



OPEN

## Comprehensive multi-cohort transcriptional meta-analysis of muscle diseases identifies a signature of disease severity

C. J. Walsh<sup>1,2</sup>, J. Batt<sup>1,2</sup>, M. S. Herridge<sup>4</sup>, S. Mathur<sup>5</sup>, G. D. Bader<sup>6</sup>, P. Hu<sup>7</sup>, P. Khatri<sup>8,9</sup> & C. C. dos Santos<sup>1,3</sup>✉

Muscle diseases share common pathological features suggesting common underlying mechanisms. We hypothesized there is a common set of genes dysregulated across muscle diseases compared to healthy muscle and that these genes correlate with severity of muscle disease. We performed meta-analysis of transcriptional profiles of muscle biopsies from human muscle diseases and healthy controls. Studies obtained from public microarray repositories fulfilling quality criteria were divided into six categories: (i) immobility, (ii) inflammatory myopathies, (iii) intensive care unit (ICU) acquired weakness (ICUAW), (iv) congenital muscle diseases, (v) chronic systemic diseases, (vi) motor neuron disease. Patient cohorts were separated in discovery and validation cohorts retaining roughly equal proportions of samples for the disease categories. To remove bias towards a specific muscle disease category we repeated the meta-analysis five times by removing data sets corresponding to one muscle disease class at a time in a “leave-one-disease-out” analysis. We used 636 muscle tissue samples from 30 independent cohorts to identify a 52 gene signature (36 up-regulated and 16 down-regulated genes). We validated the discriminatory power of this signature in 657 muscle biopsies from 12 additional patient cohorts encompassing five categories of muscle diseases with an area under the receiver operating characteristic curve of 0.91, 83% sensitivity, and 85.3% specificity. The expression score of the gene signature inversely correlated with quadriceps muscle mass ( $r = -0.50$ ,  $p\text{-value} = 0.011$ ) in ICUAW and shoulder abduction strength ( $r = -0.77$ ,  $p\text{-value} = 0.014$ ) in amyotrophic lateral sclerosis (ALS). The signature also positively correlated with histologic assessment of muscle atrophy in ALS ( $r = 0.88$ ,  $p\text{-value} = 1.62 \times 10^{-3}$ ) and fibrosis in muscular dystrophy (Jonckheere trend test  $p\text{-value} = 4.45 \times 10^{-9}$ ). Our results identify a conserved transcriptional signature associated with clinical and histologic muscle disease severity. Several genes in this conserved signature have not been previously associated with muscle disease severity.

Skeletal muscle diseases result in decreased muscle mass and muscle dysfunction thereby inducing physical disability and increased mortality<sup>1</sup>. Skeletal muscle dysfunction has been shown to contribute to decreased quality of life, increased disease morbidity and mortality in respiratory illness including for example, chronic obstructive pulmonary disease (COPD), pulmonary arterial hypertension and acute respiratory distress syndrome (ARDS)<sup>2–4</sup>. Despite the profound clinical implications, knowledge of the molecular mechanisms of muscle dysfunction, as well as objective, non-volitional methods to quantify the degree of muscle dysfunction are insufficient<sup>5,6</sup>. Understanding the pathomolecular mechanisms conserved across muscle diseases may provide vital insight to help develop therapies to ameliorate them.

<sup>1</sup>Keenan Research Center for Biomedical Science, Saint Michael’s Hospital, Toronto, ON, Canada. <sup>2</sup>Institute of Medical Sciences, University of Toronto, Toronto, ON, Canada. <sup>3</sup>Interdepartmental Division of Critical Care, University of Toronto, Toronto, ON, Canada. <sup>4</sup>Interdepartmental Division of Critical Care, University Health Network, University of Toronto, Toronto, ON, Canada. <sup>5</sup>Department of Physical Therapy, University of Toronto, Toronto, ON, Canada. <sup>6</sup>The Donnelly Center, University of Toronto, Toronto, ON, Canada. <sup>7</sup>Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, MB, Canada. <sup>8</sup>Stanford Institute for Immunity, Transplantation and Infection (ITI), Stanford University School of Medicine, Stanford, CA, USA. <sup>9</sup>Department of Medicine, Stanford Center for Biomedical Informatics Research (BMIR), Stanford University, Stanford, CA, USA. ✉email: dosSantosC@smh.ca

A growing number of studies of human muscle disease have identified dysregulated gene expression that is associated with disease severity<sup>1,7–9</sup>. These studies however, are usually limited by relatively small sample sizes without external validation from independent cohorts<sup>10</sup>. Moreover, individually these studies are not representative of biological and clinical heterogeneity observed in the real-world patient population, which substantially limits their generalizability. The vast quantity of expression profiling data in the public repositories Gene Expression Omnibus (GEO) and ArrayExpress represents novel opportunities to address these challenges by facilitating comprehensive integration of human muscle disease cohorts for meta-analysis.

We applied a multi-cohort analysis framework<sup>11,12</sup> that leverages the biological, clinical, and technical heterogeneity across independent data sets to identify a reproducible disease gene signature<sup>13,14</sup>. This approach has discovered robust signatures in organ rejection<sup>13</sup>, neurodegenerative diseases<sup>14</sup>, sepsis<sup>15</sup>, tuberculosis<sup>16</sup>, viral infections<sup>17</sup>, vaccination<sup>18</sup>, and systemic sclerosis<sup>19</sup>, many of which have been successfully validated in prospective independent cohorts<sup>20–22</sup>. We hypothesized that convergent transcriptional abnormalities occur across muscle diseases regardless of the specific muscle pathophysiology and that the relative expression of these genes is associated with the degree of muscle dysfunction. To the best of our knowledge, this is the largest systematic multi-cohort analysis investigating transcriptional changes across multiple human muscle diseases.

We identified a conserved gene signature across five muscle disease categories including muscular dystrophies, inflammatory myopathies, critical illness myopathy, and chronic systemic diseases associated with muscle dysfunction such as chronic obstructive pulmonary disorder (COPD). Importantly, we validated the discriminatory power of this signature in other diseases with muscle phenotypes that were not part of the discovery meta-analysis, cerebral palsy (CP) and amyotrophic lateral sclerosis (ALS). We found that the common muscle disease gene signature is significantly associated with clinical and histological disease severity in independent validation cohorts. Finally, we identified patterns of gene dysregulation unique to each muscle disease category relative to the others.

Portions of this manuscript, including the methods section, have been presented previously reported in a PhD thesis by the first author<sup>1</sup>. The present study has an expanded number of patient cohorts that were not included in the PhD thesis.

## Results

**Meta-analysis identifies a common gene signature of muscle diseases.** A total of 45 independent patient cohorts that profiled human muscle diseases and normal muscle controls (862 cases, 512 controls), comprising 1374 samples met criteria for inclusion (Supplementary Fig. 1, Table 1). Collectively, the cohorts represent a broad range of patient ages and peripheral muscles from both upper and lower extremities. Available phenotypic data for patient samples included in public repositories is shown in Supplementary Table 1 and summary descriptions of each study are found in Supplementary Document 1.

For the discovery cohort we ensured that there were at least three cohorts for each disease category that met our inclusion criteria. As there were only two cohorts for the MND category, this was not included in the discovery cohort as a disease category; instead, the two MND cohorts were included in the secondary validation cohort.

We chose smaller patient cohorts (< 30 samples) for the discovery meta-analysis and reserved larger patient cohorts and/or cohorts with clinical measures of muscle mass or strength or histologic assessments for validation analysis. For the discovery meta-analysis, 30 patient cohorts (348 cases, 288 controls) containing at least three cohorts from each of the five muscle disease categories were analyzed.

To identify the most robust differentially expressed (DE) genes across muscle diseases measured on multiple different microarray platforms we performed gene expression meta-analysis<sup>10,11</sup> using a “leave-one-disease-out” strategy to correct for heterogeneity in genes DE between muscle disease categories and to avoid one muscle disease influencing the overall analysis, as described before<sup>13,14</sup>. We identified 209 genes that remained significantly DE in all 5 iterations of the “leave-one-disease-out” analysis (Supplementary Table 2).

We then applied an iterative greedy forward search<sup>15</sup> to the 209 genes and identified a set of 52 genes (36 up-regulated, 16 down-regulated) that was optimized for discriminatory power termed the Common Muscle disease Module (CMDM). As expected, the 52-gene CMDM score distinguished muscle disease from healthy controls with summary area under the curve (AUC) = 0.91 (95% confidence interval [CI] 0.83–0.96) in the discovery cohorts (Fig. 1A,B).

Next, we tested the CMDM 52-gene signature in the validation dataset of 12 cohorts (N = 657 total samples; Table 1). The validation set included at least one cohort belonging to each of the five muscle disease categories. (Table 2; Supplementary Table 3). The CMDM accurately identified muscle disease samples in most cohorts (summary AUC = 0.91 [95% CI 0.77–0.97]) (Fig. 2A).

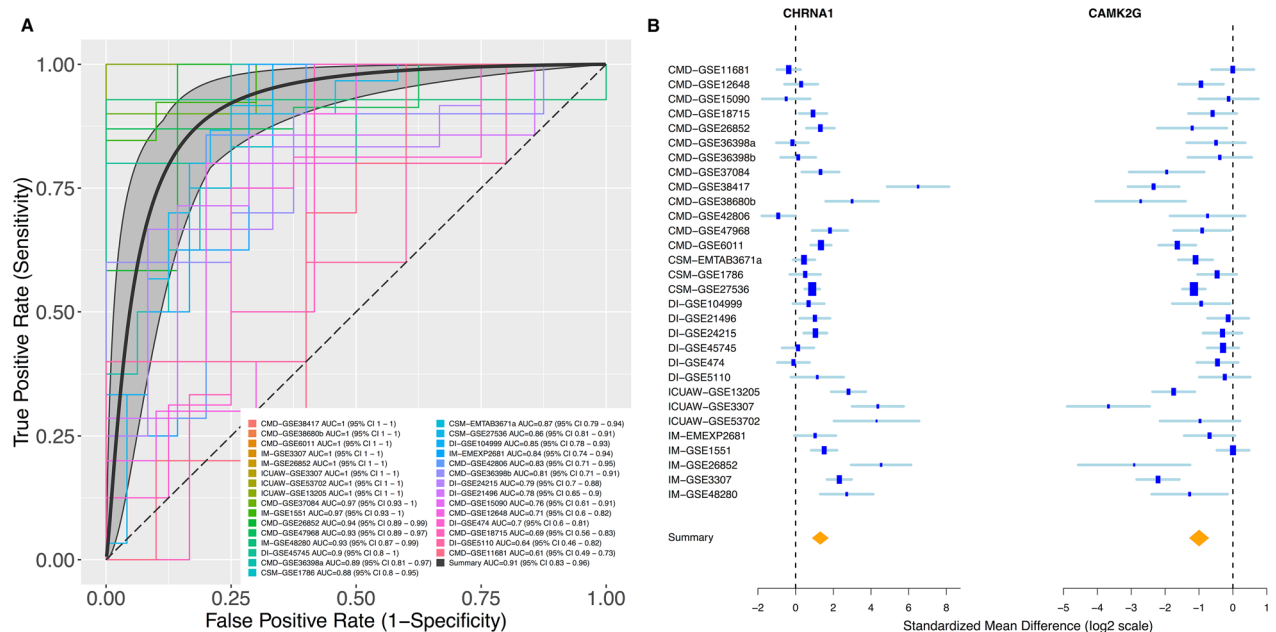
An additional three cohorts (two cohorts from amyotrophic lateral sclerosis [ALS] and one cohort from cerebral palsy [CP]) could not be classified into any of the five muscle disease categories present in the discovery and validation sets. Therefore we tested these three cohorts (N = 81 total samples; Table 1) as a secondary validation set to assess the generalizability of the CMDM. The CMDM accurately identified muscle disease samples in two of the three cohorts (ALS-GSE3307 and GSE31243); summary AUC of all three cohorts = 0.92 [95% CI 0.54–0.99]) (Fig. 2B).

Visual inspection of the heatmap of the 52 CMDM genes in Fig. 2C shows the pattern of expression is generally highly consistent between the discovery and validation set, as well as the secondary validation set, further supporting the generalizability of the CMDM to muscle diseases. One notable exception was the cohort GSE13608 in the validation analysis with AUC 0.23 (95% CI 0.14–0.32), which was reflected in the heatmap showing gene expression opposite to the majority of genes across the meta-analysis.

Disease category	Accession#	Reference	Cases	n cases	n control	Total samples	Platform
<b>Discovery</b>							
ICUAW	GSE13205	Fredriksson <sup>49</sup>	Sepsis MODS	13	8	21	GPL570
ICUAW	GSE53702	Langhans <sup>50</sup>	ICUAW	7	6	13	GPL5188
ICUAW	GSE3307	Bakay <sup>51</sup>	ICUAW	5	13 <sup>c</sup>	18 <sup>c</sup>	GPL96
Congenital	GSE15090	Arashiro <sup>52</sup>	FSHD	5	5	10	GPL570
Congenital	GSE18715	Voets <sup>a</sup>	POLG1	6	12	18	GPL570
Congenital	GSE36398a <sup>b</sup>	Rahimov <sup>53</sup>	FSHD	8	16	24	GPL6244
Congenital	GSE36398b	Rahimov <sup>53</sup>	FSHD	10	8	18	GPL6244
Congenital	GSE37084	Perfetti <sup>54</sup>	MMD	10	10	20	GPL5175
Congenital	GSE26852	Tasca <sup>55</sup>	FSHD, dysferlinopathy	12	7	19	GPL6947
Congenital	GSE47968	Nakamori <sup>56</sup>	FSHD, DM	23	8	31	GPL5188
Congenital	GSE42806	Screen <sup>57</sup>	TMD	7	5	12	GPL570
Congenital	GSE38417	Dorsey <sup>a</sup>	DMD	16	6	22	GPL570
Congenital	GSE38680b	Palermo <sup>58</sup>	GSD II	9	10	19	GPL570
Congenital	GSE11681	Saenz <sup>59</sup>	LGMD2A	10	10	20	GPL96
Congenital	GSE12648	Eisenberg <sup>60</sup>	HIBM	10	10	20	GPL96
Congenital	GSE6011	Pescatori <sup>61</sup>	DMD	23	14	37	GPL96
IM	GSE48280	Surez-Calvet <sup>62</sup>	PM, IBM, DM	14	5	19	GPL6244
IM	GSE3307	Bakay <sup>51</sup>	Juvenile DM	21	13 <sup>c</sup>	34 <sup>c</sup>	GPL96
IM	GSE1551	Greenberg <sup>63</sup>	DM	13	10	23	GPL96
IM	GSE26852	Tasca <sup>55</sup>	PM, IM, DM	7	7	14	GPL6947
IM	EMEXP2681	Bernasconi <sup>a</sup>	DM, PM	8	7	15	GPL96
Immobility	GSE45745	Barres <sup>64</sup>	Morbid obesity	5	6	11	GPL13667
Immobility	GSE21496	Reich <sup>65</sup>	Unloading	7	7	14	GPL570
Immobility	GSE5110	Urso <sup>66</sup>	Immobility	5	5	10	GPL570
Immobility	GSE24215	Alibegovic <sup>67</sup>	Immobility	12	12	24	GPL6480
Immobility	GSE104999	Rullman <sup>68</sup>	Immobility	12	12	24	GPL17692
Immobility	GSE474	Park <sup>69</sup>	Morbid obesity	16	8	24	GPL96
Chronic	GSE27536	Turan <sup>70</sup>	COPD	30	24	54	GPL570
Chronic	GSE1786	Radom-Aizik <sup>71</sup>	COPD	12	12	24	GPL96
Chronic	EMTAB3671	Kreiner <sup>72</sup>	PMR	12	12	24	GPL570
Total				348	288	636	
<b>Validation</b>							
ICUAW	GSE78929	Walsh <sup>8</sup>	ICUAW	24	8	32	GPL10558
Congenital	GSE13608	Bachinski <sup>73</sup>	DMD, MMD	59	9	68	GPL570
Congenital	GSE38680a	Palermo <sup>58</sup>	GSD II	32	7	39	GPL570
Congenital	GSE109178	Dadgar <sup>9</sup>	MD	42	6	48	GPL570
Congenital	GSE3307	Bakay <sup>51</sup>	MD	66	13 <sup>c</sup>	79 <sup>c</sup>	GPL570
Congenital	GSE10760	Osborne <sup>74</sup>	FSHD	38	60	98	GPL96
IM	GSE3112	Greenberg <sup>75</sup>	PM, IBM	29	11	40	GPL96
IM	GSE39454	Zhu <sup>76</sup>	PM, IBM, NM	31	5	36	GPL570
Immobility	GSE14901	Abadi <sup>30</sup>	Limb disuse (casting)	48	24	72	GPL570
Immobility	GSE45462	Chen <sup>77</sup>	Limb disuse (casting)	16	16	32	GPL570
Chronic	GSE34111	Gallagher <sup>78</sup>	Cancer	12	6	30	GPL570
Chronic	GSE100281	Willis-Owen <sup>79</sup>	COPD	80	15	85	GPL11532
Total				477	180	657	
<b>Secondary validation</b>							
MND	EMEXP3260	Pradat <sup>7</sup>	ALS	9	10	19	GPL96
MND	GSE3307	Bakay <sup>51</sup>	ALS	9	13 <sup>c</sup>	22 <sup>c</sup>	GPL96
Continued							

Disease category	Accession#	Reference	Cases	n cases	n control	Total samples	Platform
CP	GSE31243	Smith <sup>80</sup>	CP	20	20	40	GPL570
Total				38	43	81	

**Table 1.** Summary of public gene expression-based discovery and validation data sets used in the meta-analysis. *IM* inflammatory myositides, *MMD* myotonic muscular dystrophy, *MD* muscular dystrophies, *DMD* Duchene's muscular dystrophy, *FSHD* fascioscapulohumeral muscular dystrophy, *LGMD2A* limb-girdle muscular dystrophy type 2A, *HIBM* heritable IBM, *POLG1*, *MD*, *TMD* tibial muscular dystrophy, *IBM* inclusion body myositis, *DM* diabetes mellitus, *GSD II* glycogen storage disease type II, also called Pompe disease, *POLG1* mitochondrial DNA polymerase  $\gamma$ , *ICUAW* intensive care unit acquired weakness, *MODS* multi-organ dysfunction syndrome. <sup>a</sup>Not published yet. <sup>b</sup>Deltoid muscle samples removed as FSHD typically affects biceps. <sup>c</sup>Same healthy controls used in subcohorts of GSE3307.



**Figure 1.** Discovery of the 52-gene signature expressed across human muscle diseases. **(A)** Meta-analysis and leave-one-disease-out analysis reveal common differentially expressed genes across muscle diseases. **(B)** Representative forest plots of most up-regulated (*CHRNA1*, left) and most down-regulated (*CAMK2II*, right) across muscle disease. The x axis represents standardized mean difference (Hedges'  $g$  in z-scaled log<sub>2</sub> values) between muscle disease and controls. Summary effect sizes for each gene across all cohorts are represented as a yellow diamond.

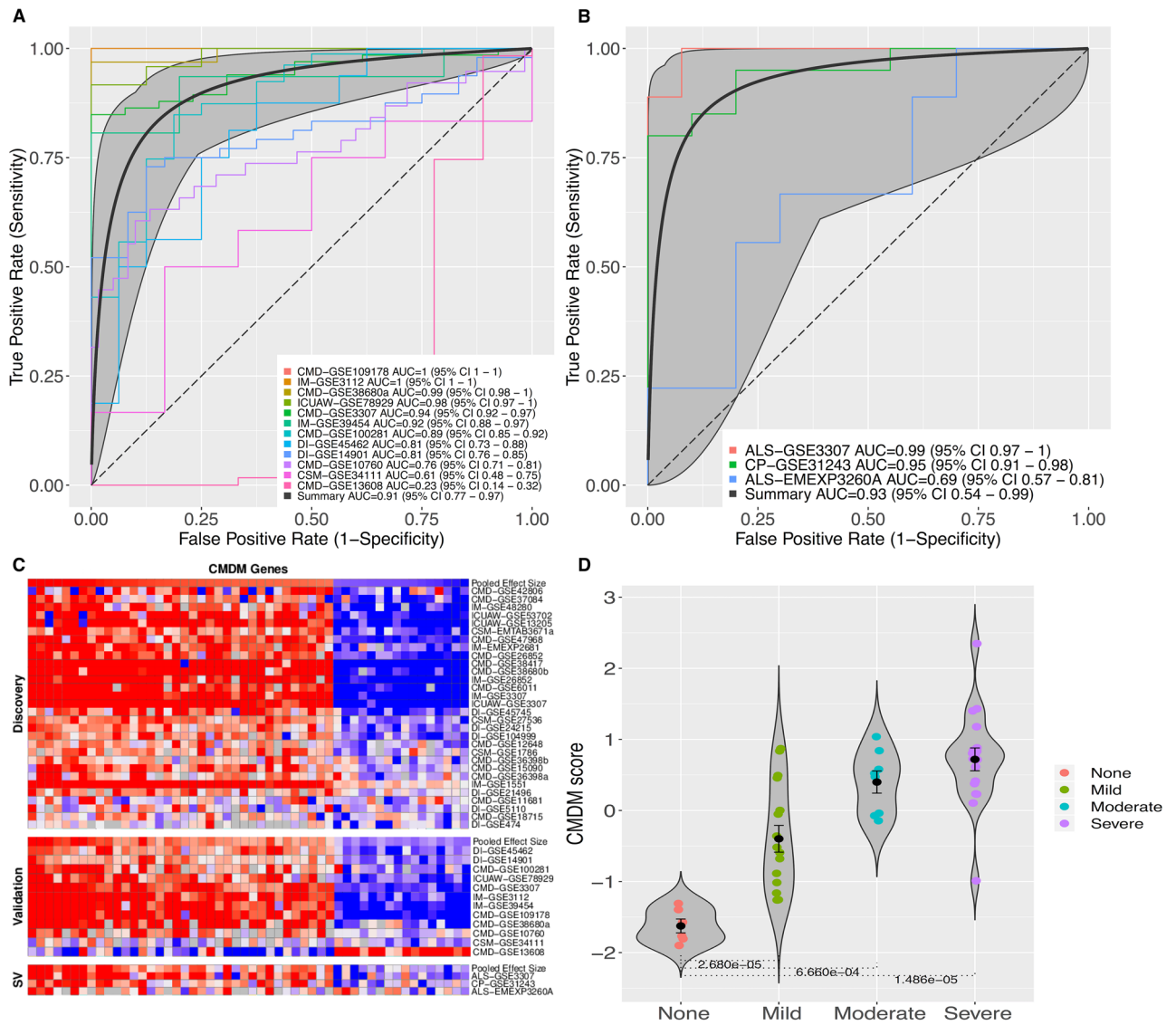
**CMDM score significantly associates with clinical and histological measures of disease severity.** When selecting differentially expressed genes using the multi-cohort analysis, we did not consider disease severity. Every sample was classified as either “control” or “case.” As muscle disease severity exists along a continuum, we hypothesized that the summary expression of the CMDM would correlate with the severity of muscle disease and clinical measures of muscle function. We calculated and then correlated CMDM scores for each cohort to measures of disease severity and extent, muscle mass, strength, and function, when this detail was provided.

Five of the cohorts reported disease severity scores and/or clinical measures of muscle mass, strength and function. The cohort GSE109178 assessed the degree of muscle fibrosis histologically in patients with dystrophic subtypes of CMD, and categorized cases into subgroups of normal, mild, moderate, or severe fibrosis (Fig. 2D, Supplementary Table 1). Cohorts GSE78929 (ICUAW), GSE34111 (CSM) and EMEXP3260 (ALS), reported strength, muscle mass and/or physical functional capacity and the ALS cohort was additionally classified as “early” vs “late” disease severity. We found CMDM summary expression scores were significantly correlated to the histologic measures of disease severity and clinical measures of muscle mass, strength and function (Table 3 and Fig. 3A–F).

**Meta-analysis highlights common mechanisms of muscle diseases.** We next sought to identify conserved pathways dysregulated across muscle diseases using meta-analysis. Gene Set Enrichment Analysis (GSEA) evaluated the enrichment of Gene Ontology (GO) terms in the complete ranked list of genes based on

Gene symbol	Summary effect size in discovery set	Summary effect size in validation set	Summary effect size in secondary validation set
CHRNA1	1.31	1.125	1.796
LGMN	1.27	1.14	1.114
MYH8	1.161	1.241	2.009
C1R	1.094	0.924	0.743
AKR1A1	1.041	0.879	2.603
CDKN1A	1.04	1.009	1.855
CILP	0.993	1.189	1.01
TNFRSF21	0.992	0.942	0.464
OSBPL8	0.981	0.827	1.309
KLHL2	0.961	0.894	2.101
TMEM208	0.946	0.47	0.83
TMEM87A	0.94	0.49	0.671
IFITM2	0.932	0.599	0.195
C3	0.892	1.083	0.801
DUSP22	0.88	0.635	0.07
DDOST	0.855	0.308	0.321
LETMD1	0.847	0.844	0.563
CETN2	0.835	0.712	0.792
GPX3	0.833	0.549	-0.252
ITPA	0.821	0.468	0.865
CLTC	0.808	0.633	0.894
SCPEP1	0.796	0.767	0.099
HEXA	0.789	0.54	-0.145
SAE1	0.787	0.549	1.234
CHI3L1	0.765	0.442	-0.44
USP3	0.76	0.386	0.619
HSP90B1	0.754	0.373	0.169
CKAP4	0.752	0.46	0.638
FST	0.747	0.923	1.546
NIP7	0.742	0.243	-0.076
PANX1	0.737	0.666	0.389
HEY1	0.737	0.37	0.828
TBC1D16	0.726	0.571	0.84
TRMT112	0.711	0.475	0.698
TPP1	0.698	0.586	-0.092
CERS2	0.673	0.158	0.793
ATP2A1	-0.696	-0.192	-1.703
CS	-0.71	-0.673	0.185
ALDOA	-0.719	-0.428	-1.847
SH2B2	-0.721	-0.724	-0.573
PTP4A1	-0.75	-0.568	-0.259
ACADSB	-0.751	-0.356	-0.814
FXYD1	-0.754	-0.534	0.013
ATP5F1D	-0.762	-0.603	-0.03
CAPN3	-0.774	-0.525	0.399
DUSP26	-0.792	-0.661	-0.388
SLC25A12	-0.834	-0.479	-0.32
TAPT1	-0.865	-0.37	-0.506
GOT2	-0.88	-0.612	-0.3
PYGM	-0.956	-0.432	-1.556
LRRC20	-0.968	-0.635	-0.208
CAMK2G	-0.996	-0.652	-0.445

**Table 2.** Common muscle disease module (CMDM) genes. Genes are listed from the largest absolute meta-effect size to the smallest (from summary effect size in discovery set).



**Figure 2.** Validation of the 52-gene set of genes expressed across human muscle diseases. **(A)** ROC curves comparing 657 patients in the validation cohort. **(B)** ROC curves comparing 81 patients in the secondary validation cohort. **(C)** Heat map shows consistent differential expression in the majority of discovery, validation, and secondary validation cohort data sets. Columns represent CMDM genes ranked from the highest to the lowest standardized mean difference (Hedges’ g in z-scaled log<sub>2</sub> values) from left to right. Rows denote data sets used in each stage of meta-analysis, arranged by unsupervised hierarchical clustering using Ward’s minimum variance method. **(D)** Violin plots of CMDM muscle disease severity scores for a cohort of congenital muscle diseases, classified by degree of fibrosis (none, mild, moderate, severe), GSE109178. Error bars show middle quartiles. P values calculated with Wilcoxon rank-sum test. Jonckheere’s trend test shows significant association (two-tailed  $p \leq 0.05$ )  $p = 4.45 \times 10^{-9}$ . Refer to Table 1 for data set information. ICUAW intensive care unit acquired weakness, IM inflammatory myopathies, DI disuse and immobility, CMD congenital muscle disorders, CSM chronic systemic diseases affecting muscle. Genes unavailable for a dataset are shown in grey.

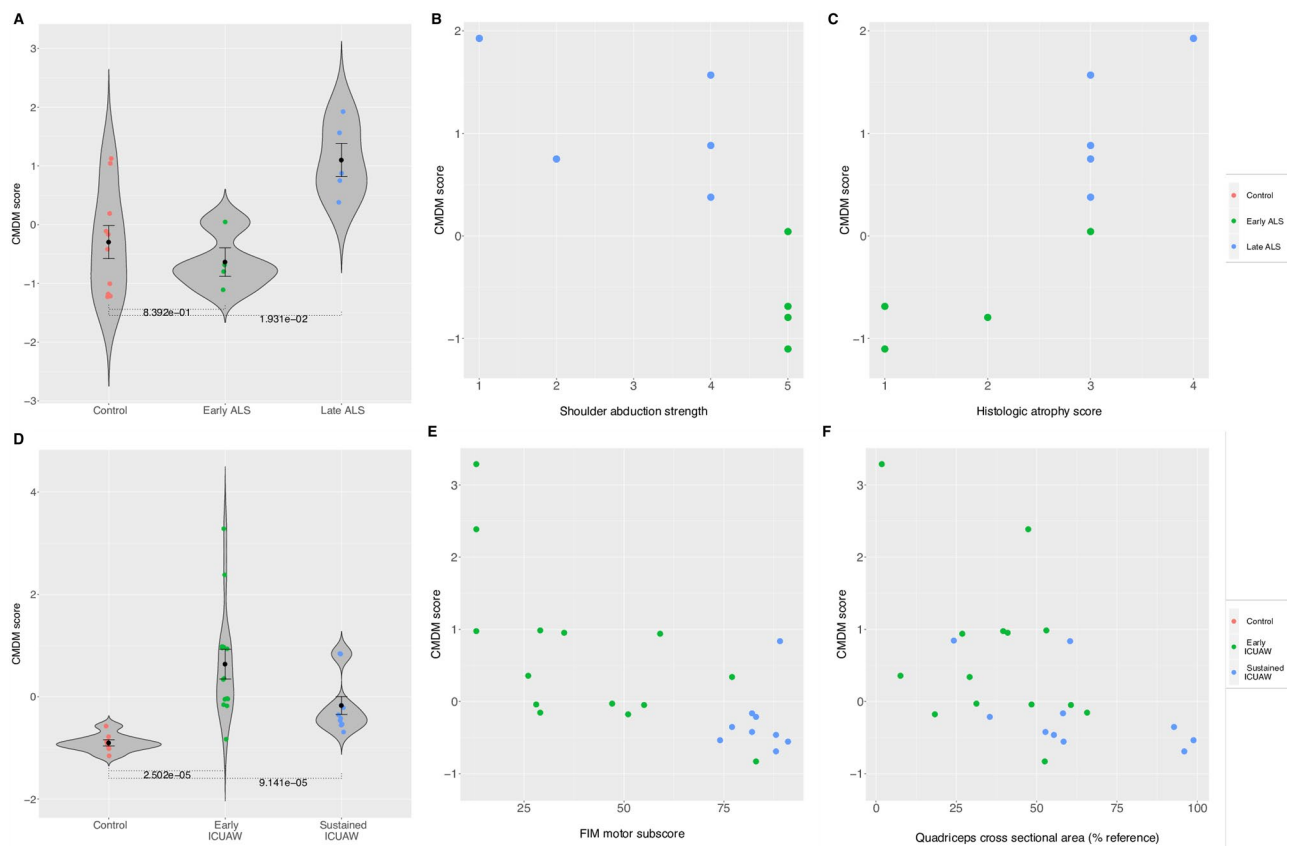
expression relative to controls from both discovery and validation cohorts combined. A total of 74 GO Biological Process (BP) terms were significantly enriched (FDR q-value < 0.05) after removing redundant GO terms (Supplementary Table 4). Twelve GO terms were down-regulated and 62 up-regulated. Networks of overlapping significantly enriched up- and down-regulated GO terms were visualized to aid in the interpretation of the GO enrichment results (Supplementary Fig. 3).

The most down-regulated and up-regulated gene sets based on enrichment score (ES) were *regulation of skeletal muscle adaptation* (ES = -0.80, q-value =  $7.9 \times 10^{-4}$ ) and *macrophage migration* (ES = 0.73, q-value =  $7.3 \times 10^{-4}$ ), respectively. Nine of the 12 down-regulated gene sets were related to mitochondrial metabolism, including *2-oxoglutarate metabolism* and *synthesis of (ubi)quinone*, a redox-active lipid that participates in several processes including mitochondrial electron transport. Four upregulated gene sets were related to collagen metabolism and



Cohort	Disease category	Clinic measure correlated to CMDM score	Correlation <i>r</i>	p-value
GSE109178	Congenital	Mild fibrosis vs no fibrosis	–	$2.68 \times 10^{-5}$
GSE109178	Congenital	Moderate fibrosis vs no fibrosis	–	$6.66 \times 10^{-4}$
GSE109178	Congenital	Severe fibrosis vs no fibrosis	–	$1.49 \times 10^{-5}$
EMEXP3260	MND	Muscle atrophy	0.88	$1.62 \times 10^{-3}$
EMEXP3260	MND	Muscle strength	–0.77	0.014
EMEXP3260	MND	Early ALS vs controls	–	0.84
EMEXP3260	MND	Late ALS vs controls	–	0.019
GSE78929	ICUAW	Functional independence measure (FIM) motor subscore	–0.59	$3.10 \times 10^{-3}$
GSE78929	ICUAW	Quadriceps muscle mass	–0.50	0.011
GSE78929	ICUAW	Early ICUAW (day 7 post-ICU discharge) vs controls	–	$2.50 \times 10^{-5}$
GSE78929	ICUAW	Sustained ICUAW (month 6 post-ICU discharge) vs controls	–	$9.14 \times 10^{-5}$
GSE34111	Chronic	Quadriceps muscle strength	–0.09	0.73

**Table 3.** Associations of common muscle disease module (CMDM) scores with clinical and histological measures of disease severity.



**Figure 3.** CMDM score significantly associates with clinical and histological severity in ALS and ICUAW. Plots of (A) CMDM scores in violin plots for controls, early ALS, and late ALS (EMEXP3260) Error bars show middle quartiles. P-values calculated with Wilcoxon rank-sum test. (B) Grading score of muscle atrophy in ALS (based on histology) versus CMDM score. (C) Shoulder abduction (muscle strength) versus CMDM score. (D) CMDM scores in violin plots for controls, early ICUAW (Day 7 post-ICU), and sustained ICUAW (month 6 post-ICU; GSE78929). (E) CMDM scores versus functional independence measure [FIM] motor subscore. (F) CMDM score versus quadriceps muscle mass. Each dot corresponds to individual samples. ICUAW ICU acquired weakness, ALS acute amyotrophic lateral sclerosis.

extracellular structure organization. Forty-four upregulated gene sets were related to immune system processes including *neutrophil activation*, *innate immune response* and *antigen processing*.

**Disease-specific patterns of gene expression changes.** We hypothesized that functional analysis of a disease category-specific gene signature, after removing genes shared with other disease categories, would

provide insights into the unique pathomolecular mechanisms underlying each individual muscle disease. Thus, for each disease category we used the meta-analysis approach to generate a rank ordered list of genes based on expression relative to controls for the combined discovery and validation cohorts. We next visually examined the location of each of the CMDM genes, within the ordered list of genes, for each disease category (Fig. 4A). The CMDM genes were more densely distributed amongst the most up- and down-regulated genes for each muscle-specific category gene list, validating that the CMDM genes are similarly dysregulated in each muscle disease.

We then utilized the “leave-one-disease-out” meta-analysis approach to iteratively generate ranked lists containing four of the five disease categories. Significantly DE genes identified in the four-disease meta-analysis gene list were then removed from the single disease category meta-analysis gene list. We removed 359, 99, 200, 484, and 365 genes from the gene lists for CMD, IM, ICUAW, DI, and CSM, respectively.

The disease-specific gene lists represent genes that are expressed more strongly in a specific muscle disease category (Fig. 4B). The gene lists were then assessed for GO term enrichment to identify disease-specific pathways (Supplementary Fig. 4A–E, Supplementary Table 5A–E). Despite removing genes significant to other disease categories, persistent down-regulation of (i) gene sets related to mitochondrial electron transport were found in IM, DI, ICUAW, and CSM cohorts and (ii) gene sets related to mitochondrial translation were found in CMD, DI, and CSM cohorts. Significant down-regulation of genes sets related to muscle contraction were identified in CMD, IM, and ICUAW. Upregulation of mRNA splicing via spliceosome was found in ICUAW and DI. Upregulation of extracellular matrix organization genes were observed in CMD and IM. Significant up regulation of NF- $\kappa$ B signaling genes was found in IM and ICUAW categories. Only ICUAW had down-regulation of genes related to cell fate specification, including SOX17, a transcription factor induced during satellite cell specification<sup>23</sup>.

**Subcellular localization analysis of CMDM genes.** We explored whether the CMDM gene signature is overrepresented in certain subcellular compartments. The majority of CMDM genes (98.1%) mapped to at least one subcellular localization. The vesicular exosome contained the greatest proportion of CMDM genes (25%) and was significantly overrepresented in CMDM signature ( $q$ -value = 0.039) (Supplementary Table 6A and 6B, Supplementary Fig. 5). One gene (*FST*) was extracellular.

## Discussion

We analyzed the reported transcriptomes of 1374 individual muscle samples collected from 45 independent patient/control cohorts classified into five categories of skeletal muscle disease and derived using multiple microarray platforms, to identify and validate a robust and reproducible gene signature of muscle disease. This analysis leveraged both biological and technical heterogeneity across multiple independent cohorts in the discovery cohort to avoid overfitting and validated the CMDM signature using cohorts containing larger sample sizes to reduce technical heterogeneity<sup>11</sup>. To assess the generalizability of the signature we examined three more cohorts that could not be classified within the five specified muscle disease categories and found the signature to be reproducible in these cohorts as well.

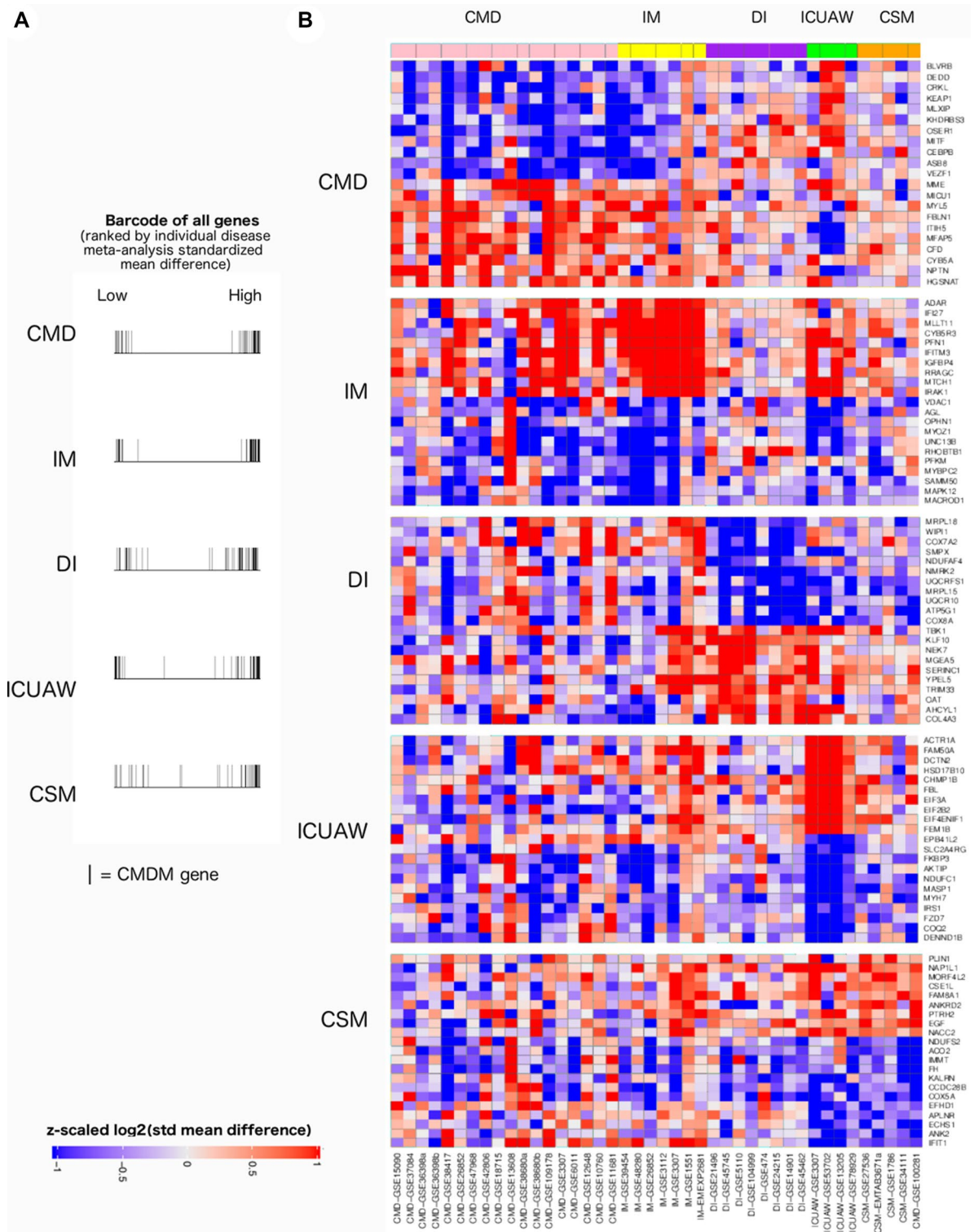
Although there are heterogeneous muscle types and diverse genetic and acquired causes of different muscle diseases, the 52-gene CMDM reflected convergent transcriptional pathways across peripheral skeletal muscles affected by disease. This implies that prior characterization of any of the genes in the CMDM may be relevant across muscle diseases. Several of the genes in the CMDM have been associated with muscle disease previously, whereas many remain unknown or poorly characterized in skeletal muscle.

Increased expression of cholinergic receptor nicotinic, alpha 1 (*CHRNA1*), the most robustly DE CMDM gene, has been recognized as a marker of severity of muscle denervation<sup>24,25</sup>. Up-regulation of *CHRNA1* has been reported to be associated with dynamic epigenetic modifications of the gene in a rat model of disuse-induced atrophy<sup>26</sup>. The most down-regulated CMDM gene *CAMK2G*, calcium/calmodulin-dependent protein kinase type II (CaMKII) subunit gamma, is involved in sarcoplasmic reticulum Ca<sup>2+</sup> transport in skeletal muscle and has been shown to remain active after exercise<sup>27</sup>. While agonists of CaMKII have been proposed as potential pharmacologic therapies of in various muscle disease<sup>28</sup>, it has remained unclear which of the CaMKII subunits is most important in the regulation of skeletal muscle adaptation, response to injury and activity, and oxidative capacity as these subunits are currently not well characterized. Given that *CAMK2G* is down-regulated across most muscle diseases in this study we propose that it may be a suitable target for future studies of potential therapeutics.

Although able to robustly identify a broad range of muscle diseases, the CMDM signature more importantly strongly correlates with clinical and histological measures of disease severity, providing persuasive evidence that the signature could have future applications as a biomarker for phenotyping muscle disease. The CMDM signature could specifically provide diagnostic information and quantify the molecular response to therapy for muscle disease. Measuring changes in CMDM scores after treatment may improve the identification of therapy responders, and using it at enrollment in therapeutic trials may aid the stratification of patients within trial arms. Furthermore this signature could also serve to phenotype patients with COPD, ICUAW and other chronic respiratory diseases based on the extent of muscle dysfunction.

We applied gene set enrichment analysis to identify functional pathways that are similarly altered across muscle diseases. As expected, genes involved in skeletal muscle skeletal adaptation and mitochondrial function were down-regulated. Coordinate down-regulation of mitochondrial genes has been described in a number of muscle diseases<sup>29–32</sup>. The predominant up-regulated functional terms were related to immune activation. Muscle damage secondary to disease induces immune activation culminating in inflammation and deposition of extracellular matrix (ECM)<sup>33,34</sup>. Skeletal muscle diseases are characterized by up-regulation of ECM genes including collagen, with progressive development of fibrosis leading to dysfunctional muscle tissue<sup>35,36</sup>. Collectively, these





**Figure 4.** Disease-specific meta-analysis. (A) Distribution of the 52 CMD genes among individual disease meta-analysis gene lists. Each line presents the presence of a CMDM gene among the 24,572 gene probes generated from disease-specific meta-analysis ranked from the most positive standardized mean difference (left) to the most negative standardized mean difference (right). (B) Disease-specific meta-analysis after removing genes differentially expressed across the other four disease categories, identifies genes more strongly expressed in a single disease. Top 10 up-regulated and top 10 down-regulated genes shown (if 20 or more genes present). ICUAW intensive care unit acquired weakness, IM inflammatory myopathies, DI disuse and immobility, CMD congenital muscle disorders, CSM chronic systemic diseases affecting muscle.

findings are consistent with literature in chronic skeletal muscle diseases proposing the convergence of final common pathways including chronic inflammation, fibrosis, oxidative stress, and mitochondrial dysfunction<sup>36</sup>.

We next removed genes similarly dysregulated across muscle disease to identify pathways altered within specific muscle diseases. This strategy identified pathways unique to ICUAW as well as those shared with other muscle diseases. Significant up-regulation of NF- $\kappa$ B signaling genes was found in IM and ICUAW. NF- $\kappa$ B has been previously shown to play a role in IM<sup>37,38</sup>, has been studied in animal models of cancer cachexia and ICUAW<sup>39–41</sup> and has been shown to be an inhibitor of skeletal myogenesis and muscle regeneration<sup>42</sup>. Remarkably, only ICUAW had down-regulation of genes related to cell fate specification. Decreased numbers of satellite cells (precursors to skeletal muscle cells) in ICUAW sustained long term, compared to healthy controls have been detected histologically<sup>43</sup>, supporting the finding of a down-regulated stem-cell gene set.

An unexpected finding of our meta-analysis was that the CMDM signature is enriched for genes targeted to the exosomal vesicle. Vesicular exosomes, cell derived vesicles containing signaling factors (including genes and microRNAs) for intercellular communication, have been found to have roles in muscle regeneration and congenital muscle diseases<sup>44</sup>. Monitoring exosomal miRNAs has been proposed as a non-invasive method for tracking muscle disease progression<sup>44,45</sup>. Future studies will assess whether plasma protein concentrations of the exosomal CMDM genes correlate with muscle severity to the same extent as their transcripts.

Our meta-analysis has limitations despite its comprehensiveness. Although most included studies attempted to select patients without co-morbidities that span more than one muscle disease category, there are potentially multiple pathologies in some of the muscle samples. Given the number of cohorts and size of the overall study, such confounding is likely to be minimal. The broad inclusion criteria applied in this study has identified a robust disease signature that reflects the heterogeneity observed in the real-world patient population. The considerable variance in gene expression profiles between the different muscle tissue sites<sup>46</sup> included in this analysis is expected to have reduced the number of significant genes, while increasing the generalizability of the significant genes detected. We primarily focused on identifying a gene signature that is conserved between several muscle disease categories and across samples. Although this is beneficial for capturing features that are consistent across multiple diseases, it is ill-suited for identifying subgroups of disease.

Based on the use of microarray data from multiple platforms, we cannot test for alterations in splicing regulation, which has been associated with several congenital muscle diseases including the most common adult onset muscular dystrophies<sup>47</sup>. Analysis of RNA-seq transcriptome data will be necessary to determine whether altered splice variants lead to muscle pathology in other disease categories. Identification of conserved epigenetic signatures of muscle disease will provide important insights into the underlying mechanisms resulting in gene transcriptomic dysregulation identified here, once future epigenome-wide association studies of various muscle diseases are available.

The cohort GSE34111 had a global expression pattern that differed markedly from the other muscle diseases and disease categories. As this cohort was the only one in the analysis that included cancer cachexia, it remains unclear whether the difference in global expression pattern reflects significant differences in the pathomechanism of cancer cachexia or technical or experimental differences in the study. Future analysis comparing peripheral muscle from patients with cancer cachexia and controls are required. Within the validation set, the chronic systemic disease and ICUAW categories each consisted of one cohort, reducing the power to detect significant effect size differences from controls within these disease categories. For this reason, disease specific pathway analysis was performed by combing both discovery and validation cohorts.

CMDM genes may be conceptually divided into those having direct etiological contribution to muscle disease and those that represent a secondary phenomenon in the development of muscle disease, include stress-related changes or cell survival mechanisms<sup>48</sup>. Further experimentation will be required to identify the CMDM genes directly contributing to disease as these genes are expected to be good candidates for novel disease modifying therapies<sup>14</sup>. CMDM genes without functional annotation can be prioritized for future experimental evaluation based on the strength of the molecular data (e.g. effect size or correlation with clinical phenotype). Direct experimentation will be necessary to determine the role of the dysregulated genes and pathways in muscle disease as either causal drivers or responses to muscle disease.

Our results identify a conserved muscle disease transcriptional signature associated with clinical and histologic disease severity, and identify numerous novel genes associated with muscle disease severity. Muscle disease specific analysis identifies pathways uniquely altered in ICUAW. Thus our findings serve as a valuable resource for interpreting disease mechanisms, connecting findings across muscle diseases, and driving novel hypotheses.

## Methods

The analysis workflow is shown in Supplementary Fig. 1.

**Data collection and pre-processing.** Two public gene expression microarray repositories (ArrayExpress, NIH GEO) were searched for human muscle disease datasets (search date: Aug 29, 2019). We found 45 independent datasets with 1374 muscle biopsies that met our inclusion criteria (Supplementary Methods).

We divided the sample cohorts into 6 disease categories for analysis: (1) inflammatory myopathies (IM), (2) ICU acquired weakness (ICUAW), (3) congenital muscle diseases (CMD), (4) chronic systemic disease affecting muscle (CSM), (5) disuse and immobility (DI), (6) motor neuron disease (MND). Next, we divided the patient cohorts into a *discovery cohort* for the initial meta-analysis and a validation cohort for the independent validation analysis. For the discovery cohort we ensured that there were at least three cohorts for each disease category that met our inclusion criteria. As there were only two cohorts for the MND category, this was not included in the discovery cohort as a disease category; instead, the two MND cohorts were included in the secondary validation cohort.

Normalization and probe expression summarization are described in Supplementary Methods. The number of studies measured for each gene are listed in Supplementary Tables 2 and 3.

**Meta-analysis.** Multicohort meta-analysis of gene expression was performed (using R package *MetaIntegrator*)<sup>12</sup> as described in the Supplementary Methods. The utility of the leave-one-disease-out approach in identifying a robust gene expression signature during acute rejection across different transplanted solid organs<sup>13</sup> and across neurodegenerative diseases<sup>14</sup> has been shown before.

**Derivation of common muscle disease module (CMDM) score.** We applied a greedy forward search as described in the Supplementary Methods section to identify a gene signature maximized for diagnostic power, termed the Common Muscle Disease Module (CMDM).

**Validation of CMDM score and correlation of the CMDM genes with clinical and histological severity.** Tukey's Biweight correlation was used to assess the association of the CMDM score with the histologic and clinical measures. Between- and within-group CMDM score comparisons were done with the Wilcoxon rank sum test.

**Muscle disease category specific meta-analysis.** To identify patterns on gene expression changes that are unique to each muscle disease category we performed meta-analysis using the combined data of the discovery and validation cohorts. We first analyzed each disease category separately, as well as the other four diseases together. Genes that were significantly differentially expressed in the four-disease category meta-analysis were then removed from the individual disease category meta-analysis, thereby removing the DE genes common to all muscle disorders, from the disease-specific gene list.

Gene ontology functional analysis identified functional themes within differentially expressed genes across muscle disease categories as described in the Supplementary Methods.

Subcellular localization analysis was performed for each gene within the CMDM as described in the Supplementary Methods.

All analyses were completed in R language for statistical computing (version 3.4.1). Significance levels were set at two-tailed  $p < 0.05$ , unless specified otherwise.

## Data availability

The datasets supporting the results of this article are available in GEO and ArrayExpress online repositories at <http://www.ebi.ac.uk/arrayexpress/> and <http://www.ncbi.nlm.nih.gov/geo/>. Data set accession numbers can be found in Table 1.

Received: 12 November 2021; Accepted: 3 May 2022

Published online: 04 July 2022

## References

- Walsh, C. J. *Transcriptional Profiling and Regulation in Survivors of Critical Illness with Muscle Weakness and Meta-analysis across Human Muscle Diseases*. TSpace <http://hdl.handle.net/1807/97727>. (University of Toronto, 2019).
- Barreiro, E., Sznajder, J. I., Nader, G. A. & Budinger, G. R. Muscle dysfunction in patients with lung diseases: A growing epidemic. *Am. J. Respir. Crit. Care Med.* **191**(6), 616–619 (2015).
- Abdulai, R. M. *et al.* Deterioration of limb muscle function during acute exacerbation of chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **197**(4), 433–449 (2018).
- Riou, M. *et al.* Skeletal and respiratory muscle dysfunctions in pulmonary arterial hypertension. *J. Clin. Med.* **9**(2), 410 (2020).
- Puthuchery, Z. A. *et al.* Rectus femoris cross-sectional area and muscle layer thickness: Comparative markers of muscle wasting and weakness. *Am. J. Respir. Crit. Care Med.* **195**(1), 136–138 (2017).
- Man, W. D. *et al.* Non-volitional assessment of skeletal muscle strength in patients with chronic obstructive pulmonary disease. *Thorax* **58**(8), 665–669 (2003).
- Pradat, P. F. *et al.* Muscle gene expression is a marker of amyotrophic lateral sclerosis severity. *Neurodegener. Dis.* **9**(1), 38–52 (2012).
- Walsh, C. J. *et al.* Transcriptomic analysis reveals abnormal muscle repair and remodeling in survivors of critical illness with sustained weakness. *Sci. Rep.* **6**, 29334 (2016).
- Dadgar, S. *et al.* Asynchronous remodeling is a driver of failed regeneration in Duchenne muscular dystrophy. *J. Cell Biol.* **207**(1), 139–158 (2014).
- Walsh, C. J., Hu, P., Batt, J. & Santos, C. C. Microarray meta-analysis and cross-platform normalization: Integrative genomics for robust biomarker discovery. *Microarrays (Basel)* **4**(3), 389–406 (2015).
- Sweeney, T. E., Haynes, W. A., Vallania, F., Ioannidis, J. P. & Khatri, P. Methods to increase reproducibility in differential gene expression via meta-analysis. *Nucleic Acids Res.* **45**(1), e1 (2017).
- Haynes, W. A. *et al.* Empowering multi-cohort gene expression analysis to increase reproducibility. *Pac. Symp. Biocomput.* **22**, 144–153 (2017).
- Khatri, P. *et al.* A common rejection module (CRM) for acute rejection across multiple organs identifies novel therapeutics for organ transplantation. *J. Exp. Med.* **210**(11), 2205–2221 (2013).
- Li, M. D., Burns, T. C., Morgan, A. A. & Khatri, P. Integrated multi-cohort transcriptional meta-analysis of neurodegenerative diseases. *Acta Neuropathol. Commun.* **2**, 93 (2014).
- Sweeney, T. E., Shidham, A., Wong, H. R. & Khatri, P. A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. *Sci. Transl. Med.* **7**(287), 287ra271 (2015).
- Sweeney, T. E., Braviak, L., Tato, C. M. & Khatri, P. Genome-wide expression for diagnosis of pulmonary tuberculosis: A multicohort analysis. *Lancet Respir. Med.* **4**(3), 213–224 (2016).
- Sweeney, T. E., Wong, H. R. & Khatri, P. Robust classification of bacterial and viral infections via integrated host gene expression diagnostics. *Sci. Transl. Med.* **8**(346), 346ra391 (2016).



18. Team H-CSP, Consortium H-I: Multicohort analysis reveals baseline transcriptional predictors of influenza vaccination responses. *Sci. Immunol.* **2**(14) (2017).
19. Lofgren, S. *et al.* Integrated, multicohort analysis of systemic sclerosis identifies robust transcriptional signature of disease severity. *JCI Insight* **1**(21), e89073 (2016).
20. Maslove, D. M. *et al.* Validation of diagnostic gene sets to identify critically ill patients with sepsis. *J. Crit. Care* **49**, 92–98 (2019).
21. Francisco, N. M. *et al.* Diagnostic accuracy of a selected signature gene set that discriminates active pulmonary tuberculosis and other pulmonary diseases. *J. Infect.* **75**(6), 499–510 (2017).
22. Mayhew, M. B. *et al.* A generalizable 29-mRNA neural-network classifier for acute bacterial and viral infections. *Nat. Commun.* **11**(1), 1177 (2020).
23. Alonso-Martin, S. *et al.* SOXF factors regulate murine satellite cell self-renewal and function through inhibition of beta-catenin activity. *Elife* **7**, 26039 (2018).
24. Anderson, D. M. *et al.* Severe muscle wasting and denervation in mice lacking the RNA-binding protein ZFP106. *Proc. Natl. Acad. Sci. U S A* **113**(31), E4494–4503 (2016).
25. von Grabowicki, Y. *et al.* Transcriptional activator TAp63 is upregulated in muscular atrophy during ALS and induces the pro-atrophic ubiquitin ligase Trim63. *Elife* **5**, 5 (2016).
26. Fisher, A. G. *et al.* Transcriptomic and epigenetic regulation of disuse atrophy and the return to activity in skeletal muscle. *FASEB J.* **31**(12), 5268–5282 (2017).
27. Rose, A. J. & Hargreaves, M. Exercise increases Ca<sup>2+</sup>-calmodulin-dependent protein kinase II activity in human skeletal muscle. *J. Physiol.* **553**(Pt 1), 303–309 (2003).
28. Chin, E. R. The role of calcium and calcium/calmodulin-dependent kinases in skeletal muscle plasticity and mitochondrial biogenesis. *Proc. Nutr. Soc.* **63**(2), 279–286 (2004).
29. Jiroutkova, K. *et al.* Mitochondrial function in skeletal muscle of patients with protracted critical illness and ICU-acquired weakness. *Crit. Care* **19**, 448 (2015).
30. Abadi, A. *et al.* Limb immobilization induces a coordinate down-regulation of mitochondrial and other metabolic pathways in men and women. *PLoS ONE* **4**(8), e6518 (2009).
31. Taivassalo, T. & Hussain, S. N. Contribution of the mitochondria to locomotor muscle dysfunction in patients With COPD. *Chest* **149**(5), 1302–1312 (2016).
32. Temiz, P., Weihl, C. C. & Pestronk, A. Inflammatory myopathies with mitochondrial pathology and protein aggregates. *J. Neurol. Sci.* **278**(1–2), 25–29 (2009).
33. Dumont, N., Bouchard, P. & Frenette, J. Neutrophil-induced skeletal muscle damage: A calculated and controlled response following hindlimb unloading and reloading. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **295**(6), R1831–R1838 (2008).
34. Madaro, L. & Bouche, M. From innate to adaptive immune response in muscular dystrophies and skeletal muscle regeneration: The role of lymphocytes. *Biomed. Res. Int.* **2014**, 438675 (2014).
35. Gillies, A. R. *et al.* High resolution three-dimensional reconstruction of fibrotic skeletal muscle extracellular matrix. *J. Physiol.* **595**(4), 1159–1171 (2017).
36. Mann, C. J. *et al.* Aberrant repair and fibrosis development in skeletal muscle. *Skelet. Muscle* **1**(1), 21 (2011).
37. Monici, M. C., Aguenouz, M., Mazzeo, A., Messina, C. & Vita, G. Activation of nuclear factor-B in inflammatory myopathies and Duchenne muscular dystrophy. *Neurology* **60**(6), 993–997 (2003).
38. Yang, C.-C., Askanas, V., Engel, W. K. & Alvarez, R. B. Immunolocalization of transcription factor NF- $\kappa$ B in inclusion-body myositis muscle and at normal human neuromuscular junctions. *Neurosci. Lett.* **254**(2), 77–80 (1998).
39. Guttridge, D. C. NF- $\kappa$ B-induced loss of MyoD messenger RNA: Possible role in muscle decay and cachexia. *Science* **289**(5488), 2363–2366 (2000).
40. Mourkioti, F. & Rosenthal, N. NF- $\kappa$ B signaling in skeletal muscle: Prospects for intervention in muscle diseases. *J. Mol. Med. (Berl)* **86**(7), 747–759 (2008).
41. Friedrich, O. *et al.* The sick and the weak: Neuropathies/myopathies in the critically ill. *Physiol. Rev.* **95**(3), 1025–1109 (2015).
42. Bakkar, N. & Guttridge, D. C. NF- $\kappa$ B signaling: A tale of two pathways in skeletal myogenesis. *Physiol. Rev.* **90**(2), 495–511 (2010).
43. Dos Santos, C. *et al.* Mechanisms of chronic muscle wasting and dysfunction after an intensive care unit stay. A pilot study. *Am. J. Respir. Crit. Care Med.* **194**(7), 821–830 (2016).
44. Murphy, C. *et al.* Emerging role of extracellular vesicles in musculoskeletal diseases. *Mol. Aspects Med.* **60**, 123–128 (2018).
45. Koutsoulidou, A. *et al.* Elevated muscle-specific miRNAs in serum of myotonic dystrophy patients relate to muscle disease progress. *PLoS ONE* **10**(4), e0125341 (2015).
46. Terry, E. E. *et al.* Transcriptional profiling reveals extraordinary diversity among skeletal muscle tissues. *Elife* **7**, 1 (2018).
47. Imbriano, C. & Molinari, S. Alternative splicing of transcription factors genes in muscle physiology and pathology. *Genes (Basel)* **9**(2), 107 (2018).
48. Cagliani, R. *et al.* Mutation finding in patients with dysferlin deficiency and role of the dysferlin interacting proteins annexin A1 and A2 in muscular dystrophies. *Hum. Mutat.* **26**(3), 283 (2005).
49. Fredriksson, K. *et al.* Dysregulation of mitochondrial dynamics and the muscle transcriptome in ICU patients suffering from sepsis induced multiple organ failure. *PLoS ONE* **3**(11), e3686 (2008).
50. Langhans, C. *et al.* Inflammation-induced acute phase response in skeletal muscle and critical illness myopathy. *PLoS ONE* **9**(3), e92048 (2014).
51. Bakay, M. *et al.* Nuclear envelope dystrophies show a transcriptional fingerprint suggesting disruption of Rb-MyoD pathways in muscle regeneration. *Brain* **129**(Pt 4), 996–1013 (2006).
52. Arashiro, P. *et al.* Transcriptional regulation differs in affected facioscapulohumeral muscular dystrophy patients compared to asymptomatic related carriers. *Proc. Natl. Acad. Sci. USA* **106**(15), 6220–6225 (2009).
53. Rahimov, F. *et al.* Transcriptional profiling in facioscapulohumeral muscular dystrophy to identify candidate biomarkers. *Proc. Natl. Acad. Sci. USA* **109**(40), 16234–16239 (2012).
54. Perfetti, A. *et al.* Genome wide identification of aberrant alternative splicing events in myotonic dystrophy type 2. *PLoS ONE* **9**(4), e93983 (2014).
55. Tasca, G. *et al.* Different molecular signatures in magnetic resonance imaging-staged facioscapulohumeral muscular dystrophy muscles. *PLoS ONE* **7**(6), e38779 (2012).
56. Nakamori, M. *et al.* Splicing biomarkers of disease severity in myotonic dystrophy. *Ann. Neurol.* **74**(6), 862–872 (2013).
57. Screen, M. *et al.* Gene expression profiling in tibial muscular dystrophy reveals unfolded protein response and altered autophagy. *PLoS ONE* **9**(3), e90819 (2014).
58. Palermo, A. T. *et al.* Transcriptional response to GAA deficiency (Pompe disease) in infantile-onset patients. *Mol. Genet. Metab.* **106**(3), 287–300 (2012).
59. Saenz, A. *et al.* Gene expression profiling in limb-girdle muscular dystrophy 2A. *PLoS ONE* **3**(11), e3750 (2008).
60. Eisenberg, I. *et al.* Mitochondrial processes are impaired in hereditary inclusion body myopathy. *Hum. Mol. Genet.* **17**(23), 3663–3674 (2008).
61. Pescatori, M. *et al.* Gene expression profiling in the early phases of DMD: A constant molecular signature characterizes DMD muscle from early postnatal life throughout disease progression. *FASEB J.* **21**(4), 1210–1226 (2007).

62. Suarez-Calvet, X. *et al.* Altered RIG-I/DDX58-mediated innate immunity in dermatomyositis. *J. Pathol.* **233**(3), 258–268 (2014).
63. Greenberg, S. A. *et al.* Interferon-alpha/beta-mediated innate immune mechanisms in dermatomyositis. *Ann. Neurol.* **57**(5), 664–678 (2005).
64. Barres, R. *et al.* Weight loss after gastric bypass surgery in human obesity remodels promoter methylation. *Cell Rep.* **3**(4), 1020–1027 (2013).
65. Reich, K. A., Chen, Y. W., Thompson, P. D., Hoffman, E. P. & Clarkson, P. M. Forty-eight hours of unloading and 24 h of reloading lead to changes in global gene expression patterns related to ubiquitination and oxidative stress in humans. *J. Appl. Physiol.* **109**(5), 1404–1415 (2010).
66. Urso, M. L., Scrimgeour, A. G., Chen, Y. W., Thompson, P. D. & Clarkson, P. M. Analysis of human skeletal muscle after 48 h immobilization reveals alterations in mRNA and protein for extracellular matrix components. *J. Appl. Physiol.* (1985) **101**(4), 1136–1148 (2006).
67. Alibegovic, A. C. *et al.* Insulin resistance induced by physical inactivity is associated with multiple transcriptional changes in skeletal muscle in young men. *Am. J. Physiol. Endocrinol. Metab.* **299**(5), E752–763 (2010).
68. Rullman, E., Mekjavic, I. B., Fischer, H. & Eiken, O. PlanHab (planetary habitat simulation): The combined and separate effects of 21 days bed rest and hypoxic confinement on human skeletal muscle miRNA expression. *Physiol. Rep.* **4**, 8 (2016).
69. Park, J. J., Berggren, J. R., Hulver, M. W., Houmard, J. A. & Hoffman, E. P. GRB14, GPD1, and GDF8 as potential network collaborators in weight loss-induced improvements in insulin action in human skeletal muscle. *Physiol. Genomics* **27**(2), 114–121 (2006).
70. Turan, N. *et al.* A systems biology approach identifies molecular networks defining skeletal muscle abnormalities in chronic obstructive pulmonary disease. *PLoS Comput. Biol.* **7**(9), e1002129 (2011).
71. Radom-Aizik, S. *et al.* Effects of exercise training on quadriceps muscle gene expression in chronic obstructive pulmonary disease. *J. Appl. Physiol.* (1985) **102**(5), 1976–1984 (2007).
72. Kreiner, F. E., Borup, R., Nielsen, F. C., Schjerling, P. & Galbo, H. Gene expression profiling in patients with polymyalgia rheumatica before and after symptom-abolishing glucocorticoid treatment. *BMC Musculoskelet. Disord.* **18**(1), 341 (2017).
73. Bachinski, L. L. *et al.* Altered MEF2 isoforms in myotonic dystrophy and other neuromuscular disorders. *Muscle Nerve* **42**(6), 856–863 (2010).
74. Osborne, R. J., Welle, S., Venance, S. L., Thornton, C. A. & Tawil, R. Expression profile of FSHD supports a link between retinal vasculopathy and muscular dystrophy. *Neurology* **68**(8), 569–577 (2007).
75. Greenberg, S. A. *et al.* Plasma cells in muscle in inclusion body myositis and polymyositis. *Neurology* **65**(11), 1782–1787 (2005).
76. Zhu, W. *et al.* Genomic signatures characterize leukocyte infiltration in myositis muscles. *BMC Med. Genomics* **5**, 53 (2012).
77. Chen, Y. W. *et al.* Molecular signatures of differential responses to exercise trainings during rehabilitation. *Biomed. Genet. Genom.* **2**, 1 (2017).
78. Gallagher, I. J. *et al.* Suppression of skeletal muscle turnover in cancer cachexia: Evidence from the transcriptome in sequential human muscle biopsies. *Clin. Cancer Res.* **18**(10), 2817–2827 (2012).
79. Willis-Owen, S. A. G. *et al.* COPD is accompanied by co-ordinated transcriptional perturbation in the quadriceps affecting the mitochondria and extracellular matrix. *Sci. Rep.* **8**(1), 12165 (2018).
80. Smith, L. R., Chambers, H. G., Subramaniam, S. & Lieber, R. L. Transcriptional abnormalities of hamstring muscle contractures in children with cerebral palsy. *PLoS ONE* **7**(8), e40686 (2012).

## Acknowledgements

We thank the many scientists who have made our research possible by making their gene expression data accessible in the public domain. We are grateful to Dr Paul Pavlidis for his thoughtful comments and to the MEND-ICU and Canadian Critical Care Translational Biology Group (<https://ccctg.ca/Members>) for their continued support. This work is supported by the Canadian Institutes of Health Research (MOP-137002 to CCDS, J.B, MH, SM; MOP-106545 to CCDS) and Physician Services Incorporated Foundation (PSI 09–21 to JB and CCDS), and Early Research Award from the Ministry of Research and Innovation of Ontario (ERA/MRI 2011). CJW is a recipient of the Canadian Thoracic Society (CTS) Research Committee of the Lung Association (CLA) student-ship (2015–2017).

## Author contributions

C.J.W.: conceived the study design, performed data collection, execution of computational analysis, analysis and interpretation of the data, and manuscript writing. S.M.: data collection, analysis and interpretation of the data. P.H.: analysis and interpretation of the data. G.B.: analysis and interpretation of the data. J.B.: analysis and interpretation of the data. C.C.D.: analysis and interpretation of the data. P.K.: development of computational methods, analysis and interpretation of the data. All authors discussed, critically reviewed, and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-15003-1>.

**Correspondence** and requests for materials should be addressed to C.C.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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





**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022