Analytical Biochemistry 601 (2020) 113779

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Subfamily-specific quantification of endogenous mouse L1 retrotransposons by droplet digital PCR

Simon J. Newkirk[a,**], Lingqi Kong[a], Mason M. Jones[a], Chase E. Habben[a], Victoria L. Dilts[a], Ping Ye[b,c], Wenfeng An[a,*]

[a] Department of Pharmaceutical Sciences, South Dakota State University, Brookings, SD, 57007, United States
[b] Department of Pharmacy Practice, South Dakota State University, Brookings, SD, 57007, United States
[c] Avera Research Institute, Sioux Falls, SD, 57108, United States

ARTICLE INFO

Keywords:
LINE-1
Droplet digital PCR
L1 retrotransposon
Mouse model
Copy number variation
Retrotransposition

ABSTRACT

Long interspersed element type 1 (LINE-1; L1) mobilizes during early embryogenesis, neurogenesis, and germ cell development, accounting for 25% of disease-causing heritable insertions and 98% of somatic insertions in cancer. To better understand the regulation and impact of L1 mobilization in the genome, reliable methods for measuring L1 copy number variation (CNV) are needed. Here we present a comprehensive analysis of a droplet digital PCR (ddPCR) based method for quantifying endogenous mouse L1. We provide experimental evidence that ddPCR assays can be designed to target specific L1 subfamilies using diagnostic single nucleotide polymorphisms (SNPs). The target and off-target L1 subfamilies form distinct droplet clusters, which were experimentally verified using both synthetic gene fragments and endogenous L1 derived plasmid clones. We further provide a roadmap for *in silico* assay design and evaluation of target specificity, ddPCR testing, and optimization for L1 CNV quantification. The assay can achieve a sensitivity of 5% CNV with 8 technical replicates. With 24 technical replicates, it can detect 2% CNV because of the increased precision. The same approach will serve as a guide for the development of ddPCR based assays for quantifying human L1 copy number and any other high copy genomic target sequences.

1. Introduction

Long interspersed element type 1 (LINE-1; L1 hereafter) accounts for 17% and 19% of the human and mouse genomes, respectively. While the vast majority of L1 copies are incapable of further mobilization due to "structural defects" [1], a small fraction of them are full-length and retain the ability to mobilize autonomously and amplify within the genome [2,3]. L1 promoters are usually silenced by multiple epigenetic pathways, most prominently, via DNA methylation [1,4,5]. However, during aberrant or programmed DNA demethylation, some L1s are expressed and may undergo retrotransposition. Recent studies have illuminated the scale and impact of active movement by L1 and other non-LTR retrotransposons (e.g., Alu and SVA) in the human genome [6]. It is noteworthy that L1 contributes to only ~25% of disease-causing heritable insertions but ~98% of somatic (or non-heritable) insertions in cancer patients [6–8]. L1 mobilization during early embryonic or germ cell development can produce heritable insertions in mice [9–11] and humans [12–15]. Among somatic organ systems, the brain appears to be the hub for active L1 mobilization [16,17]. Importantly, L1 mobilization has been detected in different types of cancers and has been shown to initiate somatic transformation in selected cases [7,18,19].

It is technically challenging to quantify new L1 insertions. Despite its prevalence in different developmental and cellular contexts mentioned above, the frequency of L1 mobilization is relatively low on a per cell basis. Recent pedigree analyses estimate one heritable L1 insertion in 8 mouse births [11] or 63 human births [15]. For somatic insertions, single-cell sequencing estimates the rate ranging from 1 to 14 insertions per neuron [20–22]. Using various enrichment and amplification strategies, multiple sequencing-based methods have been developed to detect and map specific L1 insertions, achieving the proverbial feat of "finding the needle in a haystack" [11,23–27]. However, it is nearly impossible to quantify copy number variation (CNV) due to low frequencies of *de novo* insertions against an immense baseline of

* Corresponding author.
** Corresponding author.
E-mail addresses: sjnewkirk@gmail.com (S.J. Newkirk), Lingqi.Kong@jacks.sdstate.edu (L. Kong), mason.jones@coyotes.usd.edu (M.M. Jones), chase.habben@coyotes.usd.edu (C.E. Habben), victoria.bishop@jacks.sdstate.edu (V.L. Dilts), Ping.Ye@avera.org (P. Ye), wenfeng.an@sdstate.edu (W. An).

https://doi.org/10.1016/j.ab.2020.113779
Received 16 February 2020; Received in revised form 6 May 2020; Accepted 14 May 2020
Available online 20 May 2020

preexisting copies [21]. Previously, standard real-time quantitative PCR (qPCR hereafter) has been utilized to quantify the copy number of endogenous L1s [17,28–31]. Despite being routinely used for detecting 25% CNV, qPCR assays require extensive optimization and large number of replicates for even detecting 10% CNV [32].

Here we present a droplet digital PCR (ddPCR) based method for quantifying endogenous mouse L1. ddPCR is an alternative to the "analogue" qPCR. In the commercially available Bio-Rad platform, up to 20,000 water-in-oil droplets are generated from a 20 μL mixture of sample and reagents, thermocycled to end point, and read "digitally" as either positive or negative for the target sequence [33]. Partitioning provides multiple benefits to ddPCR over qPCR, including increasing the effective concentration of low abundance targets, purifying the target of interest from interfering compounds, and improving assay precision and linearity [34,35]. In the context of L1 biology, it has been previously used to quantify the allelic abundance of specific human L1 insertions [36], the copy number of *de novo* insertions from a single-copy mouse L1 transgene [10], and the copy number of endogenous mouse L1s in a maternal care model [37]. We discuss factors involved in the primer and probe design for targeting specific mouse L1 subfamilies, evaluate the available genomic reference targets, optimize parameters (primer Tm, annealing temperature, and primer/probe concentration), define assay specificity and sensitivity, and determine the effect of technical replicates on assay precision. Our study provides a platform for applying ddPCR to quantify L1 CNV in the mouse model and is expected to serve as a guide for developing ddPCR to quantify L1 CNV in human samples.

2. Material and methods

Note, in this section, optimized parameters for subfamily specific L1 ddPCR assays are marked as '[Default]'. Additional optimization may be required for some applications.

2.1. Primer and probe design

The consensus sequences of 29 different mouse L1 subfamilies [38] were aligned in MacVector software and manually scanned to identify SNPs that were unique to one, or few subfamilies. Primers and probes were designed in order to take advantage of these subfamily specific SNPs. Primers were designed using Applied Biosystems Primer Express software and then purchased from Eurofins. The target Tm of primers was 58–62 °C and the Tm of the probe was designed 5–10 °C higher than that of primers [Default]. Probes were manually designed to be centered on one or multiple SNPs with at least one SNP centrally located within the probe [Default]. TaqMan probes containing the 3′ minor groove binding non-fluorescent quencher were ordered from Life Technologies [Default]. LNA probes containing the 3' Iowa Black quencher were from Integrated DNA Technologies (IDT). The free energy of each primer/probe to the target or off-target subfamily templates was calculated using the OligoAnalyzer tool on the IDT website. Probe and primer sequences are provided (Supplemental Table 1). Synthetic gene fragments were ordered from Twist Bioscience (Supplemental Table 2).

2.2. Genomic DNA extraction

Animal work was approved by the Institutional Animal Care and Use Committee. 10 mg of tissues from C57BL/6J (B6) mouse were first homogenized using Qiagen TissueLyser II at 20 Hz for 1 min, and then was extracted with the Qiagen DNeasy kit according to the user-developed protocol "Isolation of DNA from soft tissues using the TissueLyser and QIAamp DNA Mini Kit" (available online at Qiagen website). gDNA stored at −20 °C until time of analysis.

2.3. ddPCR workflow

gDNA concentration was quantified on Nanodrop 1000 and serially diluted with 1x TE buffer to desired concentrations. For the majority of the L1 assays presented here the input was 60 pg [Default]. The input concentration of the reference vs. HPRT assay was 400 pg. The input concentration will depend on the copy number of the target and reference and should be experimentally determined. Furthermore, since the overall Poisson error of ddPCR is dependent on the subsampling and partitioning error of both the target and reference, for assays where sensitivity is critical, the concentration should be optimized to minimize Poisson error as much as possible. Each 22 μL ddPCR reaction was set up according to manufacture recommendations and contained 11 μL 2x Supermix for Probes (without dUTP) (Bio-Rad), template DNA, primers (1800 nM each), and FAM or VIC labeled probe(s) (500 nM each) [Default]. Adequate mixing at each step of preparation is key to have the necessary consistency between reactions and within technical replicates. Reactions were prepared in semi-skirted 96-well plates (Eppendorf or Bio-Rad). Samples were incubated at room temperature for 10 min to ensure a consistent temperature during droplet generation. Droplets were generated according to manufacturer recommendations using a manual QX200 Droplet Generator with Droplet Generation Oil for Probes (Bio-Rad). Plates were sealed with perforable heat-sealing foil (Bio-Rad) and run on Eppendorf Master Cycler thermocycler. PCR reaction conditions were [95 °C 10 min; 95 °C 15 s; 60 °C 30 s] x 50; 4 °C 5 min; 98 °C 10 min; 4 °C HOLD] [Default]. Droplets were analyzed with the Bio-Rad QX200 system using the QuantaSoft Analysis Pro software (version 1.0.596). Due to the presence of rain and complex droplet clusters, manually setting thresholds was preferred over algorithmic threshold and cluster assignment tools. Thresholds were manually set in regions of minimal rain between clusters based on examination of the aggregated 2-D droplet plots and the 1-D histogram profiles from all samples to be compared. Examining the 2-D droplet plots would reveal any shifts in droplet clusters (e.g. due to competition between the duplexed PCR reactions). The 1-D histogram for each channel helps to visualize the separation ("the valley") between droplet clusters. The threshold should be set at the lowest point of the valley. Any droplets above the threshold are counted as "positive" and those below the threshold as "negative". An example for setting thresholds is provided (Supplemental Figure 3). Annealing temperature gradients were conducted simultaneously using the Eppendorf Master Cycler. EvaGreen reactions were set up, processed, and analyzed similarly as for probes using the EvaGreen Supermix and Droplet Generation Oil for EvaGreen (Bio-Rad). EvaGreen primers were at 100 nM each. PCR reaction conditions were [95 °C 5 min; (95 °C 30 s; 60 °C 1 min) x 45; 4 °C 5 min; 90 °C 5 min; 4 °C HOLD].

2.4. Analysis of the effect of technical replicates on assay precision

ddPCR reactions were prepared as above with a sufficient amount of master mix (supermix, template, primers, probe) for 40 technical replicates. 20 μL was directly proportioned from the master mix into each droplet generation well. Technical replicates were simulated by randomly assigning 1, 3, 6, 8, 16, 24, 32, and 40 of the samples to a replicate group. 1000 of these groups were processed and the average of the L1/U6 ratio and % CV of each group calculated, plotted, and compared against 1000 randomly selected individual samples. The spread of the data was plotted using PlotsOfData [39].

2.5. Establish assay sensitivity with plasmid spike-in dilution series

Endogenous L1 loci representing A_I, Tf_I, Tf_II, Tf_III, and Gf_I subfamilies were amplified from B6 mouse gDNA using PrimeStar GXL DNA polymerase (Takara). Detailed description of the cloning procedure will be reported elsewhere. Up to three bacterial clones were selected for each L1 locus and approximately 4000 copies of each plasmid

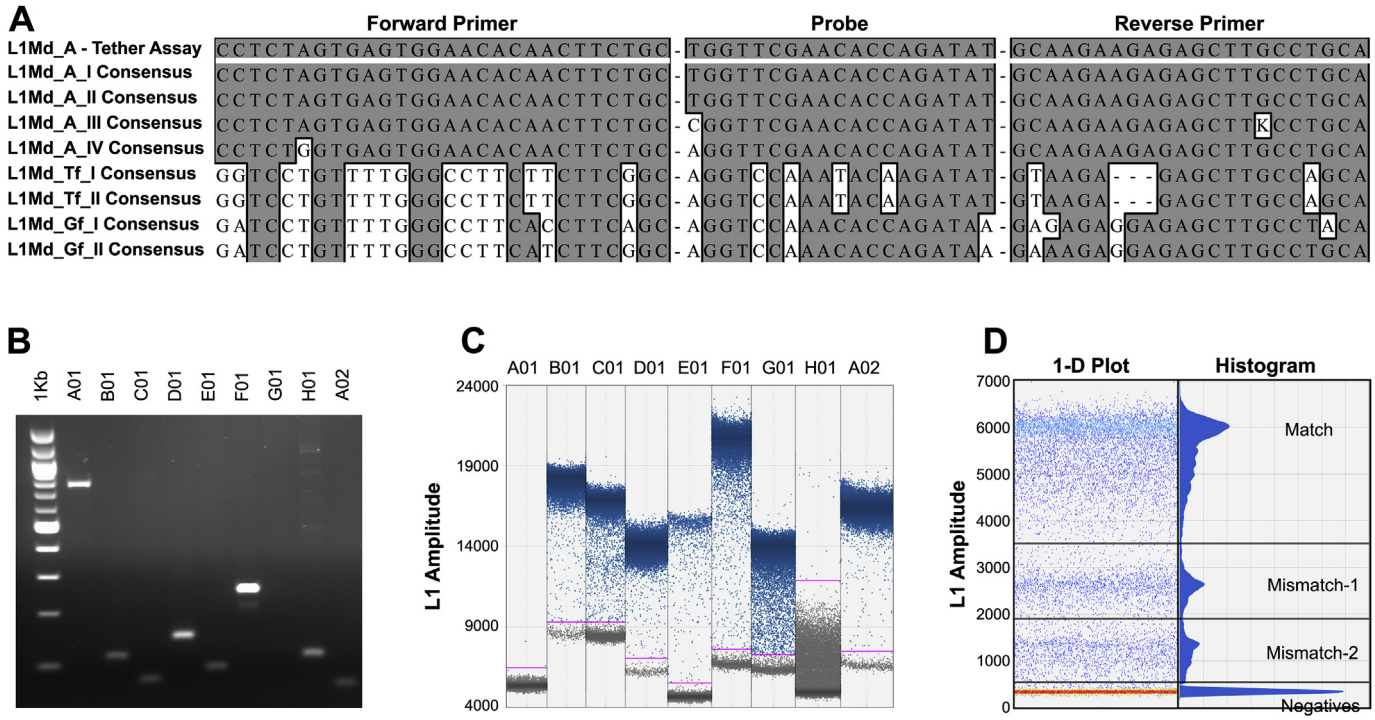


Fig. 1. The design and preliminary analysis of ddPCR primers and probe. (A) Eight of the youngest L1 subfamily consensus sequences were aligned. SNPs unique to one or two subfamilies were manually identified and used to design primers and probes that would distinguish that specific subfamily. The primers and probe for the L1Md_A-Tether assay are depicted at the top of the alignment. **(B)** Representative endpoint PCR resolved on a 1% agarose gel. Invitrogen's 1 Kb plus ladder was used as the molecular weight marker (left lane). These primers map to the ORF2 region of Tf subfamily consensus sequence. **(C)** 1-D droplet plot for EvaGreen ddPCR. Each column is a ddPCR reaction, which has a corresponding end-point PCR in panel B using the same primers. Each point on the plot represents one individual droplet. The Y-axis represents the fluorescent amplitude (i.e., the raw fluorescent readout) of each droplet. Thresholds (magenta lines) were manually set to delineate the positive droplets (blue) and negative droplets (gray). **(D)** TaqMan ddPCR assay targeting L1Md_A-Tether. In the 1-D droplet plot (left panel), each point represents one individual droplet. Droplet density is represented with a color gradient from blue to red (low density to high). Histogram (right panel) highlights the droplet distribution. In addition to the negative droplets (negatives), there are 3 distinct clusters, which are denoted as match (target subfamily; A_I/II), mismatch-1 and mismatch-2 (off-target subfamilies). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

was used to validate the specificity of L1 target assays (see section 3.5). Prior to the spike-in of the plasmid Tf_II P2, the copy number of the gDNA sample was quantified (4250 L1 copies/μL) with the L1Md_Tf-ORF2;BC1 duplex ddPCR. The plasmid was quantified by the simplex L1Md_Tf-ORF2 assay. The plasmid was diluted with 1x TE buffer to an estimated copy number of 850 copies/μL (20% of gDNA L1 copies) and then serially diluted 1:2 with TE buffer. 4 μL of the plasmid was combined with 4 μL of gDNA (gDNA + spike-in) or TE buffer (plasmid only). ddPCR reactions prepared as described above with enough master mix for 8 technical replicates. The plasmid only reaction was simplex L1Md_Tf-ORF2. gDNA + spike-in reaction was duplexed with L1Md_Tf-ORF2 and BC1 probes. The copy number of each plasmid dilution was quantified as the average of the L1 copy number among the 8 technical replicates with standard deviation. This concentration was plotted against the estimated 20% - > 0.625% increase in copy number. The ratio of each gDNA + spike-in dilution was quantified as the average of L1/BC1 ratio of the 8 technical replicates with standard error of the mean (SEM).

The % increase of each spike-in was calculated as:

$$\% \text{ increase} = 100 - [(\text{gDNA + spike-in L1/BC1 ratio})/(\text{gDNA L1/BC1 ratio})] \times 100$$

The 'estimated increase' shown on the graph was calculated as:

Estimated increase = (Estimated % increase x gDNA L1/BC1 ratio) + gDNA L1/BC1 ratio. Statistical significance calculated with both a one-tailed Student's t-test and Mann-Whitney *U* test. Both the 20% and 5% estimated spike-in samples were significant with p-value < 0.01. The 2.5% was not statistically significant with a p-value = 0.0601, and 0.0505, for *t*-test and Mann-Whitney, respectively.

2.6. *In silico* mapping and computational analysis of ddPCR assay targets

Each primer pair was aligned to the mouse genome (mm10) using Bowtie 2 (2.2.9) "–very-sensitive" mode [40]. Alignments were reported in SAM files. The SAM files were converted to BED files. The primer aligned genomic sequences as defined in the BED files were extracted from the mouse genome using bedtools getfasta. The returned hits encompassing the '*in silico* PCR amplicon' were aligned in RStudio using the multiple sequence alignment package 'msa' [41] and the default parameters for ClustalW [42]. The region corresponding to the probe was then extracted from the multiple sequence alignment dataset. The aligned probed region was visually inspected and manually adjusted to remove misalignment due to improperly placed gaps. The number of sequences that correspond to 0 up to 5 mismatches to the probe was counted and plotted as a bar chart for each probe. The nucleotide composition is visualized in RStudio using 'ggseqlogo' package [43].

2.7. ddPCR data analysis

After setting thresholds, wells to be analyzed are exported from QuantaSoft Analysis Pro software as a Microsoft Excel file. The ratio of L1 target versus the reference target is in the column "ratio". For technical replicates the average and standard deviation of the replicate ratios were calculated and statistical significance between samples determined with one-tailed Student's T-test, with p-value < 0.05 considered statistically significant. To calculate the copy number of an L1 target per cell, the copy number of the reference target was first determined in a separate TaqMan ddPCR assay, in which the reference target (e.g., U6) was duplexed with the single-copy genomic target

HPRT. The copy number of U6 was calculated to be 357 per cell. As an example, the L1Md_Tf-ORF2 assay reports an L1/U6 ratio of 16.1, which would correspond to 5748 L1 copies per cell (i.e., 16.1 L1/U6 multiplied by 357 U6/cell).

3. Results and discussion

3.1. Design ddPCR assays to differentiate active L1 subfamilies

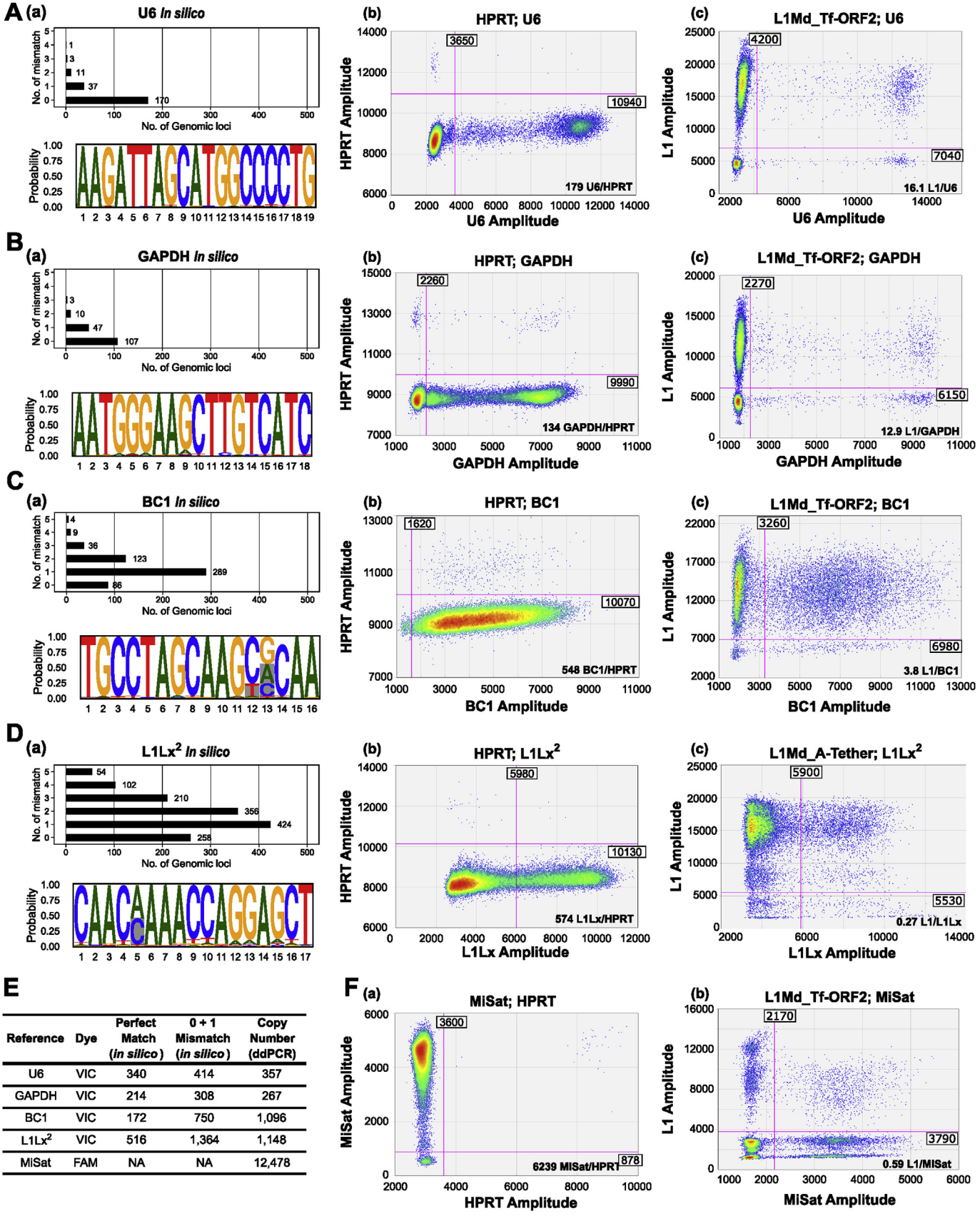
At least 29 different L1 subfamilies have inserted into the mouse genome since the split between mouse and rat [38]. The youngest four subfamilies (A_I, Tf_I, Tf_II, Gf_I) evolved within the last one million years and appear to harbor full-length elements that are capable of further retrotransposition [3,38,44,45]. Alignment of subfamily consensus sequences identified single nucleotide polymorphisms (SNPs) to differentiate between these youngest subfamilies and from older L1 subfamilies (Supplemental Figure 1). For example, against the tether region, we designed a primer pair that shows perfect match to A_I and A_II but at least one nucleotide mismatch toward the 3' end of the primer annealing site to other subfamily consensus sequences (Fig. 1A). We first tested the specificity of the primers alone with end-point PCR (Fig. 1B). Primer sets were eliminated if they had either amplicons of incorrect size (lane A01), more than one amplicon (lane F01), or weak/no amplification (lane G01). Primer sets that passed these criteria were further tested with EvaGreen-based ddPCR. EvaGreen is a DNA binding dye that preferentially binds double stranded DNA (dsDNA) with superior DNA binding profiles over the commonly used SYBR green I dye [46]. A TaqMan probe was subsequently synthesized for primer sets that showed good amplification (clear separation between the positive and negative droplet clusters) and limited "rain" (droplets with fluorescent amplitudes falling between the positive and negative droplet clusters) in EvaGreen reactions (Fig. 1C; lanes B01, D01, and A02). Since centrally located mismatches have the highest impact on the stability of oligo binding [47], when possible, the probes were positioned in regions that contain at least one centrally located SNP between the matched and mismatched targets (Fig. 1A). A total of 12 probe-based ddPCR assays ("TaqMan ddPCR" hereafter) for endogenous L1 targets were defined and evaluated (Supplemental Figure 1). A list of primers and probes is provided elsewhere (Supplemental Table 1).

In standard qPCR, both specific and nonspecific reactions contribute to the overall fluorescence and quantification, owing to either the propensity of the fluorescent dye to bind any dsDNA products, or the non-specific binding of probes in probe-based qPCR, giving rise to false positive signals. In contrast, due to template partitioning in ddPCR, nonspecific reactions can be identified as droplets with diminished fluorescence and then quantified separately [48]. In the context of endogenous L1 detection, owing to the overwhelming number of polymorphic L1 sequences present in a single genome, we expect substantial amplification from imperfectly matched templates, in the form of low fluorescence clusters. Even with ddPCR, most primer-only reactions (i.e., EvaGreen ddPCR) showed only one distinct positive fluorescent droplet cluster, suggesting limited resolution of specific versus nonspecific targets (Fig. 1C; Supplemental Figure 2A). The addition of a probe with discriminating SNPs allowed the separation of droplets into distinct clusters with varied fluorescence amplitudes (Supplemental Figure 2B). For TaqMan ddPCR assay targeting SNPs in the tether region of L1Md_A subfamilies (i.e., the L1Md_A-Tether assay; Fig. 1A), the inclusion of the probe allowed the separation of droplets into three distinct positive clusters (Fig. 1D). Some of the clusters with weaker fluorescence were later confirmed as droplets containing mismatched target sequences.

3.2. Establish reference targets for endogenous L1 quantification

By design, a simplex ddPCR reports the absolute copy number of a target in a gDNA sample [33]. However, the simplex design is inadequate for comparing relative changes of a target sequence between two samples. In this case, the amount of input material should be normalized by a reference target [33]. To accurately quantify the copy number of endogenous L1 sequences, it is imperative to establish a reliable reference target and quantify both the L1 and reference in duplex probe-based assays. Several factors dictate the choice of reference targets for endogenous L1 quantification. The first factor is the genomic copy number of the reference relative to the L1 target. Both qPCR and ddPCR are affected by subsampling error, which is caused by random effects when analyzing part of a whole (a subsample). In addition, the precision of ddPCR is affected by partitioning error, which originates from partitioning of the DNA targets into droplets [34]. Subsampling error dominates at low target concentrations (when there are fewer than 3 target molecules per droplet on average) whereas partitioning error at higher target concentrations. Thus, the closer the copy number of the genomic reference is to that of the L1 target, the lower the combined error (further discussed in section 3.4). This parameter necessitates the use of other high-copy genomic sequences as the reference target for endogenous L1 quantification (e.g., BC1, U6, GAPDH, L1Lx), and disfavors single or low copy sequences. The second factor is the genomic distribution of the reference sequences. Some high-copy genomic sequences present challenges to ddPCR owing to being tandem repeats (e.g., minor satellite repeats; MiSat hereafter). Conceptually, tandemly repeated targets would impact the randomness of partitioning and are disproportionally affected by the amount of fragmentation of the gDNA. Lastly, any competition between the reference and the L1 target should be evaluated.

We selected five different reference targets for examination (Fig. 2). U6 is a small nuclear RNA (snRNA) required for RNA splicing. The mouse genome has up to two copies of U6 genes but approximately 900 copies of processed U6 pseudogenes due to trans-mobilization by L1 [43,49–51]. We extracted potential PCR targets by paired-end mapping of the U6 primers with Bowtie2. Among the 222 hits, 170 loci (77%) had no mismatch in the U6 probe region, 37 (17%) had one SNP, and the rest (6%) had two or more SNPs (Fig. 2A, subpanel a). Sequence logo analysis revealed that the SNPs were even distributed across the length of probe target region and, overall, the target region of the U6 probe was well conserved (Fig. 2A, subpanel a). To establish the empirical copy number of the U6 reference target, we duplexed the U6 reference target with the X-linked single-copy gene, hypoxanthine guanine phosphoribosyl transferase (HPRT) (Fig. 2A, subpanel b). Quantification of this duplex ddPCR assay reported 179 copies of U6 per HPRT. The 2nd reference target is the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene, which has 186 pseudogene copies in the mouse genome, the highest among mouse protein coding genes [52]. The GAPDH primers and probe were previously used as genomic reference to quantify *de novo* insertions from an L1 transgene [53]. Our *in silico* analysis identified 167 potential targets for GAPDH primers, among which 107 (64%) showed no mismatch in the probe region, 47 (28%) had one SNP, and the rest (8%) had two or more SNPs (Fig. 2B, subpanel a). The GAPDH;HPRT duplex assay reported 134 copies of GAPDH per HPRT (Fig. 2B, subpanel b). The 3rd reference target is BC1. The mouse genome has a single copy of BC1 gene, which is highly expressed in neurons as a small noncoding RNA with a role in translational control [54], but as many as 6758 copies of BC1-related sequences per UCSC genome browser. The latter belong to a class of rodent specific short interspersed elements (SINEs), termed identifier (ID). Interestingly, it appears that all mouse ID elements originated from the single copy BC1 master gene via retrotransposition [55,56]. *In silico* analysis identified 547 copies of BC1 as potential targets for our BC1 primers. Among them, 86 (16%) had no mismatch in the probe region, 289 (53%) contained one SNP, and the remaining (32%) had two or more SNPs (Fig. 2C subpanel a). The BC1;HPRT duplex ddPCR reported 548 copies of BC1 per HPRT (Fig. 2C, subpanel b). The 4th reference target is L1Lx, an ancient and now defunct murine L1



E

Reference	Dye	Perfect Match (*in silico*)	0 + 1 Mismatch (*in silico*)	Copy Number (ddPCR)
U6	VIC	340	414	357
GAPDH	VIC	214	308	267
BC1	VIC	172	750	1,096
L1Lx^2	VIC	516	1,364	1,148
MiSat	FAM	NA	NA	12,478

(caption on next page)

Fig. 2. *In silico* analysis and ddPCR validation of high-copy reference targets. **(A)** U6. **(B)** GAPDH. **(C)** BC1. **(D)** L1Lx. **(E)** Copy number of reference targets from *in silico* analysis and from ddPCR. See subpanel legends below for details. **(F)** MiSat. Each of the 5 references tested were analyzed and validated using a similar workflow. The information is organized into subpanels. **Subpanel (a)** *In silico* mapping of reference targets. Potential PCR amplicons were extracted from the mouse genome. The bar chart reports the number of hits that contain 0 up to 5 SNPs in the designed probe region. The probe sequence conservation and location of SNPs among all extracted sequences were visualized using sequence logo. The Y-axis displays the nucleotide composition (expressed as the probability of having A, C, G, T or a gap), and the X-axis marks the position of each nucleotide in the probe. Note not all nucleotide positions add up to 1.0 due to the presence of gaps in the alignment. By default, the most dominant nucleotide at each position is contained in the probe. At positions where there are multiple apparent nucleotide variants, the non-probe variants are shaded in gray. **Subpanel (b)** 2-D droplet plot of duplex ddPCR for the reference target with HPRT. Reference;HPRT duplex ddPCR was used to empirically determine the copy number of the reference. Copies/cell is calculated by multiplying copies/HPRT by 2 since female gDNA was used. Each point on the plot represents one individual droplet. The fluorescent amplitude of each droplet is read under two fluorescent channels. HPRT amplitude is depicted in the Y-axis (FAM channel) and the reference on the X-axis (VIC channel). Thresholds (magenta lines; cutoff values are denoted by the values in the white boxes) are established in areas of minimal rain by examination of both the 2-D plots and 1-D histograms provided by the software. The copies/HPRT is noted for each reaction in the bottom corner of each plot. Droplet density is represented with a color gradient from blue to red (low density to high). **Subpanel (c)** 2-D droplet plot of duplex ddPCR for L1 target with a reference target. L1 amplitude is represented in the Y-axis, and the reference in the X-axis. Note, the MiSat reaction competes with the L1 reaction as seen by a reduction in Y-axis fluorescent amplitude of the double positive droplets. Since MiSat sequences are not included in the reference genome database the *in silico* analysis was not applicable. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

subfamily [57]. *In silico* analysis identified 1404 copies of L1Lx as potential targets for L1Lx primers, among which 258 (18%) had no mismatch, 424 (30%) had one SNP, and the remaining (52%) had two or more SNPs in the probe region (Fig. 2D, subpanel a). The L1Lx;HPRT ddPCR assay reported 574 copies of L1Lx per HPRT (Fig. 2D, subpanel b). However, this estimation is likely underreported due to the challenge in setting the threshold, which was set at a relatively higher amplitude where the valley between clusters was more distinct. Lastly, we attempted the minor satellite MiSat as a reference. MiSat are centromeric tandem repeats of a 120 bp monomer, accounting for about 1% of the mouse genome [58,59]. Since MiSat repeats are omitted from the mouse genome assembly [60] *in silico* analysis returned only a few hits (not shown). The duplex ddPCR with HPRT reported 6239 copies per HPRT (Fig. 2F, subpanel a). Taken together, in our ddPCR assays, the copy number of these five reference targets ranges from 267 to 12,478 copies per cell (Fig. 2E).

Then we compared the performance of these references when duplexed with one of our endogenous L1 targets. The duplex assays using U6 or GAPDH reference (both have the fewest copies per cell) showed a single positive cluster and minimal rain (Fig. 2A and B, subpanel c), consistent with abovementioned *in silico* analyses that the majority of the mapped hits showed no mismatch to each probe. The BC1 reference (with an intermediate copy number per cell) showed substantial amount of rain in the duplex assay (Fig. 2C, subpanel c), likely caused by a preponderance of mismatched targets. When duplexed with an L1 target, the L1Lx reference behaved similarly to BC1 in terms of copy number and rain (Fig. 2D, subpanel c), but as a reference sequence L1Lx risks including active elements in the reference quantification. This situation can lead to a false underestimation of any active L1 copy number increase if that increase is also quantified in the reference. Lastly, although the MiSat reference is the most comparable in copy number to L1 subfamilies to be quantified on a per cell basis, using the standard primer concentrations, the MiSat;L1 duplex reactions suffered competition between the two reactions (Fig. 2F, subpanel b). Such a situation created a challenge in setting thresholds. The best practice would be to consult with both 1-D and 2-D droplet plots while manually adjusting the threshold for each channel (Supplemental Figure 3). ddPCR is superior to qPCR in addressing competition in that, first, most droplets won't contain both targets simultaneously and second, competition is easier to detect with ddPCR and can be overcome by adjusting primer concentrations and/or the annealing temperature (Supplemental Figure 4; also see the following section as a general strategy). In sum, almost all reference targets come with a caveat and robust data should include the use of more than one reference target.

3.3. Optimizing ddPCR reaction conditions by adjusting annealing temperature, primer Tm, and primer/probe concentration

Similar to qPCR, ddPCR can be optimized by adjusting the annealing temperature of the reaction. Determining the optimal annealing temperature can be easily achieved by testing a temperature gradient (Fig. 3). Viewing at 1-D, there is a single positive cluster (and some rain) in U6 reference reactions. As annealing temperature increases the efficiency of the U6 reaction decreases, significantly reducing the amplitude of the positive cluster (Fig. 3A). In contrast, the L1Md_A-ORF2 target started with at least two positive fluorescent clusters at low temperature settings. As temperature rises the amplitude of the low fluorescent cluster decreases and is eventually merged with the negative cluster, while the high fluorescent cluster does not change significantly (Fig. 3B). Based on free energy calculation, we suspect that the high fluorescent cluster is derived from perfectly matched A_I target, and the low cluster is the mismatched families with a centrally located SNP in the probe annealing region (Fig. 3C). The dynamic changes are best viewed with 2-D plots, which clearly demonstrates that an optimal annealing temperature is at ~58 °C, where the positive and negative clusters have adequate separation (Fig. 3D).

Another parameter is the melting temperature of the primers. For the MiSat reference target, we tested a combination of three forward and three reverse primers (Supplemental Figure 5A; primers 1–6). The forward primers are partially overlapping and have slightly different Tm, so are the reverse primers. Minimal impact on the final fluorescent output was observed (Supplemental Figure 5B). In a separate test, we increased the primer Tm by including additional nucleotides at the 5' end, which are not expected to alter the priming specificity significantly (Supplemental Figure 5A; primers 7–10). The primers with lower Tm showed three positive clusters (Supplemental Figure 5C). In contrast, increasing the Tm of the primers enabled the separation of the high fluorescent cluster into two distinct clusters (Supplemental Figure 5D).

Aside from annealing temperature, ddPCR reactions can be further improved by optimizing the relative primer and probe concentrations. Per Bio-Rad, the standard concentration is 250 nM for probe and 900 nM for each primer (Supplemental Figure 6, blue line). Doubling the primer concentration had minimal impact on the reactions (Supplemental Figure 6, green line). Doubling the probe concentration significantly shifted both the negative and positive clusters to higher amplitudes, while maintaining the relative distance between the clusters (Supplemental Figure 6, red line). Simultaneously increasing the probe and primer concentrations further shifted the clusters to higher amplitudes, and also increased the separation of the two droplet clusters (Supplemental Figure 6, purple line). This latter condition may help to reduce variability when setting the threshold between samples or between plates.

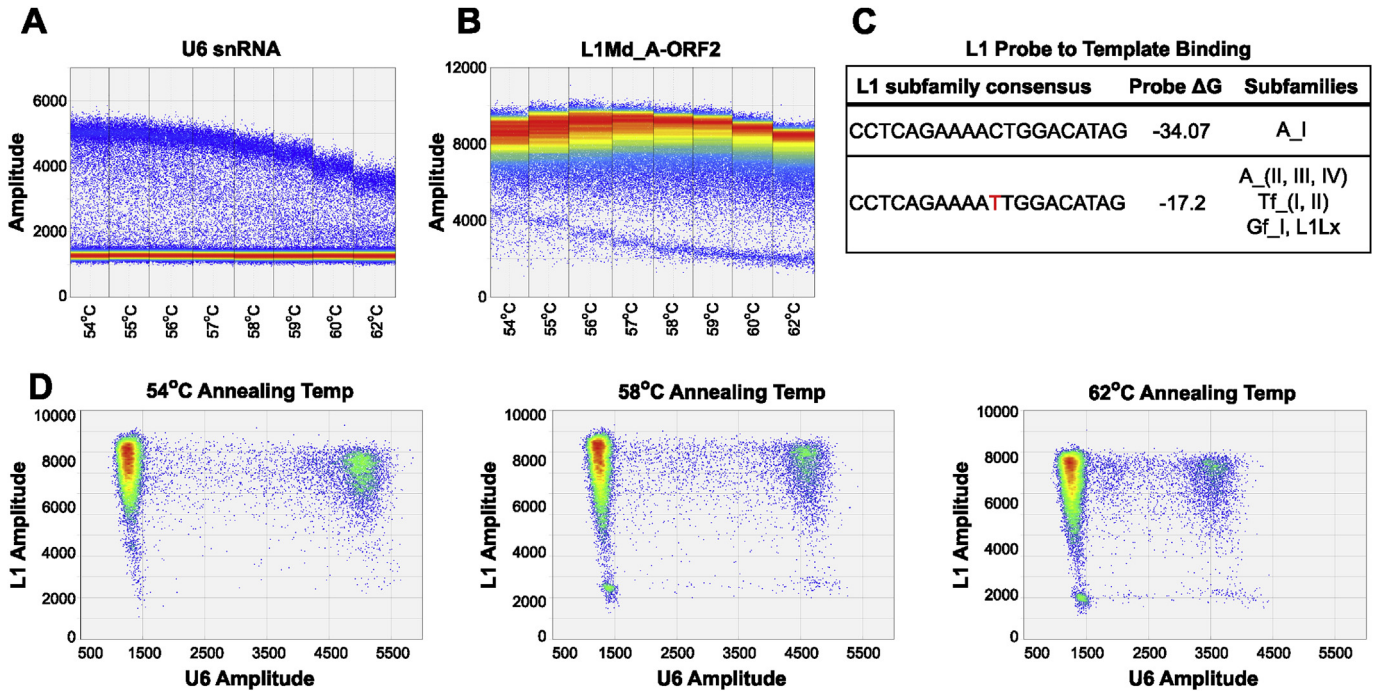


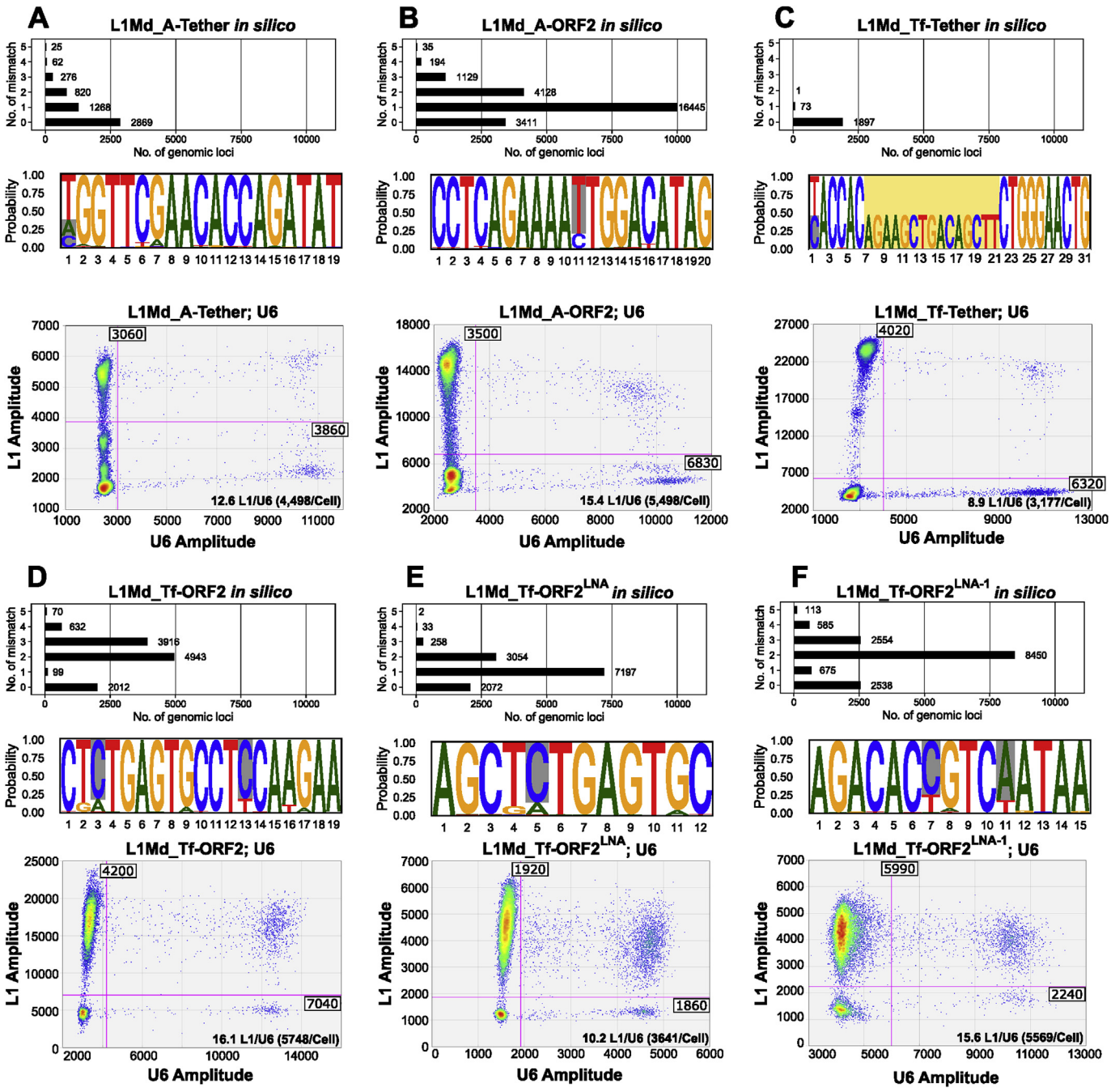
Fig. 3. Optimization of annealing temperature for duplex L1 ddPCR. **(A)** 1-D views of the U6 amplitude for L1Md_A-ORF2;U6 duplex reactions. **(B)** 1-D views of the L1 amplitude for L1Md_A-ORF2;U6 duplex reactions. While the U6 reaction efficiency decreases with higher annealing temperatures, the L1 assay shows little change in the high droplet cluster amplitude but significant decrease in the low fluorescent cluster amplitude. **(C)** Calculated binding free energy for the L1Md_A-ORF2 probe with its target and off-target subfamily templates. The discriminating SNP is marked in red. **(D)** Representative 2-D views of duplex L1Md_A-ORF2;U6 ddPCR annealing temperature gradient (left panel 54 °C, middle panel 58 °C, right panel 62 °C). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.4. Quantify active L1 subfamilies

As a first step to analyze endogenous L1 activities in the mouse genome, we developed several L1 assays that target L1Md_A and L1Md_Tf subfamilies (Supplemental Figure 1). Much like in the human genome, most L1 sequences in the mouse genome are 5′ truncated [61,62]. The mouse genome has approximately 6000–8000 full-length L1s that retain the 5′UTR [63,64]. The proportion of full-length elements varies among L1 subfamilies. In the human genome, evolutionarily younger L1 subfamilies tend to possess a higher fraction of full-length elements [65,66]. L1 subfamilies in the mouse genome follow the same trend [38], likely the consequence of post-insertion selection and other mutational processes. We designed ddPCR assays targeting either the 5′UTR or ORF2 region. Assays specific for the 5′UTR region have the advantage of targeting full-length (and potentially active) elements. Additionally, the 5′UTR is the most divergent region among mouse L1 subfamilies [38], allowing for greater subfamily specificity. However, unlike human L1 5′UTR, the 5′UTRs of mouse L1 subfamilies are composed of tandem repeats of ~200 bp monomers followed by a non-monomeric tether region [3,38,67]. Targeting the monomeric region of 5′UTR can potentially skew the target distribution in the droplets and subsequent quantification. Therefore, we chose to design probes within the non-monomeric tether region of the 5′UTR. An alternative approach is to use a reverse primer anchored in the tether or ORF1 region. In contrast, assays targeting ORF2 are more inclusive and allow for the quantification of both full-length and a subset of truncated insertions. In assays described below, each L1 target was duplexed with the U6 reference (Fig. 2).

For L1Md_A subfamilies, we designed two assays. The L1Md_A-Tether assay is designed against the tether sequences of predominantly A_I and A_II subfamilies (Supplemental Figure 1) as previously reported [30]. Paired-end mapping of the primers using Bowtie2 reported a total of 5320 hits in the mouse genome. Among these hits, 2868 had no mismatch in the probe region, 1268 had one SNP, and the rest had two or more SNPs (Fig. 4A, top panel). Sequence logo analysis revealed the nucleotide composition in the probe region. Position 1 had the highest fraction (38%) of mismatched nucleotides (Fig. 4A, middle panel). The matching nucleotide (T) was found in the intended A_I/II target subfamilies. The mismatched nucleotides (A or C) were found in consensus sequences of A_III/IV, Tf_I/II, and Gf_I/II subfamilies (Supplemental Figure 1). Moreover, several internal SNPs found in the consensus sequences of Tf_I/II and Gf_I/II subfamilies (see Supplemental Figure 1) were absent in the sequence logo (Fig. 4A, middle panel). As a result, the A-Tether assay is predicted to exclude Tf_I/II and Gf_I/II targets. This prediction was later validated by using plasmid-derived or synthetic gene fragments (discussed in section 3.5). The A-Tether;U6 duplex assay reports a ratio of L1 versus U6 at 12.6 in the sample. As the number of U6 target per cell is 357 (Fig. 2E), this readout translates into 4498 copies of L1 targets per cell (i.e., 12.6 L1/U6 x 357 U6/cell = 4498 L1/cell) (Fig. 4A).

The L1Md_A-ORF2 assay is designed to target predominantly the ORF2 of the A_I subfamily. Three SNPs were strategically placed in the forward primer (at the very 3′ end), the reverse primer (the penultimate 3' nucleotide), and the probe (Supplemental Figure 1). *In silico* mapping of the primers returned a total of 25,342 hits per haploid genome, among which 3411 (13%) had no mismatch and the majority (16,445; 65%) had one SNP (Fig. 2B, top panel). The C > T mismatch at position 11 is centrally located, and it occurs at a frequency of 0.79, corresponding to A_II/III/IV, Tf_I/II, and Gf_II consensus sequences. The C > T mismatch at position 16 is off center; it occurs at a frequency of 0.06 and corresponds to the Gf_I subfamily consensus (compare Supplemental Figure 1 with Fig. 2B, middle panel). The A-ORF2;U6 duplex assay reports 15.4 copies of L1 per U6, corresponding to 5498 copies of L1 targets per cell (Fig. 4B, bottom panel). As shown in the subsequent section, this measurement is in agreement with a scenario that single SNPs at position 11 and 16 excluded the corresponding targets from detection (i.e., A_II/III/IV, Tf_I/II, and Gf_I/II subfamily sequences).



G

Assay Name	L1 Targets	L1 Region	Copy Number (from ddPCR)	Perfect Match (from *in silico*)	Total Copies (from *in silico*)
L1Md_A-Tether	L1Md_AI, AII	Tether	4,498	5,738	10,640
L1Md_A-ORF2	L1Md_AI	ORF2	5,498	6,822	50,684
L1Md_Tf-Tether	L1Md_TfI, TfII	Tether	3,177	3,794	3,942
L1Md_Tf-ORF2	L1Md_TfI, TfII	ORF2	5,748	4,024	23,344
L1Md_Tf-ORF2^{LNA}	L1Md_TfI, TfII	ORF2	3,641	4,144	25,232
L1Md_Tf-ORF2^{LNA-1}	L1Md_TfI, TfII	ORF2	5,569	5,076	29,830

(caption on next page)

For L1Md_Tf subfamilies, we designed one assay targeting the 5′UTR tether and three assays targeting the ORF2 region of Tf_I and Tf_II subfamilies (Supplemental Figure 1). The L1Md_Tf-Tether assay is designed to target a unique region in the Tf_I/II tether sequences. Its forward primer is highly specific to Tf subfamily elements (Supplemental Figure 1). *In silico* mapping of the primers returned 3721 hits, among which 1897 (51%) matched perfectly to the designed probe, and 2% had one or more mismatch to the probe sequence. The

Fig. 4. L1 subfamily specific ddPCR assays in duplex with the U6 reference. (**A**) L1Md_A-Tether assay. (**B**) L1Md_A-ORF2 assay. (**C**) L1Md_Tf-Tether assay. (**D**) L1Md_Tf-ORF2 assay. (**E**) L1Md_Tf-ORF2^{LNA} assay. (**F**) L1Md_Tf-ORF2^{LNA-1} assay. Each L1 assay is designed to target specific L1 subfamilies as well as different regions on the L1 sequence. For each assay, the related data are organized into three subpanels. **Top subpanel:** *In silico* mapping of L1 assay targets. Potential PCR amplicons were extracted from the mouse genome. The bar chart reports the number of hits that contain 0 up to 5 SNPs in the designed probe region. **Middle subpanel:** The probe sequence conservation and location of SNPs among all extracted sequences were visualized using sequence logo. The Y-axis displays the nucleotide composition (expressed as the probability of having A, C, G, T or a gap), and the X-axis marks the position of each nucleotide in the probe. Note not all nucleotide positions add up to 1.0 due to the presence of gaps in the alignment. By default, the most dominant nucleotide at each position is contained in the probe. At positions where there are multiple apparent nucleotide variants, the non-probe variants are shaded in gray. At positions where the gap constitutes a major variant, they are highlighted by a yellow background (panel C). **Bottom subpanel:** 2-D droplet plot of duplex ddPCR for the L1 target with U6 reference. Each point on the plot represents one individual droplet. The fluorescent amplitude of each droplet is read under two fluorescent channels. The L1 amplitude is depicted in the Y-axis (FAM channel) and the U6 reference on the X-axis (VIC channel). Thresholds (magenta lines; cutoff values are denoted by the values in the white boxes) are established in areas of minimal rain by examination of both the 2-D plots and 1-D histograms provided by the software and in consultation with empirical cluster validation results. The L1/U6 ratio is noted for each reaction in the bottom corner of each plot (in parentheses followed are L1 copies per cell, which were based on the experimentally derived 357 U6/cell quantification [see Fig. 2A]). Droplet density is represented with a color gradient from blue to red (low density to high). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

remainder of the hits (47% of the total) contained a 15-bp insertion (Fig. 4C, middle panel), which are expected to be excluded from detection by the probe-based ddPCR assay. Further analysis indicated that the insertion-containing hits were derived from Tf_III subfamily elements (not shown). Indeed, the Tf-Tether;U6 duplex assay was borne out to be the most selective among all Tf-specific assays. It reports 3177 copies of L1 targets per cell (Fig. 4C, bottom panel), the lowest when compared to the Tf-ORF2 assays described below.

The L1Md_Tf-ORF2 assay was designed to distinguish Tf_I/II targets from other subfamily sequences by including two diagnostic SNPs in the probe region (Supplemental Figure 1). *In silico* mapping of the primers identified 11,672 potential targets in the mouse genome, among which 2012 (17%) had no mismatch in the probe region, 99 (1%) with one SNP, 4943 (42%) with two SNPs, and the remainder (40%) with three or more SNPs (Fig. 4D, top panel). Sequence logo analysis revealed that the two most frequent mismatches were A > C at position 3 and T > C at position 13 in the probe region (Fig. 4D, middle panel), corresponding to the consensus sequences for subfamilies A_I/II/III/IV and Gf_I/II (Supplemental Figure 1). The Tf-ORF2;U6 duplex assay reports 5748 copies of L1 per cell (Fig. 4D, bottom panel).

In an attempt to further increase the assay precision, we tested locked nucleic acid (LNA) probes, which are known to help improve mismatch discrimination [68]. Two LNA probes were designed to target Tf_I/II subfamilies (Fig. 4E and F). Unlike the L1Md_Tf-ORF2 assay mentioned above, the L1Md_Tf-ORF2^{LNA} assay targets only the first discriminating SNP that is present in the L1Md_Tf-ORF2 probe (i.e., the A > C mutation at position 3 in Tf_ORF2 probe; now shifted to position 5 in the 12-nt long Tf-ORF2^{LNA} probe) (Fig. 4E, middle panel). Mapping of the primers returned 12,616 hits, among which 2072 (16%) had no mismatch, 57% with one SNP, and 26% with two or more SNPs (Fig. 4E, top panel). The Tf-ORF2^{LNA};U6 duplex assay reports 3641 L1 targets per cell (Fig. 4E, bottom panel). The second LNA probe, L1Md_Tf-ORF2^{LNA-1}, is designed to target two discriminating SNPs within 1 kb from the end of ORF2 (Supplemental Fig. 1A and B). *In silico* mapping of the primers returned 14,915 hits, among which 2538 (17%) had no mismatch in the probe region, 5% with one SNP, 57% with two SNPs, and the remainder with three or more SNPs (Fig. 4F, top panel). Sequence logo analysis of the mapped hits revealed two dominant SNPs at position 8 and 11 (Fig. 4F, middle panel), corresponding to those in the consensus sequences of A_I/II/III/IV and Gf_I/II (Supplemental Figure 1A). The Tf-ORF2^{LNA-1};U6 duplex assay reports 5569 copies of L1 targets per cell (Fig. 4F, bottom panel). Overall, neither of the two LNA probes showed significant improvement in terms of droplet clustering or lowering the copy number measured. However, as illustrated here, the use of LNA probes allowed the use of much shorter probe with similar or better mismatch discrimination than standard DNA probes [68]. This feature can be advantageous as it provides flexibility when designing a suitable probe targeting a single SNP in the L1 sequence.

For ddPCR assays, the total error (subsampling and partitioning errors combined) is at its lowest when the concentration of target is ~1.5 copies per droplet [32]. In all of our endogenous L1 assays, the copy number of L1 targets ranges from 3177 to 5780 (Fig. 4G). In contrast, the U6 reference is measured at 357 copies per cell (Fig. 2E). To explore the influence of DNA concentration on the total error, we retrospectively analyzed 88 L1Md_Tf-ORF2^{LNA-1};U6 duplex reactions by plotting the Poisson error for the L1 target alone, for U6 target alone, and for L1/U6 targets combined as a function of the reported copy numbers (Supplemental Fig. 7A-C, respectively). As expected, for samples with lower concentration, error is predominately from U6 (15% from U6 versus 5% from L1). For samples with higher concentration, it is the reverse (5% from U6 versus > 10% from L1). Coincidently, the majority of the samples fall into the middle range of input DNA concentrations, showing the lowest overall Poisson error. This optimal range corresponds to 3000–6000 copies of L1 and 150–300 copies of U6 per μL (i.e., 3–6 copies of L1 and 0.15–0.3 copies of U6 per droplet) (Supplemental Figure 7C). Remarkably, the copy number reported from *in silico* analysis and the experimentally derived ddPCR copy number for both the references and the L1 targets are consistently similar (Fig. 2E; Fig. 4G). Not only does this highlight the importance of a robust initial *in silico* design but also the subfamily specificity and sensitivity of the ddPCR assays.

3.5. Evaluate assay specificity and identify droplet clusters using subfamily-specific synthetic targets or L1 plasmids

So far we have focused on selected L1 targets (Fig. 4). These and six additional L1 probes (see all 12 alignments in Supplemental Figure 1) generated multiple consistent and defined droplet clusters as highlighted by 1-D droplet plots and histograms (Supplemental Figure 8). Presumably, these clusters are primarily derived from different subfamily specific SNPs. To test this hypothesis, we subjected 4 endogenous L1 assays (L1Md_A-Tether, L1Md_A-ORF2, L1Md_Tf-ORF2, L1Md_Tf-ORF2^{LNA-1}) to synthetic gene fragments containing different subfamily consensus sequences for the assay target region (Fig. 5A–D; Supplemental Figure 9). The complete sequence for the gene fragments is provided elsewhere (Supplemental Table 2). As an alternative to synthetic DNA templates, plasmid DNA containing specific endogenous L1 loci were tested in an overlapping subset of 8 different L1 assays (Supplemental Figure 10). Overall, ddPCR reactions containing the perfect-match gene fragment template (lane “match”) produced droplets with a fluorescent amplitude similar to the highest droplet cluster from the gDNA (Fig. 5). On the other hand, the reactions with gene fragments containing SNPs matching off-target subfamilies (lanes “mis-1” and “mis-2”) produced droplet clusters of much lower fluorescent amplitude (Fig. 5; detailed below). Such clustering profiles were used as a guide to set the threshold for gDNA reactions so that only droplets from the intended L1 subfamily targets were counted as positive (see Fig. 4).

Specifically, for the L1Md_A-Tether assay, the mis-1 synthetic target corresponds to the Tf_I/II consensus containing 5 discriminating

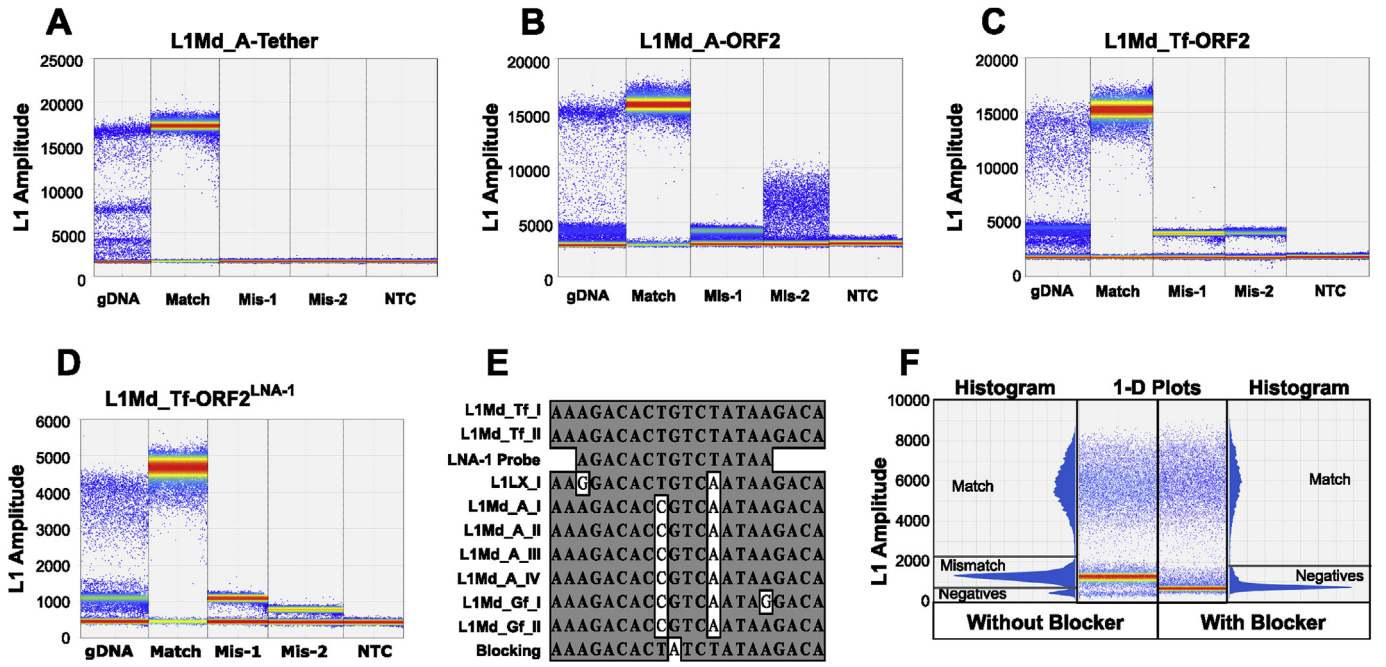


Fig. 5. Validation of droplet clusters with synthetic gene fragments. Four assays were selected for droplet cluster validation, A-Tether (**A**), A-ORF2 (**B**), Tf-ORF2 (**C**), Tf-ORF2^{LNA-1} (**D**). Synthetic gene fragments were designed for targeted subfamily (match) and two off-target subfamilies (mismatch-1 or mis-1; and mismatch-2 or mis-2), see Supplemental Fig. 9 for alignments. NTC, no template control. Droplet clusters resulting from the perfect-match gene fragment had slightly higher amplitude than those generated from gDNA, most likely due to shortening of synthetic template during design as we removed most of the intervening sequence between the primers and probes in the gene fragment. Overall, there is a highly consistent correlation between the droplet amplitude from off-target subfamily gene fragments and the low amplitude cluster(s) from the gDNA. **(E)** Alignment of the blocking oligo with 8 L1 subfamily consensus sequences across the L1Md_Tf-ORF2^{LNA-1} probe binding region. **(F)** Comparison of L1Md_Tf-ORF2^{LNA-1} droplet profiles with or without 4 μM blocking oligo (8x probe concentration). Gray dashed lines define the boundary of each droplet cluster.

mismatches in the probe region, and mis-2 synthetic target is the consensus of Gf_I/II, with 4 mismatches in the probe region (Supplemental Figure 9A; see Supplemental Figure 1 for subfamily alignment). Both produced no positive fluorescent droplets when serving as the only DNA template in the ddPCR reaction (Fig. 5A). While it is possible that the presence of multiple SNPs in the synthetic template singularly abolished its hybridization to the probe, the specificity of the A-Tether assay was aided by its forward primer, which is highly unique for A_I/II/III/IV subfamilies as all alternative subfamily (i.e., Tf_I/II and Gf_I/II) sequences are expected to be poor substrates for the forward primer (Fig. 5A; Supplemental Figure 1). Just like the perfectly matched synthetic gene fragments, two endogenous A_I elements produced a single high fluorescence cluster when supplied as plasmid DNA (Supplemental Figure 10A; A_I plasmid 1 and plasmid 2). Meanwhile, two endogenous Tf_I elements showed no discernable fluorescence above the background (Supplemental Figure 10A; Tf_I plasmid 1 clone a, b and c, and plasmid 2 clone a and b). Our testing did not uncover the identity of the two low fluorescence clusters seen in the gDNA reactions. We speculate that these signals were derived from A_III/IV subfamily sequences, which had a single T > A/C mismatch at position 1 but no mismatches in other positions in the consensus (Fig. 4A, middle panel). To make the measurement specific to A_I/II subfamily, we set the threshold at just below the "perfect match" cluster, thus excluding all other subfamilies (Fig. 4A, bottom panel).

For the L1Md_A-ORF2 assay, we tested two synthetic targets each with a single mismatch to the probe designed for A_I subfamily. The synthetic target mis-1 contains a C > T mismatch at position 11, corresponding to A_II/III/IV, Tf_I/II, and Gf_II subfamily consensus sequences (Supplemental Figure 9B; see Supplemental Figure 1 for subfamily alignment). It is centrally located in the 20-nt probe, thus producing a tight cluster of low fluorescence (Fig. 5B). A similar, low fluorescence cluster was observed using the Tf_I plasmid clones (Supplemental Figure 10B). There was a prominent cluster at the same amplitude in the droplet profile from gDNA, reflecting the relative abundance of this SNP (Fig. 4B, middle panel) and associated subfamilies in the mouse genome [38]. In contrast, the synthetic target mis-2 contains a C > T mismatch at position 16, corresponding to the Gf_I subfamily consensus sequence (Supplemental Figure 9B; see Supplemental Figure 1 for subfamily alignment). Potentially due to its more peripheral location in the probe, the fluorescence cluster formed was of higher amplitude and not as compact as that from mis-1. The gDNA lacks a distinct cluster at this position, reflecting its low abundance in the genome [3,38]. In our A-ORF2 assay, we set the threshold above the mis-1 cluster, thus excluding the mismatched targets (mainly, A_II/III/IV, Tf_I/II and Gf_II) from the measurement (Fig. 4B).

For the L1Md_Tf-ORF2 assay, the synthetic target mis-1 contains three SNPs that are unique to A_I consensus, although two of the three SNPs are shared by A_II/III/IV and Gf_I/II subfamilies. The mis-2 synthetic target contains three SNPs that are unique to the Gf_I subfamily, with G > A at position 9 centrally located (Supplemental Figure 9C; see Supplemental Figure 1 for subfamily alignment). Both synthetic targets yielded a tight, low fluorescence cluster of similar amplitude (Fig. 5C), which was also present in the gDNA reactions, suggesting these SNPs are sufficient to discriminate between A_I target and other off-target subfamilies. In parallel experiments with plasmid DNA, A_I element containing plasmids produced the characteristic low fluorescence cluster while Tf_I plasmid clones all displayed the expected high fluorescence cluster (Supplemental Figure 10C). In our Tf-ORF2 assay, we set the threshold above these low fluorescence clusters, thus excluding the mismatched targets from the measurement (Fig. 4D).

For the L1Md_Tf-ORF2^{LNA-1} assay, the synthetic target mis-1 contains two SNPs that are found in A_I/II/III/IV and Gf_I/II subfamilies (position 7 and 11). The mis-2 contains one additional A > G SNP at position 15 that is unique for the Gf_I subfamily (Supplemental Figure 9D; see Supplemental Figure 1 for subfamily alignment). Mis-1 formed a tight low fluorescence cluster that was also present in gDNA reactions.

The addition of the A > G SNP in mis-2, although at the very 3' end of the probe position, reduced the amplitude further toward the baseline fluorescence for negative droplets (Fig. 5D). When tested with plasmid-borne L1 sequences, Tf_I plasmid 1/2 and Tf_II plasmid 1 all yielded high fluorescence clusters (Supplemental Fig. 10D and E). Tf_I plasmid 3 is an interesting case. It is a full-length Tf_I element located on chromosome 7:72927724–72934575 but unexpectedly contains three SNPs in the probe region (two of them are the same found in A and Gf elements at position 7 and 11; the third one is at position 8) (Supplemental Figure 10I). Not surprisingly, unlike Tf_I plasmid 1 and 2, plasmid 3 produced a low fluorescence cluster (Supplemental Figure 10E). Among other mismatched targets, both A_I plasmid 1 and Tf_III plasmid 1 had the same two SNPs as in the mis-1 synthetic target and produced a droplet cluster similar to that from mis-1. As a comparison, A_I plasmid 2 only had one SNP at position 11. Interestingly, the fluorescence amplitude of the resulting cluster was nearly 2-fold higher than that from A_I plasmid 1 (Supplemental Figure 10D). Lastly, Gf_I plasmid 1 had identical SNP and clustering profile as synthetic target mis-2 (Supplemental Figure 10E). These results suggest that the probe-based ddPCR assay is highly sensitive to mismatches in the target sequence. In our Tf-ORF2^{LNA-1} assay, we set the threshold above these low fluorescence clusters, thus excluding the mismatched targets from the measurement (Fig. 4F).

To further validate the identity of the observed low fluorescent clusters, we designed an antisense blocking oligo that was unlabeled and carried a 3' phosphate to eliminate priming from the oligo (Fig. 5E). The blocking oligo was designed with an infrequently represented SNP to determine if it would improve the inter-cluster rain. Surprisingly, the addition of the blocking oligo eliminated the bulk of low fluorescent droplets (i.e., they coalesced into the negative droplet cluster) (Fig. 5F). This result is most likely due to higher competition for the probe binding to mismatched targets than to the perfect match template (which was unaffected by the blocking oligo). Our results suggest that clusters derived from a specific sequence may be experimentally excluded from quantification by using blocking oligos specific to that sequence.

3.6. Evaluate assay sensitivity using a spike-in dilution series and technical replicates

To establish the sensitivity for detecting relative changes in L1 copy number, we performed a spike-in dilution series. To this end, we first cloned a full-length endogenous L1 subfamily into a plasmid (Tf_II P2 from Supplemental Figure 10). The cloned L1 is a perfectly matched target for L1Md_Tf-ORF2 probe and primers. A dilution series of the plasmid was first established in separate tubes (Fig. 6A; green tubes). An aliquot of the diluted plasmid was then mixed with gDNA in another series of tubes (Fig. 6A: orange tubes). The amount of plasmid had been calibrated so that the spiked-in plasmid was predicted to introduce an increase in Tf target copy number by 20%, 5%, 2.5%, 1.25%, and 0.625%, respectively. To verify the accuracy of the initial plasmid dilution series, the amount of plasmid for each spike-in was quantified in 8 replicate reactions. The copy-number measurement from the simplex L1Md_Tf-ORF2 assay showed near perfect correlation with the projected changes in copy number (Fig. 6B). Subsequently, the spiked-in gDNA samples (along with the control non-spiked gDNA [blue tube]) were subjected in 8 technical replicates to the duplex L1Md_Tf-ORF2;BC1 assay, which reports the normalized L1 copies in the form of L1/BC1 ratio (Fig. 6C). The measured changes were largely consistent with the expected increases in L1 copy number owing to spiked-in plasmid DNA. However, because of variance in the replicate reactions, only spike-in samples with 20% and 5% expected increases showed statistical significance ($p < 0.01$ and $p < 0.0001$, respectively). The sample with 2.5% change is not statistically significant ($p = 0.06$). These results indicate that the L1Md_Tf-ORF2;BC1 duplex assay has a sensitivity to detect a 5% CNV of endogenous L1 when using 8 technical replicates per sample.

Conceptually, the digital counting nature of ddPCR circumvents the need for technical replicates. Nevertheless, technical replicates are expected to provide an estimate of the precision of the measurement [69]. To take a closer look at the potential benefits of using technical replicates, we first generated empirical data by subjecting a single gDNA sample to 40 replicate reactions. We plotted the calculated L1/BC1 ratio and the associated Poisson error (as 95% CI) for individual reactions (Fig. 6D). It is evident that, for the vast majority of the individual reactions (39 out of 40), the corresponding 95% CI would capture the expected L1 copy number in this gDNA sample. However, the reported L1/BC1 ratio does vary from one reaction to another, ranging from 3.265 to 3.516. There are two different approaches to reporting the overall copy number measurement and associated error. One is the standard mathematical calculation from all the replicates. The average L1/BC1 ratio among the 40 replicate reactions is 3.402 (95%CI [3.386–3.418]). The other approach is to first pool the droplets from all reactions together and then treat them as a single "mega-well". This latter approach generates a "merged" L1/BC1 ratio at 3.402 (95%CI [3.382–3.422]). Using our data, the two different approaches gave nearly identical copy number estimates and error. The precision of the 40 replicates, as measured by coefficient of variation (CV), is 1.54%. Obviously, it would be cost prohibitive and impractical to conduct 40 ddPCR reactions per sample. To investigate the cost-benefit ratio of the number of replicates, we performed a simulation using this empirical data set. We randomly assigned samples into different replicate groups (1, 3, 6, 8, 16, 24, 32, or 40 replicates), each iterated for 1000 times (Fig. 6E). The calculated average for the L1/BC1 ratio showed little change among different replicate groups. However, the total assay variation decreases as a function of the number of replicates. For example, the spread of the data range is reduced by 5-fold as the number of replicates increases from one (L1/BC1 ratio range = 0.291) to 40 (L1/BC1 ratio range = 0.058) (Fig. 6E), corresponding to a 6.3-fold increase in assay precision (Fig. 6F). Considering the cost of ddPCR assays, 6–8 replicates should be adequate for most applications. Nevertheless, the added precision at higher replicates could increase the assay sensitivity for endogenous L1 quantification. Indeed, when the number of technical replicates increased to 24, we were able to detect an expected 1.5% difference in L1 target copy number from a subsequent Tf plasmid spike-in (Supplemental Figure 11; $p < 0.0001$). It is important to note that these are matched samples. Other variations associated with biological samples may reduce the differentiating power of the assay.

4. Conclusions

L1 copy number quantification is key to our understanding of the regulation and impact of L1 mobilization in the genome. The quantification method described here is the first comprehensive analysis of the endogenous mouse L1 copy number using ddPCR. Sample partitioning in ddPCR affords the opportunity to differentiate heterogeneous target sequences into distinct droplet clusters. However, due to the highly repetitive nature and the abundance of polymorphic copies, the behavior of an L1 target assay can be unpredictable. We provided experimental evidence that specific clusters originated from amplification of individual matched and mismatched target sequences by using synthetic target sequences or plasmid clones (specificity validation). Our ability to identify off-target clusters enabled evidence-based differential thresholding, which ensured selective reporting of target L1 subfamilies, representing a unique advantage of ddPCR over qPCR. We further provided a roadmap for L1 assay optimization with the goal of better cluster separation. It involves several complementary strategies. The first, is in the design of L1 primers and probe, maximizing the differentiation power by targeting subfamily specific SNPs. The second, is to adjust the balance between primer Tm and annealing temperature. The third, is to selectively eliminate certain clusters with blocking

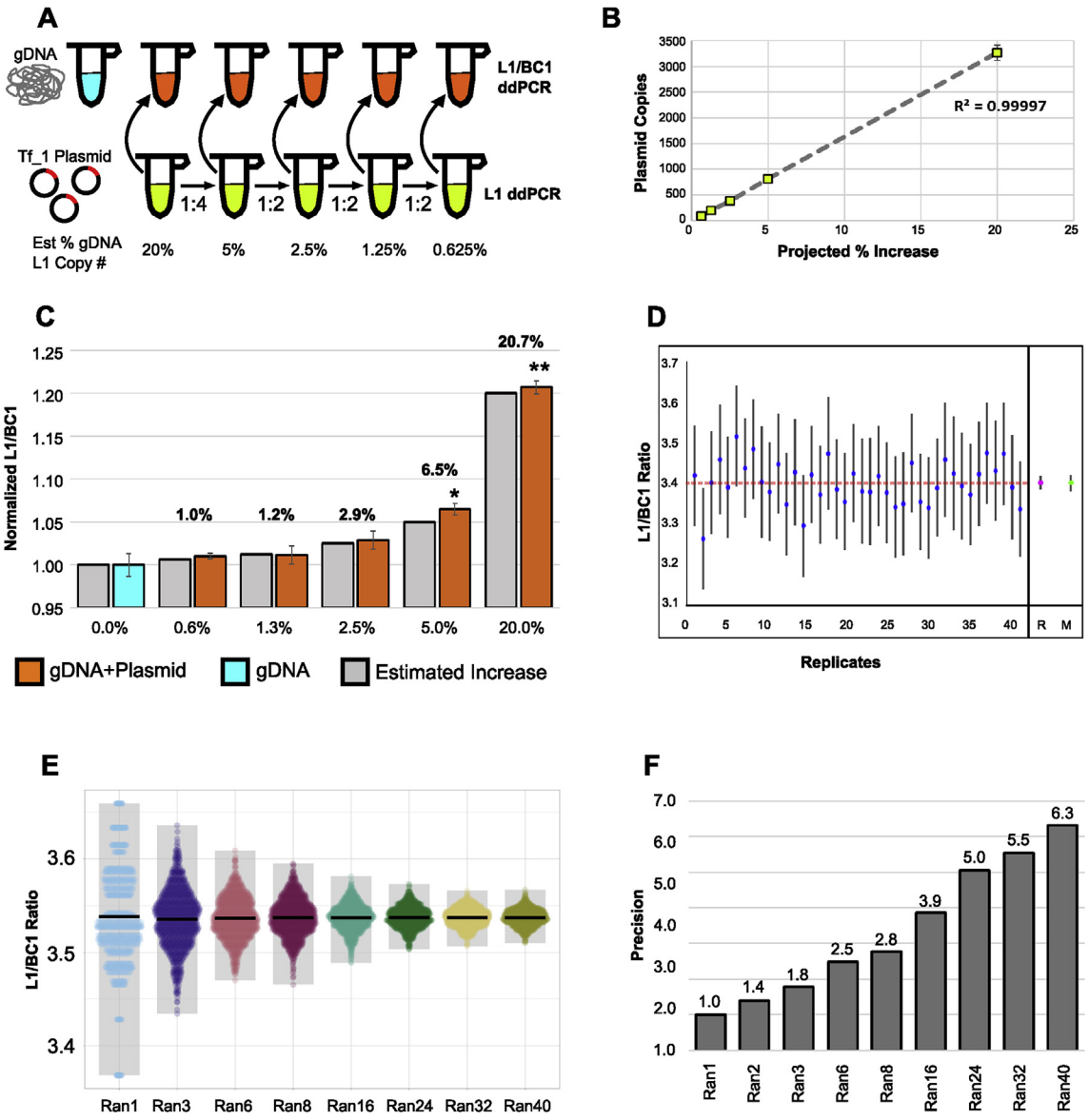


Fig. 6. Effect of technical replicates on L1 ddPCR sensitivity and precision. **(A)** Experimental setup for validating L1 ddPCR assay sensitivity. A plasmid containing the target subfamily (green tubes) was spiked into gDNA (blue tube) starting at an estimated 20% increase down to a 0.625% increase (orange tubes). **(B)** The plasmid input (green tubes) was quantified with the simplex L1Md_Tf-ORF2 assay and showed a highly linear dilution. Error bars represent standard deviations. **(C)** The L1 copy number change (orange tubes vs blue tube) was quantified with the duplex L1Md_Tf-ORF2;BC1 assay. Eight technical replicates were used for quantification of each sample. While only the two highest spike-in samples were statistically significant, all of the remaining gDNA + spike-in samples were higher than the gDNA and consistent with the estimated plasmid input trend. * p-value < 0.01; ** p-value < 0.0001 (one-tailed Student's t-test). Error bars represent standard errors of the mean. **(D)** To assess the cost/benefit of technical replicates, 40 aliquots of the same gDNA sample were quantified on one 96-well plate using the L1Md_Tf-ORF2;BC1 assay. The ratio of each data point and the 95% Poisson CI are plotted in blue. The mean of the 40 samples combined is represented by the dashed red line. The mean and 95% CI for the combined 40 samples were calculated by two distinct approaches: either considering each replicate as a separate sample (replicate or 'R'; in magenta) or combining all reactions into one 'mega-well' using the ddPCR software and deriving a total 95% Poisson CI (merged or 'M'; in green). **(E)** Simulation of the effect of technical replicate group size on assay precision. From an initial pool of 40 reactions, reactions were randomly assigned into different replicate groups (1, 3, 6, 8, 16, 24, 32, or 40 replicates), each iterated for 1000 times. Note a specific reaction could be sampled more than once. As expected, larger technical replicates increase the precision of the data by decreasing the spread and distribution of the output ratios. **(F)** Increase in assay precision for each technical replicate group. The Y-axis is the fold change normalized to the single technical replicate group. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

oligos (cluster knockout). Similar strategies were applied to the selection and testing of reference targets. In addition, we determined the assay sensitivity and precision by using a spike-in experiment (sensitivity and precision benchmark). Not only can this method be utilized to report the copy number of specific L1 subfamilies, but it is also sufficiently sensitive to detect below a 2% change in endogenous L1 copy number. As the baseline copy number for the current assay ranges 3000–6000 copies per cell, these assays may be applied to detect changes in L1 copy number at 150–300 copies per cell. The assays presented here can easily be utilized for analysis of extrachromosomal L1 cDNA in aging [70,71] and other pathophysiological processes. Among the subfamily specific L1 assays presented here we recommend the use of the A-Tether and Tf-Tether assays in duplex with U6 for the optimal quantification of the L1Md_A and L1Md_Tf subfamilies, respectively. Both tether-based assays displayed clear clustering, the lowest copy number, and high specificity towards the selected

subfamilies owing to the highly discriminating SNPs in the forward primer and probe. In addition, the tether assays could be supplemented with L1Md_A-ORF2 and L1Md_Tf-ORF^{LNA-1} assays as these two ORF2-based assays show good droplet clustering and specificity towards their respective subfamilies. As for the choice of reference probes, either U6 or GAPDH is preferred due to their favorable clustering profiles, enabling consistency across experiments. BC1 has a less desirable clustering profile but it should be considered especially if L1 assays are detecting higher copy number (e.g. > 8000/cell). Further efforts should be focused on the reduction of the L1 baseline copy number so that it can detect even smaller changes in the absolute L1 copies per cell. This goal can potentially be achieved by thorough *in silico* exploration, additional improvement of the target specificity, increase of technical replicates, and data gridding [72]. The data presented here strongly suggest that, with extensive optimization, a well-designed ddPCR assay could target a single or small number of L1 elements with unique SNPs. The same approach presented here can serve as a guide for the development of ddPCR based assay for quantifying human L1 copy number, especially given the fact that there are fewer active elements and subfamilies in the human genome [2]. While L1 quantification was the focus of this work, the design and optimization described here could be beneficial for other high copy targets or quantification of highly similar but distinct sequences (e.g. microbiomic, somatic or germline mosaicism, and gene editing applications). Lastly, while droplet sorting is not yet commercially available it has been proven experimentally sound in several cases [73]. Using ddPCR for quantification, for any of the above purposes, opens up very interesting aspects of post-quantification enrichment of select droplets and high-throughput sequencing of amplicons or templates.

Author contributions

Simon Newkirk: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data Curation, Visualization, Writing - Original Draft, Writing - Review & Editing.

Wenfeng An: Conceptualization, Methodology, Validation, Formal analysis, Data Curation, Visualization, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Ping Ye: Software, Methodology, Data Curation, Writing - Original Draft, Writing - Review & Editing.

Lingqi Kong: Resources, Validation.

Mason Jones: Resources, Validation.

Chase Habben: Resources, Validation.

Victoria Bishop: Resources, Validation.

Acknowledgments and funding

We thank Mike Ducat and Brandon Mckethan for technical advice on ddPCR. The work was supported by National Institutes of Health [grant numbers R21HD080143, R21OD017965, R15GM131263, and R03HD099412]. W.A. was supported, in part, by the Markl Faculty Scholar Fund.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ab.2020.113779.

References

[1] S.J. Newkirk, W. An, L1 regulation in mouse and human germ cells, in: G. Cristofari (Ed.), Human Retrotransposons in Health and Disease, Springer International Publishing, 2017, pp. 29–61.

[2] B. Brouha, J. Schustak, R.M. Badge, S. Lutz-Prigge, A.H. Farley, J.V. Moran, H.H. Kazazian Jr., Hot L1s account for the bulk of retrotransposition in the human population, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 5280–5285.

[3] J.L. Goodier, E.M. Ostertag, K. Du, H.H. Kazazian Jr., A novel active L1 retrotransposon subfamily in the mouse, Genome Res. 11 (2001) 1677–1685.

[4] J.W. Jachowicz, M.E. Torres-Padilla, LINEs in mice: features, families, and potential roles in early development, Chromosoma 125 (2016) 29–39.

[5] K.H. Burns, Transposable elements in cancer, Nat. Rev. Canc. 17 (2017) 415–424.

[6] P.S. Saha, W. An, Recently Mobilised Transposons in the Human Genome, eLS, John Wiley & Sons, Ltd, 2020, pp. 1–10.

[7] B. Rodriguez-Martin, E.G. Alvarez, A. Baez-Ortega, J. Zamora, F. Supek, J. Demeulemeester, M. Santamarina, Y.S. Ju, J. Temes, D. Garcia-Souto, H. Detering, Y. Li, J. Rodriguez-Castro, A. Dueso-Barroso, A.L. Bruzos, S.C. Dentro, M.G. Blanco, G. Contino, D. Ardeljan, M. Tojo, N.D. Roberts, S. Zumalave, P.A.W. Edwards, J. Weischenfeldt, M. Puiggros, Z. Chong, K. Chen, E.A. Lee, J.A. Wala, K. Raine, A. Butler, S.M. Waszak, F.C.P. Navarro, S.E. Schumacher, J. Monlong, F. Maura, N. Bolli, G. Bourque, M. Gerstein, P.J. Park, D.C. Wedge, R. Beroukhim, D. Torrents, J.O. Korbel, I. Martincorena, R.C. Fitzgerald, P. Van Loo, H.H. Kazazian, K.H. Burns, P.S.V.W. Group, P.J. Campbell, J.M.C. Tubio, P. Consortium, Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition, Nat. Genet. 52 (2020) 306–319.

[8] D.C. Hancks, H.H. Kazazian Jr., Roles for retrotransposon insertions in human disease, Mobile DNA 7 (2016) 9.

[9] H. Kano, I. Godoy, C. Courtney, M.R. Vetter, G.L. Gerton, E.M. Ostertag, H.H. Kazazian Jr., L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism, Genes Dev. 23 (2009) 1303–1312.

[10] S.J. Newkirk, S. Lee, F.C. Grandi, V. Gaysinskaya, J.M. Rosser, N. Vanden Berg, C.A. Hogarth, M.C.N. Marchetto, A.R. Muotri, M.D. Griswold, P. Ye, A. Bortvin, F.H. Gage, J.D. Boeke, W. An, Intact piRNA pathway prevents L1 mobilization in male meiosis, Proc. Natl. Acad. Sci. U. S. A. 114 (2017) E5635–E5644.

[11] S.R. Richardson, P. Gerdes, D.J. Gerhardt, F.J. Sanchez-Luque, G.O. Bodea, M. Munoz-Lopez, J.S. Jesuadian, M.H.C. Kempen, P.E. Carreira, J.A. Jeddeloh, J.L. Garcia-Perez, H.H. Kazazian Jr., A.D. Ewing, G.J. Faulkner, Heritable L1 retrotransposition in the mouse primordial germline and early embryo, Genome Res. 27 (2017) 1395–1405.

[12] H.H. Kazazian Jr., C. Wong, H. Youssoufian, A.F. Scott, D.G. Phillips, S.E. Antonarakis, Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man, Nature 332 (1988) 164–166.

[13] B. Brouha, C. Meischl, E. Ostertag, M. de Boer, Y. Zhang, H. Neijens, D. Roos, H.H. Kazazian Jr., Evidence consistent with human L1 retrotransposition in maternal meiosis I, Am. J. Hum. Genet. 71 (2002) 327–336.

[14] J.A. van den Hurk, I.C. Meij, M.C. Seleme, H. Kano, K. Nikopoulos, L.H. Hoefsloot, E.A. Sistermans, I.J. de Wijs, A. Mukhopadhyay, A.S. Plomp, P.T. de Jong, H.H. Kazazian, F.P. Cremers, L1 retrotransposition can occur early in human embryonic development, Hum. Mol. Genet. 16 (2007) 1587–1592.

[15] J. Feusier, W.S. Watkins, J. Thomas, A. Farrell, D.J. Witherspoon, L. Baird, H. Ha, J. Xing, L.B. Jorde, Pedigree-based estimation of human mobile element retrotransposition rates, Genome Res. 29 (2019) 1567–1577.

[16] A.R. Muotri, V.T. Chu, M.C. Marchetto, W. Deng, J.V. Moran, F.H. Gage, Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition, Nature 435 (2005) 903–910.

[17] J.K. Baillie, M.W. Barnett, K.R. Upton, D.J. Gerhardt, T.A. Richmond, F. De Sapio, P.M. Brennan, P. Rizzu, S. Smith, M. Fell, R.T. Talbot, S. Gustincich, T.C. Freeman, J.S. Mattick, D.A. Hume, P. Heutink, P. Carninci, J.A. Jeddeloh, G.J. Faulkner, Somatic retrotransposition alters the genetic landscape of the human brain, Nature 479 (2011) 534–537.

[18] Y. Miki, I. Nishisho, A. Horii, Y. Miyoshi, J. Utsunomiya, K.W. Kinzler, B. Vogelstein, Y. Nakamura, Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer, Canc. Res. 52 (1992) 643–645.

[19] E.C. Scott, E.J. Gardner, A. Masood, N.T. Chuang, P.M. Vertino, S.E. Devine, A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer, Genome Res. 26 (2016) 745–755.

[20] G.D. Evrony, X. Cai, E. Lee, L.B. Hills, P.C. Elhosary, H.S. Lehmann, J.J. Parker, K.D. Atabay, E.C. Gilmore, A. Poduri, P.J. Park, C.A. Walsh, Single-neuron sequencing analysis of l1 retrotransposition and somatic mutation in the human brain, Cell 151 (2012) 483–496.

[21] G.D. Evrony, E. Lee, P.J. Park, C.A. Walsh, Resolving rates of mutation in the brain using single-neuron genomics, eLife 5 (2016).

[22] K.R. Upton, D.J. Gerhardt, J.S. Jesuadian, S.R. Richardson, F.J. Sanchez-Luque, G.O. Bodea, A.D. Ewing, C. Salvador-Palomeque, M.S. van der Knaap, P.M. Brennan, A. Vanderver, G.J. Faulkner, Ubiquitous L1 mosaicism in hippocampal neurons, Cell 161 (2015) 228–239.

[23] W. An, J.S. Han, S.J. Wheelan, E.S. Davis, C.E. Coombes, P. Ye, C. Triplett, J.D. Boeke, Active retrotransposition by a synthetic L1 element in mice, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 18662–18667.

[24] D.V. Babushok, E.M. Ostertag, C.E. Courtney, J.M. Choi, H.H. Kazazian Jr., L1 integration in a transgenic mouse model, Genome Res. 16 (2006) 240–250.

[25] C.R. Huang, A.M. Schneider, Y. Lu, T. Niranjan, P. Shen, M.A. Robinson, J.P. Steranka, D. Valle, C.I. Civin, T. Wang, S.J. Wheelan, H. Ji, J.D. Boeke, K.H. Burns, Mobile interspersed repeats are major structural variants in the human genome, Cell 141 (2010) 1171–1182.

[26] A.D. Ewing, H.H. Kazazian Jr., High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes, Genome Res. 20 (2010) 1262–1270.

[27] R.C. Iskow, M.T. McCabe, R.E. Mills, S. Torene, W.S. Pittard, A.F. Neuwald, E.G. Van Meir, P.M. Vertino, S.E. Devine, Natural mutagenesis of human genomes by endogenous retrotransposons, Cell 141 (2010) 1253–1261.

[28] N.G. Coufal, J.L. Garcia-Perez, G.E. Peng, G.W. Yeo, Y. Mu, M.T. Lovci, M. Morell, K.S. O'Shea, J.V. Moran, F.H. Gage, L1 retrotransposition in human neural

progenitor cells, Nature 460 (2009) 1127–1131.
[29] A.R. Muotri, M.C. Marchetto, N.G. Coufal, R. Oefner, G. Yeo, K. Nakashima, F.H. Gage, L1 retrotransposition in neurons is modulated by MeCP2, Nature 468 (2010) 443–446.
[30] M. De Cecco, S.W. Criscione, E.J. Peckham, S. Hillenmeyer, E.A. Hamm, J. Manivannan, A.L. Peterson, J.A. Kreiling, N. Neretti, J.M. Sedivy, Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements, Aging Cell 12 (2013) 247–256.
[31] P. Vitullo, I. Sciamanna, M. Baiocchi, P. Sinibaldi-Vallebona, C. Spadafora, LINE-1 retrotransposon copies are amplified during murine early embryo development, Mol. Reprod. Dev. 79 (2012) 118–127.
[32] S. Weaver, S. Dube, A. Mir, J. Qin, G. Sun, R. Ramakrishnan, R.C. Jones, K.J. Livak, Taking qPCR to a higher level: analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution, Methods 50 (2010) 271–276.
[33] B.J. Hindson, K.D. Ness, D.A. Masquelier, P. Belgrader, N.J. Heredia, A.J. Makarewicz, I.J. Bright, M.Y. Lucero, A.L. Hiddessen, T.C. Legler, T.K. Kitano, M.R. Hodel, J.F. Petersen, P.W. Wyatt, E.R. Steenblock, P.H. Shah, L.J. Bousse, C.B. Troup, J.C. Mellen, D.K. Wittmann, N.G. Erndt, T.H. Cauley, R.T. Koehler, A.P. So, S. Dube, K.A. Rose, L. Montesclaros, S. Wang, D.P. Stumbo, S.P. Hodges, S. Romine, F.P. Milanovich, H.E. White, J.F. Regan, G.A. Karlin-Neumann, C.M. Hindson, S. Saxonov, B.W. Colston, High-throughput droplet digital PCR system for absolute quantitation of DNA copy number, Anal. Chem. 83 (2011) 8604–8610.
[34] F. Bizouarn, Introduction to digital PCR, Methods Mol. Biol. 1160 (2014) 27–41.
[35] A.S. Basu, Digital assays Part I: partitioning statistics and digital PCR, SLAS Technol. 22 (2017) 369–386.
[36] T.B. White, A.M. McCoy, V.A. Streva, J. Fenrich, P.L. Deininger, A droplet digital PCR detection method for rare L1 insertions in tumors, Mobile DNA 5 (2014) 30.
[37] T.A. Bedrosian, C. Quayle, N. Novaresi, F.H. Gage, Early life experience drives structural variation of neural genomes in mice, Science 359 (2018) 1395–1399.
[38] A. Sookdeo, C.M. Hepp, M.A. McClure, S. Boissinot, Revisiting the evolution of mouse LINE-1 in the genomic era, Mobile DNA 4 (2013) 3.
[39] M. Postma, J. Goedhart, PlotsOfData-A web app for visualizing data together with their summaries, PLoS Biol. 17 (2019) e3000202.
[40] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (2012) 357–359.
[41] U. Bodenhofer, E. Bonatesta, C. Horejs-Kainrath, S. Hochreiter, msa: an R package for multiple sequence alignment, Bioinformatics 31 (2015) 3997–3999.
[42] J.D. Thompson, D.G. Higgins, T.J. Gibson, W. Clustal, Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.
[43] A.J. Doucet, G. Droc, O. Siol, J. Audoux, N. Gilbert, U6 snRNA pseudogenes: markers of retrotransposition dynamics in mammals, Mol. Biol. Evol. 32 (2015) 1815–1832.
[44] T.P. Naas, R.J. DeBerardinis, J.V. Moran, E.M. Ostertag, S.F. Kingsmore, M.F. Seldin, Y. Hayashizaki, S.L. Martin, H.H. Kazazian, An actively retrotransposing, novel subfamily of mouse L1 elements, EMBO J. 17 (1998) 590–597.
[45] R.J. DeBerardinis, J.L. Goodier, E.M. Ostertag, H.H. Kazazian Jr., Rapid amplification of a retrotransposon subfamily is evolving the mouse genome, Nat. Genet. 20 (1998) 288–290.
[46] F. Mao, W.Y. Leung, X. Xin, Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications, BMC Biotechnol. 7 (2007) 76.
[47] J. Letowski, R. Brousseau, L. Masson, Designing better probes: effect of probe size, mismatch position and number on hybridization in DNA oligonucleotide microarrays, J. Microbiol. Methods 57 (2004) 269–278.
[48] G.P. McDermott, D. Do, C.M. Litterst, D. Maar, C.M. Hindson, E.R. Steenblock, T.C. Legler, Y. Jouvenot, S.H. Marrs, A. Bemis, P. Shah, J. Wong, S. Wang, D. Sally, L. Javier, T. Dinio, C. Han, T.P. Brackbill, S.P. Hodges, Y. Ling, N. Klitgord, G.J. Carman, J.R. Berman, R.T. Koehler, A.L. Hiddessen, P. Walse, L. Bousse, S. Tzonev, E. Hefner, B.J. Hindson, T.H. Cauly 3rd, K. Hamby, V.P. Patel, J.F. Regan, P.W. Wyatt, G.A. Karlin-Neumann, D.P. Stumbo, A.J. Lowe, Multiplexed target detection using DNA-binding dye chemistry in droplet digital PCR, Anal. Chem. 85 (2013) 11619–11627.
[49] Y. Yuan, R. Reddy, Organization of spliceosomal U6 snRNA genes in the mouse genome, Mol. Biol. Rep. 13 (1988) 159–164.
[50] A. Buzdin, S. Ustyugova, E. Gogvadze, T. Vinogradova, Y. Lebedev, E. Sverdlov, A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3' terminus of l1, Genomics 80 (2002) 402–406.
[51] J.L. Garcia-Perez, A.J. Doucet, A. Bucheton, J.V. Moran, N. Gilbert, Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase, Genome Res. 17 (2007) 602–611.
[52] Z. Zhang, N. Carriero, M. Gerstein, Comparative analysis of processed pseudogenes in the mouse and human genomes, Trends Genet. 20 (2004) 62–67.
[53] J.M. Rosser, W. An, Repeat-induced gene silencing of L1 transgenes is correlated with differential promoter methylation, Gene 456 (2010) 15–23.
[54] D. Lin, T.V. Pestova, C.U. Hellen, H. Tiedge, Translational control by a small RNA: dendritic BC1 RNA targets the eukaryotic initiation factor 4A helicase mechanism, Mol. Cell Biol. 28 (2008) 3008–3019.
[55] J. Kim, J.A. Martignetti, M.R. Shen, J. Brosius, P. Deininger, Rodent BC1 RNA gene as a master gene for ID element amplification, Proc. Natl. Acad. Sci. U. S. A. 91 (1994) 3607–3611.
[56] P.L. Deininger, H. Tiedge, J. Kim, J. Brosius, Evolution, expression, and possible function of a master gene for amplification of an interspersed repeated DNA family in rodents, Prog. Nucleic Acid Res. Mol. Biol. 52 (1996) 67–88.
[57] A.V. Furano, B.E. Hayward, P. Chevret, F. Catzeflis, K. Usdin, Amplification of the ancient murine Lx family of long interspersed repeated DNA occurred during the murine radiation, J. Mol. Evol. 38 (1994) 18–27.
[58] A.K. Wong, J.B. Rattner, Sequence organization and cytological localization of the minor satellite of mouse, Nucleic Acids Res. 16 (1988) 11645–11661.
[59] I. Kuznetsova, O. Podgornaya, M.A. Ferguson-Smith, High-resolution organization of mouse centromeric and pericentromeric DNA, Cytogenet. Genome Res. 112 (2006) 248–255.
[60] A.S. Komissarov, E.V. Gavrilova, S.J. Demin, A.M. Ishov, O.I. Podgornaya, Tandemly repeated DNA families in the mouse genome, BMC Genom. 12 (2011) 531.
[61] C.F. Voliva, C.L. Jahn, M.B. Comer, C.A. Hutchison 3rd, M.H. Edgell, The L1Md long interspersed repeat family in the mouse: almost all examples are truncated at one end, Nucleic Acids Res. 11 (1983) 8847–8859.
[62] T.G. Fanning, Size and structure of the highly repetitive BAM HI element in mice, Nucleic Acids Res. 11 (1983) 5073–5091.
[63] T. Penzkofer, T. Dandekar, T. Zemojtel, L1Base: from functional annotation to prediction of active LINE-1 elements, Nucleic Acids Res. 33 (2005) D498–D500.
[64] K. Akagi, J. Li, R.M. Stephens, N. Volfovsky, D.E. Symer, Extensive variation between inbred mouse strains due to endogenous L1 retrotransposition, Genome Res. 18 (2008) 869–880.
[65] I. Ovchinnikov, A. Rubin, G.D. Swergold, Tracing the LINEs of human evolution, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 10522–10527.
[66] A. Pavlicek, J. Paces, R. Zika, J. Hejnar, Length distribution of long interspersed nucleotide elements (LINEs) and processed pseudogenes of human endogenous retroviruses: implications for retrotransposition and pseudogene detection, Gene 300 (2002) 189–194.
[67] D.D. Loeb, R.W. Padgett, S.C. Hardies, W.R. Shehee, M.B. Comer, M.H. Edgell, C.A. Hutchison 3rd, The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retrotransposons, Mol. Cell Biol. 6 (1986) 168–182.
[68] Y. You, B.G. Moreira, M.A. Behlke, R. Owczarzy, Design of LNA probes that improve mismatch discrimination, Nucleic Acids Res. 34 (2006) e60.
[69] B.K. Jacobs, E. Goetghebeur, L. Clement, Impact of variance components on reliability of absolute quantification using digital PCR, BMC Bioinf. 15 (2014) 283.
[70] M. De Cecco, T. Ito, A.P. Petrashen, A.E. Elias, N.J. Skvir, S.W. Criscione, A. Caligiana, G. Brocculi, E.M. Adney, J.D. Boeke, O. Le, C. Beausejour, J. Ambati, K. Ambati, M. Simon, A. Seluanov, V. Gorbunova, P.E. Slagboom, S.L. Helfand, N. Neretti, J.M. Sedivy, L1 drives IFN in senescent cells and promotes age-associated inflammation, Nature 566 (2019) 73–78.
[71] M. Simon, M. Van Meter, J. Ablaeva, Z. Ke, R.S. Gonzalez, T. Taguchi, M. De Cecco, K.I. Leonova, V. Kogan, S.L. Helfand, N. Neretti, A. Roichman, H.Y. Cohen, M.V. Meer, V.N. Gladyshev, M.P. Antoch, A.V. Gudkov, J.M. Sedivy, A. Seluanov, V. Gorbunova, LINE1 derepression in aged wild-type and SIRT6-deficient mice drives inflammation, Cell Metabol. 29 (2019) 871–885 e875.
[72] B.T. Lau, C. Wood-Bouwens, H.P. Ji, Robust multiplexed clustering and denoising of digital PCR assays by data gridding, Anal. Chem. 89 (2017) 11913–11917.
[73] Y. Ding, P.D. Howes, A.J. deMello, Recent advances in droplet microfluidics, Anal. Chem. 92 (2020) 132–149.