Published in final edited form as:

Leukemia. 2018 August ; 32(8): 1874–1878. doi:10.1038/s41375-018-0193-y.

Discriminating a common somatic *ASXL1* mutation (c.1934dup; p.G646Wfs*12) from artifact in myeloid malignancies using NGS

Michael O. Alberti¹, Sridhar Nonavinkere Srivatsan², Jin Shao², Samantha N. McNulty¹, Gue Su Chang^{2,3}, Christopher A. Miller^{2,3}, Jennifer B. Dunlap⁴, Fei Yang^{4,5}, Richard D. Press^{4,5}, Qingsong Gao^{2,3}, Li Ding^{2,3,6,7}, Jonathan W. Heusel¹, Eric J. Duncavage¹, and Matthew J. Walter^{2,7}

¹Department of Pathology & Immunology, Washington University School of Medicine, St. Louis, MO, USA;

²Division of Oncology, Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO, USA;

³McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO, USA;

⁴Department of Pathology, Oregon Health & Science University, Portland, OR, USA;

⁵Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA;

⁶Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA;

⁷Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO, USA.

Mutations in the *ASXL1* (*additional sex comb-like 1*) gene are common in several myeloid malignancies, including 14% of patients with myelodysplastic syndrome (MDS); 2–23% with myeloproliferative neoplasms (MPN); 5% with acute myeloid leukemia (AML); and 27–49% with chronic myelomonocytic leukemia (CMML).¹ *ASXL1* mutations have also been identified in patients with clonal hematopoiesis of indeterminate potential (CHIP)² and are associated with adverse survival in MDS, MPN, CMML, and AML.^{3, 4}

Detection of the most common *ASXL1* mutation (NM_015338:c.1934dup; p.G646Wfs*12) has been controversial. An early report suggested this *ASXL1* duplication mutation could be the result of PCR errors as the mutation was variably present in matched germline tissue and ~25% of healthy controls.⁵ The c.1934dup mutation represents a one nucleotide expansion

CORRESPONDENCE: M.J.W. (mjwalter@wustl.edu).

AUTHOR CONTRIBUTIONS

Conception and design: MOA, EJD, MJW; methodology: MOA, SNS, EJD, CAM, QG, LD, MJW; data acquisition and analysis: MOA, SNS, JS, SNM, GSC, JBD, FY, RDP; interpretation of data: MOA, SNS, MJW; writing: MOA and MJW with all authors contributing to and approving the final manuscript; project supervision: JWH, EJD, MJW.

Publisher's Disclaimer: This is a PDF file of an unedited peer-reviewed manuscript that has been accepted for publication. Springer Nature are providing this early version of the manuscript as a service to our customers. The manuscript will undergo copyediting, typesetting and a proof review before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers apply.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

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

of a contiguous (or ‘homopolymer’) repeat of eight guanine nucleotides (GGGGGGGG) in exon 12 of *ASXL1*, and it is therefore possible that it represents an artifact by slipped-strand mispairing of the DNA polymerase during PCR amplification. While this artifact may occur, subsequent reports using Sanger sequencing identified the c.1934dup mutation in blood or marrow from patients but not in matched germline samples⁶ or healthy controls.⁷ More recently, it was reported that this mutation could also be identified by quantitative PCR (qPCR).⁸ As patient care decisions in AML and MDS are being made increasingly based on the results of gene-panel based testing and as next-generation sequencing (NGS) methodologies represent the new standard for clinical testing,⁴ a clear approach is needed to accurately identify the *ASXL1* c.1934dup variant as an artifact or a somatic mutation. Therefore, we examined the c.1934dup variant in data generated using Illumina sequencing from (1) paired tumor (bone marrow) and normal (skin) DNA from patients with MDS ($n=62$), and (2) tumor-only DNA from 921 patients (161 diagnosed with a hematological malignancy).

MDS patients were consented for sequencing studies on a protocol approved by the Human Research Protection Office at Washington University. Genomic DNA was extracted from paired bone marrow and skin (a surrogate for normal DNA) and enriched for all coding exons of a panel of 285 recurrently mutated genes in myeloid disease, including *ASXL1*, as previously described.⁹ Each of the 62 patients was previously identified to have a hotspot mutation in one of three spliceosome genes: *SF3B1* ($n=8$), *SRSF2* ($n=31$), and *U2AF1* ($n=23$) (Figure 1). Captured libraries were sequenced on a HiSeq 2500 (Illumina, San Diego, CA, USA) instrument, and sequencing data were aligned and analyzed for detection of single nucleotide variants and insertions/deletions (INDELs) with matched normal DNA (see Supplementary Methods). Manual review of sequencing read alignments in the region of *ASXL1* containing the eight guanine homopolymer repeat (chr20:31,022,441–31,022,450; GRCh37/hg19) for all patients revealed 19 INDELs with at least 50x total coverage in both the marrow and skin samples, with 5 sequencing reads containing variant alleles and 0.05 (5%) variant allele fraction (VAF) in the marrow. Of these INDELs, 17 were identified as c.1934dup, in addition to one instance each of c.1932_1933insC and c.1932_1933delinsT. We queried the alignment files for all 62 paired samples using the bioinformatics tool bam-readcount (see Supplementary Methods) to assess the fraction of reads containing INDELs of varying lengths (e.g., delGG [6G], delG [7G], dupG [9G], dupGG [10G]). Using this approach, we consistently found rare delG and dupG reads in nearly every skin and marrow sample, however in most cases these reads comprised <5% of the total counted reads and were equally represented by delG and dupG reads (Figures 1a–b). In contrast, we found that the c.1934dup variants identified by our standard pipeline were present in 5% of the counted reads from the marrow, as expected, and generally <5% of the counted reads from the skin (Figures 1a–b), further indicating that these c.1934dup variants were *bona fide* somatic mutations. In addition, cases with a c.1934dup mutation contained a higher fraction of insertion (dupG) reads compared to deletion (delG) reads (INS-DEL read fractions >0.03; Figure 1c) and did not show a strand bias (plus-strand bias 0.35–0.65; Figure 1d), in contrast to the skin and marrow samples from non-mutated patients. While this homopolymer repeat in *ASXL1* produces a small background level of both delG and dupG INDEL reads using the Illumina platform, somatic mutations are distinguishable from

background and are easily identified using a set of simple metrics, including VAF, strand bias, and insertion versus deletion (INS-DEL) reads.

Using these data, we observed an enrichment of *ASXL1* mutations with *U2AF1* or *SRSF2* mutations in our cohort of MDS patients with spliceosome gene mutations, as previously reported.¹⁰ However, *U2AF1* mutations coding for the p.Q157(P/R/H) amino acid substitution co-occurred at a higher frequency with *ASXL1* mutations, regardless of the c.1934dup variant, compared to *U2AF1* mutations coding for p.S34(F/Y) (Figures 1a; $p=0.0272$, Fisher's exact test). This enrichment for *ASXL1* mutations with *U2AF1* p.Q157 mutations is consistent across several MDS sequencing cohorts,^{10–13} suggesting a unique cooperation between the type of *U2AF1* mutation and *ASXL1* (Supplementary Figure S1 and Figure 2a; $p<0.0001$, Fisher's exact test).

Next, we validated that these dupG variants were somatic and not sequencing artifacts using Sanger sequencing as an orthogonal platform and DNA from paired marrow-skin samples from c.1934dup cases, the two non-dupG *ASXL1* INDEL cases (c.1932_1933insC, c.1932_1933delinsT), and 17 other randomly selected cases that did not show evidence of the c.1934dup mutation by NGS (see Supplementary Methods). All NGS-identified c.1934dup mutations (and two other INDELs) interrogated were identified in the marrow but not the paired skin samples using Sanger sequencing (Supplementary Figure S2A and Figure 2b). In addition, we found no evidence of convincing INDELs in the skin or marrow from any of the 17 non-mutated control samples (Supplementary Figure S2A and Figure 2b). However, we did observe a low nonreproducible background level of sequencing artifact after the eighth guanine nucleotide present in samples tested (Supplementary Figure S2 and Figure 2c). Finally, subjects with c.1934dup mutations also had VAFs that were highly reproducible from independent NGS library preparations (Figure 2d), again supporting that these mutations are not artifacts.

We extended these findings using bam-readcount to assess the c.1934dup variant in 921 cases diagnosed with a hematological malignancy ($n=161$), a solid tumor ($n=595$), or suspected somatic overgrowth spectrum disorder (SOMA) ($n=165$) that were submitted to the Genomics and Pathology Services (GPS) at Washington University. For GPS cases, genomic DNA was extracted from submitted tumor tissue (FFPE, bone marrow, peripheral blood, or fresh tissue) and enriched for all coding exons of a panel of 130 or 178 recurrently mutated cancer genes, including *ASXL1*, as previously described.¹⁴ Captured libraries were sequenced on a HiSeq 2500 (Illumina) instrument and sequencing data were analyzed using bam-readcount to identify *ASXL1* c.1934dup variants.¹⁴ While we could find low levels of both delG and dupG reads in nearly every sample, as discussed above, these reads typically comprised <5% of the total counted reads (Supplementary Figure S3). We identified 13 *ASXL1* c.1934dup mutations that were present in 5% of the total counted reads from the tumor, and as expected, all but two of these were identified in myeloid malignancy samples (Supplementary Figure S3). Six additional *ASXL1* INDELs were identified, including c.1933_1934del ($n=1$), c.1934del ($n=3$), and c.1933_1934dup ($n=2$), and confirmed by manual review. Overall, 84.6% of the c.1934dup mutations were identified in myeloid malignancy patients, and 9.4% (11 of 117) of patients with myeloid malignancies had a c.1934dup mutation (Supplementary Figure S3A). In contrast, only 0.3% (2 of 595 patients) with solid

tumors (Supplementary Figure S3B) and no SOMA patients (Supplementary Figure S3C) had evidence of the c.1934dup mutation, as expected.

Collectively, these data demonstrate that the most common *ASXL1* variant is a *bona fide* somatic mutation and can generally be easily identified and discriminated from background ‘noise’ in Illumina sequencing data using a set of simple metrics, including VAF threshold, strand bias, and insertion versus deletion reads. As gene-panel based testing becomes more common in the clinical setting, appropriate platform-dependent calling of the common *ASXL1* c.1934dup variant should be incorporated to ensure uniform reporting of this variant. In this regard, analysis of semiconductor-based sequencing data that is known to have reduced sequencing accuracy at genomic loci with homopolymer repeats of the same nucleotide, such as the c.1934dup mutation, may need its own optimization that is distinct from the metrics we have described here.¹⁵ Consistent with this, while the background levels of dupG or delG reads from a small number of semiconductor-based sequencing alignment files made it difficult to call this variant using the Illumina-based metrics described here (Supplementary Figure S4), the variant was called correctly using the Torrent Variant Caller (Thermo Fisher), and confirmed as a true variant by Sanger sequencing. Collectively, the data emphasize that platform-dependent variant calling parameters need to be optimized to identify this mutation. The true frequency of the *ASXL1* c.1934dup variant in myeloid malignancies and CHIP may be greater than previously reported.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We wish to thank David Spencer, Chad Storer, and Catherine Cottrell for helpful discussions; and Felicitas Thol, Torsten Haferlach, Elli Papaemmanuil, Olivier Kosmider, and R. Coleman Lindsley for discussions and data related to their publications. Support was provided to MJW through a SPORE in Leukemia grant (P50CA171963) from the National Institutes of Health/National Cancer Institute (NIH/NCI), the Edward P. Evans Foundation, the Lottie Caroline Hardy Trust, and a Leukemia and Lymphoma Society Scholar Award. Support for procurement of human samples was provided by an NIH/NCI grant (P01CA101937). Technical assistance was provided by the Alvin J. Siteman Cancer Center Tissue Procurement Core supported by an NCI Cancer Center Support Grant (P30CA91842).

REFERENCES

1. Gelsi-Boyer V, Trouplin V, Adelaide J, Bonansea J, Cervera N, Carbuccion N, et al. Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. *Br J Haematol* 2009; 145: 788–800. [PubMed: 19388938]
2. Steensma DP, Bejar R, Jaiswal S, Lindsley RC, Sekeres MA, Hasserjian RP, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* 2015; 126: 9–16. [PubMed: 25931582]
3. Gelsi-Boyer V, Brecqueville M, Devillier R, Murati A, Mozziconacci MJ, Birnbaum D. Mutations in ASXL1 are associated with poor prognosis across the spectrum of malignant myeloid diseases. *J Hematol Oncol* 2012; 5: 12. [PubMed: 22436456]
4. Dohner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Buchner T, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 2017; 129: 424–447. [PubMed: 27895058]

5. Abdel-Wahab O, Kilpivaara O, Patel J, Busque L, Levine RL. The most commonly reported variant in ASXL1 (c.1934dupG;p.Gly646TrpfsX12) is not a somatic alteration. *Leukemia* 2010; 24: 1656–1657. [PubMed: 20596031]
6. Metzeler KH, Becker H, Maharry K, Radmacher MD, Kohlschmidt J, Mrozek K, et al. ASXL1 mutations identify a high-risk subgroup of older patients with primary cytogenetically normal AML within the ELN Favorable genetic category. *Blood* 2011; 118: 6920–6929. [PubMed: 22031865]
7. Thol F, Friesen I, Damm F, Yun H, Weissinger EM, Krauter J, et al. Prognostic significance of ASXL1 mutations in patients with myelodysplastic syndromes. *J Clin Oncol* 2011; 29: 2499–2506. [PubMed: 21576631]
8. Yannakou CK, Jones K, McBean M, Thompson ER, Ryland GL, Doig K, et al. ASXL1 c.1934dup;p.Gly646Trpfs*12-a true somatic alteration requiring a new approach. *Blood Cancer J* 2017; 7: 656. [PubMed: 29242575]
9. Uy GL, Duncavage EJ, Chang GS, Jacoby MA, Miller CA, Shao J, et al. Dynamic changes in the clonal structure of MDS and AML in response to epigenetic therapy. *Leukemia* 2017; 31: 872–881. [PubMed: 27740633]
10. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* 2014; 28: 241–247. [PubMed: 24220272]
11. Thol F, Kade S, Schlarmann C, Loffeld P, Morgan M, Krauter J, et al. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood* 2012; 119: 3578–3584. [PubMed: 22389253]
12. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* 2013; 122: 3616–3627; quiz 3699. [PubMed: 24030381]
13. Lindsley RC, Saber W, Mar BG, Redd R, Wang T, Haagenson MD, et al. Prognostic Mutations in Myelodysplastic Syndrome after Stem-Cell Transplantation. *N Engl J Med* 2017; 376: 536–547. [PubMed: 28177873]
14. Carter JH, McNulty SN, Cimino PJ, Cottrell CE, Heusel JW, Vigh-Conrad KA, et al. Targeted Next-Generation Sequencing in Molecular Subtyping of Lower-Grade Diffuse Gliomas: Application of the World Health Organization's 2016 Revised Criteria for Central Nervous System Tumors. *J Mol Diagn* 2017; 19: 328–337. [PubMed: 28042970]
15. Feng W, Zhao S, Xue D, Song F, Li Z, Chen D, et al. Improving alignment accuracy on homopolymer regions for semiconductor-based sequencing technologies. *BMC Genomics* 2016; 17 Suppl 7: 521. [PubMed: 27556417]

Figure 1.

The *ASXL1* c.1934dup (p.G646Wfs*12) mutation is readily discriminated from artifact using paired tumor-normal NGS data from a cohort of 62 MDS patients. **(a)** Fraction of sequencing reads supporting delG/GG (red), dupG/GG (green), or reference (gray) alleles in paired normal (top) and tumor (bottom) tissues. **(b)** Number of sequencing reads supporting reference or variant alleles in tumor tissue. **(c)** Difference in fraction of reads supporting insertion alleles (dupG+dupGG) and deletion alleles (delG+delGG) from either normal (orange squares) or tumor (blue circles) tissues. **(d)** Fraction of reads on plus-strand that

support reference (gray circles) or dupG (green squares) alleles in tumor tissue. Data for each patient are aligned in a single column for **a-d**. Patients are organized by *U2AF1* and *ASXL1* mutation status and ranked from left to right by the absolute difference of tumor and normal INS-DEL fractions (**c**).

Figure 2.

Orthogonal evidence confirming the *bona fide* somatic nature of the *ASXL1* c.1934dup mutation and association of *ASXL1* and *U2AF1* p.Q157 mutations. **(a)** *ASXL1* mutations co-occur more frequently with *U2AF1* p.Q157(P/R/H) mutations compared to *U2AF1* p.S34(F/Y) in 270 *U2AF1*-mutant MDS patients combined from four studies^{10–13} and the data presented herein (see also Supplementary Figure 1). **(b)** Examples of Sanger sequencing results of paired normal and tumor tissues from a patient with the c.1934dup mutation (UPN 148296) or without the mutation ('Non-mutated', UPN 122256) (see also Supplementary Figure 2). **(c)** Example of the variance between technical replicate sequence traces derived from a single non-mutated patient (UPN 311093) sample. **(d)** Correlation of *ASXL1* c.1934dup VAFs from two independent NGS library preparation and sequencing replicates for MDS patients with the c.1934dup mutation (green triangles) or negative for the mutation (grey circles). In **b–c**, the filled arrowhead (▲) indicates the location of the dupG.