Comprehensive transcriptomic analysis shows disturbed calcium homeostasis and deregulation of T lymphocyte apoptosis in inclusion body myositis

Mridul Johari[1,2]*, MSc, Anna Vihola[1,2,3], PhD, Johanna Palmio[4], MD, PhD, Manu Jokela[3,5], MD, PhD, Per Harald Jonson[1,2], PhD, Jaakko Sarparanta[1,2], PhD, Sanna Huovinen[6], MD, Marco Savarese[1,2], PhD, Peter Hackman[1,2], PhD, Bjarne Udd[1,2,4,7], MD, PhD

[1]Folkhälsan Research Center, Helsinki, Finland
[2]Department of Medical Genetics, Medicum, University of Helsinki, Finland
[3]Neuromuscular Research Center, Department of Genetics, Fimlab Laboratories, Tampere, Finland
[4]Neuromuscular Research Center, Department of Neurology, Tampere University and University Hospital, Tampere, Finland
[5]Division of Clinical Neurosciences, Department of Neurology, Turku University Hospital, Turku, Finland
[6]Department of Pathology, Fimlab Laboratories, Tampere University Hospital, Tampere, Finland
[7]Department of Neurology, Vaasa Central Hospital, Vaasa, Finland

*Corresponding Author
Mridul Johari, MSc
Folkhälsan Research Center,
Department of Medical Genetics, Medicum, University of Helsinki
Biomedicum, Haartmaninkatu 8, PO Box 63, FI-00014 Helsinki, Finland
Email: mridul.johari@helsinki.fi
Tel (office): +358 2941 25629

Keywords: inclusion body myositis, calcium, differential expression, differential splicing, T cells

Abstract

Objective: Inclusion body myositis (IBM) has an unclear molecular etiology due to the co-existence of characteristic cytotoxic T-cell activity and degeneration of muscle fibers. Using in-depth gene expression and splicing studies, we aimed at understanding the different components of the molecular pathomechanisms in IBM.

Methods: We performed RNA-seq on RNA extracted from skeletal muscle biopsies of clinically and histopathologically defined IBM (n=24), tibial muscular dystrophy (n=6), and histopathologically normal group (n=9). In a comprehensive transcriptomics analysis, we analyzed the differential gene expression, differential splicing and exon usage, downstream pathway analysis, and the interplay between coding and non-coding RNAs (micro RNAs and long non-coding RNAs).

Results: We observe dysregulation of genes involved in calcium homeostasis, particularly affecting the T-cell activity and regulation, causing disturbed Ca^{2+} induced apoptotic pathway of T cells in IBM muscles. Additionally, LCK/p56, which is an essential gene in regulating the fate of T-cell apoptosis, shows altered expression and splicing usage in IBM muscles

Interpretation: Our analysis provides a novel understanding of the molecular mechanisms in IBM by showing a detailed dysregulation of genes involved in calcium homeostasis and its effect on T-cell functioning in IBM muscles. Loss of T-cell regulation is hypothesized to be involved in the consistent observation of no response to immune therapies in IBM patients. Our results show that loss of apoptotic control of cytotoxic T cells could indeed be one component of their abnormal cytolytic activity in IBM muscles.

Introduction

Inclusion body myositis (IBM) is a late-onset, acquired muscle disease with unclear etiology, and the poorly understood molecular pathogenesis is under debate due to several factors. The CD8+ T-cell infiltration and overexpression of class I MHC antigens in all muscle fibers indicate an autoimmune cascade and are, in fact, the most consistent finding together with the degeneration of myofibers. However, IBM largely remains refractory to immunosuppressive drugs [1], and comprehensive clinical trials have generally been ineffective [2]. A partial clinical and histopathological overlap with other rimmed-vacuolar (RV) myopathies [3] including accumulations of similar proteins in the RVs [4] support a degenerative pathophysiology. Accumulation/aggregation of these misfolded proteins suggests that IBM could be a protein aggregate disease with immune-mediated cytotoxic inflammation as a resulting secondary feature [5]. However, there is a significant variance in nature and the number of accumulated proteins observed in the IBM muscle biopsies [6]. Similar aggregates observed in HIV-associated IBM [7] suggest that protein aggregation can still be a downstream effect of immune dysfunction. Additionally, the occurrence of rare familial cases [8] and a strong association with immune MHC locus 8.1 ancestral haplotype [9, 10] support a possible genetic predisposition for IBM.

Analysis of tissue-specific mRNAs and subsequent RNA-seq based transcriptomics studies focused on understanding the expression of genes, participating pathways, and networks can increase our understanding of underlying pathomechanisms. Prior studies have investigated the differential gene expression in IBM muscles for both the inflammatory and the degenerative pathology [11-17]. However, no study has attempted a comprehensive analysis of RNA-seq data combining differential gene expression, differential exon, and splicing usage along with an in-depth analysis of the relation between dysregulation of coding and regulatory RNAs in IBM muscles.

Our study used RNA extracted from muscle biopsies of IBM patients, of non-myositis RV-myopathy disease group, and a histopathologically group. We first studied the differential expression of coding, long non-coding RNAs (lncRNAs), and micro RNAs (miRNAs) and then evaluated their possible interplay. Additionally, we studied the transcriptome-wide differential exon and splicing usage. We observed a significant association with genes involved in various calcium-related pathways and identified disturbed calcium regulation specific to T cells in IBM muscles, highlighting the relevance of calcium homeostasis for T-cell activity in IBM muscles. In particular, we identified calcium-induced T lymphocyte apoptosis to be disturbed in IBM muscles.

Materials and methods

Patients and skeletal muscle biopsies

Muscle biopsies (predominantly Tibialis anterior or Vastus lateralis) from 24 Finnish patients diagnosed with clinically and pathologically defined IBM according to the ENMC criteria [18] were included. The age of onset was 60 ± 11 years (median ± SD), and the age at muscle biopsy was 70 ± 9 years. Additionally, muscle biopsies from six patients with genetically diagnosed Tibial muscular dystrophy (TMD, caused by heterozygous FINmaj mutation the titin gene) [19] were included. In the TMD cohort, the age of onset was 49 ± 11 years, and age at biopsy 54 ± 14 years. Nine muscle biopsies from individuals that underwent leg amputation for reasons other than a muscle disease [20] were also included. These nine biopsies did not show pathologically defined muscle degeneration or inflammation. Age at sampling for amputees was 70 ± 11 years. All muscle biopsies were snap-frozen and stored at -80 °C. Muscle biopsies were collected at the Tampere Neuromuscular Research Center, Tampere University Hospital, Finland.

RNA extraction, selection, and library preparation

Muscle tissue homogenization steps were performed using SpeedMill PLUS (Analytik Jena AG, Germany). RNA was extracted with Qiagen RNeasy Plus Universal Mini Kit (Qiagen, Hilden, Germany) and treated with Invitrogen TURBO DNAse buffer (ThermoFisher Scientific, MA, USA) according to the manufacturers' instructions. RNA was quantified and qualitatively assessed using High Sensitivity RNA ScreenTape (Agilent Technologies, CA, USA) on Agilent 4200 TapeStation system (Agilent Technologies).

Library preparations and sequencing were performed at Oxford Genomics Center, University of Oxford. For PolyA+ RNA selection, the NEBNext Ultra II Directional RNA Library Prep kit (E7760) for Illumina (NEB, Beverly, MA, USA) was used to prepare strand-specific RNA-seq libraries. Libraries were multiplexed and sequenced on HiSeq4000: 75bp paired-end sequencing (Illumina, CA, USA), and an average of ~47 million reads per sample were produced. Samples with enough RNA were used for library preparation for small RNA (< 200 nt) selection (18 IBM, nine amputees, and four TMD). NEBNext Small RNA Library Prep Set (E7330) for Illumina was used per the manufacturer's instructions (NEB). Libraries were multiplexed and sequenced on HiSeq2500: 50bp single-end sequencing (Illumina), and an average of ~10 million reads per sample were produced.

RNA-seq data pre-processing, QC, and alignment

Adapter sequences and low-quality bases were removed with fastp [21]. Trimmed sequences were then mapped with STAR 2.7.0d [22] (STAR, RRID: SCR_004463) with index generated from Gencode.v29 human reference (release date 05.2018, based on ENSEMBL GRCh38.p12) and comprehensive gene annotation (primary assembly) using the STAR two-pass method according to the guidelines from the ENCODE project for alignment of long RNA (>200 nt) and small RNA (<200 nt) data.

RNA-seq quantification and differential gene expression analysis

Uniquely mapped fragments were summarized and quantified (referred to as counts) by featureCounts [23] (featureCounts, RRID: SCR_012919) using Gencode.v29 primary comprehensive gene annotation, which lists 58,780 RNAs including 19,969 protein-coding, 16,066 non-coding, and 22,745 other types of RNAs (primary gene expression analysis). Separate quantification of counts for lncRNA (lncRNA analysis) was done using long non-coding RNA gene annotation from Gencode.v29 (a subset of the primary annotation). Quantification of counts for miRNAs (miRNA analysis) in 31 samples was done using miRBase human miRNA annotation (Release 22.1 October 2018) [24]. Differential gene expression analysis was performed with DESeq2 [25] (v1.26.0) (DESeq2, RRID: SCR_015687) in Rstudio (v1.2.5019) (RStudio, RRID: SCR_000432) based on R (v3.6.3) (R Project for Statistical Computing, RRID: SCR_001905). Counts were normalized with variance stabilizing transformation function within DESeq2. A principal component analysis (PCA) was performed on the gene expression data of the IBM samples compared to amputee and TMD groups. Further, pairwise comparisons between cohorts were performed using the Wald test. Log_2 fold changes (LFC) were shrunk using 'ashr' adaptive shrinkage estimation [26], and results were generated with default independent filtering for increasing power. Only genes with LFC values larger than ±1.5 and a Benjamini-Hochberg adjusted p-value of ≤0.01 were considered further. Genes specifically dysregulated in IBM muscles were considered for downstream analysis.

Pathway analysis

Ingenuity Pathway Analysis (IPA, QIAGEN Inc.) (Ingenuity Pathway Analysis, RRID: SCR_008653) was used for pathway analysis and enrichment analysis of the obtained differential gene expression data. Using Ingenuity Pathways Knowledge Base (Ingenuity Pathways Knowledge Base, RRID: SCR_008117), IPA mapped and annotated genes to the pathways and predicted activation state based on the direction of changes comparing it with the change in the database.

144 Differential splicing analysis

145 To investigate differential usage of exons and splicing, independent of the differential gene
146 expression analysis, we used QoRTS [27] java-based application (v1.3.6) (QoRTs, RRID:
147 SCR_018665) to prepare counts from exons and splice junctions (known and novel) from the aligned
148 data. Downstream analysis of this data was performed using JunctionSeq [28] (v1.16.0) in R.
149 JunctionSeq results produce a q-value (based on FDR) on gene-level analysis, which considers that
150 one or more exon/junction in this gene is differentially used. A conservative q-value threshold of 0.01
151 was used to select significant observations. IBM-specific differentially expressed genes and
152 differentially spliced genes were compared (Fig. 1). Statistical over-enrichment analysis for Gene
153 ontology terms in categories: Molecular function, biological process, and cellular component, was
154 performed on results obtained from QoRTs/JunctionSeq using clusterProfiler [29] (clusterProfiler,
155 RRID: SCR_016884). Gene sets were compared using UpSet plot [30].

156 # Results

157 Expression signature in IBM muscles

158 Fig. 1a shows the summarized workflow of the methodology. The PCA shown in Fig. 1b explains the
159 differences between the three cohorts. Pairwise comparisons were performed to reduce the potential
160 confounding effects of groups, which identified 2,288 and 302 genes specifically up- or down-
161 regulated in the IBM cohort, respectively (Fig. 1c). Non-coding RNA analyses resulted in 497
162 lncRNAs upregulated, 106 lncRNAs downregulated, 140 miRNAs upregulated, and 126 miRNAs
163 explicitly downregulated in the IBM cohort compared to other groups. These IBM-specific
164 dysregulated RNAs were used for downstream pathway analysis using IPA workflow. The top 15
165 genes dysregulated specifically in IBM muscles, with their functional annotations and normalized
166 expression in the different cohorts, are shown in Fig. 2.

167 Pathway analysis

168 We performed IPA workflow analysis on IBM-specific dysregulated genes to better understand the
169 pathways and the upstream regulators associated with the observed expression dysregulation. Out of
170 these, 2,588 genes, 596 lncRNAs, and 257 miRNAs mapped to the Ingenuity database. From the
171 primary gene expression analysis, IPA identified 91 pathways as significantly altered. Table 1 shows
172 a summary of the IPA results with the top identified pathways.

173 The top upstream regulators in both miRNA and lncRNA analysis are shown in table 2 and table 3,
174 respectively. We identified an increased expression of the lncRNA *DNM3OS* (DNM3 antisense RNA)

and *MIAT* (Myocardial infarction associated transcript) from these analyses. IPA suggested this dysregulation may be due to *JDP2* (Jun Dimerization Protein 2) and *TARDBP* (TAR DNA Binding Protein), acting as an upstream regulator of *DNM3OS* and *MIAT* respectively (table 4).

Dysregulation of calcium-related pathways in IBM muscles

IPA identified calcium-induced T lymphocyte apoptosis as one of the most significant pathways dysregulated in IBM muscles (table 1). Our IBM-specific dataset contained 69 genes with significant dysregulation out of the 232 genes annotated in this pathway. A part of this pathway, including the major players, is shown in Fig. 3. Another pathway outside the top results identified that 29 genes (29/208, p = 7.05E-03) significantly dysregulated in our dataset are also involved in calcium signaling. These results prompted us to investigate further for calcium related issues in cellular signaling, and we found that IPA also detects dysregulation of the following processes, mobilization of Ca^{2+} (80 genes), the release of Ca^{2+} (33 genes), quantity of Ca^{2+} (51 genes) and flux of Ca^{2+} (51 genes), as significantly disturbed in IBM muscles (table 4).

Altered exon usage and splicing pattern in IBM muscles

To explore IBM-specific exon usage, we performed an independent transcriptome-wide differential splicing analysis in our three cohorts. We obtained a list of 1,271 differentially spliced genes in IBM from our differential splicing analysis. These transcripts either showed IBM-specific increased usage of a known junction or a known exon or contained a novel exon-exon junction resulting in an alternative isoform. To understand the diverse portfolio of mature mRNAs created from pre-mRNAs, we used gene ontology over-enrichment analysis on these 1,271 differentially spliced genes and identified the first splicing signature specific to IBM muscles. To understand the different classes over-represented in these genes, we performed statistical over-enrichment analysis using clusterProfiler for all three GO categories as seen in Fig. 4 a,b,c. Our analysis showed an enrichment of genes involved in the structure and organization of actin filaments assembly in IBM muscles and, interestingly, proteins involved in mRNA processing and metabolism.

We then compared the list of differentially spliced genes with differentially expressed genes in our analysis and found an overlap of 79 genes (Fig. 1d). Next, we wanted to observe the overlap between six different sets of genes, namely IBM specific differentially spliced genes, calcium-induced T Lymphocyte apoptosis, Mobilization of Ca2+, Flux of Ca2+, Quantity of Ca2+, and Release of Ca2+ (Fig. 4d). We observed 10 genes to be associated with calcium-related processes; *HLA-DPA1*, *HLA-DPB1,* and *HLA-DQB1* are associated with calcium-induced T Lymphocyte apoptosis, *ANXA1* is associated with mobilization, flux, and release of Ca^{2+}, *CCL4* is associated with mobilization, flux,

and quantity of Ca^{2+}, *GRK3* and *RARRES2* are associated with mobilization, *SH3KBP1* with flux, and *ITGAM* with the quantity of Ca^{2+}. In particular, one specific differentially spliced gene, *LCK,* is part of all six sets.

Fig. 5a shows the gene expression of *LCK* in three cohorts, with expression in IBM muscles being significantly higher than the others (log_2FC = +2.86, padj=3.50E-11, ranking = 355/2590). Additionally, Fig. 5b shows the differential splicing pattern observed in *LCK* in all three groups. The highlighted E016 corresponds to an alternative exon (chr1:32274818-32274992, GRCh38).

Discussion

In this study, we aimed to identify a more detailed IBM-specific molecular signature, using different RNA-seq based methods that can help us explore the inflammatory and degenerative parts in depth. Antigen-driven T-cell cytotoxicity is the most reproducible and plausible part of the complex molecular pathomechanism in IBM. However, it remains unknown what antigen drives this IBM-specific immune cascade.

As part of the RV pathology, accumulated proteins or the unfolded protein response have been hypothesized to prompt an immune reaction [5]. A recent unbiased proteomics study dissected these RVs in IBM [31]. Interestingly, the protein encoded by one of our top differentially expressed genes, *MYL4,* is also detected in the RVs in IBM along with *ANXA1,* which is both differentially expressed and differentially spliced in IBM muscles. In our study design, we considered TMD, another RV muscle disease but without immune involvement, to understand if there are any RV-specific antigens in IBM muscles. Additionally, using age-matched histopathologically normal muscles from amputees, we aimed to understand if general inflammatory signatures can be replicated and studied in more detail using additional methods such as non-coding RNAs and differential splicing studies. Consequently, our strong study design and robust methodology helped us replicate findings from previous studies [11-17] and identify essentially new calcium-related issues in IBM muscles and their link with the altered T-cell cytotoxicity in IBM muscle fibers.

We found that several genes contributing to calcium homeostasis are differentially expressed in IBM muscles resulting in dysregulation of several critical pathways, specifically, calcium-induced T lymphocyte apoptosis and related Nur77 signaling. Ca^{2+} is a universal second messenger in T cells, and it is known to regulate proliferation and differentiation of T cells and T-cell effector functions [32]. The complexity and duration of Ca^{2+} signals and resultant cytoskeletal rearrangements determine the fate of T cells in response to an antigen [33]. On one hand, a short-term increase in intracellular Ca^{2+} concentration results in the cytolytic activity of T cells; on the other hand, prolonged

239 elevation results in proliferation, differentiation, and maturation of näive T cells into Th1, Th2, and
240 Th17 subtypes and the production of cytokines[32].

241 Ca^{2+} signaling is known to optimize the interaction between T cells and antigen-presenting cells [33].
242 The binding of antigen/MHC complexes ($CD8^+$-MHC class I/$CD4^+$-MHC class II) to T-cell receptors
243 (TCR) activates Src-family protein tyrosine kinases, e.g., LCK and FYN at the cytoplasmic side of
244 the TCR/CD3 complex. Additionally, activation of ZAP-70, a tyrosine kinase associate protein,
245 results in the phosphorylation and activation of the intracellular enzyme phospholipase C-γ1 (PLC-
246 γ1) [32, 33]. PLC-γ1 hydrolyses phosphatidylinositol 4,5-biphosphate (PIP2) to produce two other
247 second messengers, inositol 1,4,5-triphosphate (IP3) and diacylglycerol (DAG). IP3 binds to its
248 receptor (IP3R) on the endoplasmic reticulum (ER) membrane to promote rapid release of Ca^{2+} from
249 ER to the cytosol [32]. However, this release of Ca^{2+} is insufficient for antigen-derived T-cell fate
250 but results in depletion of intracellular Ca^{2+} triggering a rapid influx of Ca^{2+} through activation and
251 opening of Ca^{2+} release-Ca^{2+} activated channels (CRAC) on the plasma membrane formed by
252 different STIM1/ORAI1 combinations [34]. The duration of Ca^{2+} influx is vital for activating the
253 calcineurin-dependent nuclear factor activate T cells (NFAT) transcription pathway [32]. In the
254 cytoplasm, calcineurin removes excess phosphate residues from the N terminus of NFAT, promoting
255 its entry into the nucleus. Disruptions in NFAT signaling can cause several phenotypes, including
256 cardiovascular, musculoskeletal, and immunological diseases [35]. Meanwhile, DAG, another
257 secondary messenger, activates protein kinase C (PKC), which in turn activates the nuclear factor
258 kappa B (NFκB). The duration and complexity of Ca^{2+} signals drive the NFAT/NFκB signaling and
259 determine downstream T-cell activation.

260 The genes in the NR4A family (NR4A1/Nur77, NR4A2/Nurr1, NR4A3/Nor1) act as critical
261 molecular switches in cell survival and inflammation. Human *NR4A1* encodes for a homolog of a
262 mouse protein called Nur77, a zinc transcription factor expressed as an early gene in T cells upon
263 antigen-TCR interaction. In addition to being a transcriptional activator, Nur77 has an apoptotic role
264 in T regulatory fate [36] and other non-genomic proapoptotic functions via mitochondrial interactions
265 with Bcl-2 [37]. T cells deficient in Nur77 have been shown to have high proliferation, enhanced T-
266 cell activation, and increased susceptibility for T-cell-mediated inflammatory diseases [38]. The
267 expression of Nur77 is Ca^{2+} dependent and is controlled by the myocyte enhancer factor 2 (MEF2)
268 transcription factor [39], whose DNA-binding and transcriptional activity is enhanced by Calcineurin.
269 Another calcium-dependent transcription factor, CABIN1, acts as a transcriptional repressor of
270 MEF2, thus keeping the Nur77 promoter silent in the absence of a TCR signal [40]. The interaction
271 between CABIN1 and calcineurin is influenced by intracellular Ca^{2+} and PKC activation, resulting in

hyperphosphorylation of CABIN1 and its subsequent transcription repressing activity. An increase in intracellular Ca^{2+} concentration activates the interaction of the calmodulin family of genes (CALM) with CABIN1, triggering the dissociation of MEF2 from Cabin and MEF2 to become transcriptionally active [41]. In the nucleus, NFAT interacts with MEF2 and enhances its transcriptional activity by recruiting the co-activator p300 for the transcription of Nur77.

In our dataset, 69 genes mapping to the calcium-induced T Lymphocyte apoptosis and 72 genes mapping the Nur77 signaling in T Lymphocytes are differentially expressed in IBM muscles. As seen in Fig. 3, several essential genes like *ZAP70*, *LCK*, different subunits of Protein Kinase C, and *ATP2A1* which encodes for SERCA, are significantly changed in IBM muscles. Additionally, we also observed genes associated with the mobilization of Ca^{2+}, the release of Ca^{2+}, the quantity of $Ca^{2+,}$ and the flux of Ca^{2+} as significantly dysregulated in IBM muscles, indicating a possible widespread disturbance with the handling of calcium entry and release in cells. In T cells, especially, this disturbance could dramatically impact their activation, differentiation, and most likely, the regulation of T-cell apoptosis will be disturbed.

Apoptosis in T cells is necessary to resolve their inflammatory activity, and defective or delayed apoptosis may contribute to the pathogenesis of inflammatory diseases [42]. In this scenario, loss of apoptotic control could be one mechanism explaining the lack of immune-suppressive therapeutic effect in IBM [43].

The diversity of the skeletal muscle proteome is, among others, dependent on the diversity of exon usage in pre-mRNAs [44]. From our transcriptome-wide splicing analysis within the differentially expressed genes, we identified 79 genes, out of which ten are associated with different calcium-related functions. Amongst these, LCK is a T lymphocyte-specific protein tyrosine kinase involved in downstream events of antigen–TCR interaction. LCK/p56 is essential in transducing signals leading to apoptotic cell death in mature T cells [45], and its activity is tightly regulated to protect against hyperactivation of T cells and autoimmunity, thus maintaining T-cell homeostasis [46]. Moreover, LCK also selectively influences the flux and release of calcium in cells [47]. In our analysis, LCK is both differentially expressed and differentially spliced in IBM muscles. Disturbed T-cell apoptosis and the dysregulation of LCK in IBM muscles provide novel insights into the molecular mechanisms of IBM. Considering the crucial regulatory activity of LCK, it might be a potential therapeutic target for IBM patients.

We also observe dysregulation of several non-coding RNAs in our study. Previously, Hamann and colleagues have discussed lncRNAs in the context of IBM [13]. The benefits of our study design,

304 especially the homogenous molecular pathology and the larger sample size, let us dig deeper into the
305 dysregulation of lncRNAs specific to IBM muscles. We identified that JDP2 (DNA binding
306 transcription factor) and TARDBP/TDP-43 (DNA and RNA binding protein) might have altered
307 regulator activity since their downstream non-coding partners (*DNM3OS* and *MIAT,* respectively) are
308 significantly overexpressed IBM muscles. Additionally, both these proteins are specific to RNA
309 polymerase II (RNA Pol II), facilitating transcription and pre-mRNA maturation. Alteration in RNA
310 or DNA binding proteins (expression or localization) associated with the activity of the spliceosome
311 machinery can directly affect the downstream events. Since TDP-43 is accumulated in RVs, one
312 possibility is that the unavailability of TDP-43 can affect its transcription and splicing activities. The
313 normal expression of *TARDBP* we observe in IBM patients is expected and is in coherence with the
314 previous reports [48]. In inherited muscle diseases, damaging variants in the disease-associated gene
315 can result in mislocalization and accumulation of mutant protein in the muscle fibers. Previous studies
316 have reported rare exonic variants in genes, including *VCP* and *SQSTM1* in IBM [49, 50]. However,
317 in our cohort of IBM patients, there were no rare exonic *TARDBP, VCP,* or *SQSTM1* variants [9] that
318 could suggest a possible association with abnormal protein turnover and accumulation/aggregation.
319 Therefore, further evidence to suggest the pathogenic role of variants in such genes and their
320 downstream effect on pathological protein accumulation in IBM is still missing. However, the
321 potential downside of TDP-43 not being available for its traditional roles, such as effective splicing,
322 because of the aggregation is noteworthy. Further evidence of possible dysregulation of splicing in
323 IBM muscles comes from our differential splicing results where proteins involved in mRNA
324 processing, transcription, and regulation are enriched, suggesting that additional studies are needed
325 to understand the possible impact of dysregulated mRNA processing in IBM muscles.

326 Previously, Pinal-Fernandez and colleagues observed that calcium-induced T lymphocyte apoptosis
327 was a significant IBM-specific dysregulated pathway in their extensive analysis of different
328 inflammatory myopathies but did not comment further on the possible importance [17]. Additionally,
329 using a smaller sample size, Amici and colleagues identified calcium signaling as one of the
330 significant disturbed pathways in IBM muscles and hypothesized its potential major role [14].
331 Furthermore, previous gene expression studies in IBM have analyzed data primarily from microarrays
332 [11, 12, 15]. Only recently paired-end reads RNA-seq have been used in IBM studies [13, 14, 16, 17].
333 While microarray-based analyses are comparable for differential expression studies, RNA-seq based
334 methodologies are superior for in-depth transcriptome analysis. In our study design, we used matched
335 muscle biopsies and state-of-the-art RNA-seq analysis tools. Our analyses show novel molecular
336 events in IBM muscles which increase our understanding of IBM and provide valuable additions to

improve the therapeutic interventions considering the disturbed calcium homeostasis, dysregulation of LCK, and associated deregulation of apoptotic control of T cells in IBM muscles.

Acknowledgments

The authors would like to thank the patients and their families. We thank the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (funded by Wellcome Trust grant reference 203141/Z/16/Z) for the generation and initial processing of the sequencing data. We would like to thank CSC – IT Center for Science, Finland, for its computational resources. We thank Merja Soininen, Talha Qureshi and Eini Penkkimäki for their technical assistance. We would like to thank Peter-Bram 't Hoen for his critical review of this manuscript.

Declarations

Funding

This work was supported by the Folkhälsan Research Foundation, Doctoral program in Integrative Life Science (ILS) and Doctoral school in Health Sciences (DSHealth), University of Helsinki (MJ), the Päivikki and Sakari Sohlberg Foundation (MJ), the Biomedicum Helsinki Foundation (MJ), Finska läkaresällskapet (BU/MJ), the Finnish Medical Foundation (JP), the Paulo foundation (MS), the Jane and Aatos Erkko Foundation (PH) and the Sigrid Jusélius Foundation (BU).

Author Contributions

MJ: Conceptualization of the study, funding acquisition, data analysis and curation, methodology, project administration, visualization, writing the original draft, review, and editing of the manuscript.

AV: data analysis, methodology, writing the original draft.

JP: Patient samples and data collection, writing the original draft.

MJ: Patient samples and data collection, review, and editing of the manuscript.

PH.J.: Conceptualization of the methodology, review, and editing of the manuscript.

JS: Data analysis, review, and editing of the manuscript.

SH: Patient samples and data collection, review, and editing of the manuscript.

MS: Conceptualization of the study and methodology, review and editing of the manuscript.

PH: Funding acquisition, project administration, and review and editing of the manuscript.

BU.: Conceptualization of the study, funding acquisition, supervision, project administration, review, and editing of the manuscript.

Conflicts of interest

The authors report no conflicts of interest.

Ethical approval

The study was performed in line with the principles of the Declaration of Helsinki. Ethical approval for this study falls under HUS:195/13/03/00/11. Informed consent from the patients was obtained at the time of sample collection.

Data availability

Raw counts and normalized DESeq2 counts from polyA+ RNAs and miRNAs are available in GEO as superseries GSE151758.

References

[1]. Ikenaga C, Kubota A, Kadoya M*, et al.* Clinicopathologic features of myositis patients with CD8-MHC-1 complex pathology. *Neurology*. 2017 **89:** 1060-1068. https://doi.org/10.1212/WNL.0000000000004333

[2]. Dalakas MC, Koffman B, Fujii M, Spector S, Sivakumar K, Cupler E. A controlled study of intravenous immunoglobulin combined with prednisone in the treatment of IBM. *Neurology*. 2001 **56:** 323-327. https://doi.org/10.1212/wnl.56.3.323

[3]. Cai H, Yabe I, Sato K*, et al.* Clinical, pathological, and genetic mutation analysis of sporadic inclusion body myositis in Japanese people. *J Neurol*. 2012 **259:** 1913-1922. https://doi.org/10.1007/s00415-012-6439-0

[4]. Askanas V, Engel WK. Sporadic inclusion-body myositis: a proposed key pathogenetic role of the abnormalities of the ubiquitin-proteasome system, and protein misfolding and aggregation. *Acta Myol*. 2005 **24:** 17-24.

[5]. Askanas V, Engel WK. Sporadic inclusion-body myositis and hereditary inclusion-body myopathies: current concepts of diagnosis and pathogenesis. *Curr Opin Rheumatol*. 1998 **10:** 530-542. https://doi.org/10.1097/00002281-199811000-00005

[6]. Greenberg SA. Theories of the pathogenesis of inclusion body myositis. *Curr Rheumatol Rep*. 2010 **12:** 221-228. https://doi.org/10.1007/s11926-010-0102-5

[7]. Hiniker A, Daniels BH, Margeta M. T-Cell-Mediated Inflammatory Myopathies in HIV-Positive Individuals: A Histologic Study of 19 Cases. *J Neuropathol Exp Neurol*. 2016 **75:** 239-245. https://doi.org/10.1093/jnen/nlv023

[8]. Amato AA, Shebert RT. Inclusion body myositis in twins. *Neurology*. 1998 **51:** 598-600. https://doi.org/10.1212/wnl.51.2.598

[9]. Johari M, Arumilli M, Palmio J*, et al.* Association study reveals novel risk loci for sporadic inclusion body myositis. *Eur J Neurol*. 2017 **24:** 572-577. https://doi.org/10.1111/ene.13244

[10]. Rothwell S, Chinoy H, Lamb JA*, et al.* Focused HLA analysis in Caucasians with myositis identifies significant associations with autoantibody subgroups. *Annals of the Rheumatic Diseases*. 2019 **78:** 996-1002. https://doi.org/10.1136/annrheumdis-2019-215046

[11]. Greenberg SA, Sanoudou D, Haslett JN, *et al.* Molecular profiles of inflammatory myopathies. *Neurology*. 2002 **59:** 1170-1182. https://doi.org/10.1212/wnl.59.8.1170

[12]. Eisenberg I, Eran A, Nishino I, *et al.* Distinctive patterns of microRNA expression in primary muscular disorders. *Proc Natl Acad Sci U S A*. 2007 **104:** 17016-17021. https://doi.org/10.1073/pnas.0708115104

[13]. Hamann PD, Roux BT, Heward JA, *et al.* Transcriptional profiling identifies differential expression of long non-coding RNAs in Jo-1 associated and inclusion body myositis. *Sci Rep*. 2017 **7:** 8024. https://doi.org/10.1038/s41598-017-08603-9

[14]. Amici DR, Pinal-Fernandez I, Mazala DA, *et al.* Calcium dysregulation, functional calpainopathy, and endoplasmic reticulum stress in sporadic inclusion body myositis. *Acta Neuropathol Commun*. 2017 **5:** 24. https://doi.org/10.1186/s40478-017-0427-7

[15]. Greenberg SA, Pinkus JL, Kong SW, Baecher-Allan C, Amato AA, Dorfman DM. Highly differentiated cytotoxic T cells in inclusion body myositis. *Brain*. 2019 **142:** 2590-2604. https://doi.org/10.1093/brain/awz207

[16]. Pinal-Fernandez I, Casal-Dominguez M, Derfoul A, *et al.* Identification of distinctive interferon gene signatures in different types of myositis. *Neurology*. 2019 **93:** e1193-e1204. https://doi.org/10.1212/WNL.0000000000008128

[17]. Pinal-Fernandez I, Casal-Dominguez M, Derfoul A, *et al.* Machine learning algorithms reveal unique gene expression profiles in muscle biopsies from patients with different types of myositis. *Ann Rheum Dis*. 2020 **79:** 1234-1242. https://doi.org/10.1136/annrheumdis-2019-216599

[18]. Rose MR, Group EIW. 188th ENMC International Workshop: Inclusion Body Myositis, 2-4 December 2011, Naarden, The Netherlands. *Neuromuscul Disord*. 2013 **23:** 1044-1055. https://doi.org/10.1016/j.nmd.2013.08.007

[19]. Hackman P, Vihola A, Haravuori H, *et al.* Tibial muscular dystrophy is a titinopathy caused by mutations in TTN, the gene encoding the giant skeletal-muscle protein titin. *Am J Hum Genet*. 2002 **71:** 492-500. https://doi.org/10.1086/342380

[20]. Huovinen S, Penttila S, Somervuo P*, et al.* Differential isoform expression and selective muscle involvement in muscular dystrophies. *Am J Pathol*. 2015 **185:** 2833-2842. https://doi.org/10.1016/j.ajpath.2015.06.018

[21]. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018 **34:** i884-i890. https://doi.org/10.1093/bioinformatics/bty560

[22]. Dobin A, Davis CA, Schlesinger F*, et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013 **29:** 15-21. https://doi.org/10.1093/bioinformatics/bts635

[23]. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014 **30:** 923-930. https://doi.org/10.1093/bioinformatics/btt656

[24]. Griffiths-Jones S. The microRNA Registry. *Nucleic Acids Res*. 2004 **32:** D109-111. https://doi.org/10.1093/nar/gkh023

[25]. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014 **15:** 550. https://doi.org/10.1186/s13059-014-0550-8

[26]. Stephens M. False discovery rates: a new deal. *Biostatistics*. 2016 **18:** 275-294. https://doi.org/10.1093/biostatistics/kxw041

[27]. Hartley SW, Mullikin JC. QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. *BMC Bioinformatics*. 2015 **16:** 224. https://doi.org/10.1186/s12859-015-0670-5

[28]. Hartley SW, Mullikin JC. Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq. *Nucleic Acids Res*. 2016 **44:** e127. https://doi.org/10.1093/nar/gkw501

[29]. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012 **16:** 284-287. https://doi.org/10.1089/omi.2011.0118

[30]. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: Visualization of Intersecting Sets. *IEEE Trans Vis Comput Graph*. 2014 **20:** 1983-1992. https://doi.org/10.1109/TVCG.2014.2346248

[31]. Güttsches A-K, Brady S, Krause K*, et al.* Proteomics of rimmed vacuoles define new risk allele in inclusion body myositis. *Annals of Neurology*. 2017 **81:** 227-239. https://doi.org/10.1002/ana.24847

[32]. Joseph N, Reicher B, Barda-Saad M. The calcium feedback loop and T cell activation: how cytoskeleton networks control intracellular calcium flux. *Biochim Biophys Acta*. 2014 **1838:** 557-568. https://doi.org/10.1016/j.bbamem.2013.07.009

[33]. Lewis RS. Calcium signaling mechanisms in T lymphocytes. *Annu Rev Immunol*. 2001 **19:** 497-521. https://doi.org/10.1146/annurev.immunol.19.1.497

[34]. Qu B, Al-Ansary D, Kummerow C, Hoth M, Schwarz EC. ORAI-mediated calcium influx in T cell proliferation, apoptosis and tolerance. *Cell Calcium*. 2011 **50:** 261-269. https://doi.org/10.1016/j.ceca.2011.05.015

[35]. Crabtree GR, Olson EN. NFAT signaling: choreographing the social lives of cells. *Cell*. 2002 **109 Suppl:** S67-79. https://doi.org/10.1016/s0092-8674(02)00699-2

[36]. Calnan BJ, Szychowski S, Chan FK, Cado D, Winoto A. A role for the orphan steroid receptor Nur77 in apoptosis accompanying antigen-induced negative selection. *Immunity*. 1995 **3:** 273-282. https://doi.org/10.1016/1074-7613(95)90113-2

[37]. Thompson J, Winoto A. During negative selection, Nur77 family proteins translocate to mitochondria where they associate with Bcl-2 and expose its proapoptotic BH3 domain. *J Exp Med*. 2008 **205:** 1029-1036. https://doi.org/10.1084/jem.20080101

[38]. Liebmann M, Hucke S, Koch K*, et al.* Nur77 serves as a molecular brake of the metabolic switch during T cell activation to restrict autoimmunity. *Proceedings of the National Academy of Sciences*. 2018 **115:** E8017-E8026. https://doi.org/10.1073/pnas.1721049115

[39]. Youn HD, Chatila TA, Liu JO. Integration of calcineurin and MEF2 signals by the coactivator p300 during T-cell apoptosis. *The EMBO journal*. 2000 **19:** 4323-4331. https://doi.org/10.1093/emboj/19.16.4323

[40]. Youn HD, Liu JO. Cabin1 represses MEF2-dependent Nur77 expression and T cell apoptosis by controlling association of histone deacetylases and acetylases with MEF2. *Immunity*. 2000 **13:** 85-94. https://doi.org/10.1016/s1074-7613(00)00010-8

514 [41]. Palmer E. Negative selection--clearing out the bad apples from
515 the T-cell repertoire. *Nat Rev Immunol*. 2003 **3:** 383-391.
516 https://doi.org/10.1038/nri1085
517 [42]. Raza K, Scheel-Toellner D, Lee CY*, et al.* Synovial fluid leukocyte
518 apoptosis is inhibited in patients with very early rheumatoid arthritis.
519 *Arthritis Res Ther*. 2006 **8:** R120. https://doi.org/10.1186/ar2009
520 [43]. Greenberg SA. Inclusion body myositis: clinical features and
521 pathogenesis. *Nat Rev Rheumatol*. 2019 **15:** 257-272.
522 https://doi.org/10.1038/s41584-019-0186-x
523 [44]. Nakka K, Ghigna C, Gabellini D, Dilworth FJ. Diversification of the
524 muscle proteome through alternative splicing. *Skeletal Muscle*. 2018 **8:** 8.
525 https://doi.org/10.1186/s13395-018-0152-3
526 [45]. Yu X-Z, Levin SD, Madrenas J, Anasetti C. Lck Is Required for
527 Activation-Induced T Cell Death after TCR Ligation with Partial Agonists.
528 *The Journal of Immunology*. 2004 **172:** 1437-1443.
529 https://doi.org/10.4049/jimmunol.172.3.1437
530 [46]. McNeill L, Salmond RJ, Cooper JC*, et al.* The differential
531 regulation of Lck kinase phosphorylation sites by CD45 is critical for T cell
532 receptor signaling responses. *Immunity*. 2007 **27:** 425-437.
533 https://doi.org/10.1016/j.immuni.2007.07.015
534 [47]. Thompson JL, Shuttleworth TJ. A plasma membrane-targeted
535 cytosolic domain of STIM1 selectively activates ARC channels, an
536 arachidonate-regulated store-independent Orai channel. *Channels (Austin,*
537 *Tex)*. 2012 **6:** 370-378. https://doi.org/10.4161/chan.21947
538 [48]. Cortese A, Plagnol V, Brady S*, et al.* Widespread RNA metabolism
539 impairment in sporadic inclusion body myositis TDP43-proteinopathy.
540 *Neurobiol Aging*. 2014 **35:** 1491-1498.
541 https://doi.org/10.1016/j.neurobiolaging.2013.12.029
542 [49]. Weihl CC, Baloh RH, Lee Y*, et al.* Targeted sequencing and
543 identification of genetic variants in sporadic inclusion body myositis.
544 *Neuromuscul Disord*. 2015 **25:** 289-296.
545 https://doi.org/10.1016/j.nmd.2014.12.009
546 [50]. Gang Q, Bettencourt C, Machado PM*, et al.* Rare variants in
547 SQSTM1 and VCP genes and risk of sporadic inclusion body myositis.

548 *Neurobiol Aging*. 2016 **47:** 218 e211-218 e219.
549 https://doi.org/10.1016/j.neurobiolaging.2016.07.024

550

Figures legends

Fig. 1 a) Workflow and methodology used in this study. b) Principal component analysis of gene expression results showing the pairwise comparison between different groups: IBM, TMD and Amputees. c) IBM-specific differentially expressed genes were determined by comparing IBM cases with amputee and TMD groups. d) Comparison between IBM specific differentially expressed genes (cyan) and IBM specific differentially spliced genes (magenta).

Fig. 2 a) Top 15 differentially expressed genes specific to IBM muscles. Log_2 fold change (log2FC) of IBM versus amputees calculated by DEseq2 after shrinkage estimations. '+'/'-' sign denotes the direction of change, i.e., positive log2FC values indicate overexpressed genes in IBM muscles, and negative log2FC values indicate underexpressed genes in IBM muscles. The p-value of significance and adjusted p-value using the Benjamini-Hochberg corrections and associated GO terms are shown for each gene. Genes marked with * are also observed as significantly dysregulated in Hamann et al. [13] b) Normalized gene expression in the different cohorts is presented as boxplots. Median and quartile values are shown, with whiskers reaching up to 1.5 times the interquartile range. Individual expression levels are shown as jitter points. The raincloud plots illustrate the distribution of data in each cohort. The scaled Y-axis shows log normalized counts.

Fig. 3 The calcium-induced T lymphocyte apoptosis pathway with gene expression changes observed in IBM compared to groups. Created with BioRender.com

Fig. 4 Statistical over-representation tests were performed on a list of differentially spliced RNAs, using clusterProfiler for a) Biological Processes, b) Cellular component, and c) Molecular function. d) An UpSet plot is shown comparing six different sets, namely, IBM specific differentially spliced (1,271 genes), mobilization of Ca^{2+} (80 genes), calcium-induced T lymphocyte apoptosis (69 genes), the flux of Ca^{2+} (51 genes), quantity of Ca^{2+} (51 genes), and release of Ca^{2+} (33 genes). Dots and lines represent subsets of different lists. The horizontal bar graph (wine color) represents the size of each set, while the vertical histogram (black) represents the number of RNAs in each subset. The 10 RNAs that are both differentially expressed and differentially spliced are shown with a red circle with their gene names (black).

Fig. 5 a) Normalized LCK expression in the different cohorts (as explained in Fig. 2b). b) Altered isoform expression of *LCK* using JunctionSeq showing estimated normalized mean read-pair count for each exon and splice junctions in the different cohorts (left) as well as for the whole *LCK* gene (right). The significantly alternatively spliced feature, E016 (pink), corresponds to chr1:32274818-32274992 (GRCh38). The alternative *LCK* transcripts used in the JunctionSeq analysis are shown below with their corresponding ENSEMBL identifiers.

Tables

Table 1: Top 10 dysregulated canonical pathways identified by IPA. The significance of the identified pathway is shown with a p-value and the number of differentially expressed genes observed in the IBM-specific dataset compared to the number of genes present in the database for each pathway.

Ingenuity Canonical Pathways	p-value	Number of genes in dataset / Number of genes in database
Dendritic Cell Maturation	5.72E-31	106/357
T Cell Receptor Signaling	2.30E-25	97/355
T Cell Exhaustion Signaling Pathway	3.12E-25	94/338
Cdc42 Signaling	7.63E-24	88/315
iCOS-iCOSL Signaling in T Helper Cells	4.10E-23	81/280
CD28 Signaling in T Helper Cells	1.13E-22	82/290
OX40 Signaling Pathway	4.03E-21	68/222
Calcium-induced T Lymphocyte Apoptosis	1.29E-20	69/232
Nur77 Signaling in T Lymphocytes	3.07E-20	72/253
Role of NFAT in Regulation of the Immune Response	4.57E-19	87/360

Table 2: From the miRNA analysis, upstream binding partners are shown along with their target miRNA. A p-value and associated GO terms are shown.

Upstream regulator	p-value	GO terms and annotations
AGO2	7.89E-23	RNA polymerase II complex binding
SSB	4.74E-19	RNA binding
TP53	7.59E-09	Transcription regulatory region sequence-specific DNA binding
RNA polymerase III	4.09E-06	Synthesis of small RNA, RNA polymerase activity

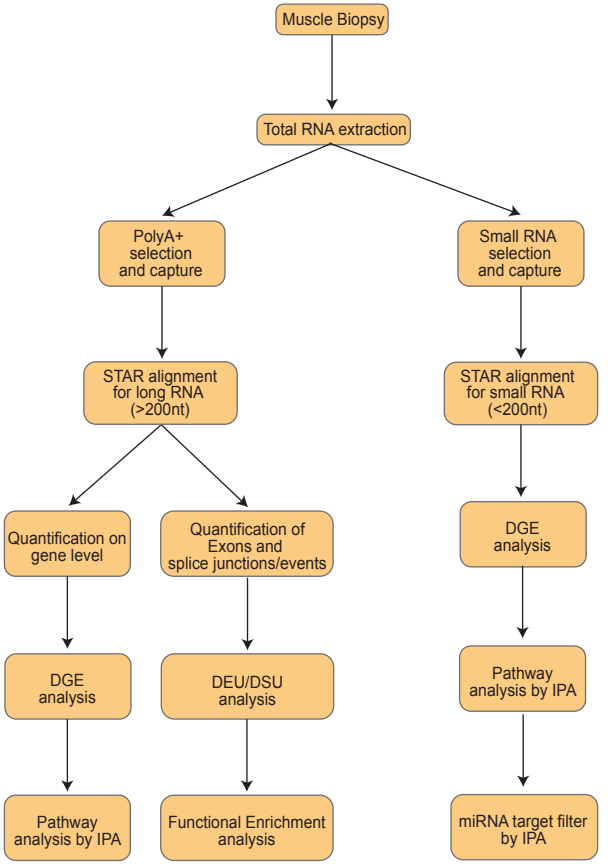
Table 3: From the long non-coding RNA analysis, upstream binding partners are shown along with their target lncRNA. A p-value and associated GO terms are shown

Upstream regulator	Target molecule in dataset	p-value	GO terms and annotations
JDP2	*DNM3OS*	4.10E-03	DNA-binding transcription factor activity, RNA polymerase II-specific
miR-338-3p	*NR2F1-AS1*	4.10E-03	Negative regulation of gene expression; negative regulation of IL-6 production; negative regulation of cytokine production involved in inflammatory response
miR-150-5p	*MIAT*	4.10E-03	mRNA binding involved in posttranscriptional gene silencing
TARDBP	*MIAT*	2.03E-02	RNA polymerase II cis-regulatory region sequence-specific DNA binding
mir-150	*MIAT*	2.63E-02	mRNA binding involved in posttranscriptional gene silencing
PGF	*DNM3OS*	3.43E-02	Protein binding, signal transduction
FUS	*RMRP*	3.43E-02	mRNA binding, mRNA stabilization
DDX58	*EGOT*	4.61E-02	double-stranded RNA binding

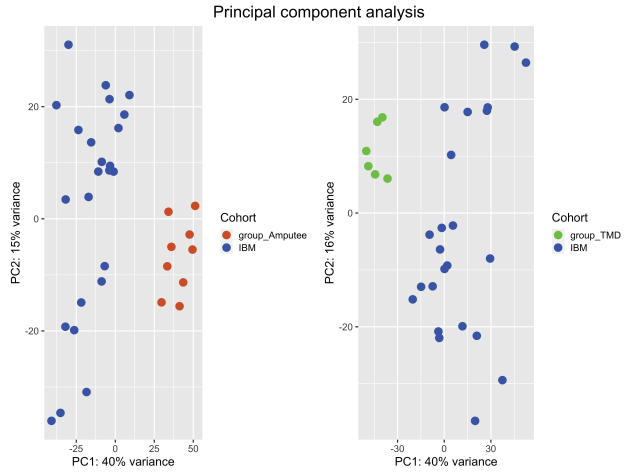
Table 4: In cell signaling processes, different pathways associated with calcium homeostasis are shown along with their p-value and a prediction state.

Functional annotations	p-value	Predicted activation state
Mobilization of Ca^{2+}	5.04E-26	Increased
Flux of Ca^{2+}	4.68E-13	Increased
Quantity of Ca^{2+}	4.34E-09	Increased
Release of Ca^{2+}	5.00E-09	Increased

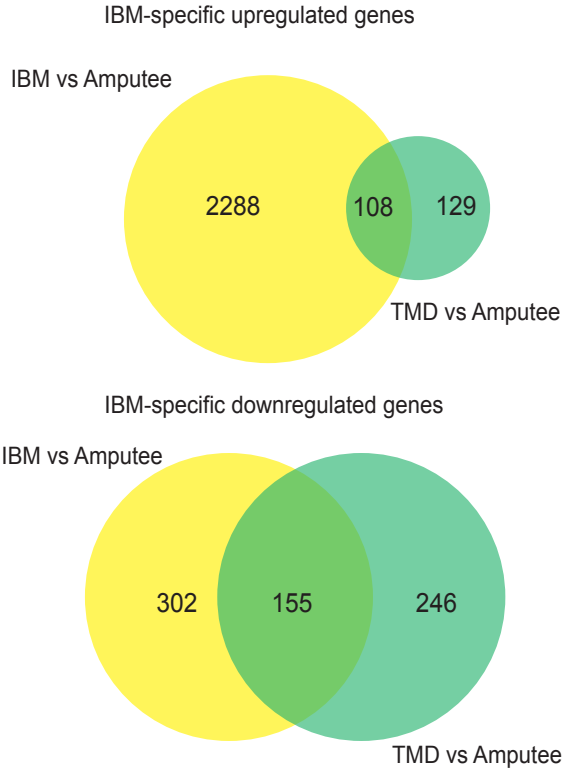
a


Muscle Biopsy
Total RNA extraction
PolyA+ selection and capture
Small RNA selection and capture
STAR alignment for long RNA (>200nt)
STAR alignment for small RNA (<200nt)
Quantification on gene level
Quantification of Exons and splice junctions/events
DGE analysis
DGE analysis
DEU/DSU analysis
Pathway analysis by IPA
Pathway analysis by IPA
Functional Enrichment analysis
miRNA target filter by IPA

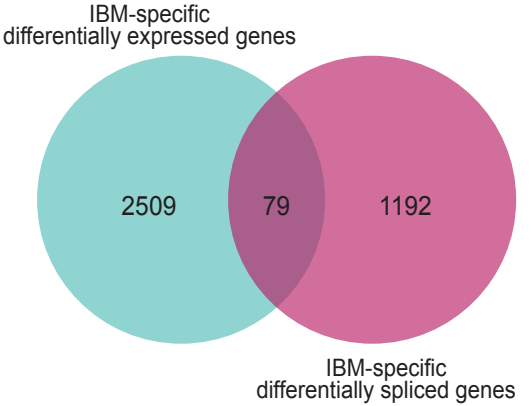
b


Principal component analysis
PC2: 15% variance
PC1: 40% variance
Cohort
group_Amputee
IBM
PC2: 16% variance
PC1: 40% variance
Cohort
group_TMD
IBM

c


IBM-specific upregulated genes
IBM vs Amputee
2288
108
129
TMD vs Amputee
IBM-specific downregulated genes
IBM vs Amputee
302
155
246
TMD vs Amputee

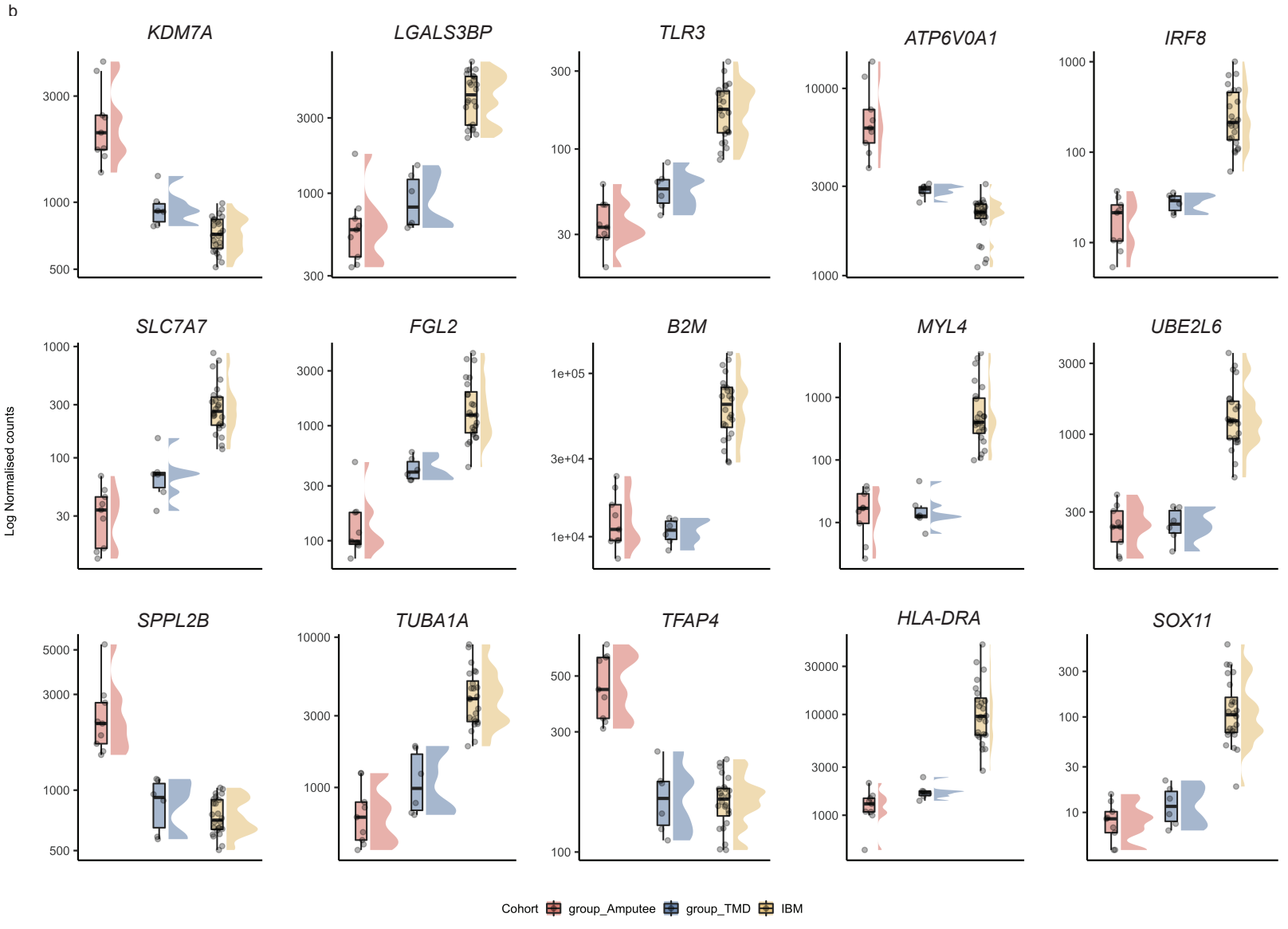
d

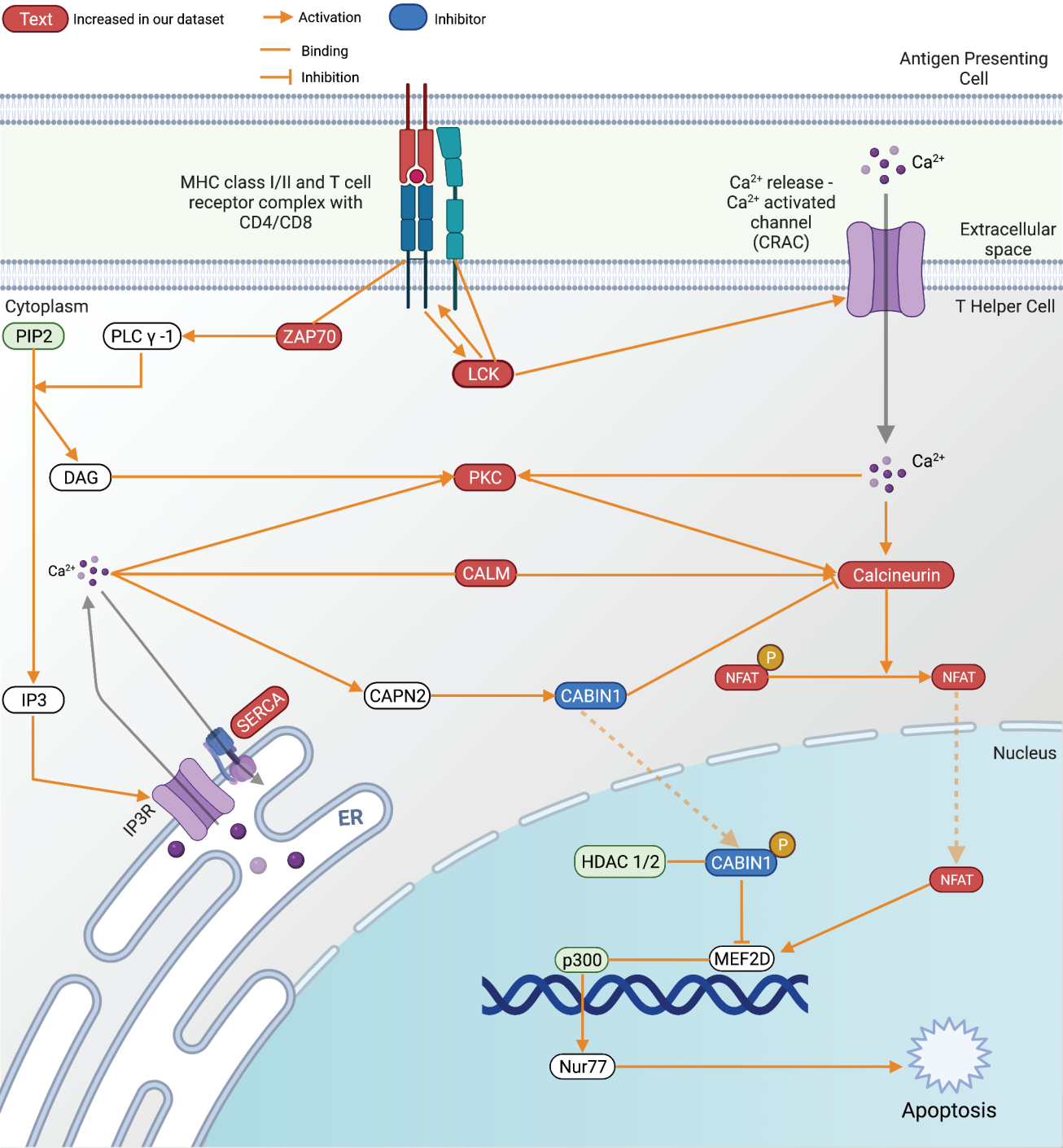

IBM-specific differentially expressed genes
2509
79
1192
IBM-specific differentially spliced genes

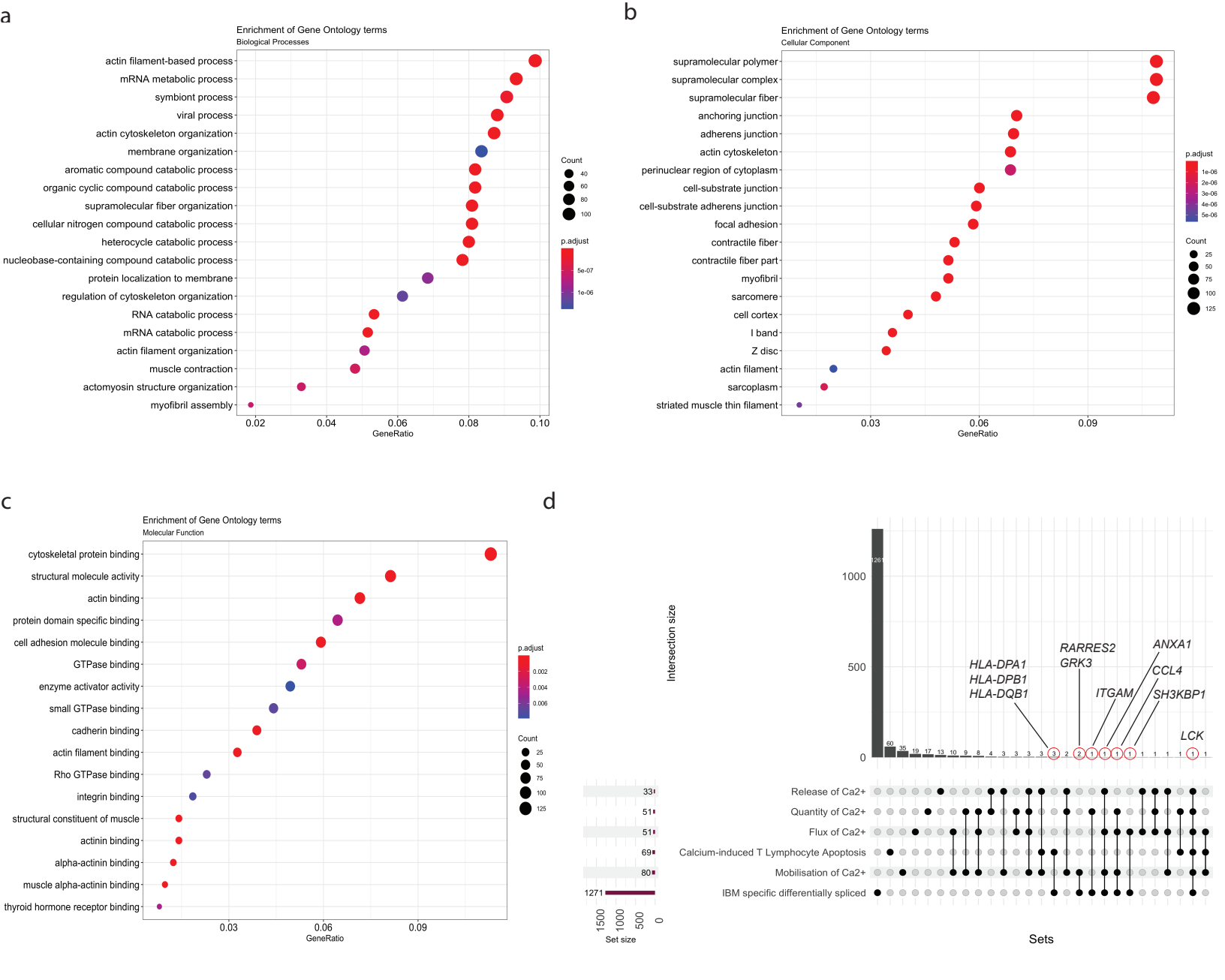
a

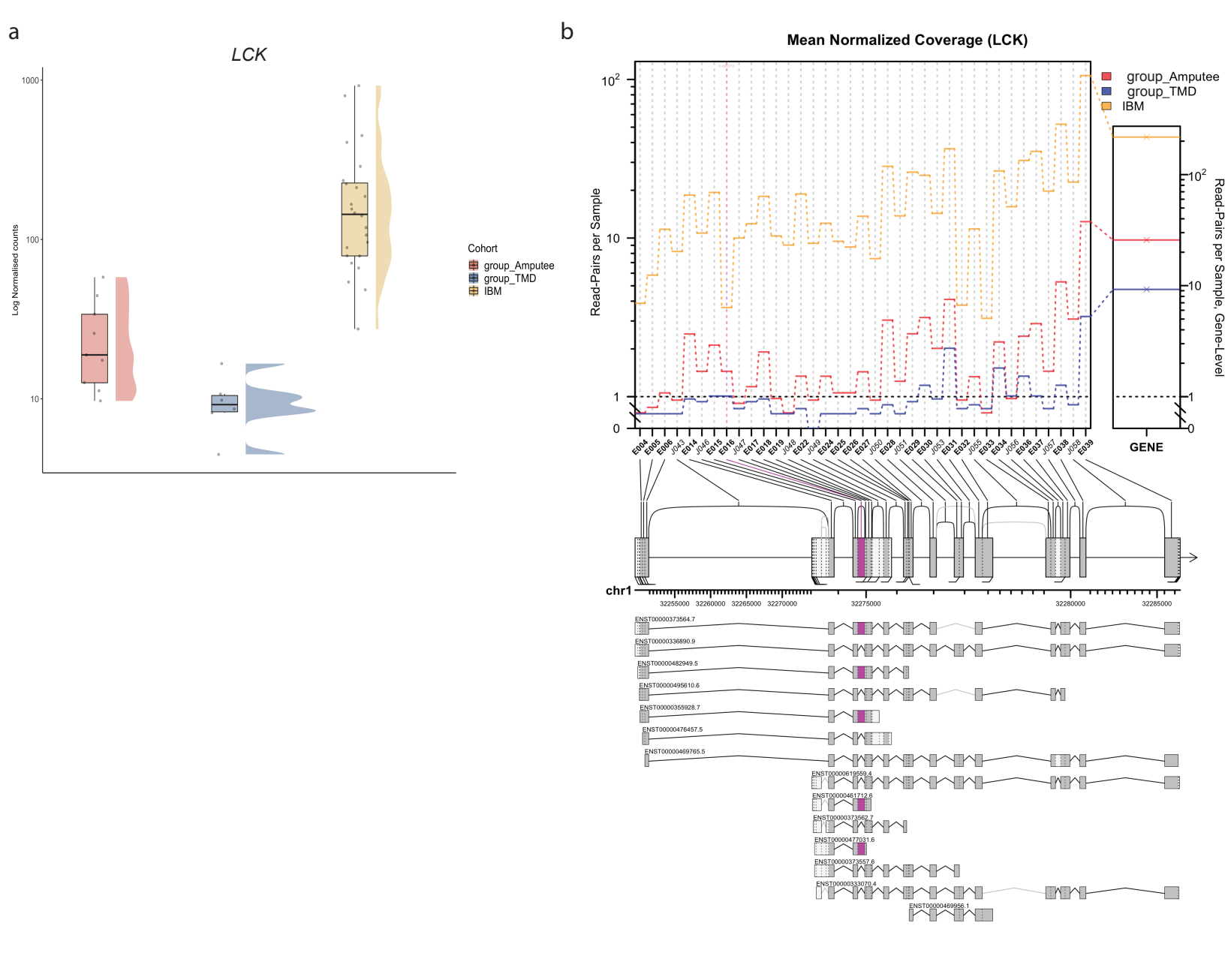
Symbol	Mean count	log2FC	pvalue	padj	GO terms
KDM7A	1147.06	-1.69	4.07E-32	2.39E-28	Iron ion binding and oxidoreductase activity
LGALS3BP*	2922.28	2.59	5.91E-31	2.89E-27	Scavenger receptor activity
TLR3	128.42	2.29	3.98E-28	1.30E-24	Regulation of dendritic cell cytokine production
ATP6V0A1	3393.79	-1.76	3.01E-27	8.83E-24	ATPase binding and proton-transporting ATPase activity
IRF8*	199.90	3.89	4.66E-27	1.24E-23	DNA-binding transcription factor activity, RNA polymerase II-specific
SLC7A7	216.91	3.16	2.08E-26	5.09E-23	Amino acid transmembrane transporter activity
FGL2	1102.10	3.27	3.14E-26	6.58E-23	T cell activation via T cell receptor contact with antigen bound to MHC molecule on antigen presenting cell
B2M*	46381.30	2.29	4.71E-26	9.22E-23	Positive regulation of T cell mediated cytotoxicity
MYL4*	610.22	5.44	5.92E-26	1.09E-22	Actin monomer binding, Calcium ion binding
UBE2L6*	986.03	2.47	1.86E-25	3.04E-22	Ubiquitin-protein transferase activity
SPPL2B	1161.90	-1.66	2.99E-25	4.39E-22	Protein homodimerization activity
TUBA1A*	2951.15	2.50	4.36E-25	5.81E-22	GTPase activity, structural molecule activity
TFAP4	234.41	-1.52	4.32E-25	5.81E-22	Transcription regulatory region sequence-specific DNA binding
HLA-DRA	8658.66	3.25	7.29E-25	9.00E-22	Antigen processing and presentation of endogenous peptide antigen via MHC class II
SOX11	96.02	4.03	1.97E-24	2.23E-21	DNA-binding transcription factor activity, RNA polymerase II-specific

b




Text
Increased in our dataset
Activation
Inhibitor
Binding
Inhibition
Antigen Presenting Cell
MHC class I/II and T cell receptor complex with CD4/CD8
Ca2+ release - Ca2+ activated channel (CRAC)
Ca2+
Extracellular space
T Helper Cell
Cytoplasm
PIP2
PLC γ -1
ZAP70
LCK
DAG
PKC
Ca2+
Ca2+
CALM
Calcineurin
IP3
SERCA
CAPN2
CABIN1
NFAT
P
NFAT
Nucleus
IP3R
ER
HDAC 1/2
CABIN1
P
NFAT
p300
MEF2D
Nur77
Apoptosis


a
Enrichment of Gene Ontology terms
Biological Processes
actin filament-based process
mRNA metabolic process
symbiont process
viral process
actin cytoskeleton organization
membrane organization
aromatic compound catabolic process
organic cyclic compound catabolic process
supramolecular fiber organization
cellular nitrogen compound catabolic process
heterocycle catabolic process
nucleobase-containing compound catabolic process
protein localization to membrane
regulation of cytoskeleton organization
RNA catabolic process
mRNA catabolic process
actin filament organization
muscle contraction
actomyosin structure organization
myofibril assembly
0.02
0.04
0.06
0.08
0.10
GeneRatio
Count
40
60
80
100
p.adjust
5e-07
1e-06
b
Enrichment of Gene Ontology terms
Cellular Component
supramolecular polymer
supramolecular complex
supramolecular fiber
anchoring junction
adherens junction
actin cytoskeleton
perinuclear region of cytoplasm
cell-substrate junction
cell-substrate adherens junction
focal adhesion
contractile fiber
contractile fiber part
myofibril
sarcomere
cell cortex
I band
Z disc
actin filament
sarcoplasm
striated muscle thin filament
0.03
0.06
0.09
GeneRatio
p.adjust
1e-06
2e-06
3e-06
4e-06
5e-06
Count
25
50
75
100
125
c
Enrichment of Gene Ontology terms
Molecular Function
cytoskeletal protein binding
structural molecule activity
actin binding
protein domain specific binding
cell adhesion molecule binding
GTPase binding
enzyme activator activity
small GTPase binding
cadherin binding
actin filament binding
Rho GTPase binding
integrin binding
structural constituent of muscle
actinin binding
alpha-actinin binding
muscle alpha-actinin binding
thyroid hormone receptor binding
0.03
0.06
0.09
GeneRatio
p.adjust
0.002
0.004
0.006
Count
25
50
75
100
125
d
Intersection size
1000
500
0
1261
60
35
19
17
13
10
9
8
4
3
3
3
3
2
2
1
1
1
1
1
1
1
1
1
HLA-DPA1
HLA-DPB1
HLA-DQB1
RARRES2
GRK3
ITGAM
ANXA1
CCL4
SH3KBP1
LCK
Release of Ca2+
Quantity of Ca2+
Flux of Ca2+
Calcium-induced T Lymphocyte Apoptosis
Mobilisation of Ca2+
IBM specific differentially spliced
33
51
51
69
80
1271
1500
1000
500
0
Set size
Sets


a
LCK
Log Normalised counts
1000
100
10
Cohort
group_Amputee
group_TMD
IBM
b
Mean Normalized Coverage (LCK)
Read-Pairs per Sample
Read-Pairs per Sample, Gene-Level
group_Amputee
group_TMD
IBM
GENE
chr1
32255000
32260000
32265000
32270000
32275000
32280000
32285000
ENST00000373564.7
ENST00000336890.9
ENST00000482949.5
ENST00000495610.6
ENST00000355928.7
ENST00000476457.5
ENST00000469765.5
ENST00000619559.4
ENST00000461712.6
ENST00000373562.7
ENST00000477031.6
ENST00000373557.6
ENST00000333070.4
ENST00000469956.1