LETTERS TO THE EDITOR

Cytarabine preferentially induces mutation at specific sequences in the genome which are identifiable in relapsed acute myeloid leukaemia

Leukemia (2015) 29, 491–494; doi:10.1038/leu.2014.284

Cytarabine (1- β -D-arabinofuranosylcytosine; Ara-C) is an effective remission induction treatment for acute myeloid leukemia (AML).^{1,2} However, the majority of patients relapse within 3 years, often with aggressive AML² characterised by the gain of novel base substitution mutations in genes including *IDH2* and *TET2*.^{3,4} Chemotherapy used to treat AML is predicted to induce mutations detectable in relapsed disease, although direct evidence for this is lacking. The cytotoxicity of Ara-C is mediated in part by incorporation into replicating DNA leading to inhibition of chain extension. However, even at concentrations that induce significant cell death, arabinofuranosylcytosine triphosphate, the reactive metabolite of Ara-C, is primarily incorporated into DNA at internucleotide positions,^{5,6} suggesting that this nucleoside analogue is likely mutagenic and could contribute to the aetiology of novel somatic mutations acquired during AML relapse.

To assess the mutagenicity of Ara-C, mutation frequency (Mf) was determined in TK6 cells *in vitro* at two widely used reporter genes, *thymidine kinase* (TK) and *hypoxanthine-guanine phosphoribosyltransferase* (HPRT). The procedure for performing *in vitro* quantitative mutation assays was developed by Liber and Thilly⁷ and involves treatment of cells under defined conditions, followed by incubation in non-selective medium for a time sufficient to allow phenotypic expression of the chosen genetic marker. Cells were then plated in the presence of selective agent to determine the number of mutant cells able to form colonies and in the absence of selective agent in order to determine the intrinsic cloning efficiency of the cells. Mf was calculated following adjustment for cloning efficiency and is described in detail in Supplementary Methods.

A significant Ara-C dose-dependant increase in Mf was observed in surviving cells at both loci, although Ara-C induced fewer mutations in surviving cells compared to the powerful mutagen MNU at equitoxic doses (Figure 1a). Ara-C-induced Mf was next determined in isogenic cells deficient in MSH2,⁸ a component of DNA mismatch repair (MMR) that when lost confers sensitivity to mutagenesis by agents that induce base:base mispairs.⁹ MSH2-deficient cells were more susceptible to Ara-C-induced mutation than their MMR-proficient counterparts (Figures 1a and b and Supplementary Figure S1), demonstrating a significant influence of both Ara-C dose and MMR status on Mf ($P < 0.001$). These data confirm the mutagenicity of this anti-leukaemic cytosine analogue and suggest it to be a base substitution mutagen.

In order to determine whether Ara-C preferentially induces mutation at specific sequences in the genome, a mutation spectrum at the HPRT locus was generated. Cells were exposed to 30 nM Ara-C (or vehicle) for 16 hours and mutant colonies were generated as described by Liber and Thilly.⁷ A single mutant colony from each treated cell population was expanded

in standard growth medium supplemented with 2 μ g/ml 6-thioguanine and RNA/DNA extracted for subsequent characterisation (Supplementary Methods). This procedure was performed up to 150 times for Ara-C and vehicle, ensuring full independence of mutant colonies. The mutation spectra from Ara-C-treated and vehicle-treated cells were significantly different ($P = 0.002$), with the proportion of base substitutions higher in Ara-C-treated cells (27 vs 19%; Figure 1c and Supplementary Table S1). Strikingly, 7 of 27 (25.9%) base substitution mutations from Ara-C-treated clones affected the central G:C base-pair in 5'-TpGpA^{3'}/5'-TpCpA^{3'} sequences (Figure 1d and Supplementary Table S2), including 6 single base substitutions and 1 tandem base substitution at five different sites in the HPRT gene. Using Monte-Carlo simulation, the probability of 7 or more mutations at the central position of any one of the 32 possible trinucleotide sequences occurring by chance was calculated as 4.7×10^{-4} . Mutation at the central position of 5'-TpGpA^{3'}/5'-TpCpA^{3'} sequences was observed only once in vehicle-treated control cells (1/25, 4.0%) and none of the other 31 possible trinucleotide sequences were represented more than twice in the Ara-C mutation spectrum (Figure 1d).

Ara-C induces significant perturbation of the DNA helix which could cause mutation at neighbouring base positions.¹⁰ Consistent with this phenomenon, five additional base substitution mutations in Ara-C-treated cells were identified at non-central positions in 5'-TpGpA^{3'}/5'-TpCpA^{3'} sequences, but none were observed in vehicle-treated control cells (Figure 1e and Supplementary Table S2). In total, 12 of 27 (44.4%) base substitution mutations in Ara-C-treated cells occurred in 5'-TpGpA^{3'}/5'-TpCpA^{3'} sequences, compared to only 1 of 25 (4.0%) in control cells. Using Monte-Carlo simulation, the probability of 12 or more mutations at any of the three base positions in any one of the 32 possible trinucleotide sequences occurring by chance was calculated as 2.6×10^{-4} . These data demonstrate that Ara-C preferentially induces mutation in 5'-TpGpA^{3'}/5'-TpCpA^{3'} sequences, and particularly at the central G:C position consistent with its function as a cytosine analogue.

We next looked for mutation at 5'-TpGpA^{3'}/5'-TpCpA^{3'} sequences in AML patients treated with Ara-C. Whole genome sequencing data from eight AML patients³ were analysed as described in Supplementary Methods. A total of 3587 and 452 somatic mutations were analysed from presentation and relapse disease, respectively. The frequency of mutation at the G:C position in 5'-TpGpA^{3'}/5'-TpCpA^{3'} sequences was significantly higher in relapsed AML (22/452, 4.9%) compared to chemotherapy-naïve presentation AML (111/3587, 3.1%; odds ratio (OR) 2.2, 95% confidence interval (CI) 1.4–3.6; $P = 0.003$; Table 1 and Supplementary Table S3), consistent with an Ara-C-induced mutation signature. It should be noted that Ara-C makes a relatively small contribution to relapse mutation load and that other mechanisms might be more important for driving the acquisition of novel mutations in relapsed samples. Specifically, of the 452 novel base substitution mutations seen at relapse, the

excess mutations at the G:C position in $5'\text{TpGpA}^3/5'\text{TpCpA}^3$ sequences, and potentially attributable to Ara-C, constitute less than 2% of all novel relapse-specific mutations.

In order to exclude the possibility that the specific acquisition of mutation at $5'\text{TpGpA}^3/5'\text{TpCpA}^3$ sequences is intrinsic to disease progression, sequencing data were also analysed from

three patients at diagnosis of myelodysplastic syndrome (MDS) who subsequently progressed to AML.¹¹ These patients had not received any therapy other than supportive care for MDS (Supplementary Methods). The frequency of G:C mutations at $5'\text{TpGpA}^3/5'\text{TpCpA}^3$ sequences was similar in MDS (34/1122; 3.0%) and presentation AML (13/428, 3.0%; OR

1.1, 95% CI 0.5–2.3, $P = 0.882$; Supplementary Table S4), further implicating Ara-C as the cause of mutation at these sites in relapse AML.

It will also be important to establish the contribution of other remission induction agents to mutation in relapsing AML. In addition to $5'$ -TpGpA $3'$ / $5'$ -TpCpA $3'$, the frequency of mutation

Table 1. Frequency of somatic mutation at the central base of trinucleotide sequences in AML patients at presentation (before Ara-C treatment) and relapse (post-Ara-C treatment)

Trinucleotide sequence ^a	n (%) ^b		OR ^c	95% CI ^c	P-value ^c
	Pres	Rel			
$5'$ -ApApA $3'$ / $5'$ -TpTpT $3'$	97 (2.7)	12 (2.7)	0.89	0.47–1.66	0.709
$5'$ -ApApC $3'$ / $5'$ -GpTpT $3'$	55 (1.5)	6 (1.3)	0.81	0.34–1.94	0.636
$5'$ -ApApG $3'$ / $5'$ -CpTpT $3'$	79 (2.2)	9 (2.0)	0.9	0.44–1.84	0.766
$5'$ -ApApT $3'$ / $5'$ -ApTpT $3'$	117 (3.3)	15 (3.3)	0.92	0.53–1.62	0.78
$5'$ -ApCpA $3'$ / $5'$ -TpGpT $3'$	209 (5.8)	28 (6.2)	0.97	0.64–1.47	0.876
$5'$ -ApCpC $3'$ / $5'$ -GpGpT $3'$	118 (3.3)	8 (1.8)	0.51	0.25–1.07	0.074
$5'$ -ApCpG $3'$ / $5'$ -CpGpT $3'$	408 (11.4)	39 (8.6)	0.9	0.63–1.29	0.567
$5'$ -ApCpT $3'$ / $5'$ -ApGpT $3'$	137 (3.8)	19 (4.2)	0.94	0.57–1.56	0.816
$5'$ -CpApA $3'$ / $5'$ -TpTpG $3'$	39 (1.1)	4 (0.9)	0.7	0.24–1.99	0.499
$5'$ -CpApC $3'$ / $5'$ -GpTpG $3'$	32 (0.9)	5 (1.1)	1.33	0.50–3.55	0.567
$5'$ -CpApG $3'$ / $5'$ -CpTpG $3'$	74 (2.1)	9 (2.0)	0.94	0.45–1.94	0.861
$5'$ -CpApT $3'$ / $5'$ -ApTpG $3'$	95 (2.6)	10 (2.2)	0.65	0.33–1.29	0.217
$5'$ -CpCpA $3'$ / $5'$ -TpGpG $3'$	115 (3.2)	14 (3.1)	0.86	0.79–1.53	0.615
$5'$ -CpCpC $3'$ / $5'$ -GpGpG $3'$	98 (2.7)	10 (2.2)	0.68	0.35–1.34	0.271
$5'$ -CpCpG $3'$ / $5'$ -CpGpG $3'$	221 (6.2)	28 (6.2)	1.34	0.87–2.05	0.184
$5'$ -CpCpT $3'$ / $5'$ -ApGpG $3'$	162 (4.5)	16 (3.5)	0.65	0.38–1.10	0.11
$5'$ -GpApA $3'$ / $5'$ -TpTpC $3'$	53 (1.5)	7 (1.5)	0.82	0.36–1.85	0.633
$5'$ -GpApC $3'$ / $5'$ -GpTpC $3'$	38 (1.1)	3 (0.7)	0.56	0.17–1.83	0.334
$5'$ -GpApG $3'$ / $5'$ -CpTpC $3'$	43 (1.2)	8 (1.8)	1.34	0.61–2.96	0.469
$5'$ -GpApT $3'$ / $5'$ -ApTpC $3'$	53 (1.5)	8 (1.8)	1.33	0.60–2.91	0.482
$5'$ -GpCpA $3'$ / $5'$ -TpGpC $3'$	112 (3.1)	14 (3.1)	0.92	0.51–1.64	0.773
$5'$ -GpCpC $3'$ / $5'$ -GpGpC $3'$	81 (2.3)	16 (3.5)	1.43	0.81–2.53	0.221
$5'$ -GpCpG $3'$ / $5'$ -CpGpC $3'$	233 (6.5)	19 (4.2)	0.8	0.49–1.32	0.389
$5'$ -GpCpT $3'$ / $5'$ -ApGpC $3'$	144 (4.0)	10 (2.2)	0.5	0.26–0.96	0.037
$5'$ -TpApA $3'$ / $5'$ -TpTpA $3'$	60 (1.7)	18 (4.0)	1.9	1.08–3.33	0.025
$5'$ -TpApC $3'$ / $5'$ -GpTpA $3'$	40 (1.1)	5 (1.1)	0.99	0.38–2.57	0.979
$5'$ -TpApG $3'$ / $5'$ -CpTpA $3'$	39 (1.1)	17 (3.8)	3.43	1.86–6.32	< 0.001
$5'$ -TpApT $3'$ / $5'$ -ApTpA $3'$	90 (2.5)	8 (1.8)	0.58	0.28–1.21	0.146
$5'$-TpCpA$3'$/$5'$-TpGpA$3'$	111 (3.1)	22 (4.9)	2.17	1.30–3.61	0.003
$5'$ -TpCpC $3'$ / $5'$ -GpGpA $3'$	105 (2.9)	11 (2.4)	0.88	0.46–1.70	0.713
$5'$ -TpCpG $3'$ / $5'$ -CpGpA $3'$	175 (4.9)	13 (2.9)	0.7	0.38–1.26	0.23
$5'$ -TpCpT $3'$ / $5'$ -ApGpA $3'$	154 (4.3)	41 (9.1)	2.55	1.73–3.76	< 0.001
Total	3587 (100)	452 (100)			

Abbreviations: CI, confidence interval; n, total number of mutants analysed; Pres, presentation; Rel, relapse; OR, odds ratio. The Ara-C target sequence ($5'$ -TpCpA $3'$ / $5'$ -TpGpA $3'$) is highlighted in bold. For this triplet sequence, binomial logistic regression analysis of data from eight patients with exclusion of a single case at each iteration, retained significance in all instances ($P < 0.050$), demonstrating that the association was not solely due to any one patient. ^aMutations at the central base position in any given trinucleotide sequence (the mutated base and 1 base 5' and 3' of the mutated base) were grouped with mutations in their complement sequence. ^bData is presented as the number of base substitution mutations at the central position of each trinucleotide from a total of 3587 at presentation and 452 at relapse from a total of 8 patients. Numbers in parentheses indicate mutation at each trinucleotide as a percentage of the total in the pooled dataset. ^cOR, 95% CI and associated P-values were calculated using binomial regression analysis of data from the 8 individual patients comprising the dataset (see Supplementary Table S3).

Figure 1. Mutagenicity of Ara-C *in vitro*. *Mf* at *TK* and *HPRT* loci was assessed in DNA MMR-proficient TK6 lymphoblastoid cells (a) or isogenic MMR-deficient *MSH2i* cells (b) following either 16 h exposure to Ara-C (blue squares) or 4 h exposure to MNU (orange triangles). Data represents the mean and standard error of three independently treated cell populations and is presented as *Mf* observed in control vehicle-treated cells (0% cytotoxicity) and at drug doses causing upto 95% cytotoxicity (see Supplementary Figure. 1). Exponential curves are displayed for each dataset. The associated R^2 values represent the square of the Pearson product moment correlation coefficient, indicating the 'goodness of fit' of the data. (c) spectrum of independent mutations observed at the *HPRT* locus following 16 h exposure to 30 nM Ara-C (inducing ~95% cytotoxicity) or vehicle control in B-lymphoblastoid TK6 cells. Coloured sections represent different structural classes of mutation observed. Numbers in parentheses indicate the proportion of mutants in each class. Total number of mutants analysed (n) is shown above the bars. P-value was calculated by Fisher's exact test comparing the distribution of structural mutations (large deletions, small deletions, base substitutions and partial gene duplications) derived from Ara-C treated cells with those derived from vehicle-treated cells. (d and e) the sequence context of *HPRT* base substitution mutations from control vehicle-treated and Ara-C-treated cells (see Supplementary Table S2). Heatmaps display the number of base substitution mutations at the central position (d) or any position (e) of all possible trinucleotide sequences in vehicle- (top) or Ara-C-treated (bottom) cells. It was not possible to discern whether base substitution mutation was initiated due to a mis-insertion on the plus or minus strand. As such, for the purposes of this analysis, mutations in any given trinucleotide sequence (the mutated base and 1 base 5' and 3' of the mutated base) were grouped with mutations in their complement sequence. For vehicle-treated cells, the sequence context of 25 single base substitutions was analysed. For Ara-C treated cells, the sequence context of 25 single base substitutions and 1 tandem base substitution (scored at both positions) were analysed.

at three other trinucleotides was also significantly higher in relapsed AML compared to presentation disease (Table 1), suggesting the presence of mutation signatures from other chemotherapy agents.

In summary, these data demonstrate that Ara-C preferentially induces mutation at $5'$ -TpGpA $3'$ / $5'$ -TpCpA $3'$ sequences which are significantly elevated in relapse disease after exposure to Ara-C-containing regimens. Given the relationship between Ara-C dose and mutagenicity reported here, a consideration of chemotherapy-induced mutagenicity could be important when developing strategies for treating AML that maximise the likelihood of remission whilst minimising the risk of mutation in surviving cells which could contribute to evolution of relapse disease.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We are grateful to Ding *et al.*³ and Walter *et al.*¹¹ for making their primary whole genome sequencing data available for analysis. This work was supported by grants from Leukaemia and Lymphoma Research (#11006 and #13044, to JMA).

AUTHOR CONTRIBUTIONS

SEF and JMA designed and performed research, analysed data and wrote the paper. MC analysed data. JAI designed research. JMA conceived the project.

SE Fordham, M Cole, JA Irving and JM Allan
Newcastle Cancer Centre, Northern Institute for Cancer Research,
Paul O'Gorman Building, Newcastle University,
Newcastle-upon-Tyne, UK
E-mail:james.allan@newcastle.ac.uk

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

OPEN

MARIMO cells harbor a *CALR* mutation but are not dependent on JAK2/STAT5 signaling

Leukemia (2015) **29**, 494–497; doi:10.1038/leu.2014.285

Mutations in *calreticulin* (*CALR*) were recently described to be present in the majority of patients with a *JAK2*-unmutated myeloproliferative neoplasm (MPN).^{1,2} This discovery has had rapid clinical impact, and testing for *CALR* has been embedded in national and international diagnostic guidelines.^{3–5} However, a human MPN-derived cell line harboring a *CALR* mutation has not been reported and the mechanisms by which mutated-*CALR* results in an MPN remain unclear.

To begin to investigate the pathogenetic consequences of mutant *CALR*, we searched for patient-derived cell lines harboring *CALR* mutations. None were identified by exome sequencing of 1015 cell lines, including 37 derived from hematopoietic neoplasms.¹ We therefore looked for cell lines derived from patients with leukemic transformation of a preceding MPN. Given that *CALR* and *JAK2* mutations are almost completely mutually exclusive,^{1,2,6} we focused

REFERENCES

- Lowenberg B, Pabst T, Vellenga E, van Putten W, Schouten HC, Graux C *et al.* Cytarabine dose for acute myeloid leukemia. *N Engl J Med* 2011; **364**: 1027–1036.
- Burnett A, Wetzler M, Lowenberg B. Therapeutic advances in acute myeloid leukemia. *J Clin Oncol* 2011; **29**: 487–494.
- Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 2012; **481**: 506–510.
- Chou WC, Chou SC, Liu CY, Chen CY, Hou HA, Kuo YY *et al.* TET2 mutation is an unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics. *Blood* 2011; **118**: 3803–3810.
- Graham FL, Whitmore GF. Studies in mouse L-cells on the incorporation of 1-beta-D-arabinofuranosylcytosine into DNA and on inhibition of DNA polymerase by 1-beta-D-arabinofuranosylcytosine 5'-triphosphate. *Cancer Res* 1970; **30**: 2636–2644.
- Zahn RK, Muller WE, Forster W, Maidhof A, Beyer R. Action of 1-beta-D-arabinofuranosylcytosine on mammalian tumor cells. 1. Incorporation into DNA. *Eur J Cancer* 1972; **8**: 391–396.
- Liber HL, Thilly WG. Mutation assay at the thymidine kinase locus in diploid human lymphoblasts. *Mutat Res* 1982; **94**: 467–485.
- Fordham SE, Matheson EC, Scott K, Irving JA, Allan JM. DNA mismatch repair status affects cellular response to Ara-C and other anti-leukemic nucleoside analogs. *Leukemia* 2011; **25**: 1046–1049.
- de Wind N, Dekker M, Berns A, Radman M, te Riele H. Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyper-recombination, and predisposition to cancer. *Cell* 1995; **82**: 321–330.
- Gao YG, van der Marel GA, van Boom JH, Wang AH. Molecular structure of a DNA decamer containing an anticancer nucleoside arabinosylcytosine: conformational perturbation by arabinosylcytosine in B-DNA. *Biochemistry* 1991; **30**: 9922–9931.
- Walter MJ, Shen D, Ding L, Shao J, Koboldt DC, Chen K *et al.* Clonal architecture of secondary acute myeloid leukemia. *New Engl J Med* 2012; **366**: 1090–1098.

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/>

on four such lines known to lack a *JAK2* mutation (MONO-MAC-6, MARIMO, GDM-1 and ELF-153), and also tested a further 52 other predominantly myeloid cell lines (Supplementary Table 1). Mutation screening was done by Sanger sequencing as previously described,¹ and details of other methods are in the Supplementary Information.

The only cell line found to harbor a *CALR* mutation was MARIMO, originally derived from a 68-year-old female with AML-M2, and an antecedent history of ET.⁷ MARIMO is negative for *JAK2V617F* and *MPL* exon 10 mutations (data not shown) and carries a heterozygous 61-basepair (bp) deletion in *CALR* exon 9 (c.1099_1159del; L367fs*43), which, like all other reported *CALR* mutations, results in a +1-bp shift in the reading frame and thus generates a novel C terminus (Figure 1a). In patients, the commonest two *CALR* mutations, accounting for 85% of cases, are a 52-bp deletion (type 1; c.1099_1150del; L367fs*46) and a 5-bp insertion (type 2; c.1154_1155_ins;