

Supplementary Table 2. The 30 largest clusters are summarized

Cluster ID	Size	Silhouette	Mean (Year)	Cluster name
0	455	0.959	1993	Gene location
1	307	0.955	2002	Inheritable neuromuscular disorder
2	293	0.976	2005	Auditory specific gene expression pattern
3	257	0.967	1987	Physical mapping
4	237	0.981	1993	DNA rearrangement
5	220	0.970	2003	Nuclear periphery
6	203	0.964	2002	Telomere position effect
7	188	0.939	2011	Emerging player
8	185	0.980	2015	Neuromuscular morphogenesis
9	183	0.979	2003	Genetic disease
10	178	0.992	2013	Musculoskeletal disorder
11	177	0.957	2016	DUX4 expression
12	177	0.965	2000	Transcriptional derepression
13	161	0.988	2009	Long-term regulation
14	158	0.963	1997	Chromosomal repeat
15	157	0.968	1997	Coats disease
16	149	0.993	1991	Complete genomic map
17	149	0.987	1994	Neurological disease
18	148	0.989	2014	Pathogenesis diagnosis
19	142	0.980	2015	Transcriptomic approaches
20	134	0.979	2009	Modeling diseases
21	126	0.992	2012	Prenatal diagnosis
22	124	0.978	2014	Genomic variation
23	120	0.979	2013	Congenital myopathies
24	119	0.970	2010	Facioscapulohumeral muscular dystrophy region gene
25	117	0.993	2005	Aerobic exercise training
26	113	0.997	2010	Human disease
27	113	0.997	1997	Mitochondrial disease
28	113	0.997	2003	Clinical neurophysiology
29	111	0.994	1995	Molecular genetics

*Cluster are numbered starting from 0 with the largest size cluster. Clusters with a lower silhouette value are not numbered.