UC San Diego
UC San Diego Previously Published Works

Title

Refining the serum miR-371a-3p test for viable germ cell tumor detection.

Permalink

https://escholarship.org/uc/item/15v5t0m6

Journal

Scientific Reports, 13(1)

Authors

Lafin, John
Scarpini, Cinzia
Amini, Armon
et al.

Publication Date

2023-06-29

DOI

10.1038/s41598-023-37271-1

Peer reviewed

scientific reports

OPEN

Refining the serum miR-371a-3p test for viable germ cell tumor detection

John T. Lafin[1], Cinzia G. Scarpini[2], Armon Amini[1], Bendu Konneh[1], Jeffrey M. Howard[1], Thomas Gerald[1], Michelle Nuno[3], Jin Piao[3], Anna Savelyeva[1], Zhaohui Wang[4], Jeffrey Gagan[4], Liwei Jia[4], Cheryl M. Lewis[4], Sarah Murray[5], Yun C. Sawa[6], Vitaly Margulis[1], Solomon L. Woldu[1], Douglas W. Strand[1], Nicholas Coleman[2,7], James F. Amatruda[8,9], A. Lindsay Frazier[10], Matthew J. Murray[11,2] & Aditya Bagrodia[1,6✉]

Circulating miR-371a-3p has excellent performance in the detection of viable (non-teratoma) germ cell tumor (GCT) pre-orchiectomy; however, its ability to detect occult disease is understudied. To refine the serum miR-371a-3p assay in the minimal residual disease setting we compared performance of raw (Cq) and normalized (ΔCq, RQ) values from prior assays, and validated interlaboratory concordance by aliquot swapping. Revised assay performance was determined in a cohort of 32 patients suspected of occult retroperitoneal disease. Assay superiority was determined by comparing resulting receiver-operator characteristic (ROC) curves using the Delong method. Pairwise t-tests were used to test for interlaboratory concordance. Performance was comparable when thresholding based on raw Cq vs. normalized values. Interlaboratory concordance of miR-371a-3p was high, but reference genes miR-30b-5p and cel-miR-39-3p were discordant. Introduction of an indeterminate range of Cq 28–35 with a repeat run for any indeterminate improved assay accuracy from 0.84 to 0.92 in a group of patients suspected of occult GCT. We recommend that serum miR-371a-3p test protocols are updated to (a) utilize threshold-based approaches using raw Cq values, (b) continue to include an endogenous (e.g., miR-30b-5p) and exogenous non-human spike-in (e.g., cel-miR-39-3p) microRNA for quality control, and (c) to re-run any sample with an indeterminate result.

Correct staging in early stage germ cell tumor (GCT) patients is critical for identifying patients best served with surveillance versus primary management with retroperitoneal lymph node dissection (RPLND), chemotherapy, or radiotherapy[1, 2]. In patients with clinical stage I (CS I) GCT, up to 97% of seminomas and 60% of non-seminomas recur on surveillance without marker elevation[3–5]. Additionally, 26% of patients with negative serum tumor markers (STM) and cross-sectional imaging undergoing RPLND are found to have viable tumor[6]. Consequently, the performance characteristics of current STM introduces substantial risk of under- and over-treatment.

The superior performance of circulating microRNAs (miRNAs), particularly miR-371a-3p, to detect GCT is well documented. An agreed, protocolized standard for definition of positive and negative miR-371a-3p results is lacking. The absence of a standard protocol in combination with the inherent sensitivity of the test has contributed to interlaboratory heterogeneity, making comparisons difficult and limiting widespread clinical adoption[7].

We address these issues by performing interlaboratory sample exchange experiments and re-evaluating analytic pipelines for calling results. In addition to positive and negative calls, we identify an indeterminate range, which we then validate in an independent patient cohort undergoing primary RPLND. These changes improve

[1]Department of Urology, University of Texas Southwestern Medical Center, Dallas, USA. [2]Department of Pathology, University of Cambridge, Cambridge, UK. [3]Department of Population and Public Health Sciences, University of Southern California, Los Angeles, USA. [4]Department of Pathology, University of Texas Southwestern Medical Center, Dallas, USA. [5]Department of Pathology, University of California San Diego, San Diego, USA. [6]Department of Urology, University of California San Diego, Suite 1-200, 9400 Campus Point Drive, La Jolla, CA 92037, USA. [7]Department of Histopathology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [8]Cancer and Blood Disease Institute, Children's Hospital Los Angeles, Los Angeles, USA. [9]Departments of Pediatrics and Medicine, Keck School of Medicine, University of Southern California, Los Angeles, USA. [10]Dana-Farber/Boston Children's Cancer and Blood Disorders Center, Boston, USA. [11]Department of Pediatric Hematology and Oncology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. ✉email: Bagrodia@health.ucsd.edu

assay performance, particularly specificity and negative predictive value (NPV), which upon clinical implementation will reduce potential over-treatment of patients without true minimal residual disease.

Methods

Patient population.   Thirty-two chemotherapy-naïve patients underwent primary RPLND for clinical stage I or II GCT. Serum was obtained immediately prior to RPLND. Bilateral full-template or extended modified template nerve-sparing RPLND was per surgeon discretion. Baseline clinicopathologic data were collected. Samples were classified as either 'Control' (pure teratoma or no GCT), or 'Viable GCT' [seminoma or nonseminomatous GCT (NSGCT)].

All experimental protocols were approved by an Institutional Review Board at The University of Texas Southwestern Medical Center (STU 102010-051). Informed consent was obtained from all subjects and/or their legal guardians prior to their inclusion in the study. The authors confirm that all methods described in this manuscript were performed in accordance with the relevant guidelines and regulations.

MiRNA isolation and quantification.   RNA extraction and serum miRNA quantification by qPCR (quantitative polymerase chain reaction) were performed as described[8]. Primers and probes used are detailed in Supplementary Table 1. To calculate relative quantification (RQ), the ΔΔCq method was used, with the mean of four normal control human male serum samples (males between age 18–45 years) used as reference.

Concordance studies.   Serum aliquots were shipped between the two research laboratories of Cambridge, UK and University of Texas Southwestern, US priority overnight on dry ice. Upon receipt, sample inspection confirmed that none had thawed. Each site followed an identical protocol to yield raw Cq and normalized (ΔCq and RQ) values, which were then compared against one another.

Cq vs. RQ performance.   Raw (Cq values) and normalized (ΔCq and RQ values) data from two studies previously published from our group were utilized[9,10]. Optimal thresholds were calculated for each metric using the Youden index[11] and sensitivity, specificity, and area under the receiver-operating characteristic curve (AUC) were calculated.

Establishment and assessment of an indeterminate range.   All runs included in our two previous reports[9,10], including any technical replicate runs undertaken, were pooled and grouped based on histology (Control or Viable GCT). An indeterminate range was defined as the 95% confidence interval of the distribution of the first (lower Cq, higher apparent abundance) raw Cq peak, rounded to whole numbers (down at the lower bound and up at the upper bound) and subsequently formally assessed for change in assay performance.

Statistical analysis.   Statistical significance for intergroup differences of clinicopathologic data was determined using the Kruskal–Wallis test with Dunn's post-hoc test. Concordance was assessed by a pairwise t-test. Performance characteristics, including sensitivity, specificity, NPV, positive predictive value (PPV), accuracy, and AUC were calculated using R version 4.1.2 with the pROC package (version 1.18.0) and tidyverse metapackage (version 1.3.1)[12–14]. AUC values were compared using the roc.test function in pROC with default parameters. Two-tailed $p < 0.05$ was statistically significant.

Results

Thresholding on Cq simplifies the serum miR-371a-3p test without affecting assay performance.   The requirement for a normal control serum sample in each assay run for normalization is costly and adds another potential source of variation. To determine if assay normalization is required, we examined our previously published data from samples taken pre-orchiectomy[10] and pre-RPLND[9]. We examined four metrics with varying levels of normalization- Cq (raw value), ΔCq (Cq normalized to internal control miR-30b-5p), corrected ΔCq (ΔCq corrected with an external control cel-miR-39-3p), and RQ (corrected ΔCq of sample normalized to corrected ΔCq of normal serum).

Calculated sensitivity and specificity were both greater than 0.9 in all cases and did not change appreciably across any of the metrics tested, Table 1. AUC was 0.97–0.99 for all four metrics, and none were statistically different from one another (all $p > 0.05$). These results suggest that normalization to endogenous or exogenous controls, or normal healthy serum, does not impact the performance of the serum miR-371a-3p assay.

To examine interlaboratory variation, we conducted a concordance study between the two laboratories. Aliquots of 24 serum samples were exchanged, and both sites ran identical protocols. miR-371a-3p Cq was highly concordant, with a mean difference of < 0.5 cycles between sites ($p = 0.251$) (Fig. 1). The exogenous non-human spike-in control cel-miR-39-3p was discordant ($p = 0.002$), likely due to separate preparations of highly concentrated standards. Surprisingly, the endogenous control, miR-30b-5p, was also discordant ($p < 0.001$). These results suggest that this normalization process introduces additional variation and contributes to interlaboratory heterogeneity. We therefore recommend use of raw Cq values for cutoffs for the serum miR-371a-3p test going forwards.

Identification and establishment of an indeterminate range.   The serum miR-371a-3p test is extremely sensitive, due in part to the pre-amplification step used prior to qPCR, which also exposes to risk of false positives. This risk is already heightened by the need to open PCR tubes following pre-amplification to set-up the qPCR, which may inadvertently spread amplification products. The inclusion of a water ('no template') control (NTC) sample initiated at the reverse transcription step is recommended to combat this—a positive

	Cq	ΔCq	RQ
Orchiectomy (n = 69)			
Threshold	32.75	18.03	23.46
AUC	0.98	0.98	0.98
Sensitivity	0.93	0.91	0.93
Specificity	1.00	1.00	1.00
1° RPLND (n = 24)			
Threshold	36.35	21.92	20.59
AUC	0.97	0.97	0.97
Sensitivity	1.00	1.00	0.92
Specificity	1.00	0.92	0.92

Table 1. Performance metrics of raw and normalized values of serum miR-371a-3p test from previously published results[9,10].

qPCR result on NTC suggests such upstream contamination. However, we noted occasional cases where known control samples would yield an inconsistent/stochastic positive result despite a negative NTC sample result on the same qPCR run. Repeating these samples from the reverse transcription step usually yielded the anticipated negative result. In contrast, repeating runs on samples from patients with pathologically verified disease typically returned similar Cq values. Examples of repeated runs for pathologic negative and positive samples are presented in Supplementary Fig. 1.
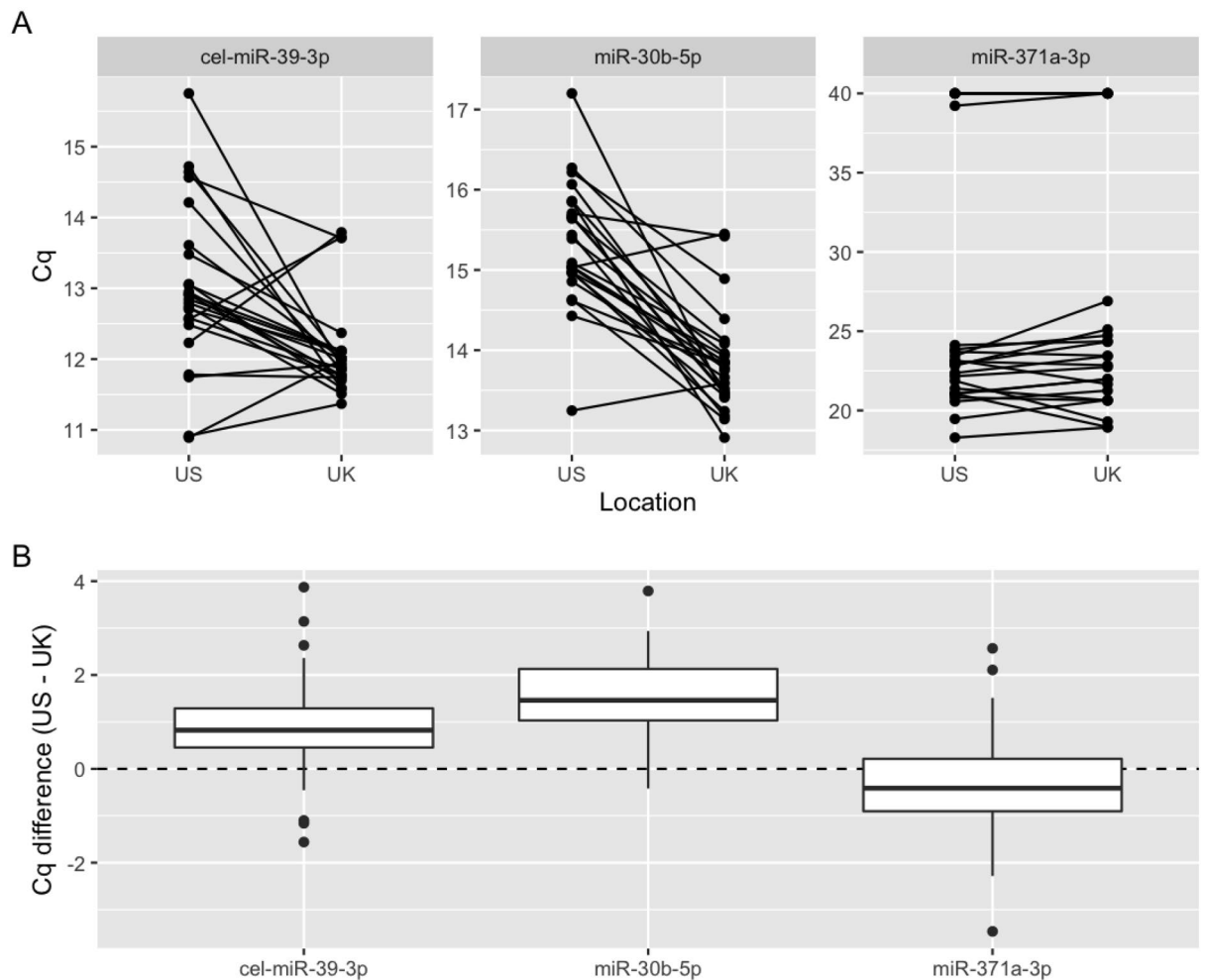
Figure 1. Raw miR-371a-3p Cq is concordant across laboratory sites. (**A**) Comparison of raw Cq values for cel-miR-39-3p (external spike-in control), miR-30b-5p (internal reference gene), and miR-371a-3p at the US and UK sites. (**B**) Boxplot of Cq differences (US − UK). n = 24.

To investigate the above observation, we aggregated a total of 150 runs from our previously published studies[9,10]. We examined the distribution of Cq values split by group, Control vs. Viable GCT, Supplementary Fig. 2A. Individual sample Cq values are displayed in Supplementary Fig. 2B. The samples in the Viable GCT group show a broad distribution with a mean Cq and standard deviation (SD) of 26.4 ± 4.33. This wide distribution is expected given the heterogenous population with differing amounts of disease burden. However, the distribution of Cq values in the Control group appeared to be bimodal, with the mean Cq of the first peak at 32.2 ± 1.53, and the mean Cq of the second peak at 39.8 ± 0.7. The mean of the second peak is anticipated, as undetected samples are assigned Cq of 40. We were surprised that approximately 25% of all runs in the Control group fell into the first peak. Two separate research laboratories (Cambridge, UK; UTSW, Dallas, US) and one clinical laboratory (Department of Pathology, UTSW, Dallas, USA) all independently reported this observation, indicating that this is unlikely to be due to technical errors. We have not found any reliable predictor for this assay behavior; it appears to be an entirely stochastic and non-predictable event. This suggests that as currently applied, the qPCR-based serum miR-371a-3p assay has an approximately 25% chance to misclassify any true negative as positive.

Mitigation of this misclassification is critical prior to clinical implementation of the test. We reasoned that defining an 'indeterminate' range based on the first distribution and repeating the qPCR for any sample that fell into that range would reduce misclassification from ~ 25 to ~ 6% (0.25 × 0.25 = 0.0625). Based on our established assay pipeline, we defined the indeterminate range as Cq 28–35, which approximates the mean of the first Cq peak ± 2 SDs in the controls. We then interrogated our aggregated data again to simulate how application of this revised methodology might improve viable GCT classification. To simulate the original methodology, the first chronological run per sample was selected. To simulate our revised methodology, the first chronological run per sample was selected unless its result fell into the indeterminate range (28 < Cq < 35). If so, the second chronological run was selected. Any sample that remained indeterminate after the second run was classified 'indeterminate' and removed from performance calculations. With this model, the original method had 81 runs. In the revised method, nine samples (11.1%) had two indeterminate results and were classified as truly indeterminate, leaving 72 runs. Two of these nine samples were in the Control group, and the remaining seven were in the Viable GCT group. We then compared the resulting Cq distributions (Fig. 2A,B). Application of the revised methodology prevented six false positives with accuracy improved from 0.85 to 0.93, and AUC from 0.909 to 0.954 (Fig. 2C,D and Supplementary Table 2). False positives in the Control group declined from 8/23 (34.8%) to 2/23 (8.7%), supporting the observation that this event is stochastic in nature.

Application of revised methodology to an updated primary RPLND dataset.
Improved performance of the serum miR-371a-3p test would allow for both early detection of recurrence and avoidance of unnecessary treatment. The detection of minimal residual disease (MRD) therefore carries great clinical significance in this context. As serum miR-371a-3p Cq is correlated with tumor burden, detection of MRD demands the greatest performance of this test. We therefore expanded a cohort of chemotherapy naïve patients receiving primary RPLND and compared the performance of the original and revised methodology.

Patient characteristics are summarized in Table 2. Thirty-two patients receiving primary RPLND were included in the present analysis. Most patients were clinical stage (CS) II (62.5%); 37.5% were CSI. At RPLND, nine patients (28.1%) had no viable tumor, 12 patients (37.5%) had pure seminoma, and 11 patients (34.4%) had non-seminomatous GCT. Pathologic stage (PS) was PS I in 28.1% and PS II in 71.9%.

The median Cq for the Control group was 40 under the original and revised methodology. Median Cq for the Viable GCT group shifted from 27.7 under the original methodology to 26.2 under the revised methodology. After applying the revised method, eight samples remained truly indeterminate, which were removed from further analysis, Fig. 3A,B. Three of these samples were in the Control group, all of which harbored pure teratoma. The remaining five indeterminate samples were in the Viable GCT group. The AUC was 0.898 (95% CI 0.79–1.00) with the original method and 0.934 (95% CI 0.84–1.00) with the revised method, Fig. 3C. Application of the revised methodology improved most other metrics, including specificity (0.80–0.92) and PPV (0.83–0.92) (Fig. 3D and Supplementary Table 3).

Discussion
We report the use of raw circulating miR-371a-3p Cq values, instead of normalized data, for optimal assay performance with excellent interlaboratory concordance. qPCR assays are extensively and routinely used in clinical laboratories and often report results using raw Cq. Introduction of a normalization procedure increases costs and hampers translation into routine clinical testing. Due to the very high sensitivity of the circulating miRNA assay for viable GCT, we believed that additional normalization would be necessary to control for variation between runs. However, results from identical samples run in two independent laboratories suggest normalization may be harmful. The addition of these normalization procedures introduces additional technical variation due to the discordance of reference genes (cel-miR-39-3p and miR-30b-5p) without performance benefits.

Other groups used raw data in their assessments and retained high performance[15,16]. However, assays used by these groups differ materially (e.g., the use of plasma extracts, detection by droplet digital PCR (ddPCR), and/or no pre-amplification). Since the largest miRNA studies to date, including a commercially available assay (miRdetect), were conducted with a serum qPCR-based method with pre-amplification, we felt it important to replicate these studies using this particular methodology.

Critically, we have identified and established an indeterminate range to maintain assay performance of the circulating miR-371a-3p test. This arises from the observation in three separate laboratories that any given negative sample has an approximately 25% random or stochastic chance to return a spurious positive result. The existence of this reproducibility issue is further supported by an independent study reporting the existence
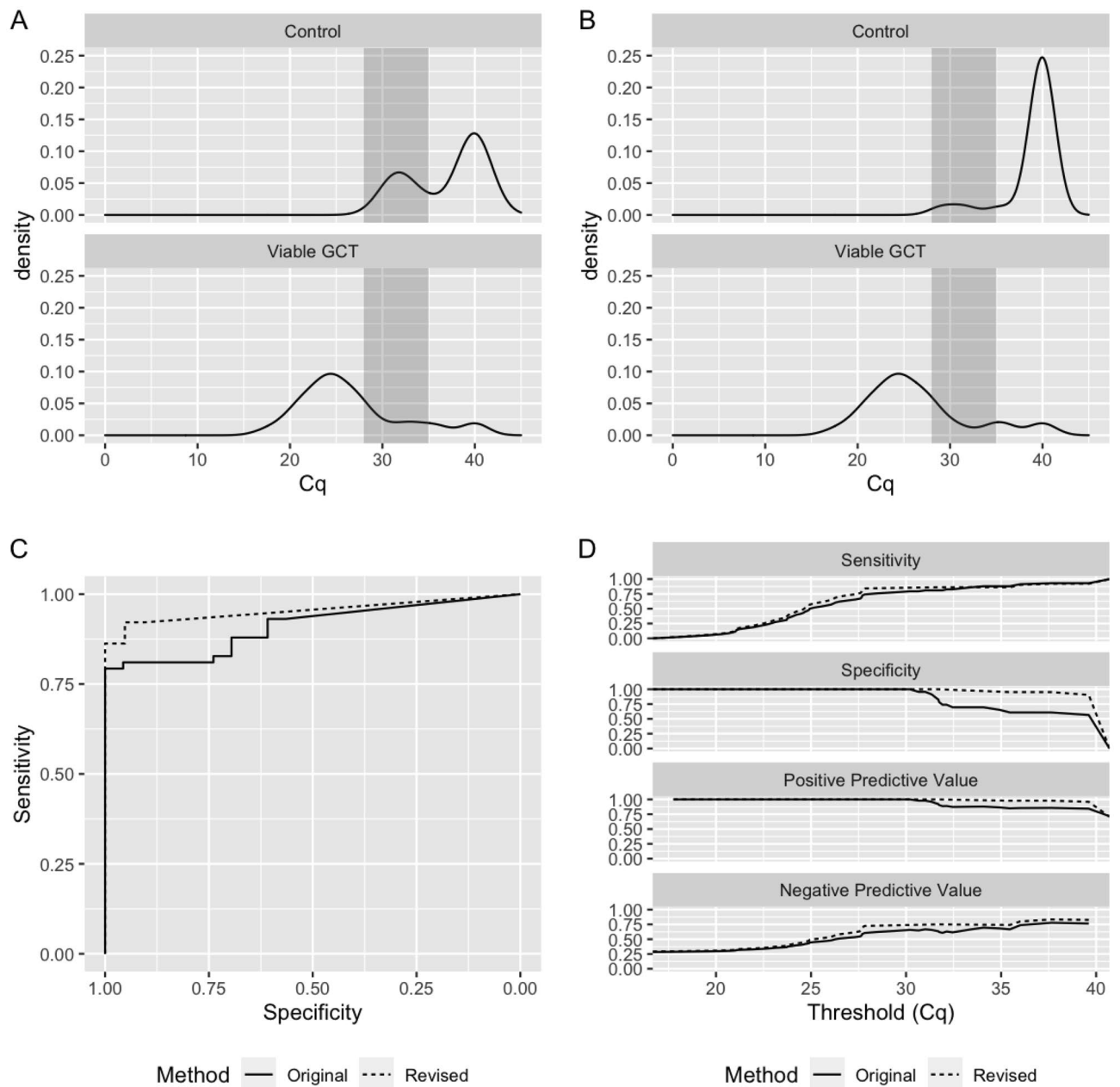
Figure 2. Establishment of indeterminate range in the serum miR-371a-3p test. (**A,B**) Density plots of serum miR-371a-3p Cq values (n = 81) under original (**A**) or revised (**B**) method. Shaded region corresponds to indeterminate range (Cq 28–35). (**C**) ROC plot showing performance of original (solid line) and revised (dashed line) methodology. (**D**) Comparison of sensitivity, specificity, positive predictive value, and negative predictive value across all measured Cq thresholds using original (solid line) and revised (dashed line) methodology.

of an indeterminate range in normalized values[17]. Additionally, Christiansen et al. recently reported that the inclusion of the pre-amplification step improved sensitivity but also led to more false positives[18]. Dropping the assay cutoff below the first distribution would lead to an unacceptable drop in sensitivity. Instead, we elect to define an indeterminate range and rerun any indeterminate extract (Fig. 4). We have observed that upon repeat, most true positive samples will maintain a Cq value very close to the first run, while most true negative samples will yield a negative result. Because outcomes for viable GCT tend to be positive even in the case of recurrence, we recommend classification of any sample that returns an indeterminate result twice as a true indeterminate. In this clinical scenario, there is comparatively greater patient cost to over-treat than under-treat. Application of our revised method to an expanded cohort of patients with MRD improved specificity and PPV, demonstrating that these changes could prevent over-treatment. Although we found that the range of 28–35 was appropriate for our data, we recommend each laboratory to determine their own range, as this may vary slightly due to technical differences.

Because many groups use a similar or identical protocol for this test, the question arises as to why this indeterminate range has not previously been described in detail. One contributing factor may be that larger

Age	Years	Median (IQR)	28 (23.5–35.0)
Race/ethnicity	White Non-Hispanic	N (%)	18 (56.3)
	Hispanic		13 (40.6)
	Other		1 (3.1)
Body Mass Index	kg/m^2	Median (IQR)	27.4 (23.7–29.8)
Primary tumor size	cm	Median (IQR)	3.3 (2.2–6.0)
Primary histopathology	Seminoma	N (%)	9 (28.1)
	Non-seminoma		21 (65.6)
	Burnt-out primary		2 (6.3)
Primary tumor lymphovascular invasion (LVI)	No	N (%)	21 (65.6)
	Yes		11 (34.4)
Primary tumor rete testis invasion (RTI)	No	N (%)	20 (62.5)
	Yes		12 (37.5)
pT stage	pT0	N (%)	2 (6.3)
	pT1		14 (43.8)
	pT2		16 (50.0)
cN stage	cN0	N (%)	12 (37.5)
	cN1		15 (46.9)
	cN2		4 (12.5)
	cN3		1 (3.1)
S stage	S0	N (%)	23 (71.9)
	S1		9 (28.1)
Clinical stage (CS)	CS I	N (%)	12 (37.5)
	CS II		20 (62.5)
RPLND histopathology	Benign	N (%)	9 (28.1)
	Seminoma		12 (37.5)
	Non-seminoma		11 (34.4)
pN stage	pN0	N (%)	9 (28.1)
	pN1		11 (34.4)
	pN2		11 (34.4)
	pN3		1 (3.1)
Pathologic stage (PS)	PS I	N (%)	9 (28.1)
	PS II		23 (71.9)
Previously published [9]	Yes	N (%)	24 (75)
	No		8 (25)

Table 2. Patient characteristics of minimal residual disease dataset.

retrospective non-blinded studies using this serum qPCR-based assays are focused on testicular GCT rather than retroperitoneal disease. Because circulating miR-371a-3p levels are dependent upon tumor burden, circulating miR-371a-3p is anticipated to be weakly positive in the context of MRD, rendering cutoff selection difficult. For example, the median Cq value for Viable GCT patients in our orchiectomy cohort[10] was 26.6, below the indeterminate range. However, the median Cq for our original primary RPLND cohort[9] was 29.3, within the indeterminate range. Additionally, a small number of spurious positive results in a control group may be written off as technical error and/or potential contamination, and the qPCR run repeated several times, subsequently yielding negative results. This enforces the utility of blinding technicians and analysts when conducting assays.

Conclusion

We recommend three important modifications to serum miR-371a-3p assay protocols going forwards: (1) revise the test by applying cutoffs to raw Cq values instead of normalized values; (2) include endogenous (e.g., miR-30b-5p) and exogenous (e.g., cel-miR-39-3p) controls for quality control purposes; (3) include an indeterminate range to enhance specificity. These changes reduce the complexity and cost of the test while improving performance, particularly with regards to the detection of MRD. We believe the present work regarding reproducibility and thresholding provides a substantial step towards the clinical implementation of the serum miR-371a-3p assay for management of patients with viable GCT disease.
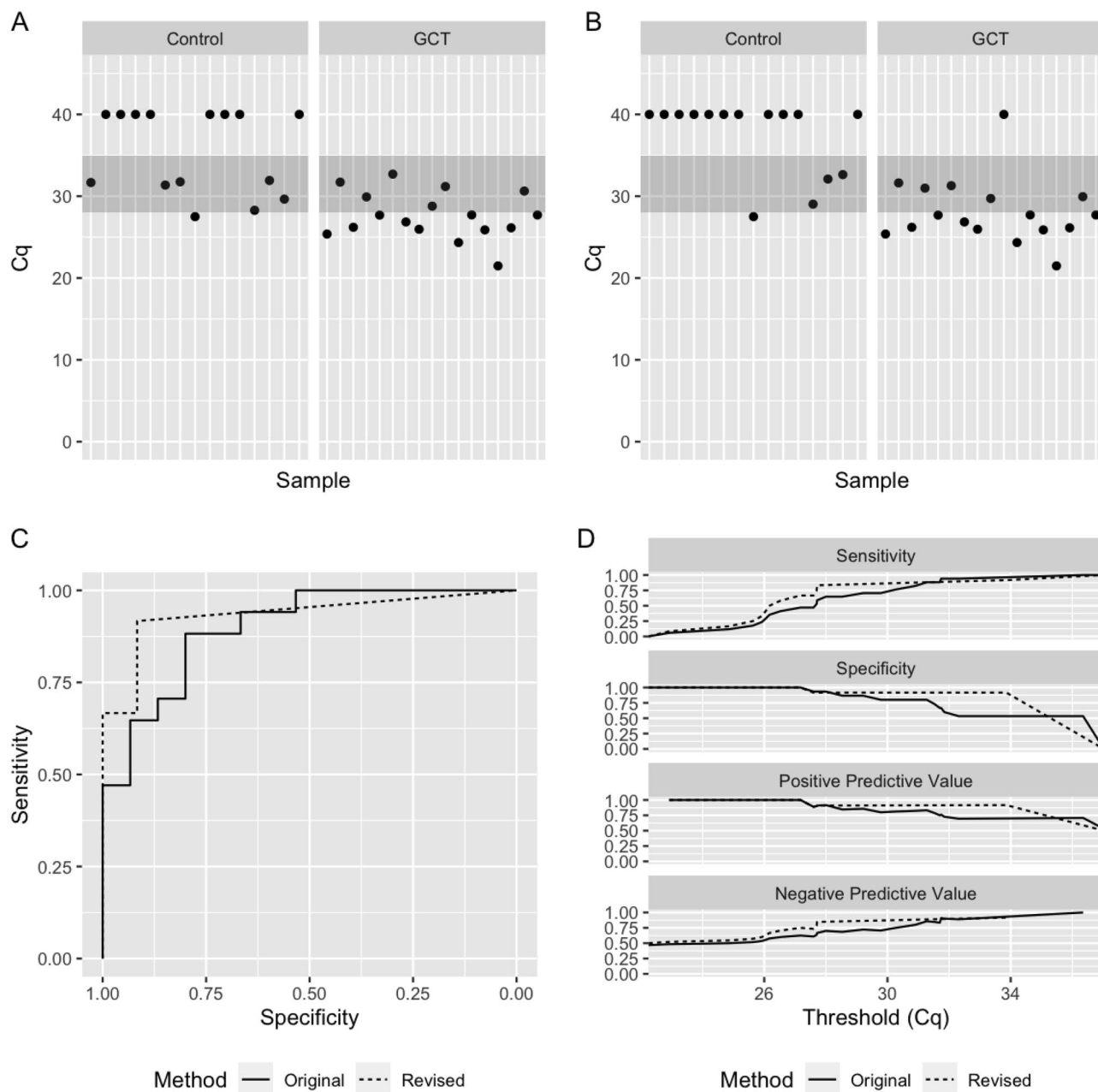
Figure 3. Revised method improves performance of serum miR-371a-3p test in patients with minimal residual disease. (**A,B**) Cq values of serum miR-371a-3p using original (**A**) and revised (**B**) methodology. (**C**) ROC plot showing performance of original (solid line) and revised (dashed line) methodology. (**D**) Comparison of sensitivity, specificity, positive predictive value, and negative predictive value across all measured Cq thresholds using original (solid line) and revised (dashed line) methodology.
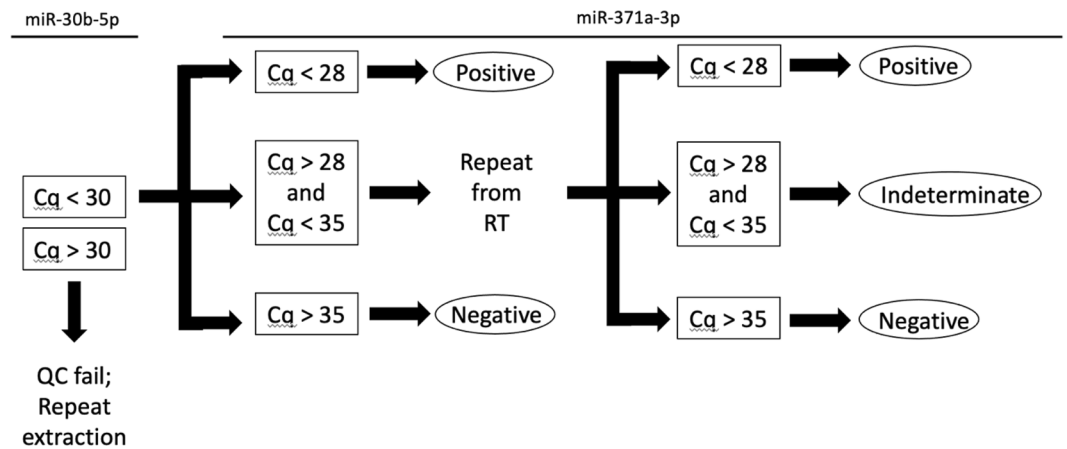
Figure 4. Decision-making flowchart for revised serum miR-371a-3p method. First, results of miR-30b-5p are evaluated for quality control. If Cq > 30, insufficient RNAs may have been isolated from the serum sample, and this sample should be reextracted. If Cq < 30, proceed to evaluation of miR-371a-3p. If Cq < 28, accept positive result. If Cq > 35, accept negative result. If 28 < Cq < 35, assay should be repeated from reverse transcription (RT) step. If Cq < 28 or Cq > 35, accept results as above. If 28 < Cq < 35 again, report indeterminate and recommend short interval follow up.

Data availability
The data analyzed for this publication are available upon reasonable request from the corresponding author.

References
1. Saoud, R. M. *et al.* Impact of non-guideline-directed care on quality of life in testicular cancer survivors. *Eur. Urol. Focus* **7**, 1137–1142. https://doi.org/10.1016/j.euf.2020.10.005 (2021).
2. Wymer, K. M. *et al.* Mildly elevated serum alpha-fetoprotein (AFP) among patients with testicular cancer may not be associated with residual cancer or need for treatment. *Ann. Oncol.* **28**, 899–902. https://doi.org/10.1093/annonc/mdx012 (2017).
3. Ehrlich, Y., Brames, M. J., Beck, S. D., Foster, R. S. & Einhorn, L. H. Long-term follow-up of Cisplatin combination chemotherapy in patients with disseminated nonseminomatous germ cell tumors: Is a postchemotherapy retroperitoneal lymph node dissection needed after complete remission?. *J. Clin. Oncol.* **28**, 531–536. https://doi.org/10.1200/jco.2009.23.0714 (2010).
4. Chakiryan, N. H. *et al.* Reliability of serum tumor marker measurement to diagnose recurrence in patients with clinical stage I nonseminomatous germ cell tumors undergoing active surveillance: A systematic review. *J. Urol.* **205**, 1569–1576. https://doi.org/10.1097/ju.0000000000001685 (2021).
5. Kollmannsberger, C. *et al.* Patterns of relapse in patients with clinical stage I testicular cancer managed with active surveillance. *J. Clin. Oncol.* **33**, 51–57. https://doi.org/10.1200/jco.2014.56.2116 (2015).
6. Beck, S. D., Foster, R. S., Bihrle, R. & Donohue, J. P. Significance of primary tumor size and preorchiectomy serum tumor marker level in predicting pathologic stage at retroperitoneal lymph node dissection in clinical Stage A nonseminomatous germ cell tumors. *Urology* **69**, 557–559. https://doi.org/10.1016/j.urology.2006.12.011 (2007).
7. Liu, Q., Lian, Q., Lv, H., Zhang, X. & Zhou, F. The diagnostic accuracy of miR-371a-3p for testicular germ cell tumors: A systematic review and meta-analysis. *Mol. Diagn. Ther.* **25**, 273–281. https://doi.org/10.1007/s40291-021-00521-x (2021).
8. Murray, M. J. *et al.* A pipeline to quantify serum and cerebrospinal fluid microRNAs for diagnosis and detection of relapse in paediatric malignant germ-cell tumours. *Br. J. Cancer* **114**, 151–162. https://doi.org/10.1038/bjc.2015.429 (2016).
9. Lafin, J. T. *et al.* Serum MicroRNA-371a-3p levels predict viable germ cell tumor in chemotherapy-naïve patients undergoing retroperitoneal lymph node dissection. *Eur. Urol.* **77**, 290–292. https://doi.org/10.1016/j.eururo.2019.10.005 (2020).
10. Badia, R. R. *et al.* Real-world application of pre-orchiectomy miR-371a-3p test in testicular germ cell tumor management. *J. Urol.* https://doi.org/10.1097/ju.0000000000001337 (2020).
11. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3**, 32–35. https://doi.org/10.1002/1097-0142(1950)3:1%3c32::Aid-cncr2820030106%3e3.0.Co;2-3 (1950).
12. R Foundation for Statistical Computing, R: A language and environment for statistical computing, Vienna. https://www.R-project.org/ (2019).
13. Robin, X. *et al.* pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **12**, 77. https://doi.org/10.1186/1471-2105-12-77 (2011).
14. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
15. Nappi, L. *et al.* Developing a highly specific biomarker for germ cell malignancies: Plasma miR371 expression across the germ cell malignancy spectrum. *J. Clin. Oncol.* **37**, 3090–3098. https://doi.org/10.1200/jco.18.02057 (2019).
16. Myklebust, M. P. *et al.* Serum miR371 in testicular germ cell cancer before and after orchiectomy, assessed by digital-droplet PCR in a prospective study. *Sci. Rep.* **11**, 15582. https://doi.org/10.1038/s41598-021-94812-2 (2021).
17. Ye, F. *et al.* Analytical validation and performance characteristics of molecular serum biomarkers, miR-371a-3p and miR-372-3p, for male germ cell tumors, in a clinical laboratory setting. *J. Mol. Diagn.* **24**, 867–877. https://doi.org/10.1016/j.jmoldx.2022.04.007 (2022).
18. Christiansen, A. J. *et al.* Impact of differing methodologies for serum miRNA-371a-3p assessment in stage I testicular germ cell cancer recurrence. *Front. Oncol.* **12**, 1056823. https://doi.org/10.3389/fonc.2022.1056823 (2022).

Author contributions

J.T.L., C.G.S., N.C., A.L.F., M.J.M., and A.B. conceived the study. J.T.L., C.G.S., A.A., B.K., and Z.W. performed the study. J.M.H., T.G., V.M., S.L.W., A.B., L.J., and C.M.L. provided clinical support. M.N. and J.P. provided statistical support. J.G., N.C., M.J.M., and A.B. supervised the execution of the project. N.C., M.J.M., J.F.A., A.L.F., and A.B. provided support. J.T.L., C.G.S., Y.C.S., N.C., M.J.M., A.L.F., J.F.A., and A.B. wrote the main text. A.S., S.M., D.W.S. conducted a critical review and drafted a manuscript. All authors revised and approved the final manuscript.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-37271-1.

Correspondence and requests for materials should be addressed to A.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.