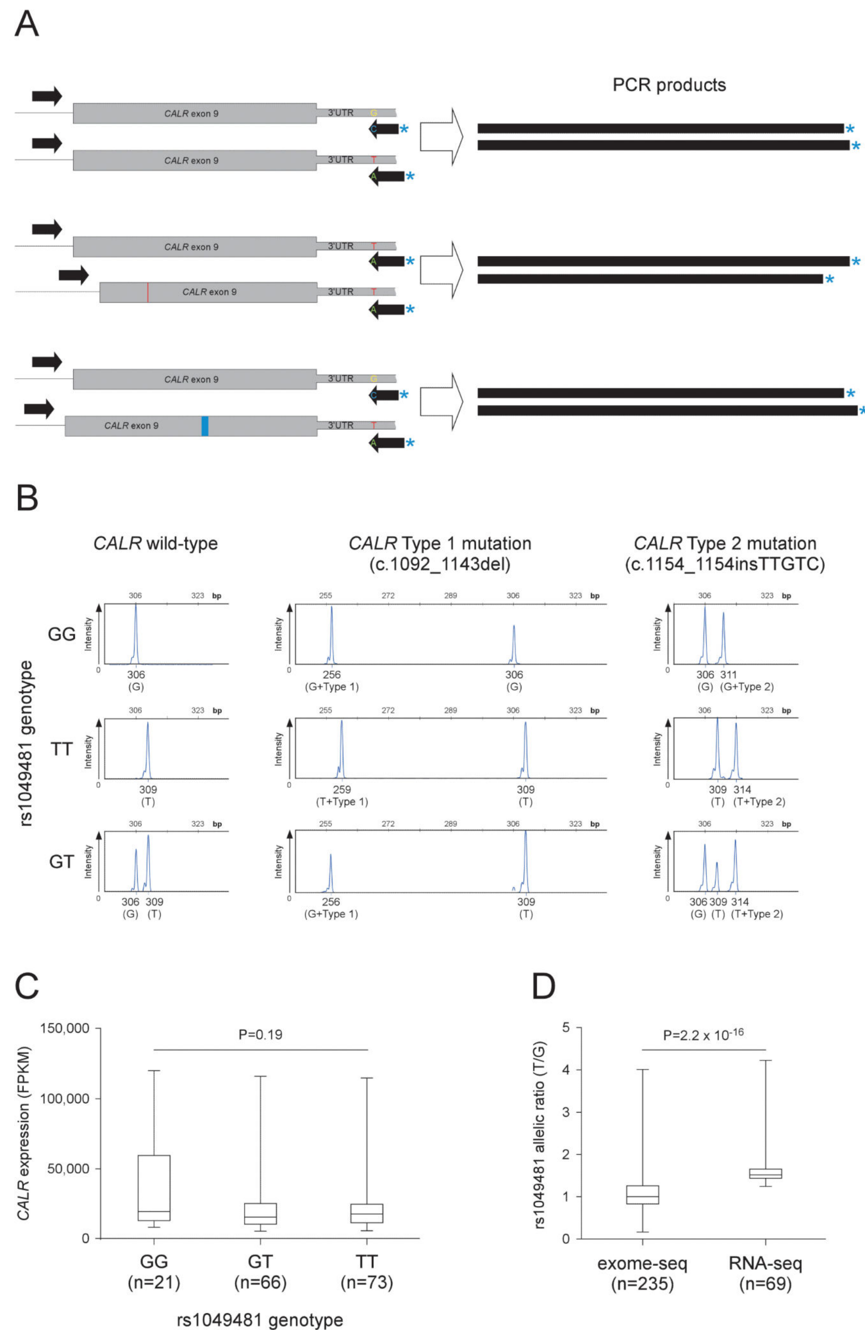

Figure 1. Investigation of potential allelic imbalances in *CALR* mutational acquisition and gene expression

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 size. In addition, 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*

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

mutations. Below each peak, size, corresponding genotype (in brackets), and *CALR* mutational status are indicated. Co-existence of mutated and wild-type CALR on the T allele suggests incomplete clonality (bottom right panel). **C**. Relationship between rs1049481 genotypes and *CALR* mRNA expression levels. RNA-Seq data on peripheral blood from 173 acute myeloid leukemia (AML) patients were downloaded from The Cancer Genome Atlas (TCGA) project (LAML dataset). 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 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. Shown are allelic read depth ratios at rs1049481 (T/G) of heterozygous samples which passed genotyping quality filtering. Whereas the T/G ratio of exome sequencing calls is around 1, allelic ratios at the RNA level significantly deviate from allelic balance ($p=2.2\times10^{-16}$; Mann-Whitney-U test).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

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