

Sau3AI (5' GATC)

↓

```

Bsr1 : catcacctgggtgatcatcattgcagagatacgtcacaataccccctgtaggtggggcctagacaagag--g 67
Bsr2 : catcacttaggtgatcagtgccagatgtgttttcagaactccata-gtagactgaacctagagaatgggta 68
Bsr3 : cataacttcgggtgatcagtgccagagatatgtcacaatatccccctgtagaaaaagcctgaaattgattta 69
Bsr4 : catcacctag-tgatcagtgtagagatatgtta-aaattctcgtgtagacagagcctagacaattgtta 67
Bsr5 : aatcacctcagagatcagtgccagagatatgtaccagtgtccccctgtaggcagtgccctagacaagagttg 69
Bsr6 : catcacctgtttgatcagtgccagagatatctcacaaagcccc-tataagccaaaccttgacaagggtta 68
Bsr7 : cttcacctgggtgatcagtgccagtgatgatgtcacaaaatccccctgtagacagagcctagacaagagttta 69
Bsr8 : cctgcctgggctgatcagtgccagggataagtcataaaagcctcctgtaggcagagtgtaggcaagtgttc 69
Bsr9 : catcacctgggtgatcagtgccagagatatgtcacaatatccccctctaggcagagtatatagagaagagttcc 69
Bsr10 : catcacctgggggatcagtgccagagatatgtca-aaacgctcctgtaggctgaacctagacaggagttta 68
Bsr11 : catcacctgggtgatctgtgcagagctatgtca-aaacgccccctgtaggcagagcctagatgagtggtta 68
Bsr12 : catcacctgggtgatcagtgccagagatatgtca-aaacgccccctgtaggcagagcctagatgagtggtta 68
Bsr13 : catcacctgggtgatcagtgccagagatatgtca-aaacgccccctgtaggcagagcctagatgagtggtta 68
Bsr14 : catcacctgggtgatcagtgccagagatatgtca-aaacgccccctgtaggcagagcctaggaagacttc 68

```

Figure 1S. BSR homology. Exhaustive pairwise alignment using neighbour-joining phylogeny analysis by Clone Manager7 software shows the high homology of 12 monomers of the β -satellite repeats (BSR) inside the 4qA allele.

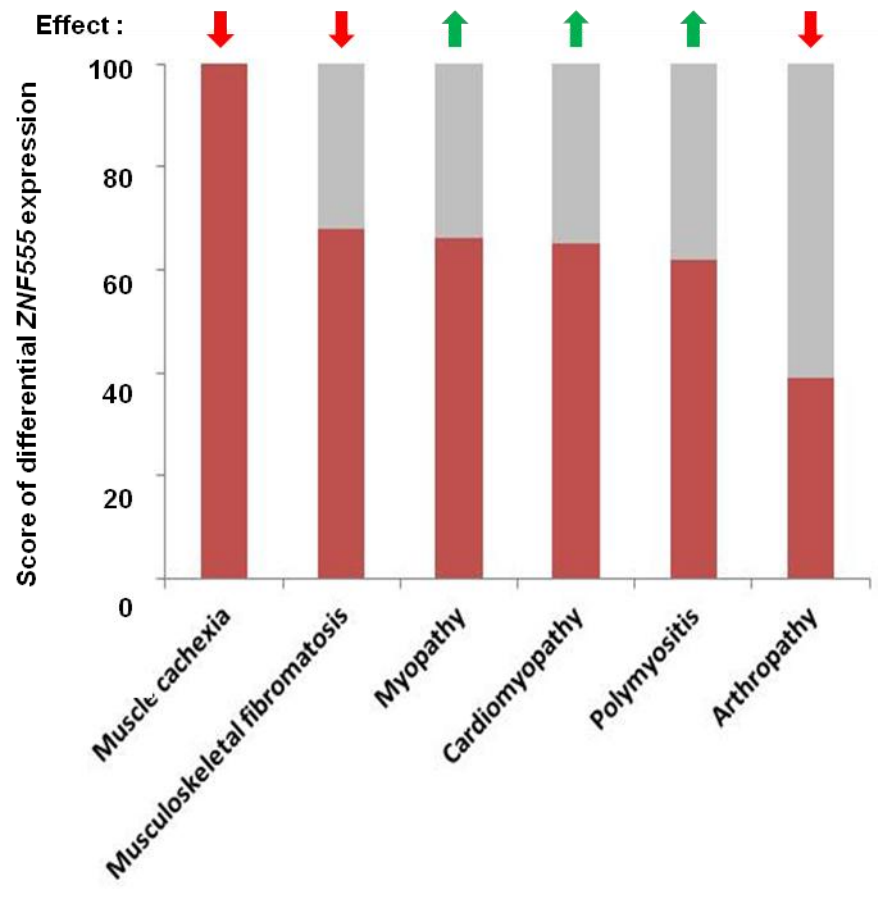
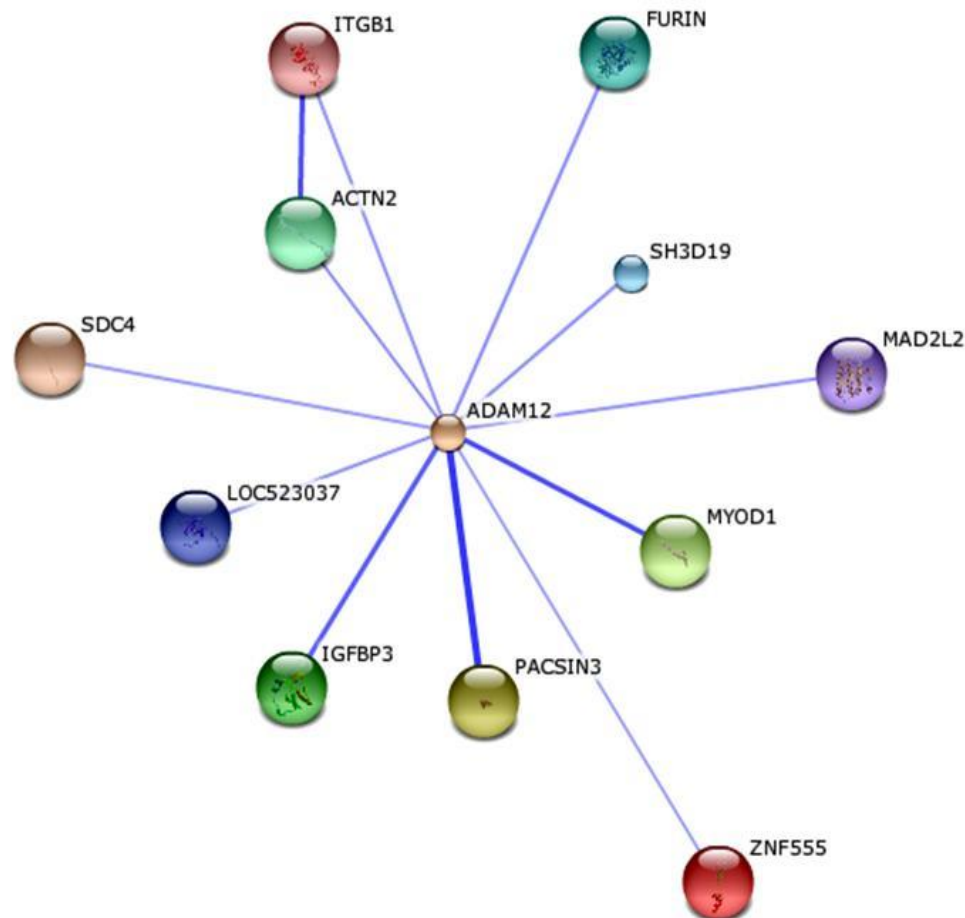


Figure 3S. Analysis of *ZNF555* RNA expression under different pathologies showed the highest score for musculoskeletal and connective tissue diseases. Data were extracted from Disease Atlas (<http://www.nextbio.com>). Disease Atlas ranks diseases, traits, conditions, and surrogate endpoints associated with a gene. Results are grouped by disease and ranked according to statistical significance. The Illumina (Illumina acquired NB) NextBio analysis tool was used to analyze *ZNF555* at different pathologies. NextBio is a curated and correlated repository of experimental data derived from an extensive set of public sources (e.g., ArrayExpress and GEO) that allows the user to compare patterns of gene expression between thousands of genomic signatures derived from published datasets. Statistical analysis is carried out using rank-based enrichment analysis to compute pairwise correlation scores of the uploaded dataset and all studies contained in NextBio. The statistical analysis method used by NextBio is referred to as a “Running Fischer” (51). A numerical score of 100 corresponds to the most significant result for *ZNF555* gene expression and the scores of the other results were normalized to the top-ranked result. Query executed on date. All samples were compared with healthy control.



ADAM12 - A Disintegrin And Metalloproteinase Domain 12

- PACSIN3 - PREDICTED-misc_RNA (PACSIN3), miscRNA
- MYOD1 - Myoblast determination protein 1; Involved in muscle differentiation (myogenic factor)
- IGFBP3 - Insulin-like growth factor-binding protein 3 Precursor (IGF-binding protein 3)(IGFBP-3) (IBP-3)
- ACTN2 - Alpha-actinin-2 (Alpha-actinin skeletal muscle isoform 2)(F-actin cross-linking protein)
- FURIN - Furin Precursor (EC 3.4.21.75)(Paired basic amino acid residue-cleaving enzyme)(PACE)
- SH3D19 - SH3 domain containing 19
- LOC523037 - SH3 multiple domains 1
- MAD2L2 - MAD2 mitotic arrest deficient-like 2
- ITGB1 - Integrin beta-1 Precursor (Fibronectin receptor subunit beta)(Integrin VLA-4 subunit beta)
- SDC4 - syndecan 4

Figure 4S. ADAM12 is a predicted functional partner of ZNF555. The analysis was performed using to the database of known and predicted protein-protein Interactions (<http://string-db.org/>).

Table 1S. Constructs used in the present study.

Vector	Insertion Fragment	Size, bp	Chromosome Start/End, +Strand
p4qA	4qA	1464	Pirozhkova I et al., PLoS One, 2008
p4qA_1	Fr 1	564	Un_gl000228 : 115554 / 116117
p4qA_2	Fr 2	435	Un_gl000228 : 115119 / 115553
p4qA_3	Fr 3	468	Un_gl000228 : 114650 / 115117
p4qA_4	Fr 4=4qAe	67	Un_gl000228 : 116051 /116117
p4qA_5	Fr 5	203	Un_gl000228 : 115915 /116117
p4qA_6	Fr 6	63	Un_gl000228 : 115553 /115615
p4qA_7	Fr 7	106	Un_gl000228 : 114650 /114749
pANT1	ANT1 promoter	715	4q:186063741 /186064455
pFRG1	FRG1 promoter	588	Petrov A et al., Genome Res, 2008
p4qAe_ANT1	4qAe ANT1 promoter	67 715	Un_gl000228 : 116051 /116117 4q:186063741 /186064455
p4qAe_FRG1	4qAe FRG1 promoter	67 577	Un_gl000228 : 116051 /116117 Petrov A et al., Genome Res, 2008

Table 2S. List of primers used in the present study.

Sequence ID	Forward	Reverse
<i>ANT1</i> promoter	ACCCAAGCATGATATGG	TTGACTACTGCTGGAGTG
<i>FRG1</i> promoter	GCTTGATATTGTTGGTGAGT	GACAACCGACTTCTACAAT
<i>FRG2</i> promoter	GTTGTTGTTGAGCCTGG	CCTAGAAGGTCACCGAA
<i>DUX4</i>	ACGGAGACTCGTTTGGA	TGGCCCTTCGATTCTGA
<i>DUX4c</i>	TGGCCCTTCGATTCTGA	GTGGAGGTGGTAGGTCTTT
<i>4qAe</i>	TCCCCTGTAGGCAGAGA	CACTGATAACCCAGGTGA
<i>ZNF555</i>	CCGCCTGCCCCTAGCGGTCC	TGAGCAGAATCCAGCAAAGCCCA
<i>ANT1</i>	GCCAGCAAACAGATCAGTGC	CCCCTCCAGAAGGAGAGGAA
<i>FRG1</i>	GAAACCCGGAAGTGGA ACTCT	CCTGACAGCCTACGTCTCTG
<i>GAPDH</i>	ACCACAGTCCATGCCATCAC	CCAGTGAGCTTCCCGTTCAG

Table 3S. List of proteins identified by analyzing the specific for *4qAe* band (with cold competitor) obtained from DIG-EMSA experiment on TE671 cells. The results presented in this table were obtained by NanoLC-ESI-MS/MS for peptide mass fingerprint (PMF). The gray highlighted proteins were detected only in experimental samples.

MW	ID Swissprot	Entry Name
70898	P11142	Heat shock cognate 71 kDa protein – Homo sapiens
84660	P07900	Heat shock protein HSP 90-alpha - Homo sapiens
121371	A5A3E0	ANKRD26-like family C member 1B - Homo sapiens
42911	P49903	Selenide, water dikinase 1 - Homo sapiens
38170	O75436	Vacuolar protein sorting-associated protein 26A - Homo sapiens
73063	Q8NEP9	Zinc finger protein 555 - Homo sapiens

Supplementary Table 4S. List of proteins identified by analyzing the control (Ctl) band without DNA obtained from DIG-EMSA experiment on TE671 cells. The results presented in this table were obtained by NanoLC-ESI-MS/MS for peptide mass fingerprint (PMF).

MW	ID Swissprot	Entry Name
84660,2	P07900	Heat shock protein HSP 90 - 75 - Homo sapiens
70898,4	P11142	Heat shock cognate 71kDa protein - Homo sapiens
51804,8	Q02790	FK506-binding protein 4 - Homo sapiens
32575,2	P06748	Nucleophosmin - Homo sapiens
36053,4	P04406	Glyceraldehyde-3-phosphate dehydrogenase - Homo sapiens
121371,3	A5A3E0	ANKRD26-like family C member 1B - Homo Sapiens
94300,5	P34932	Heat shock 70 kDa protein 4 - Homo sapiens
164940,4	P31327	Carbamoyl-phosphate synthase [ammonia], mitochondrial precursor - Homo sapiens
27692,8	P30048	Thioredoxin-dependent peroxide reductase, mitochondrial precursor - Homo sapiens
50151,9	P68363	Tubulin alpha-1B chain - Homo sapiens
61055	P10809	60 kDa heat shock protein, mitochondrial precursor - Homo sapiens