

Muscle single-cell analysis reveals that RNA foci accumulation is linked to muscle dysfunction in patients with myotonic dystrophy type I

Judit Núñez-Manchón¹, Alfonsina Ballester-Lopez², Ian Linares-Pardo³, Emma Koehorst⁴, Ana Maria Cobo⁵, Jaume Col-Cantí³, Adolfo Lopez de Munain⁶, Alejandro Lucia⁷, Darren Monckton⁸, Sarah Cumming⁸, Jonathan Magaña⁹, Laura Palomo¹⁰, Francesc Solé¹⁰, Guillem Pintos-Morell¹¹, Giuseppe Lucente³, Miriam Almendrote³, Alba Ramos-Fransi³, Alicia Martínez-Piñeiro³, and Gisela Nogales-Gadea¹²

¹Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol

²Fundacio Institut d'Investigacio en Ciències de la Salut Germans Trias i Pujol

³Hospital Universitari Germans Trias i Pujol

⁴Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol

⁵Hôpital Marin de Hendaye

⁶Biodonostia Health Research Institute

⁷Universidad Europea de Madrid

⁸University of Glasgow

⁹National Institute of Rehabilitation Luis Guillermo Ibarra Ibarra

¹⁰José Carreras Leukaemia Foundation

¹¹Hospital Vall d'Hebron

¹²Foundation Institute of Research in Health Sciences Germans Trias i Pujol

July 29, 2020

Abstract

Single cell analysis has numerous potential applications, for instance in the context of myotonic dystrophy type I (DM1). This disease is caused by a CTG expansion in the dystrophin myotonic-protein kinase (DMPK) gene, with CTG expansion length being heterogeneous not only between tissues but also between cells in the same tissue. We studied muscle pathophysiology from five patients with DM1 at single cell level, including RNA foci (i.e., nuclear DMPK-transcribed aggregates that trap proteins) and splicing alterations (caused by trapping of important splicing regulator proteins by the RNA foci). Single cell myoblasts were heterogeneous in RNA foci load, with most showing 0 to 3 foci. The percentage of myoblasts carrying foci differed among patients, ranging from 29 to 99%. We found a significant direct correlation between mean and median number of RNA foci in myoblasts and patients' endurance capacity ($r=-0.975$, $p=0.005$). Although the expression of DMPK and MBNL1 transcripts was variable among the myoblasts, no relationship was found between RNA foci number and their expression in individual cells. CTG size in muscle correlated with patients' disease onset. In summary, single cell analysis poses some technical challenges but allows an in-depth analysis of heterogeneous molecular alterations in patients with DM1.

Introduction

Single cell analysis is a recent, fast expanding technology that allows a deep characterization of cellular diversity in tissues (Kulkarni et al., 2019; Ståhlberg & Kubista, 2018). This method has led to the identi-

fication of new cell types as well as to the discovery of cell-specific functions and the recognition of their maturation state (Perchet et al., 2017; Ståhlberg & Kubista, 2018). It is useful for studying heterogeneous cell populations, especially in diseases with genetic variability (Rantalainen, 2018; Ren et al., 2018). In these cases, indeed, an in-depth understanding of all the cell populations that form an affected organ/tissue and of their characteristics and interactions are important issues in the context of personalized medicine.

One of the disorders that can be studied using single cell analysis is myotonic dystrophy type I (DM1). This autosomal dominant genetic condition is characterized by genetic instability and is caused by a CTG repeat expansion in the non-coding 3' untranslated (UTR) region of the myotonic dystrophy protein kinase (*DMPK*) gene (Bird, 1993). More than 50 CTG repeats are considered pathogenic and repeat length can vary from 50 to thousands of repeats in affected patients (Bird, 1993). Furthermore, the size of the CTG repeat also varies between tissues within the same patient (Ashizawa et al., 1993; Lavedan et al., 1993; Mahadevan et al., 1992) as well as during a patient's lifetime (Wong et al., 1995), and produces anticipation (Harper et al., 1992).

CTG repeat expansions produce toxic RNAs that trigger some of the patient's symptoms. When the pathological repeats are transcribed *DMPK* transcripts carry CUG expansions that sequester important cellular proteins, altering their levels and functionalities (Miller et al., 2000). These RNA and protein complexes are known as "RNA foci" and they are located in the nucleus of DM1 cells (Mykowska et al., 2011). One of the proteins sequestered by the RNA foci is the splicing regulator muscleblind-like 1 (*MBNL1*). Proteins aberrantly spliced due to MBNL1 sequestration include insulin receptor (INSR), sarcoplasmic reticulum Ca(2+)-ATPase 1 (ATP2A1), chloride channel 1 (CLCN1), and MBNL1 – which regulates its own splicing (Konieczny et al., 2017). The altered transcription of these genes in DM1 has been related to several symptoms of the disease. Thus, INSR and CLCN1 misregulation are associated with insulin resistance (Renna et al., 2019) and myotonia (Charlet-B. et al., 2002), respectively, and ATP2A1 might impair calcium homeostasis in skeletal muscle (Kimura et al., 2005).

The skeletal muscle is one of the most affected tissues in patients with DM1, but their clinical manifestations are highly heterogeneous (Bundey, 1982). Muscle impairment can be assessed using several scales: the muscle research council (MRC) scale, which evaluates muscle power (Compston, 2010), muscle impairment rating (MIRS) scale (which assesses the degree of distal to proximal muscle involvement) (Mathieu et al., 2001), the 6-minute walking distance (6MWD, an index of endurance [or 'aerobic'] capacity (Butland et al., 1982)), or the modified Rankin Scale (mRS, an indicator of disability in patients (Van Swieten et al., 1988)).

In DM1 there is a need for single cell data. Scientific evidence indicates that every DM1 muscle cell may contain a different number of CTG repeats in its genome, and therefore could potentially behave differently. In this study, we aimed to analyze the diversity of the muscle cells that compose the DM1 muscle, and analyze how frequently alterations such as RNA foci and splicing occur, and if they can be correlated with the muscle tissue function in patients. By doing so, we attempted to answer the following questions: are all the cells in the muscle tissue affected equally? Is there a correlation between RNA foci and splicing alteration at the single-cell level? Can the average of single-cell data in a patient be correlated with the severity of muscle-related symptoms?

Results

Patients showed heterogeneity in muscle affectation and disability measured by different scales

Five clinically heterogeneous patients with DM1 were included in this study, whose age of onset ranged between 15 and 48 years. Their muscle dysfunction was studied using several different scales (**Table 1**). By using the mRS, we determined the degree of disability and/or dependence in the daily activities of these DM1 patients. In our cohort, P3 (with a mRS score of 1) had no significant disability and had no limitation in daily life activities despite the presented symptoms. Three patients, P1, P2 and P5 (mRS score of 2) presented with a slight disability, meaning they were unable to do activities they were able to in the past, but they remained independent. Moderate disease disability, (mRS score of 4) was found in P4, who was unable to walk and to attend to self-care. With regard to the MIRS scale, P3 showed minimal signs, P2 and

P5 had distal weakness, and P1 and P4 mild-to-moderate proximal weakness. We also evaluated the same patients using the MRC scale, to measure strength in the biceps muscle (the muscle from which we obtained muscle cells). P2, P3 and P5 had preserved muscle strength (*i.e.*, MRC score = 5). Yet, P1 and P4 had an MRC of 4, indicating decreased muscle strength, a result consistent with the more proximal weakness that these two patients had when assessed with the MIRS scale. The same patients were submitted to a 6MWD test and P5 showed values within the normal range (Casanova et al., 2011). Values below the normal range were found for 6MWD in P1, P2 and P3. Although having better scores for mRS, MIRS and biceps MRC than P1, P2 walked 100 meters less. P4 was not able to perform the 6MWD test, since she was unable to walk. All scales showed heterogeneity among the patients in muscle function and degree of disability.

CTG expansion is enlarged from muscle to patient myoblast

To determine the degree of similarity in CTG expansion size between the isolated myoblasts and the muscle they originated from, a small pool PCR analysis was performed. We measured the progenitor allele in four of the five patients included in the study (Table 1). The myoblast pools were measured for the progenitor allele. In the case of P3 and P4, who showed clonal amplification of the expansion, only one CTG expanded band could be detected in the small pool PCR analysis. When comparing the size of the expanded allele in the myoblast pool and muscle from the same patient, no correlation was found. However, there were some myoblasts with expansion size very similar to the muscle tissue from which they derive. In all patients, myoblasts showed a tendency to have longer expanded alleles compared to the muscle tissue from which they were isolated. The CTG expansion size in muscle correlated with the age of symptom onset in the patients ($p=0.05$). No other correlation were found between CTG size and clinical or molecular data.

Single cell analysis of RNA foci shows a heterogeneous behavior between and within patients' myoblasts, with several molecular alterations linked to muscle function

For these experiments, we developed a fast protocol of RNA foci staining, in order to preserve RNA for posterior analysis. Two approaches were used to perform RNA foci single cell quantification: single cell Fluidigm C1 platform; and sorting cytometry.

In the first approach, previously stained, single myoblasts were captured into the C1 IFC chip and were visible by bright field microscopy (**Figure 1A**). 4',6-diamidino-2-phenylindole (DAPI) fluorescent signal was also visible in captured myoblasts. However, we failed to detect RNA foci fluorescent signal (**Figure 1B**), but the latter could be detected in the same cells and with the same staining when using standard immunocytochemistry (**Figure 1C**). We found incompatibility of the focal distance of the microscope with the thickness of the chip, which was much higher and made it impossible to focus on such small particles as the RNA foci.

Due to the problems found in the first approach, we started a second approach based on sorting cytometry. By this technique, we were able to isolate and count foci from 39 to 72 single myoblasts (depending on the patient) of 120 myoblasts that were sorted for every patient (**Table S1**). The wells from the 96-well plates used to sort the myoblasts were analyzed with a fluorescent microscope. The number of RNA foci per myoblast was manually annotated from right to left and top to bottom in the plates.

We found evidence of heterogeneity in myoblast foci number both between and within patients (**Figure 2**, **Table 1**). The percentage of DM1 myoblasts with RNA foci was high in three of the cell lines (>80%). However, the distribution of RNA foci per myoblast was very different in every patient (**Figure 3**). In all myoblast cell lines we could find some myoblasts without foci, indicating a dynamic process of foci accumulation, that reached a maximum of 7 RNA foci in some myoblasts. However, when looking at all the DM1 myoblasts analyzed in this study, we concluded that most of them were carrying between 0 to 3 RNA foci (20% had 0 RNA foci, 21% 1 RNA foci, 24% 2 RNA foci, 22% 3 RNA foci, 10% 4 RNA foci, and 2% 5-7 RNA foci). No foci was observed in the myoblasts of healthy controls. Although no correlation with CTG repeat number was found, P5 (who was the carrier of the shortest CTG expansion in myoblasts) was showing the highest number of myoblasts without RNA foci. Regarding muscle function and disability, P4 (the most severely affected patient) had more myoblasts carrying numerous foci than the rest of the patients

and less myoblasts without RNA foci. We found a correlation between 6MWD and the mean number of foci per cell ($p=0.017$) and foci median ($p=0.005$).

Single cell expression levels of *MBNL1* and *DMPK* are highly variable and no correlation was found at single cell level

After RNA foci visualization, RNA was obtained from every single myoblast for transcript analysis. Even though we could detect by qPCR all the transcripts of interest in myoblast pools (*DMPK*, *MBNL1*, *INSR* and *ATP2A1*), detection rate at single cell level was very low for most transcripts and only the exon 7 exclusion *MBNL1* splicing variant (normal variant) and *DMPK* expression levels could be analyzed. No correlation at the single cell level was found between the number of foci and *DMPK* or the *MBNL1* exon 7 exclusion isoform at single cell level (**Table S1**).

We found that the exon 7 exclusion *MBNL1* variant was significantly ($p = 0.01$) more expressed in control than in patient's myoblasts, whereas no differences were found in *DMPK* expression levels between controls and patients ($p = 0.1$) (**Figure 4**). *DMPK* expression levels significantly differed ($p < 0.001$) among patients, but all patient myoblasts showed similar expression levels for the *MBNL1* exon 7 exclusion isoform ($p = 0.8$). We found variability in the expression of *DMPK* and *MBNL1* exon 7 exclusion isoform among the different control individuals ($p < 0.001$) (**Figure 5**).

To allow for a better understanding of *DMPK* expression differences between patients and to determine if transcript levels were derived from the expanded or the normal allele, we searched for a single nucleotide polymorphism (SNP) that allowed differentiating expanded from normal transcripts in single cells. When testing our patients for the presence of the informative SNP, described in exon 10 (Korneluk, 1993), we found all of them to be homozygous, and we could therefore not use this SNP for the intended purpose.

Discussion

Single cell analysis is a novel technology that can help to answer important questions related to DM1. The first question is whether all skeletal muscle cells are affected equally by the disease. In this effect, it is known that somatic mosaicism affects muscle tissue regarding the size of the CTG expansion (Thornton et al., 1994). In our study, we could demonstrate that the number of RNA foci number is very different among DM1 myoblasts and that the expression of *DMPK* and *MBNL1* differs considerably at the single cell level. So, the myoblasts of the same DM1 muscle tissue are differently affected by the disease. A second important question is whether there are correlations between RNA foci and splicing alterations at the single cell level. In this regard, we found no correlation between RNA foci number and expression of the exon 7 exclusion isoform of *MBNL1*. We also found no correlation between RNA foci and *DMPK* expression at the single cell level. However, we could not study important splicing alterations found in pools of DM1 cells, due to the inability to detect expression of those transcripts at the single cell level. Another important question is whether the average number of RNA foci per cell in a patient correlates with the patient's muscle affection. We found correlation between RNA foci mean/median and 6MWD. Thus, patients having myoblasts carrying more foci showed a more severe manifestation of the disease, with subsequent impairment in walking ability.

Foci accumulation is likely to be a dynamic process in patient-derived myoblasts. In this effect, we found heterogeneity in both presence and number of foci in patient-derived myoblasts. RNA foci vary from zero to seven, indicating a dynamic process of RNA foci accumulation being active in all cell lines. A random process of forming and degrading RNA foci has been described in DM1 cellular models (Querido et al., 2011). However, in our patient-derived myoblasts, the proportion of myoblasts with RNA foci that every patient was carrying was very variable. For example, P3 and P5 (*i.e.*, the less severely affected patients) had overall the highest proportion of myoblasts without foci (71% and 19%, respectively). In contrast, the most severely affected patient in our series, P4, was carrying 2 and 3 RNA foci in 25% and 48% of the myoblasts, respectively.

The RNA foci number was low in DM1 myoblasts (with 90% of the myoblasts showing 0 to 3 RNA foci), as previously described (Gudde et al., 2016). An in-depth analysis is needed to unravel the processes that

contribute to the heterogeneity found in RNA foci load and what is the explanation for the fact that a cell carrying the CTG expansion might still have no RNA foci. From the point of view of a spliceopathy disease, getting rid of the RNA foci seems to be key in preserving muscle functionality. In previous studies, one of the main determinants of the RNA foci load was the CTG repeat size (Botta et al., 2008). Although we were unable to find such type of correlation in our study, the patient carrying the shortest CTG expansion had the lowest RNA foci load.

RNA foci load seems to be linked to muscle dysfunction in DM1. We found indeed correlations between 6MWD and foci mean/median. The correlations for RNA foci median and MIRS, biceps MRC and mRS scales were not significant. Many factors could have potentially contributed to our inability to find more correlations with the clinical data, but it must be noted that our study cohort was very small, largely due to the need to perform a muscle biopsy – an invasive and unpleasant procedure – in the patients as well as to research budget constraints. In addition, the heterogeneous manifestation of the disease and therefore the heterogeneous pool of patients must be kept in mind, as well as the fact that the majority of scales we used are categorical (*i.e.* , MIRS, MRC or mRS) which, in comparison to continuous scales (6MWT) are more challenging to use in correlation studies. In the first design of our study with DM1 experts, we decided to evaluate these scales and we did not consider using a test to measure biceps strength in a more quantitative way. 6MWD was the only physical measure that was related to RNA foci number, but caution must be taken when interpreting these results, since the data was obtained from biceps-derived myoblasts. Additional studies with a larger patients' cohorts are needed to test the validity of these results and to analyze whether other correlations can be found with the categorical scales.

The splicing analysis was very challenging at single myoblast level since many transcripts could not be detected. The transcripts with higher expression levels in myoblasts were *DMPK* and the normal *MBNL1* isoform. Although we were able to detect and analyze these two transcripts, this was only possible in less than 50% of the studied myoblasts. Regarding *INSR* and *ATP2A1*, we were able to detect the transcripts in myoblast cell pools. However, this was not translated to the single cell level, even though a customized qPCR method was used. A possible explanation for our inability to find the transcripts by qPCR is a diminished RNA integrity. At the start of the experiments, we wanted to use the automated system fluidigm, which would decrease the processing time and therefore preserve RNA integrity. However, due to incompatibility of the fluidigm chip with the RNA visualization, we had to rely on sorting cytometry, thereby increasing the time of cell manipulation and microscopic cell visualization, with a higher risk of compromising RNA integrity. One example of the difficulty of using the sorting cytometry is that we could only recover information from 40 to 70 myoblasts of the 120 myoblasts that were originally sorted.

In previous studies analyzing cell pools, RNA foci load correlated with splicing expression patterns (Botta et al., 2008). Yet, we were unable to replicate these results at the single cell level. Overall, we could only find that the *MBNL1* normal isoform was more expressed in controls and a wide variety in the expression levels of *DMPK* was found among the participants (patients and controls). Previous studies analyzing *DMPK* expression levels in patients are contradictory, and it is not clear if they are affected by the disease (Furling et al., 2001, 2003; Gudde et al., 2016). For a deeper insight into *DMPK* expression we tried to analyze *DMPK* expression by differentiating expanded from normal transcripts with a previously described polymorphism present in exon 10 of the *DMPK* gene (Korneluk, 1993). Unfortunately, this was not possible, as all of our patients were homozygous for the polymorphism. Other techniques, such as Northern Blot, which have been used for this purpose, were not applicable in this case, as we were working with single cells.

Expansion size can differ between muscle tissue and myoblast primary cultures derived from the muscle in question. As we expected and as other authors have explained before, DM1 cultures are submitted to mitotic drive (Khajavi et al., 2001). In two patients we were able to detect only one CTG expansion size, due to clonal expansion. Our results showed that CTG expansion size differs between samples, meaning that some showed greater difference compared to the muscle tissue than others. Moreover, myoblasts had a tendency towards the expansion of the progenitor allele, which has been also observed previously (Monckton et al., 1995; Wong et al., 1995). It is therefore recommended to size the expansion in DM1 cellular models since,

in culture they differ from the patients' tissue from which they are derived.

CTG expansion size in muscle correlated with age of disease onset in our patients. We used small pool PCR, a technique widely used for measuring CTG repeats (Gomes-Pereira et al., 2004). Previous relationships between CTG size in blood and age of disease onset have been reported (Cumming et al., 2019; Morales et al., 2012; Overend et al., 2019) and recently this correlation has been found applying the same technique in DNA isolated from saliva (Corrales et al., 2019). Sizing the CTG expansion in muscle is always difficult since muscle tissue has larger expansions than other tissues (Lavedan et al., 1993; Thornton et al., 1994). However, when analyzing these samples with small pool PCR, finding expansion size is easier since few molecules are amplified. By using this method, we could find that CTG muscle expansion size is also a good indicator of disease severity.

The single cell study we did here presented many challenges. For example, it can only be done with cells that can be sorted, meaning that differentiated cells such as myotubes or neurons, which are key cells to study in this disease, cannot be studied. In addition, although a quick FISH staining protocol was applied the process of staining and the time of visualization probably affects and compromises RNA material. An approach without staining may benefit splicing analysis, but in this case, one will lose the information regarding how many foci the myoblast cell contains. One way or another, with the current technical limitations of this analysis we lose precious molecular data. Finally, the diverse clinical presentation of DM1 patients and the currently available scales to evaluate muscle function further (most being categorical type of scales) complicates single cell study. The vast heterogeneity found in DM1 patients, both at a cellular and a symptomatic level, asks for a further refinement of the current single cell technologies available and further development of continuous scales to study muscle function. This will give us valuable insight into the diversity of DM1 pathology at a single cell level, and how this correlates to the clinical manifestations of the disease.

To our knowledge, this is the first study performed at a single cell level in DM1. In addition, we analyzed the highest number of myoblasts per patient compared to other studies and for the first time attempted to analyze alternative splicing defects in single cells. We determined that the number of RNA foci was heterogeneous in DM1 myoblasts and submitted to a dynamic process, allowing some DM1 myoblasts to be free of RNA foci, even though they were carrying the CTG expansion. This RNA foci load compromises muscle functionality evaluated through 6MWD. The heterogeneity found at the single cell level in the patients could be a key factor for treatment development and efficacy. Differences in alterations between cells could mean that different concentrations of future therapies would be needed. It will be key to find the equilibrium between treating all cells and avoiding toxic effects.

Material and methods

Human subjects

This study was approved by the Ethics Committee of the University Hospital *Germans Trias i Pujol* and was performed in accordance with the Declaration of Helsinki for Human Research. Written informed consent was obtained from all participants. Muscle biopsies were obtained from the *biceps brachii* of five patients with confirmed DM1. For comparison, we included as controls non-DM1 patients of the traumatology department, who needed to undergo minor surgeries. Patients and controls were age and sex-matched. We recorded clinical data for each DM1 patient, including MIRS, MRC, 6MWD and mRS, the latter rates global disability and is mostly used for strokes, even though it can be applied to other causes of neurological impairment.

Cell culture

Myoblasts were isolated from the biopsied tissue by muscle explants and then purified by CD56 magnetic separation. Early passage myoblasts were grown until 60-70% confluent on 0.1% gelatin-coated flasks in proliferation medium containing DMEM supplemented with 10% FBS, 22% M-199, PSF 1x, insulin 1.74 μ M, L-glutamine 2 mM, FGF 1.39 nM and EGF 0.135 mM. Cells for DNA extraction and FISH staining were obtained between passages P2 and P6.

DNA isolation and Small pool PCR analysis

For each participant, sizing of the expanded allele was done twice: 1) in the biopsy tissue from which myoblasts were isolated, and 2) in the same myoblast cultures that were used for FISH staining and RNA isolation. In the biopsy tissue, DNA extraction was performed following the phenol:chloroform:isoamyl alcohol method. In the myoblast cultures, DNA was extracted using a previously described protocol for blood DNA extraction (Miller, Dykes, & Polesky, 1988). To determine expansion size in these DNAs, we first performed a long PCR using the primers DM-C and DM-DR described elsewhere (Gomes-Pereira et al., 2004; Monckton et al., 1995; Salinas-Rios et al., 2011) and PCR Master Mix (Thermo Fisher Scientific; MA). We supplemented the reaction with 69 mM 2-mercaptoethanol, 5% DMSO, and Taq polymerase (Sigma-Aldrich; Gillingham, UK) at 1 unit per 10 μ L. The annealing temperature was 63.5°C. The long PCR products were resolved by electrophoresis on a 1% agarose gel and hybridized by Southern Blot as previously described (Gomes-Pereira et al., 2004; Monckton et al., 1995). Autoradiographic images were scanned and the progenitor allele in each patient was estimated by comparison against the molecular weight ladder, using GelAnalyzer 19.1 software.

Fluorescence in situ hybridization (FISH)

To visualize the RNA foci in single myoblasts before RNA analysis, a quick FISH staining method was developed. The aim of this method was to preserve RNA integrity as much as possible. In brief, myoblast pellets were resuspended in formamide and incubated for 20 min at 37°C. Myoblasts were then incubated with the hybridization buffer containing AT0488 labelled (CAG)₁₀ probe 0.01 μ M and 30% formamide in 2XSSC buffer for 20 min at 37°C. Myoblasts were then washed with 30% formamide 2x SSC, 1x SSC and 1x PBS and incubated for 5 min with 5 μ l of DAPI mounting solution. Finally, myoblasts were resuspended with 500 μ l of 1x PBS.

Myoblast isolation and fluorescence microscope visualization

For this purpose, two different approaches were tested: 1) single cell C1 Fluidigm platform (Fluidigm, South San Francisco, CA) and 2) sorting cytometry. The first approach, single cell C1 Fluidigm platform, was especially suitable for our purpose because isolated cells can be visualized through microscopy and after, RNA extraction and cDNA synthesis can be made in an automated system that works with thermal cycling and pneumatic controls, minimizing the variation between conditions and expending only small amounts of reagents in every plate. To make the analysis with the C1 Fluidigm single cell platform, myoblasts were resuspended in C1 suspension reagent and loaded into the C1 IFC chip, that contains 96 single cell capture sites. After this process, myoblasts were captured in single and isolated chambers in the IFC and then observed by fluorescent microscopy at 40X to count foci number. Unfortunately, we failed to detect foci fluorescent signal (see results section), so we decided to do this process through our second approach, sorting cytometry. For this second approach, cells were sorted using FACSARIA II flow cytometer (BD Biosciences, San Jose, CA) into 96-Well Optical-Bottom Plates. For each participant, we sorted a total of 120 FISH stained myoblasts distributed in two plates. The cytometer could not detect the fluorescent signal of the FISH staining (probably because RNA foci fluorescence signal is small and because of their nuclear localization). Therefore, sorting was only used to isolate individual myoblasts. RNA foci signal was observed by fluorescent microscopy. In order to spend the least time possible, since these same cells were later used for RNA analysis, and because there were two plates to analyze per participant, two independent observers simultaneously visualized one plate each, in similar fluorescence microscopes. They had the same criteria for annotating RNA foci number in every single myoblast by naked eye visualization.

RNA isolation, cDNA synthesis, preamplification and quantitative PCR (qPCR)

RNA from single myoblasts was isolated using GenElute Single Cell RNA Purification Kit (Sigma-Aldrich, San Luis, MO) and cDNA was synthesized using TruScript First Strand cDNA Synthesis Kit (Norgen Biotek Corp., Thorold, ON, Canada), following the manufacturer's protocol in both. Transcripts of interest were pre-amplified with gene specific Taqman assays using TaqMan PreAmp Master Mix (ThermoFisher, Waltham, MA). This pre-amplification step is used in most single cell studies. Transcripts of interest

included *DMPK*, *MBNL1* exon 7 inclusion (aberrant) and exclusion isoforms, *ATP2A1* exon 22 inclusion and exclusion (aberrant) isoforms and *INSR* exon 11 inclusion and exclusion (aberrant) isoforms. The design of *MBNL1*, *INSR* and *ATP2A1* Taqman assays was done with the Assay Design Tool (ThermoFisher), while for *DMPK* we used a previously described Taqman assay (Pandey et al., 2015). We analyzed the expression of *DMPK* and *INSR*, *ATP2A1* and *MBNL1* splicing variants by qPCR using Taqman Fast Advanced Master Mix (ThermoFisher) and the same Taqman assays as in pre-amplification (ThermoFisher). No endogenous transcript was used for normalization because in single cell analysis housekeeping genes can be misleading as their expression varies from cell to cell (Ståhlberg & Kubista, 2018). Data analysis was performed as previously described (Ståhlberg et al., 2013).

Single nucleotide polymorphism analysis

To differentiate *DMPK* transcript qPCR expression originating from the expanded allele versus the one of the wild type allele, the presence of a previously described informative SNP in exon 10 of *DMPK* gene (Korneluk, 1993) was studied. If patients with DM1 are informative for the SNP, specific qPCR designed primers could be used for detection of *DMPK* expression originating from the expanded allele and specific qPCR designed primers could be used for detection of *DMPK* expression originating from the wild type allele.

To evaluate the presence of the aforementioned SNP at the patient DNA level, we used primers 5'-CTGCAGAAGGTTTAGAAAGAGC-3' (forward) and 5'-CATCCTGTGGGGACACCGAGG-3 (reverse) (Korneluk, 1993) with the following conditions: 94 °C for 30 s, 60 °C for 30 s and 72°C for 30 s for 30 cycles. Purified products were sequenced with BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster city, CA) using primer forward. Sequences were analyzed with Chromas version 2.6.2., using NG_-009784.1 as *DMPK* reference sequence.

Statistical analysis

The statistical analysis was performed with Medcalc statistical software (Medcalc, Ostend, Belgium). The significance level was set at 0.05. We analyzed correlations using Spearman's coefficient at three different levels: patients; myoblast pools and single myoblasts. We used this test for correlating patients' clinical data (age at onset, mRS, MIRS, biceps MRC and 6MWT) and: foci mean, foci median and percentage of cells with foci, *DMPK* and *MBNL1* expression means and muscle CTG expansion size. We also studied correlations using Spearman's coefficient in myoblast pools: between mean, median and percentage of cells with foci and expansion size in myoblasts, between *DMPK* and *MBNL1* expression and expansion size in myoblasts, between expansion size in myoblasts pools and expansion size in muscle tissue in patients. At single cell level, we studied the correlation between RNA foci number and *DMPK* and *MBNL1* expression. The Mann Whitney-*U* test was used to compare *DMPK* and *MBNL1* exon 7 exclusion expression between controls and DM1 patients. The Kruskal-Wallis test for inter-individual comparisons of *DMPK* and *MBNL1* exon 7 exclusion expression, both in patients and in controls.

Data availability statement

Data available in article supplementary material.

Acknowledgements

G.N.G declares grants from Instituto de Salud Carlos III (Grant Numbers: PI15/01756; P18/00713), Madrid, Spain and AFM Telethon (Trampoline grant number #21108), France. A.L declares grants from the Spanish government granting agency Instituto de Salud Carlos III, Madrid, Spain. J.N.M is funded by AFM Telethon Trampoline Grant #21108. A.B.L is funded by an FI Agaur fellowship FI.B 01090 and by SGR 1520. E.K is funded by the "La Caixa" Foundation (ID 100010434), fellowship code LCF/BQ/IN18/11660019, co-funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement n°713673. I.L.P is funded by CP14/00032 and by SGR 1520. G.N.G is supported by a Miguel Servet research contract (ISCIII CD14/00032, CPII19/00021 and FEDER). G.L. was supported by a Rio Hortega contract (ISCIII CM16/00016 and FEDER). J.J-M was supported by the Science, Technology and Innovation of Mexico City (Grant # PICSA 12-162). This work was supported in part by a grant

from the 2017 SGR288 and SGR 1520 (GRC) Generalitat de Catalunya; economic support from CERCA Programme/Generalitat de Catalunya and Fundació Internacional Josep Carreras.

G.P-M reports personal honoraria from Shire-Takeda, Amicus, Kyowa-Kirin, and Sanofi-Genzyme, outside the submitted work. The remaining co-authors declare no competing interests.

We gratefully acknowledge the participants in this research, the patient's associations and the core facilities of the IGTP and Single Cell Unit from IJC. We acknowledge our funding agencies and Ruben Artero's team for their advice on studying *DMPK* expression.

References

- Ashizawa, T., Dubel, J. R., & Harati, Y. (1993). Somatic instability of ctg repeat in myotonic dystrophy. *Neurology* , 43 (12), 2674–2678. <https://doi.org/10.1212/wnl.43.12.2674>
- Bird, T. D. (1993). GeneReviews® Myotonic Dystrophy Type 1: Last Update: October 3, 2019. In *GeneReviews®* . University of Washington, Seattle. <http://www.ncbi.nlm.nih.gov/pubmed/20301344>
- Botta, A., Rinaldi, F., Catalli, C., Vergani, L., Bonifazi, E., Romeo, V., Loro, E., Viola, A., Angelini, C., & Novelli, G. (2008). The CTG repeat expansion size correlates with the splicing defects observed in muscles from myotonic dystrophy type 1 patients. *Journal of Medical Genetics* , 45 (10), 639–646. <https://doi.org/10.1136/jmg.2008.058909>
- Bunday, S. (1982). Clinical evidence for heterogeneity in myotonic dystrophy. *Journal of Medical Genetics* , 19 (5), 341–348. <https://doi.org/10.1136/jmg.19.5.341>
- Butland, R. J. A., Pang, J., Gross, E. R., Woodcock, A. A., & Geddes, D. M. (1982). Two-, six-, and 12-minute walking tests in respiratory disease. *British Medical Journal* , 284 (6329), 1607–1608. <https://doi.org/10.1136/bmj.285.6337.295>
- Casanova, C., Celli, B. R., Barria, P., Casas, A., Cote, C., De Torres, J. P., Jardim, J., Lopez, M. V., Marin, J. M., Montes De Oca, M., Pinto-Plata, V., & Aguirre-Jaime, A. (2011). The 6-min walk distance in healthy subjects: Reference standards from seven countries. *European Respiratory Journal* , 37 (1), 150–156. <https://doi.org/10.1183/09031936.00194909>
- Charlet-B., N., Savkur, R. S., Singh, G., Philips, A. V., Grice, E. A., & Cooper, T. A. (2002). Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Molecular Cell* , 10 (1), 45–53. [https://doi.org/10.1016/S1097-2765\(02\)00572-5](https://doi.org/10.1016/S1097-2765(02)00572-5)
- Compston, A. (2010). Aids to the investigation of peripheral nerve injuries. Medical Research Council: Nerve Injuries Research Committee. His Majesty's Stationery Office: 1942; pp. 48 (iii) and 74 figures and 7 diagrams; with aids to the examination of the peripheral nervous . *Brain : A Journal of Neurology* , 133 (10), 2838–2844. <https://doi.org/10.1093/brain/awq270>
- Corrales, E., Vasquez, M., Zhang, B., Santamaria-Ulloa, C., Cuenca, P., Krahe, R., Monckton, D. G., & Morales, F. (2019). Analysis of mutational dynamics at the *DMPK* (CTG)_n locus identifies saliva as a suitable DNA sample source for genetic analysis in myotonic dystrophy type 1. *PLoS ONE* , 14 (5). <https://doi.org/10.1371/journal.pone.0216407>
- Cumming, S. A., Jimenez-Moreno, C., Okkersen, K., Wenninger, S., Daidj, F., Hogarth, F., Littleford, R., Gorman, G., Bassez, G., Schoser, B., Lochmuller, H., van Engelen, B. G. M., Monckton, D. G., & OPTIMISTIC Consortium. (2019). Genetic determinants of disease severity in the myotonic dystrophy type 1 OPTIMISTIC cohort. *Neurology* , 93 (10), e995–e1009. <https://doi.org/10.1212/WNL.0000000000008056>
- Furling, D., Lam, L. T., Agbulut, O., Butler-Browne, G. S., & Morris, G. E. (2003). Changes in myotonic dystrophy protein kinase levels and muscle development in congenital myotonic dystrophy. *American Journal of Pathology* , 162 (3), 1001–1009. [https://doi.org/10.1016/S0002-9440\(10\)63894-1](https://doi.org/10.1016/S0002-9440(10)63894-1)

- Furling, D., Lemieux, D., Taneja, K., & Puymirat, J. (2001). Decreased levels of myotonic dystrophy protein kinase (DMPK) and delayed differentiation in human myotonic dystrophy myoblasts. *Neuromuscular Disorders* , 11 (8), 728–735. [https://doi.org/10.1016/S0960-8966\(01\)00226-7](https://doi.org/10.1016/S0960-8966(01)00226-7)
- Gomes-Pereira, M., Bidichandani, S. I., & Monckton, D. G. (2004). Analysis of unstable triplet repeats using small-pool polymerase chain reaction. *Methods in Molecular Biology (Clifton, N.J.)* , 277 , 61–76. <https://doi.org/10.1385/1-59259-804-8:061>
- Gudde, A. E. E. G., Gonzalez-Barriga, A., van den Broek, W. J. A. A., Wieringa, B., & Wansink, D. G. (2016). A low absolute number of expanded transcripts is involved in myotonic dystrophy type 1 manifestation in muscle. *Human Molecular Genetics* , 25 (8), 1648–1662. <https://doi.org/10.1093/hmg/ddw042>
- Harper, P. S., Harley, H. G., Reardon, W., & Shaw, D. J. (1992). Review article: Anticipation in myotonic dystrophy: New light on an old problem. In *American Journal of Human Genetics* (Vol. 51, Issue 1, pp. 10–16). Elsevier. [/pmc/articles/PMC1682874/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/1682874/)
- Khajavi, M., Tari, A. M., Patel, N. B., Tsuji, K., Siwak, D. R., Meistrich, M. L., Terry, N. H., & Ashizawa, T. (2001). “Mitotic drive” of expanded CTG repeats in myotonic dystrophy type 1 (DM1). *Hum Mol Genet* , 10 (8), 855–863. <https://doi.org/10.1093/hmg/10.8.855>
- Kimura, T., Nakamori, M., Lueck, J. D., Pouliquin, P., Aoike, F., Fujimura, H., Dirksen, R. T., Takahashi, M. P., Dulhunty, A. F., & Sakoda, S. (2005). Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca²⁺-ATPase in myotonic dystrophy type 1. *Human Molecular Genetics* , 14 (15), 2189–2200. <https://doi.org/10.1093/hmg/ddi223>
- Konieczny, P., Stepniak-Konieczna, E., Taylor, -Lukasz, K., Sznajder, J., & Sobczak, K. (2017). Autoregulation of MBNL1 function by exon 1 exclusion from MBNL1 transcript. *Nucleic Acids Research* , 45 (4), 1760–1775. <https://doi.org/10.1093/nar/gkw1158>
- Korneluk, R. G. (1993). *Effect of the myotonic dystrophy {DM} mutation on mRNA levels of theDMgene* . 4 (july), 233–238. <https://doi.org/10.1038/ng0793-233>
- Kulkarni, A., Anderson, A. G., Merullo, D. P., & Konopka, G. (2019). Beyond bulk: a review of single cell transcriptomics methodologies and applications. *Current Opinion in Biotechnology* , 58 , 129–136. <https://doi.org/10.1016/j.copbio.2019.03.001>
- Lavedan, C., Hofmann-Radvanyi, H., Shelbourne, P., Rabes, J. P., Duros, C., Savoy, D., Dehaupas, I., Luce, S., Johnson, K., & Junien, C. (1993). Myotonic dystrophy: Size- and sex-dependent dynamics of CTG meiotic instability, and somatic mosaicism. *American Journal of Human Genetics* , 52 (5), 875–883. <http://www.ncbi.nlm.nih.gov/pubmed/8098180>
- Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O’Hoy, K., & et, al. (1992). Myotonic dystrophy mutation: an unstable CTG repeat in the 3’ untranslated region of the gene. *Science (New York, N.Y.)* , 255 (5049), 1253–1255. <https://doi.org/10.1126/science.1546325>
- Mathieu, J., Boivin, H., Meunier, D., Gaudreault, M., & Begin, P. (2001). Assessment of a disease-specific muscular impairment rating scale in myotonic dystrophy. *Neurology* , 56 (3), 336–340. <https://doi.org/10.1212/WNL.56.3.336>
- Miller, J. W., Urbinati, C. R., Teng-Ummuay, P., Stenberg, M. G., Byrne, B. J., Thornton, C. A., & Swanson, M. S. (2000). Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *The EMBO Journal* , 19 (17), 4439–4448. <https://doi.org/10.1093/emboj/19.17.4439>
- Monckton, D. G., Wong, L. J. C., Ashizawa, T., & Caskey, C. T. (1995). Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. *Human Molecular Genetics* , 4 (1), 1–8. <https://doi.org/10.1093/hmg/4.1.1>

- Morales, F., Couto, J. M., Higham, C. F., Hogg, G., Cuenca, P., Braidia, C., Wilson, R. H., Adam, B., del Valle, G., Brian, R., Sittenfeld, M., Ashizawa, T., Wilcox, A., Wilcox, D. E., & Monckton, D. G. (2012). Somatic instability of the expanded CTG triplet repeat in myotonic dystrophy type 1 is a heritable quantitative trait and modifier of disease severity. *Human Molecular Genetics* , 21 (16), 3558–3567. <https://doi.org/10.1093/hmg/dds185>
- Mykowska, A., Sobczak, K., Wojciechowska, M., Kozłowski, P., & Krzyzosiak, W. J. (2011). CAG repeats mimic CUG repeats in the misregulation of alternative splicing. *Nucleic Acids Research* , 39 (20), 8938–8951. <https://doi.org/10.1093/nar/gkr608>
- Overend, G., Legare, C., Mathieu, J., Bouchard, L., Gagnon, C., & Monckton, D. G. (2019). Allele length of the DMPK CTG repeat is a predictor of progressive myotonic dystrophy type 1 phenotypes. *Human Molecular Genetics* , 28 (13), 2245–2254. <https://doi.org/10.1093/hmg/ddz055>
- Pandey, S. K., Wheeler, T. M., Justice, S. L., Kim, A., Younis, H. S., Gattis, D., Jauvin, D., Puymirat, J., Swayze, E. E., Freier, S. M., Bennett, C. F., Thornton, C. A., & MacLeod, A. R. (2015). Identification and characterization of modified antisense oligonucleotides targeting DMPK in mice and nonhuman primates for the treatment of myotonic dystrophy type 1s. *Journal of Pharmacology and Experimental Therapeutics* , 355 (2), 329–340. <https://doi.org/10.1124/jpet.115.226969>
- Perchet, T., Chea, S., Hasan, M., Cumano, A., & Golub, R. (2017). Single-cell gene expression using multiplex RT-qPCR to characterize heterogeneity of rare lymphoid populations. *Journal of Visualized Experiments* , 119 (119), 1–11. <https://doi.org/10.3791/54858>
- Querido, E., Gallardo, F., Beaudoin, M., Menard, C., & Chartrand, P. (2011). Stochastic and reversible aggregation of mRNA with expanded CUG-triplet repeats. *Journal of Cell Science* , 124 (10), 1703–1714. <https://doi.org/10.1242/jcs.073270>
- Rantalainen, M. (2018). Application of single-cell sequencing in human cancer. *Briefings in Functional Genomics* , 17 (4), 273–282. <https://doi.org/10.1093/bfpg/elx036>
- Ren, X., Kang, B., & Zhang, Z. (2018). Understanding tumor ecosystems by single-cell sequencing: Promises and limitations 11 Medical and Health Sciences 1112 Oncology and Carcinogenesis 06 Biological Sciences 0604 Genetics. *Genome Biology* , 19 (1), 1–14. <https://doi.org/10.1186/s13059-018-1593-z>
- Renna, L. V., Bose, F., Brignonzi, E., Fossati, B., Meola, G., & Cardani, R. (2019). Aberrant insulin receptor expression is associated with insulin resistance and skeletal muscle atrophy in myotonic dystrophies. *PLoS One* , 14 (3), e0214254. <https://doi.org/10.1371/journal.pone.0214254>
- Salinas-Rios, V., Belotserkovskii, B. P., & Hanawalt, P. C. (2011). DNA slip-outs cause RNA polymerase II arrest in vitro: Potential implications for genetic instability. In *Nucleic Acids Research* (Vol. 39, Issue 17, pp. 7444–7455). <https://doi.org/10.1093/nar/gkr429>
- Stahlberg, A., & Kubista, M. (2018). Technical aspects and recommendations for single-cell qPCR. *Molecular Aspects of Medicine* , 59 , 28–35. <https://doi.org/10.1016/j.mam.2017.07.004>
- Stahlberg, A., Rusnakova, V., Forootan, A., Anderova, M., & Kubista, M. (2013). RT-qPCR work-flow for single-cell data analysis. *Methods* , 59 (1), 80–88. <https://doi.org/10.1016/j.ymeth.2012.09.007>
- Thornton, C. A., Johnson, K., & Moxley, R. T. (1994). Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes. *Annals of Neurology* , 35 (1), 104–107. <https://doi.org/10.1002/ana.410350116>
- Van Swieten, J. C., Koudstaal, P. J., Visser, M. C., Schouten, H., & Van Gijn, J. (1988). Interobserver agreement for the assessment of handicap in stroke patients. *Stroke* , 19 (5), 604–607. <https://doi.org/10.1161/01.STR.19.5.604>

Wong, L. J., Ashizawa, T., Monckton, D. G., Caskey, C. T., & Richards, C. S. (1995). Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. *American Journal of Human Genetics*, 56 (1), 114–122. <http://www.ncbi.nlm.nih.gov/pubmed/7825566>

Tables

Table 1: Clinical characteristics, expansion size and foci analysis of DM1 patients in this study.

Patient	CLINICAL DATA	CLINICAL DATA	CLINICAL DATA	CLINICAL DATA	CLINICAL DATA	CLINICAL DATA	CLINICAL DATA	EXPANSION SIZE	EXPANSION SIZE	FOCI ANALYSIS	FOCI ANALYSIS	FOCI ANALYSIS	FOCI ANALYSIS
	Age of onset	Age at sampling	Gender	mRS	MIRS	Biceps MRC	6MWT	Muscle	Myoblasts	No. of cells studied	% cells with foci	Mean	Med
P1	15*	36	F	2	4	4	348	691	823	72	97,2%	2,3	2
P2	48	54	M	2	3	5	251	ND	ND	72	98,6%	2,6	2
P3	36	41	F	1	2	5	368	208	1686	63	81%	1,4	1
P4	27	39	F	4	4	4	NP	469	561	39	97,4%	2,7	3
P5	36	41	M	2	3	5	519	135	150	66	28,8%	0,6	0

mRS: modified Rankin Scale, MIRS: Muscular Impairment Rating Scale, MRC: Muscle Power Scale, 6MWT: 6 Minute Walking Test, NP: not possible to asses, ND: not determined, *Interrogating P1, she is self-secure that her symptoms started at 36, but the neurological team thinks that the disease started during adolescence, since she had many signs at 36, including oval pallor and temporal atrophy that needed to appear early in the patient’s life.

Figure legends

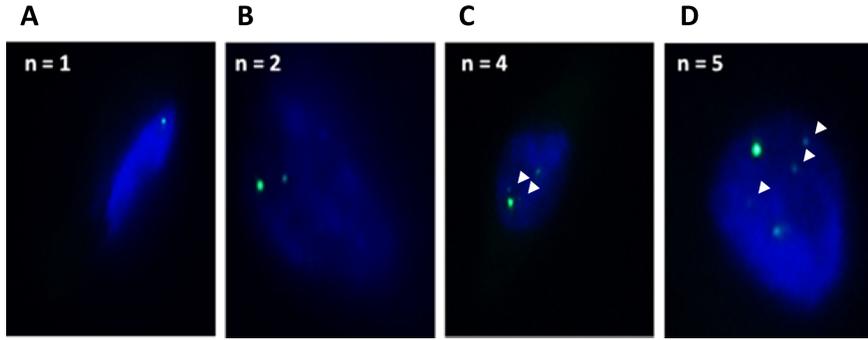
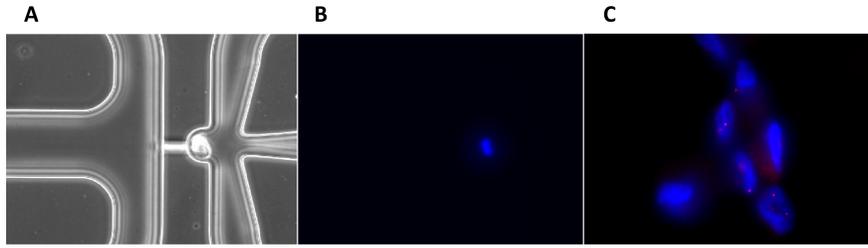
Figure 1 : RNA foci staining in myoblasts both on coverslips and in C1 IFC chip. **A**) Myoblasts stained in suspension and visualized on a coverslip with fluorescence (standard immunocytochemistry), where nuclei are stained in blue and RNA foci in red. **B**) Myoblast stained in suspension, captured in the chip and visualized with the brightfield channel, in which a myoblast can be observed. **C**) Myoblasts stained in suspension, captured in the chip and visualized with fluorescence. Note that DAPI (blue) can be observed but not RNA foci (red).

Figure 2 : RNA foci heterogeneity in the cells of one DM1 patient (n = number of foci). Some of the foci that are more difficult to see in these images are indicated by white triangles. Foci staining in green and nuclei in blue.

Figure 3 . Distribution of RNA foci per myoblast in every patient of this study. 0 = myoblasts with no RNA foci, 1 = myoblasts with 1 RNA foci, 2= myoblasts with 2 RNA foci, etc.

Figure 4 : Column graph of single cell *DMPK* (n = 245 for controls and n = 130 for patients) and exon 7 exclusion *MBNL1* (n= 232 for controls and 147 for patients) splicing variant expression levels between patients and controls. Bars indicate standard deviation.

Figure 5 : Grouped scatter plots for *DMPK* and *MBNL1* exon 7 exclusion isoform single cell data in patients and controls of our study. Bars indicate standard deviation.



Distribution of RNA foci

